UNIVERSITETI I EVROPËS JUGLINDORE
SOUTHEAST EUROPEAN UNIVERSITY
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА

SHKENCAT DHE TEKNOLOGJITË BASHKËKOHORE
CONTEMPORARY SCIENCES AND TECHNOLOGIES
СОВРЕМЕНИ НАУКИ И ТЕХНОЛОГИИ

DOCTORAL STUDIES – THIRD CYCLE

DATA MINING APPROACH IMPROVING DECISION-MAKING COMPETENCY ALONG THE BUSINESS DIGITAL TRANSFORMATION JOURNEY

CANDIDATE:

Hyrmet Mydyti

MENTOR:

Prof. Dr. Arbana Kadriu

Tetovo, June 2022

Declaration

I state that this doctorate thesis comes from the results of my own research at the Department of Contemporary Sciences and Technologies (CST), at the Southeast European University (SEEU).

References for other studies as well as reports of any other scholar are properly referenced. This doctoral dissertation, complete or partial, was not presented or submitted for any other degree.

Name: HYRMET MYDYTI

Signed:

____

Date: 10 June 2022

Proofreading

I, Kujtim Ramadani, translator/interpreter/proof-reader from English into Albanian and Macedonian and vice versa, certify that the PhD thesis "DATA MINING APPROACH IMPROVING DECISION-MAKING COMPETENCY ALONG THE BUSINESS DIGITAL TRANSFORMATION JOURNEY" has been proofread by me and most shortcomings have been eliminated. The paper in question meets the language standards to be defended as a PhD thesis.

Ramadani

Date: 10 June 2022

Abstract

Data mining is an efficient means to enhance the experience of customers and streamline the decision-making throughout the business digital transformation journey. The benefits of data mining, a key aspect of advanced analytics, head to a higher business productivity due to predictive analytics. The thesis aims to provide a model by comprehending views of scholars on the digital transformation, of digital technologies, of advanced analytics techniques, on the implementation of data mining in various areas of activity, and on the views of certain benefits and challenges of the implementation of data mining.

A qualitative research study is developed to comprehend the challenges and issues of home appliances after-sales service business, a type of business chosen as a sample for our research purposes. The research has focused the analysis in studying the impact of service quality on customer satisfaction, a nonfinancial performance measure for managerial decision-making. Hereof, effective strategic solutions are proposed in accordance with the issues and challenges.

Moreover, the primary objective of the quantitative research study is to deliver a case study - practical research in applying data mining techniques. The comparative research of a variety of algorithms helps and enables to analyze information and predict repairs in home appliances after-sales services business. The results provide evidence how data mining techniques affect decision-making. Furthermore, the quantitative research investigates the relationship between service quality attributes and customer satisfaction, a nonfinancial performance measure for managerial decision-making. The results indicate a strong relationship between the time taken to resolve the complaint and customer satisfaction in after-sales service businesses.

In line with the quantitative research study, modern platforms were selected as trends in system building, as the most appropriate applications for the recommendation system, as a portion of data mining system functionalities, as a result of this research. Finally, the evaluation of the impact on decision-making presents how the application of data mining techniques in business can improve efficiency and business productivity along their digital transformation journey.

Abstrakt

Data mining është një mjet efikas për të përmirësuar përvojën e klientëve dhe për të lehtësuar vendimmarrjen përgjatë rrugëtimit të transformimit digjital të biznesit. Përfitimet e data mining, një aspekt kyç i analitikës së avancuar, çojnë drejt një produktiviteti më të lartë të biznesit për shkak të analitikës parashikuese. Teza synon të ofrojë një model duke kuptuar pikëpamjet e studiuesve mbi transformimin digjital, të teknologjive digjitale, të teknikave të avancuara analitike, mbi zbatimin e data mining në fusha të ndryshme të veprimtarisë dhe mbi përfitimet dhe sfidat e zbatimit të data mining.

Një studim kërkimor kualitativ është zhvilluar për të kuptuar sfidat dhe çështjet e biznesit të shërbimit pas shitjes së pajisjeve shtëpiake, një tip biznesi i zgjedhur si mostër për qëllime tona studimore. Hulumtimi ka fokusuar analizën në studimin e ndikimit të cilësisë së shërbimit në kënaqësinë e klientit, një metrikë e performancës jofinanciare për vendimmarrjen menaxheriale. Këtu propozohen zgjidhje strategjike efektive në përputhje me problematikat dhe sfidat.

Për më tepër, objektivi kryesor i studimit kërkimor kuantitativ është të ofrojë një studim rasti - praktik në aplikimin e teknikave të data mining. Studimi krahasues i një sërë algoritmesh ndihmon dhe mundëson analizimin e informacionit dhe parashikimin e riparimeve në biznesin e shërbimeve pas shitjes së pajisjeve shtëpiake. Rezultatet ofrojnë dëshmi se si teknikat e data mining ndikojnë në vendimmarrje. Për më tepër, hulumtimi kuantitativ studion marrëdhënien midis atributeve të cilësisë së shërbimit dhe kënaqësisë së klientit, një metrikë e performancës jofinanciare për vendimmarrjen menaxheriale. Rezultatet tregojnë se ekziston një lidhje e fortë midis kohës së marrë për zgjidhjen e ankesës dhe kënaqësisë së klientit në bizneset e shërbimit pas shitjes.

Në përputhje me studimin kërkimor kuantitativ, si rezultat i këtij studimi u zgjodhën platformat moderne si tendenca në ndërtimin e sistemit, si aplikacionet më të përshtatshme për sistemin rekomandues, si pjesë e funksionaliteteve të sistemit të data mining. Së fundi, evaluimi i ndikimit në vendimmarrje paraqet se si aplikimi i data mining të të dhënave në biznes mund të përmirësojë efikasitetin dhe produktivitetin e biznesit përgjatë rrugëtimit të tyre të transformimit digjital.

Резиме

Дата мајнинг е ефикасно средство за подобрување на искуството на клиентите и насочување на донесувањето одлуки во текот на деловната дигитална трансформација. Придобивките од дата мајнинг, клучен аспект на напредната аналитика, се насочуваат кон повисока деловна продуктивност поради предвидливата анализа. Тезата има за цел да обезбеди модел за разбирање на ставовите на научниците за дигиталната трансформација, дигиталните технологии, напредните аналитички техники, имплементацијата на дата мајнинг во различни области на активност и за придобивките и предизвиците од имплементацијата на дата мајнинг.

Со цел да се разберат предизвиците и проблемите на бизнисот со услуги по продажбата за домашни апарати, развиена е квалитативна студија на случај. Истражувањето ја докажа анализата во проучувањето на влијанието на квалитетот на услугата врз задоволството на клиентите, нефинансиска мерка за извршување на менаџерските одлуки. Притоа, се предлагаат ефективни стратешки решенија во согласност со прашањата и предизвиците.

Покрај тоа, примарна цел на квантитативното истражување е да се спроведе студија на случај - практично истражување во примената на техниките за дата мајнинг. Компаративното истражување на различни алгоритми овозможува да се анализираат информациите и да се предвидат поправки во бизнисот со услуги по продажба на апарати за домаќинство. Резултатите обезбедуваат докази како техниките за дата мајнинг влијаат врз донесувањето одлуки. Понатаму, квантитативното истражување ја истражува врската помеѓу атрибутите за квалитет на услугата и задоволството на клиентите, мерка за нефинансиски перформанси за донесување менаџерски одлуки. Резултатите покажуваат дека постои силна врска помеѓу времето потребно за решавање на жалбата и задоволството на клиентите во бизнисите со услуги по продажба.

Во согласност со квантитативното истражување, модерните платформи беа избрани како трендови во градењето на системот, како најсоодветни апликации за системот за препораки, како дел од функционалностите на системот за дата мајнинг, како резултат на ова истражување. Конечно, евалуацијата на влијанието врз донесувањето одлуки прикажува како примената на

техниките за дата мајнинг во бизнисот може да ја подобри ефикасноста и деловната продуктивност во текот на нивната дигиталната трансформација.

Acknowledgements

This PhD thesis derives from several years of research and experience in the field of digitization of processes and modernization of systems and was conducted by the candidate, Hyrmet Mydyti.

I extend my gratitude to the mentor, Professor Arbana Kadriu, for her productive supervision and advice in preparing this PHD thesis.

I thank my family for their outstanding support and inspiration throughout my PHD study. In particular, my deepest appreciation goes to my husband, my mother, and my father.

More importantly, this thesis is dedicated to my lovely daughter Dastida who inspired me throughout my studies.

Table of contents

List of Tables

List of Figures

List of Equations

1 INTRODUCTION

Today, the digital transformation of businesses and organizations has become very significant because it ultimately impacts various aspects of the business as the opportunities for the adoption of emerging technologies such as big data analytics, social media, mobile technologies, internet of things, cloud technologies, etc., grow (Morakanyane et al., 2017). Digital transformation is the most recent and greatest initiative many businesses are taking to enhance and revolutionize.

Managers have objectives and interests to build possibilities for their businesses to align their strategy with the digital world, by incorporating new technologies into their business models, by placing greater emphasis on process and operations management (Reis et al., 2018).

Advanced data analytics and artificial intelligence are driving the in-depth study and exploration, and advancement for the digital transformation of businesses. Advanced analytics and artificial intelligence are powerful digital technologies that drive business, drive information analytics, predict and monitor business processes (West and Allen, 2018). The analytical systems and smart applications applied by businesses demonstrate the significance of producing results to improve decision-making and productivity, efficiency, and effectiveness (Bughin et al., 2017).

Businesses are making every effort to adjust their strategies to the digital reality, and more essentially assesses their businesses' success of becoming digital (Rachinger et al., 2019; Verhoef et al., 2019). Businesses apply various technologies to carry out analytics, such as mere reports and a dashboard, which are just past performance review and investigation reports (Bumblauskas et al., 2017). Companies are definitely using technology as a data analytics tool within the digital pathway (OECD-BEIS, 2018). Businesses share the goal of using digital technologies as a chance to enhance decision-making competency as the organization continues digital journey (Schwertner, 2017).

Digital transformation capabilities rely on micro basis to navigate innovation ecosystems, re-engineer internal structures, and improve digital maturity (Warner and Wäger, 2019). Maturity models measure the level of digital transformation and identify where the business is at with the transformation process (Patel and Patel, 2019). Digital maturity models are composed of dimensions and criteria (Teichert, 2019). In addition, agile challenges the traditional hierarchical approach of decision-making in question, with its bottom-up approach and the importance of understanding decision-making as a shared process (Moe et al., 2012). The implementation of the digital strategy dimension is eased by the representation of vision, governance, planning, and management processes (Valdez-de-Leon et al., 2016). The improvement of customer experience dimension is achieved through analytics-based segmentation, the use of socially informed knowledge and improved digital sales practices, the application of predictive marketing and simplified customer processes, the use of digital customer service to ensure consistency accross channels and digital self-service (Hie, 2019). The technology dimension refers to the capabilities to plan, deploy and integrate digital activities effectively (Valdez-de-Leon et al., 2016). The operations dimension includes the execution and development of processes and tasks through the application of digital transformation technologies to boost strategic governance and enhance the efficiency and efficacy of the business. The organization dimension is linked to the evolution of communications, culture, structure, training, and knowledge management within the business, and making it a digital actor (Valdez-de-Leon et al., 2016).

Data mining, as an intersection of statistics and machine learning, involves the sorting through a vast quantity of data housed in repositories, and business databases to detect correlations, patterns, and trends and establish relations to solve problems by analyzing data (Rohanizadeh and Bameni, 2009; Palmer et al., 2011). Data mining will help businesses reduce costs and improve the customer experience during their digital transformation (Sima et al., 2020; Chen et al., 2015). Businesses will focus more on the client by moving their services forward and gaining time for their customers by stepping up their proceedings (Dias et al., 2017).

Digitization facilitates enormous improvements in aftermarket services; through the capabilities of prescriptive and predictive analytics, companies are reducing lifecycle expenditures and optimizing expenditures (Rudnick et al., 2020). The rapid development of technology, strong competitiveness and growing benefits mean that after-sales services are a reform of commercial traditions. As a result, the after-sales service strategy is affected and after-sales services ensure that if clients are satisfied of their services (Othman et al., 2020). In this regard, predictive and prescriptive analytics, deep learning, and artificial intelligence are of increasing importance (Gimpel et al., 2018).

Businesses are able to recognize customer behavior, enhance production abilities, facilitate empirical evidence business decision-making, and lower the lag of novelty by conducting advanced analytics (Gimpel et al., 2018). The customer experience goes beyond the point of sale, so digital technologies are transforming after-sales service strategies (Gimpel et al., 2018).

The main benefit of implementing the data mining approach in businesses is the positive effect on customer experience and streamlining the decision-making as the business continues digital transformation.

The issue of the data mining approach, providing links between businesses and customers was identified as very important to this thesis. This thesis intends to help overcome some of the barriers that emerged in the absence of application of the data mining system.

Moreover, this thesis outlines the research methodology to investigate the requirements of the after-sales services business and the implementation of the most suitable technologies, algorithms and tools of data mining approach, investigation of the relationship between service quality attributes and nonfinancial performance measure for managerial decision-making, implementation areas of data mining and building a recommendation system, as a portion of data mining system functionalities - as an outcome.

Data mining algorithms comparison implementation and analysis on finding the key attributes of service quality impacting customer satisfaction are key steps of quantitative research study. The recommendation system contributes to this thesis, used to streamline the decision-making process, and provide reliable predictions, particularly for customers, improve the process of creating new experiences, as well as to enhance the productivity of the business. The system acts as a bridge between businesses and their peers.

Furthermore, the qualitative study with top management respondents of after-sales service businesses, produces recommendations for after-sales services businesses and suggest steps of activities to be taken in order to increase the level of service quality. The outcome benefits include reduction in cost and effort, and improvement in performance and service quality of the after-sales service of home appliances.

1.1 Problem Statement

Businesses use various technologies to carry out analytics, inspect and investigate past performance. Clearly, businesses are using technologies as a way to perform data analytics as part of digital transformation.

The businesses do not utilize data mining tools to conduct advanced analytics. Traditional processing of decision-making is retarding and time consuming compared to advanced analytics that discover deeper insights and make predictions. Hence, there is a lack of mining of data and, consequently, the provision of strategic and actionable information for better decision-making along the business digital transformation journey.

Moreover, after-sales service businesses have concerns while attempting to offer qualitative service which impacts nonfinancial performance measures and attempting to be digitally competitive. The challenges and issues that after-sales service business, in particular home appliances, cope with, in the Kosovo market are not tackled.

Thus, the aim of this thesis is to help overcome some of the barriers that appeared because of the lack of data mining approach application in the perspective of digital transformation and the lack of tackling the concerns and issues that home appliances businesses are coping with.

By studying the potential research gaps, we state a problem that we target:

How data mining approach brings strategic and actionable information for better decision-making by decision-makers along the business digital transformation journey?

Our thesis aims to study this problem by covering two types of research, qualitative and quantitative study. As a result, based on the studies and research performed, we aim to propose models that can help top management and decision makers solve different situations they cope with in order to improve business performance and financial performance measures.

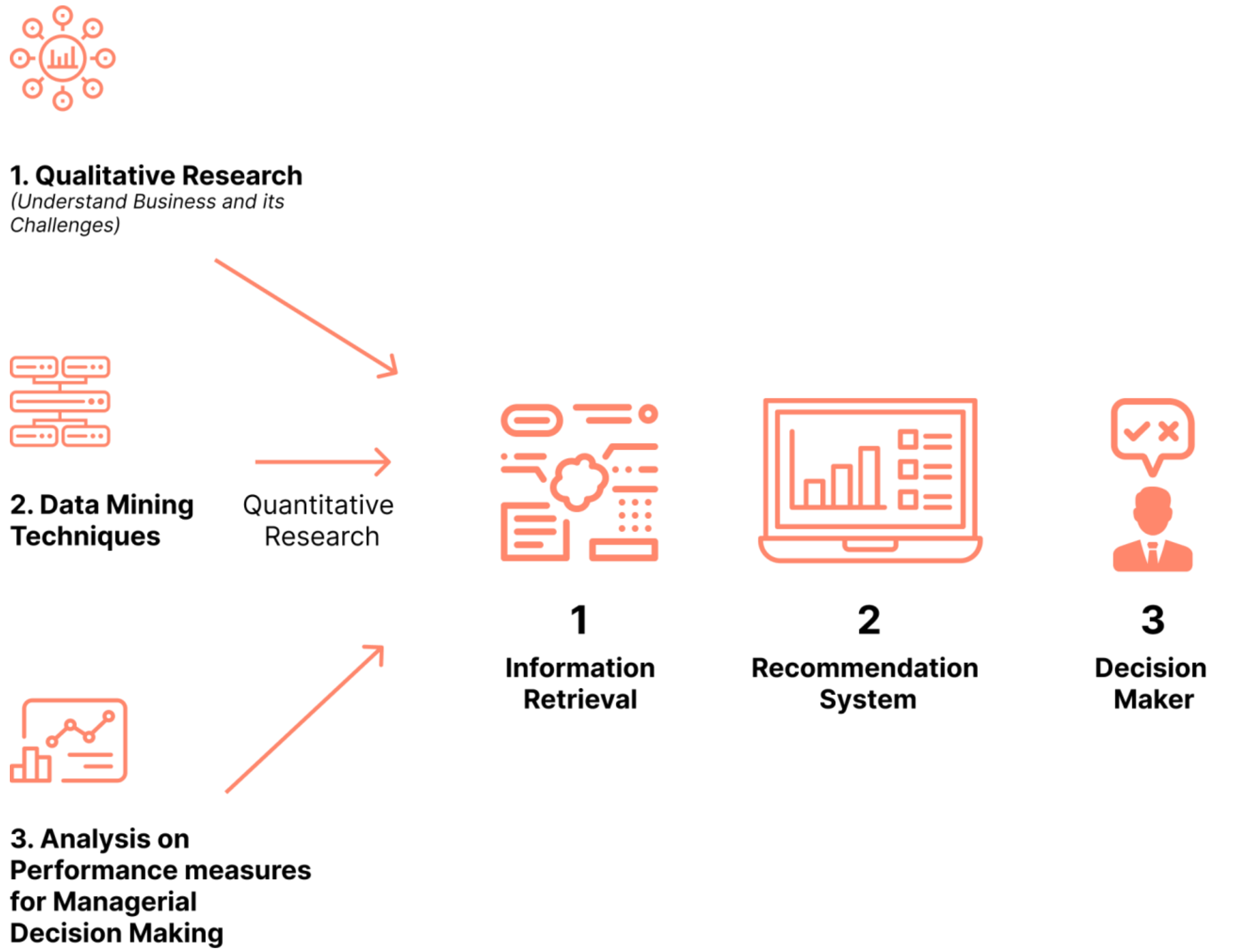

Figure 1: Overall Workflow

Figure 1 provides a graphical presentation of the methodology workflow. The first part comprises three types of research studies such as: 1) qualitative case study to understand and identify the weak areas of performance and challenges of the business, 2) analysis and practical comparative study of the application of data mining techniques, and 3) analysis on service quality attributes, a nonfinancial performance measure for managerial decision-making. The research covers proposal and implementation of recommendation system, as a portion of data mining system functionalities, which will be used to predict and digitize manual processes of data presentation for the business. As a result, the third part presents outcomes which will help the decision makers and management to formulate proper strategies to improve the business performance and enhance customer satisfaction.

1.2 Research Aims and Objectives

The goal of the study is to explore how implementing a data mining approach in businesses result to a better decision-making and greater effectiveness as the businesses continue their digital journey. The theoretical contribution and practical implications consist in comprehending the insights and the ideas of researchers on data mining along the business digital transformation. By practical involvement, our thesis analyses how implementing a data mining approach in a home appliance after-sales service business result in productivity improvements through prediction analysis. The results of this study guide businesses in improving their decision-making process, optimizing their strategic objectives, and adapting data mining approaches along the path of digital transformation.

The aim of this study is to analyze business expectations and implement an after-sales service data mining application – recommendation system. The aims of the research are as follows:

- To study different perceptions, insights, perspectives and areas of intelligent applications and advanced analytics implementation, along with several important aspects of data mining in driving digital transformation
- To identify important factors towards the expectation of the analytical applications, intelligent applications and the descriptive, predictive, and prescriptive models in the field of after-sales services
- To identify key attributes of service quality offering home appliances after-sales services businesses to impact customer satisfaction - nonfinancial performance measure
- To build and implement a recommendation system, as a portion of data mining system functionalities and to determine the reliability of the recommendation system by selecting the best trends and data mining technologies and tools
- To identify home appliances after-sales services business issues, challenges, and concerns that impact service quality and customer satisfaction
- To improve decision-making and optimize strategic goals as businesses continue their digital journey
- To reinforce linkages between businesses and customers
- To strengthen relationships among businesses.

1.3 Research Questions

Data mining for most businesses is not carried out. Several possible issues could be brought to the fore as follows:

- The increasing need of digital transformation of businesses and the necessity to implement digital technologies,
- Difficulties of the businesses and customers due to the lack of the data mining system, and
- Manual processes.

In order to address the aforementioned issues, this study raises the following key research questions:

- Research Question RQ1: Which techniques, trends and tools can influence a data mining system? Which business areas have used and benefited from data mining techniques aiming to improve decision-making along the business digital transformation?
- Research Question RQ2: What challenges are faced by the businesses? What are some of the challenges we find solvent or hard to overcome, and how do we come up with effective solutions?
- Research Question RQ3: How does a model support an after-sales service business that is considering using data mining to improve decision-making along the business digital transformation? How to analyze and design the platforms that can improve the analytics of the data?

These research questions were aligned with the research aim and objectives and are being applied in the development of a data mining system and in the facilitation of the digital transformation processes for the businesses.

1.4 Research Hypotheses

It has been used to test three hypotheses that were extracted from the theoretical framework and from practical implementation.

Three hypotheses were derived and tested:

Hypothesis H1: The data mining approach improves decision-making competency and decision-making processes in after-sales service business

Hypothesis H2: Service quality impacts customer satisfaction in after-sales service business

Hypothesis H3: Business strategic decision-making in after-sales service impacts after-sales service quality

If **hypothesis H1** is true, the implicit conclusion would be that decision makers of businesses need to consider the implementation of a data mining techniques as a possibility to enhance decision-making competency throughout the business digital transformation.

If **hypothesis H2** is true, the implicit conclusion would be that service quality impacts customer satisfaction in after-sales service business.

If **hypothesis H3** is true, the implicit conclusion is that decision-making competency impacts after-sales service quality.

1.5 Research Methodology

The performed research study includes the following steps: related work, implementation of case study approach, experiments and results, discussion of findings and conclusions. Figure 3 provides a graphical presentation of the methodology workflow.

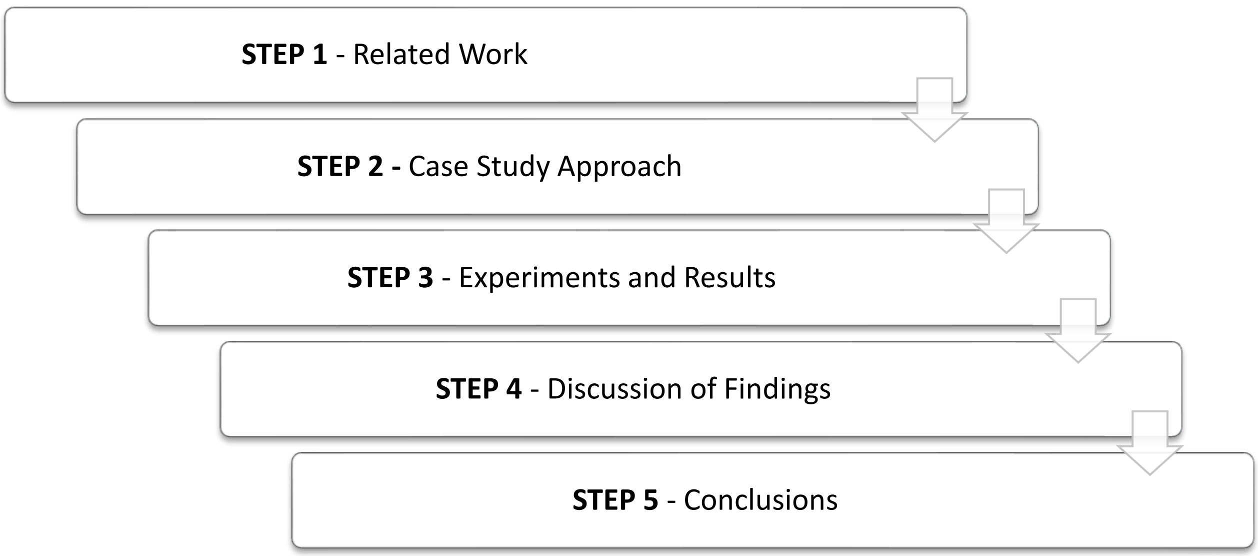

Figure 2: Research Methodology Workflow

The methodology of analysis alleviates the provision of conclusions in this context. The methodology followed by implementing the case study approach alleviates the process of answering research questions.

Step 1 provides a related work. The consulted papers in achieving the answer of Step 1 have been included. Moreover, the research question one *"RQ1: Which techniques, trends and tools can influence a data mining system? Which business areas have used and benefited from data mining techniques aiming to improve decision-making along the business digital transformation?"* has been answered by using the related work - approach.

Step 2 provides a case study approach. The case study approach covers two types of research, qualitative and quantitative study. The approach comprises analysis of how the application of the data mining approach to a business, results in better decision-making and productivity through predictive business analysis as the business continues its digital journey. Moreover, the information gathered by the qualitative research, enabled us to discover and seek new insights to facilitate the design of the strategy of the business and study the actual situation, mainly the challenges, of the home appliances after-sales service business in the Kosovo market.

Step 3 provides the research experiments and the achieved results. Through the *Step 3*, research question two, *"RQ2: What challenges are faced by the businesses? What are some of the challenges we find solvent or hard to overcome, and how do we come up with effective solutions?"* and research question three *"RQ3: How does a model support an after-sales service business that is considering using data mining to improve decision-making along the business digital transformation? How to analyze and design the platforms that can improve the analytics of the data?"* have been answered using the data mining approach.

Step 4 provides the discussion of findings.

Step 5 provides the research conclusion of the study.

1.6 Structure of the Thesis

The thesis contains seven chapters.

The first chapter gives a thorough overview to the research, a description of the problem statement, the research aims, the research questions, the research hypothesis, the research methodology and the importance of the study.

The second chapter has been divided into ten sections based on the topic of the related work as a) methodology, b) perspectives on the digital transformation, c) identification of digital transformation technologies, d) analysis of papers on the most advanced and big data analytics techniques, e) perspectives on the data mining technologies, algorithms and tools, f) perspectives on data mining application in different business areas and digital transformation, g) benefits of applying data mining, h) challenges and limitation of applying data mining, i) evaluation of data mining model outcomes, j) home appliances after-sales services and k) performance measurement of after-sales service business activity.

The third chapter presents research methodology. The research methodology has been divided into two research methodologies as qualitative and quantitative study.

The fourth chapter proceeds by carrying out the practical implication in the area of after-sales service, analysis, findings, experiments and results.

The fifth chapter provides discussions of findings.

The sixth chapter provides the research conclusion of the study.

Finally, the last chapter provides bibliography, references cited, in this thesis.

1.7 Importance of Thesis

The aim of this thesis of a data mining approach is to facilitate the process of comprehension, encouragement and motivation of businesses, adaptation of strategies, products and services, and individuals into a new digital era.

In addition, the primary objective of this research is to investigate the application of data mining algorithms in home appliances after-sales services business. This study gives concrete guidance on information analysis, repair and service quality prediction.

Furthermore, this thesis is to contribute to elaborating all parts of the application of the data mining approach and presenting the application as a proposal for other entities with similar concerns. With this case implementation, we will be in a position to provide a hands-on model for implementing the data mining techniques for analyzing information, predicting and monitoring processes in business. We will also be able to see how the data mining approach improves decision-making and increases efficiency and effectiveness through predictive business analytics.

The study of the relationship between service quality and customer satisfaction, a nonfinancial performance measure for managerial decision-making, makes understanding the impact that leads to an adequate application of decisions by decision makers of home appliances after-sales service businesses.

The recommendations as contributions of qualitative research will enable strategizing effective solutions for the after-sales services businesses in accordance with the issues and challenges indicated that quality of service is impacting customer satisfaction.

Finally, the data mining application – recommendation system, as a powerful mechanism, will provide an approach that uses digital technologies to digitize the process that improves the operation and performance of businesses. The recommendation system, as a portion of data mining system functionalities, as a solution, will perform automated tasks, such as those that are part of decision-making, will improve incomes, drive improvement, and change in business functions.

2 RELATED WORK

The researchers have provided different insights regarding data mining. In this section of the thesis, the relevant research papers have been dealt with. Following the introduction, the methodology of data mining research and studies is included, followed by perspectives on digital transformation, digital transformation technologies, on advanced analytics techniques, on data mining applications in business areas, benefits and challenges of applying data mining, evaluation of data mining model outcomes, home appliances after-sales services, performance measurement of after-sales service business activity.

2.1 Methodology

The related work mainly answers the research question RQ1:

- **Which techniques, trends and tools can influence a data mining system? Which business areas have used and benefited from data mining techniques aiming to improve decision-making along the business digital transformation?**

Figure 4 provides a graphical presentation of the related work workflow.

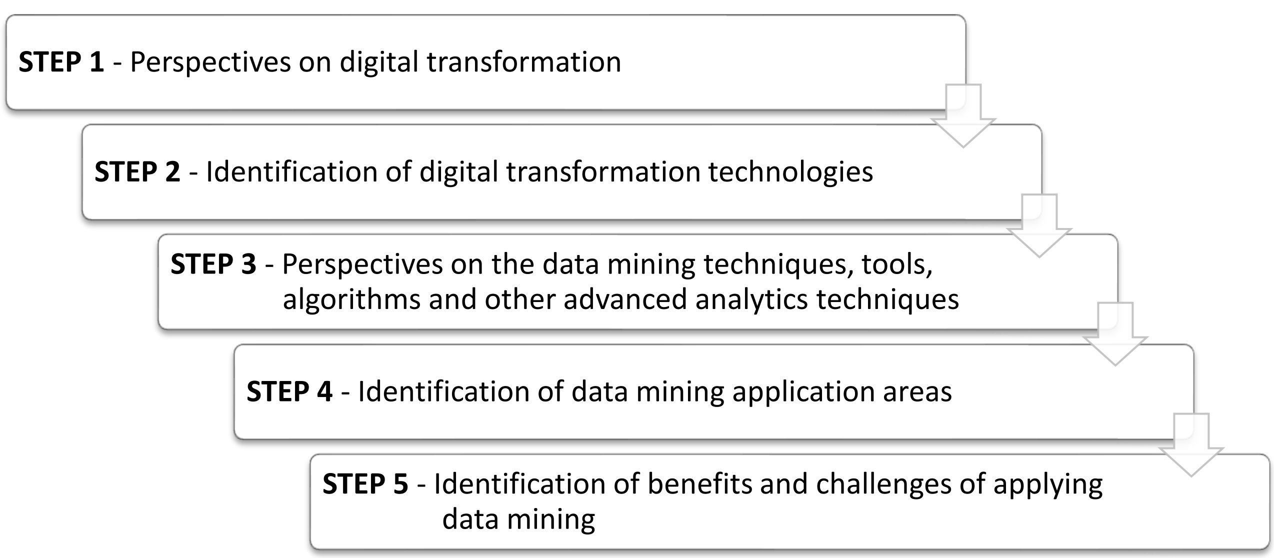

Figure 3: Related Work Workflow

The performed study includes the following steps: perspectives on the digital transformation, digital transformation technologies, data mining techniques, tools and algorithms and other advanced analytics techniques, data mining application areas and benefits and challenges of applying data mining.

Moreover, the related work includes study of digital transformation, focusing the importance of the dimensions of digital transformation assessing progress - digital maturity model and digital transformation driving framework. Research papers consulted on the identification of the dimensions of the digital transformation in the digital maturity model have been incorporated. Furthermore, related work includes an analysis on digital technologies in an endeavor to denote the importance of technologies of digital transformation of the businesses. Furthermore, related work includes a perspective on data mining technologies, algorithms, and tools with the aim of highlighting their significance for the business digital transformation.

2.2 Perspectives on Digital Transformation

2.2.1 Digital Transformation

Reis et al. (2018) describe the digital transformation as the practice of novel technologies, such as social media, mobile, analytics or embedded devices. Digital technologies bring important business enhancements, such as enhancing the customer experience, mitigating processes, or creating new models. Digital transformation is about more than simply digitizing resources and outcomes in terms of value and revenue generated by digital assets.

Digital transformation fosters the combination of technologies and capabilities into new business models (Reis et al., 2018). The digital maturity model, encompassing business dimensions, as a way to empower digital transformation by assessing the state of the business transformation path is an essential component (Deloitte, 2018). Agile has designed a framework that boosts business digital transformation.

Deloitte (2018) highlights that one of the components hindering the communications sector from making greater strides in digital transformation is the absence of an obvious and industry-driven roadmap. The digital maturity model is an effective approach to drive the transformational journey. Moreover, China

Mobile stresses that the digital maturity model will be very useful and, as such, reinforces decision-making.

Morakanyane et al. (2017) comprise two major drivers of digital transformation, digital capabilities and technologies. Digital technologies are fundamental to all efforts and provide chances for businesses. These transformative possibilities can transform business models and processes, and the client experiences. Finally, businesses can reap the benefits of digital transformation.

Orfanidis (2018) refers to the digital capabilities such as analytics, integrated business and IT, unified data and processes, and effective provision of solutions as a basis for building digital capabilities in the business. Morakanyane et al. (2017) refer to the digital technologies such as social media, mobile technologies, internet of things, cloud technologies, big data analytics, etc. as technologies that businesses embrace to enhance their day-to-day operations.

Digital transformation is regarded as a progressive process as it evolves over time, and that leads to a major transform in the business. The primary impact of digital transformation is to build values. Both the business and its customers achieve this value (Morakanyane et al., 2017; Dai et al., 2020).

2.2.1.1 Assessing progress of the business digital transformation

The digital maturity model enables decision-makers to present their perspective on digital strategic issues in order to assess and implement required changes in target sectors (Boström and Celik, 2017).

The results show that the digital maturity model is to evaluate digital capabilities mainly in five common business dimensions, such as customers, strategies, technologies, operations, organization, and culture. Table 1 presents the papers consulted for the identification of the most common digital transformation dimensions.

Table 1: Dimensions of digital transformation

Industry /Field of Study	Strategies	Customers	Technologies	Operations	Organizations and culture
Telecommunications (Valdez-de-Leon, 2016)	√	√	√	√	√
SME (Williams et al., 2019)	√		√		√
Business (Felch et al., 2019)	√	√	√	√	√
Corporate (Eremina et al., 2019)	√	√	√	√	√
Telecommunications (Deloitte, 2018)	√	√	√	√	√

The strategy, in the perspective of digital transformation, is set within the overall business strategy and emphasizes how the business is transforming or functioning to grow competitively with digital vision in the facility (Valdez-de-Leon et al., 2016). The customer experience focuses on the necessity to address customer needs and advantages as the basis for developing digital service offerings (Paschou et al., 2019). Technology contributes to the success of the digital strategy by supporting the creation, processing, storage, security and sharing of data to meet customer requests at low expense (Hie, 2019). The operations dimension emphasizes the capabilities that strengthen the delivery of services. Increased maturity in this dimension ascertains a more digitalized, automated, and flexible process (Srivastava et al., 2015). The organization and culture dimension defines a culture of governance processes and talented people to bolster growth through the digital maturity curve, and the flexibility required to meet the goals of increase and innovation (Paschou et al., 2019).

2.2.1.2 Driving framework in business' digital transformation

The modern agile approach emerged as a driving framework of the business digital transformation. Agile is a one-of-a-kind tool that drives the digital transformation as it streamlines the use of technology to manage business operations. Agile methodology is the key to revolutionizing the entire business (Balashova and Gromova, 2018; Betz et al., 2016). Gunasekaran et al. (2019) point out that agile methodology delivers stability and flexibility, continual progress, reduced risk, excellent communication and commitment, transparency and quality.

O'Regan (2017) presents agile as a popular software development method that runs opportunities to assess the direction of a project through the development lifecycle. Agile meets customers' needs better than traditional methodologies like the waterfall model, and their supporters believe it leads to better quality, higher productivity, and quicker time to market and better customer satisfaction. Agile has a robust collaboration style, and continuous shifts to demands are regarded as normal in the world of agility. It consists of controls to handle changes to the requests, and decent communication, and timely and consistent feedback are critical elements of the process.

2.2.2 The future of digital transformation and advanced analytics in business

Davenport et al. (2020) argue that multiple sections, such as business models, customer service choices and behaviors will be impacted and changed by artificial intelligence. For example, artificial intelligence will make it possible for e-commerce retailers to predict and determine customer preferences and send products to customers with no formal order, as customers have the opportunity of returning what is not needed; accordingly, artificial intelligence will advance business strategies and models, and customer behaviors and will be applied in fields like analytics and predictive behavior.

Makridakis (2017) states the influence of digital transformation of businesses will be substantial, leading to businesses that are well connected to decision-making based on big data analysis and increasing global competition between businesses. Generally, artificial intelligence technologies will affect the way businesses operate. Moreover, the digital transformation of businesses has a major effect on various dimensions of society, such as our lives and jobs, purchases, entertainment, and employment trends.

McCormick et al. (2016) assume that artificial intelligence technologies will be taken up in analytics practices, providing users outstanding accessibility to powerful and actionable information. Digital technologies, artificial intelligence, big data, and internet of things, will make data more accessible to businesses, widen the data types available for analysis, and progress the sophistication of the knowledge and information derived from it. An insight-driven enterprise leverages and implements data and analytics at every change to differentiate their products and services.

2.3 Digital Transformation Technologies

Ziyadin et al. (2019), mention digital technologies that alter business models, like social networks, mobile, big data, the internet of things, and other novelties such as blockchain. Wiesböck and Hess (2018) identify social, mobile, analytics, and cloud-based technologies as digital technologies. Hausberg et al. (2019) focus on the broad technological fields that enable digital transformation. These fields are cyber-physical systems, the internet of things, big data, artificial intelligence, cloud computing as well as augmented and virtual reality. Schwertner (2017) highlights that mature digital businesses are focusing on the integration of technologies, such as social, mobile, analytics, big data, and cloud, for business transformation. The usage of analytics enables decision-making, and the dynamics of decision-making must be responsive to changing demands and technologies (Teichert, 2019). Telegescu (2018) stresses that companies have opportunities on their way to digital transformation to take advantage of the emerging tools of the digital economy deliver. Lastly, the findings demonstrate that big data analytics is one of the drivers of business digital transformation. Table 2 presents the papers consulted to identify potential digital transformation technologies.

Table 2: Digital transformation accelerators

Authors	**Big Data Analytics**	**IoT**	**Cloud**	**Mobile**	**Social Networks**
Ziyadin *et al.* (2019)	√	√		√	√
Wiesböck and Hess (2018)	√		√	√	√
Hausberg *et al.* (2019)	√	√	√		
Schwertner (2017)	√	√	√	√	√
Telegescu (2018)	√	√	√	√	

2.3.1 Cloud Technologies

Neware and Khan (2018) define cloud computing as a template that provides adequate on-demand network access to a configurable shared computing pool which can be quickly delivered and distributed with a minimal managerial work to interact with service providers.

Mazumdar and Alharasheh (2019) highlight the main attributes of cloud computing, including on-demand self-service via a secure portal, scalability, and elasticity, pay-as-you-go, omnipresent access and resource pooling unrelated to location. The self-service attributes upon request are presented as a stand-alone service, with no interaction with service providers in the context of making the network, server, and storage capacities available. Scalability and elasticity attributes are presented as an elastic resource increase or reduction to sustain cost efficiency.

Neware and Khan (2018) introduce three main service models. IaaS (Infrastructure as a Service) is regarded as an on-demand service virtualization delivering a virtual machine, a storage infrastructure and a network. PaaS (Platform as a Service) provides an environment to create an application providing a virtual machine, an operating system, an application and a development structure. SaaS (Software as a Service) provides an on-demand cloud-based foundation for end-user software with a complete suite.

In a study, Quinn et al. (2014) demonstrate how cloud technology has helped improve decision-making relative to previous systems. The authors assert that cloud computing offers advantages for decision-making, lowers costs and makes it easier to administer systems.

Lastly, Benlian et al. (2018) highlight how, in social and economic terms, cloud computing is a catalyst for the digital transformation of industries. Cloud computing delivers the infrastructure that has fueled other major digital trends such as mobile computing, internet of things, big data, and artificial intelligence, thereby disrupting existing business models, and contributing to digital transformation.

2.3.1.1 Digital Transformation and Cloud Computing for Empowering Digital Business or Organization

Ozdogan et al. (2017), in their research, concentrate on digital agriculture, and cloud computing is identified as a key element in the digital agricultural revolution. The study used the case approach to assess the present situation of the implementation of digital agriculture in Turkey. Cloud computing has been identified as a vital infrastructure for smart agricultural applications like scalable computing, applications, data access and storage services. And with cloud computing, large-scale data storage can be done at low capital costs and immediate access to the information.

In addition, Alonso et al. (2016) provide a framework to cloud-based public utilities (and the supporting developments), based on a review of stakeholders' (citizens, public sector and the cloud sector) attitudes to cloud computing. The approach outlined provides the means to implement transformational government through the adoption and use of cloud-based technologies, supporting a modernized public administration based on ICTs and enabling digital transformation into the so-called transformational cloud-based government.

Moreover, Abdelaal et al. (2018), in their study, introduced the present situation of academic research on company digitalization strategies. It explains why businesses are digitally changing and demonstrates the internal motives of companies and their external drivers. The content of digital transformation is structured and differentiated in terms of the changing business dimensions and the sectors affected by the change. Lastly, several aspects related to the content of formulating the strategy for the digital age were identified, and feedback on the strategy progression was provided. In this regard, it has been pointed out that the intelligent factories and domestic endeavors such as Industrie 4.0, perhaps the world's biggest manufacturing trend, need cloud-based systems and optimization of networked machine production.

Furthermore, Serran W. (2018), in his study, described the notion of Digital as a Service (DaaS) that virtualizes intelligent infrastructure and cities at four levels: systems, transmission, server and management. Any comprehensive digitization can be performed regardless of the physical infrastructure involved within a cloud environment. DasS has the potential to enable virtual digital infrastructure (VDI) interoperability. Digital as a Service would facilitate this by fully interconnecting, integrating and virtualizing its space, services, and structure (3S).

On the other hand, Genzorová et al. (2019) outline how transportation businesses can modify their models as a result of digitization. The case study using an agile approach addresses the design process for recording employee time for the transportation sector, value, or supply chain of transportation services. For the benefit of developing what enables digital businesses, authors regard cloud-based technologies as an important asset and have the power to affect the customer experience. In addition, human resources have been identified as playing an active role in assisting employees deal with the impact of technology in the workplace. The recommendation continues that employees should be actively engaged in digital transformation, not just as a public.

Lastly, Hagberg et al. (2016), elaborate a digitization framework in the retailer-consumer interface which comprises four components: exchanges, actors, offerings, and settings. Based on the preceding literature, it outlines and illustrates the way in which digital transformation affects each of these components. Moreover, it has examined how exchanges are advanced by changes in communication, transactions and distribution; how players are advanced by blending humans with digital technologies, the enhanced blurring of borders, and new players, roles and relations; how parameters are changed to embrace and combine traditional and new parameters; and the way offers are advanced by changes in products and services, extensions of offers and alternative methods of pricing. As a result, the framework helped advance retail literature beyond digitization discussions.

2.3.2 Internet of Things

Pflaum and Gölzer (2018) assert that intelligent products, which are at the core of the internet of things, are behind the future digital transformation of businesses and alter their business model.

Zeinab and Elmustafa (2017) illustrate the internet of things as an innovative technology, which offers numerous applications for connecting things with things and people with things via the internet. The applications made possible by the internet of things are smart healthcare, houses, power, cities, and environments. The authors point out two major issues facing internet of things, and they include: co-existence with various networks and the size of big data on the internet of things. The mixture of multi-resource data and the internet of things enables the development of applications and services that can integrate situational and contextual awareness within decision-making processes and can generate more intelligent applications and advanced services.

Coetzee and Eksteen (2011) predict the future of the internet of things, the effect of the internet of things on society, and provide the challenges and point out that trust and privacy are probably to be the main barriers to the adoption of the internet of things. The major challenges are classified as: a) privacy, identity management, security, and access control; b) standardization and interoperability; and c) data flooding. Successful application include environmental monitoring, smart environments, retailers, logistics and supply chain management, and healthcare. The internet of things is presented as a potential for societal, environmental, economic impact and as support to digital transformation processes.

2.3.3 Big data analytics

Elgendy and Elragal (2014) designate big data as not only as large datasets, but also of a wide variety and velocity, and difficult to manage with conventional tools and techniques. Most importantly, decision makers can get useful information from this big data, ranging from daily measurements that can be provided through big data analytics. Big data analytics are recognized as an application of advanced analytics algorithms on big data. A number of advanced analytics were presented, including social media analytics, social network analysis, text mining, sentiment analysis and advanced data visualization. Big data analytics provide opportunities in fields, such as customer intelligence, fraud detection, and supply chain management. Furthermore, the advantages accrue to various areas and industries, including health care, retail trade, telecommunications, manufacturing, etc.

Memon et al. (2017) present big data analytics as an examination method on big data to discover concealed trends, elusive relations and other key data which can help to improve decisions. Moreover, Hadoop appears as a distributed, free processing environment which handles data processing and storing big data. Predictive analytics is highlighted as a later operation, using a variety of measurable, display, information-mining, and machine learning and verifiable information strategies, and enables experts to predict customer behavior and other developments in the future.

Dremel et al. (2017) argue that business digital transformation entails building big data analytics capabilities and deploying the establishment is a difficult process. Big data analytics enable evidence-based decisions. Businesses are taking advantage of the influence of big data analytics, and big data analytics support factual decision-making and empower new digital services.

2.3.3.1 An Overview of Worldwide Big Data Analytics Technology Use

The purpose of this overview is to elaborate on areas of big data analytics technology implementation worldwide.

Ridge, M. et al. (2015) evaluate the use of big data analytics in retail industries. The results showed that retailers do not use big data analytics because of the emphasis on the full exploitation of existing structured data prior to drawing on unstructured and semi-structured data. The authors argue that some retailers are using big data analytics to understand their customers, and consequently notify decision makers about pricing and marketing. Moreover, some retailers use big data analytics technology to accelerate the processing of large volumes of structured data and to provide information in a cost-efficient manner. Consequently, big data analytics claims its application has found in the retail industry to help analyze unstructured data by customer sentiment analysis, price optimization and inventory management. In this regard, the privacy of information and scalability of algorithms are barriers to implementing big data analytics. In this context, various analytic techniques and technologies that help analyze big data are studied. Furthermore, IBM BigInsights and SAP HANA are also tackled as products to assist with decision-making.

The author, Uyoyo (2014), in his paper outlines the unique characteristics that distinguish big data from conventional datasets. Besides, the author discusses the use of big data analytics in electronic commerce and in a variety of technologies. Despite the challenges, many companies are making progress in adopting big data in their e-commerce/data driven strategies. This study, in order to improve businesses competitive advantage, presents case studies on how top e-commerce traders such as Amazon.com, Walmart Inc, and Adidas are implementing big data analytics. The research focuses on the social media, predictive and mobile analytics. By implementing these analytics, we can obtain competitive advantage and business values, increase sales and profits, improve customer satisfaction, and most of all, raise brand awareness and strengthen reputation. And, the technology behind social media analytics includes text mining and sentiment analysis by using machine learning. Differently, the technologies behind predictive data analytics are statistical models and machine learning algorithms for pattern identification and learning from previous information. In addition, technologies behind mobile analytics that promote location-based marketing include the use of RFID tags, Bluetooth and GPS.

Moreover, authors, Avinash and Akarsha (2017), explore the way in which the usage of big data analytics is considered as a value factor that can drive e-commerce businesses to a competitive advantage, and drive precise and timely decision-making. The paper brings statistics on the achievements of the usage of big data analytics by e-commerce businesses, such as 73% sales growth for businesses that apply predictive analytics, 45% of internet buyers can purchase from a webpage providing customized referrals, etc. Additionally, the applications of big data are elaborated, such as: customization, dynamic pricing, customer behavior prediction, supply chain visibility, and fraud management. The resources of generating big data are listed as 1) social media and networking sites, 2) transactions, banking activities, 3) electronic gadgets and 4) sensors and network devices. In conclusion, by using big data analytics, we can say that the online behavior of consumers can be studied, and their interests can be predictable.

The authors, Srivastava et al. (2015), in their paper aim at understanding the way big data analytics become useful in the banking sector, respectively in these segments such as customer spending profile, usage of channels, client segment and profiling, cross-selling products, sentiment and feedback analysis, and fraud handling. The major sectors where the financial institutions mentioned use big data are customer centric, risk management and transaction. As a conclusion, the authors drew some outcomes from their study including: 1) numerous techniques used by financial institutions gather customer data for sentiment analysis, via social media sites to a variety of market research channels, 2) ways of identification of potential customers for selling financial products, 3) the way of implementation banks apply to strengthen data security and prevent any type of attack.

Moreover, authors, Sultan and Bechter (2019), analyze possible applications and implications of big data analytics for Islamic banks and their regulatory institutions. Risk management is one area where big data analytics finds application. Big data analytics identify sanctioned, blacklisted individuals in real time and at a minimal cost. In addition, through analytics, customer behavior can be analyzed and predicted. Another area of application is the systemic risk assessment. The findings elaborate the impact of benefits of big data analytics application on quicker and more effective decision-making, and market segmentation, risk management and design of new products. Finally, big data analytics helps the banking sector to provide an overview of risks.

The authors, O'Donovan et al. (2015), in their study, presented a review of research on big data in manufacturing to stimulate a better understanding of a new and pervasive area. Manufacturing is considered as the center of data revolution and pretends to turn traditional manufacturing services into very intelligent services. With intelligent services, real-time data-driven manufacturing provides precise and on-time decision-making and positively impacts the organization as a whole. Descriptive, predictive and prescriptive are the key big data analytics underlined. The study, besides providing answers to research questions, provided an outstanding platform for advanced research and investigation in the area.

2.4 Advanced and Big Data Analytics Techniques Empowering Business Decision-Making

Elgendy and Elragal (2016) introduce how big data analytics techniques can be applied, in the regards of digital transformation, to leverage business change and to enhance decision-making by revealing hidden information and valuable knowledge. The primary objective of the business decision makers is to improve decision-making and insights by implementing big data analytics techniques.

Rehman et al. (2019), in their study, provide several advanced analytics techniques, including text analytics, machine learning, data mining and statistical and natural language processing techniques. Prabhu et al. (2019) describe advanced analytics techniques, including machine learning, text mining and data mining. Vivekananth and Baptist (2015) include seven big analytics techniques, such as association rule learning, classification tree analysis, genetic algorithms, machine learning, regression analysis, sentimental analysis, and social network analysis. Sadiku et al. (2018) use big data analytics techniques in analyzing big data and in their paper include techniques such as, a) data mining, b) web mining, c) machine learning, d) social network analysis, and e) visualization methodologies. Galetsi et al. (2020) states that the analysis of their study depends on big analytics, such as modelling, simulation, machine learning, visualization, data mining and others. Finally, the results show that 1) data mining, 2) machine learning, and 3) natural language processing are the key techniques of advanced analytics and big data for business digital transformation. Table 3 presents the papers consulted for the identification of potential advanced analytics and big data techniques.

Table 3: Analysis on advanced analytics techniques

Authors	**Machine Learning**	**Text Mining (NLP)**	**Data Mining**	**Social network analysis**	**Visual Analytics**	**Web Mining**	**Statistics**
Vivekananth and Baptist (2015)	√	√	√	√			
Sadiku *et al.* (2018)	√	√	√		√	√	
Galetsi *et al.* (2020)	√	√	√	√	√	√	√
Rehman *et al.* (2019)	√	√	√				√
Prabhu *et al.* (2019)	√	√	√				

2.4.1 Data mining

Reddy (2011), the author argues that data mining is a process of knowledge discovery through the analysis of vast quantities of data from a variety of perspectives and transforming it into useful information concerning different spheres of human life including business, education, medicine, science, etc.

Data mining enables businesses to predict future trends. Data mining is emphasized as a blend of algorithms and notions from machine learning, artificial intelligence, data management and statistics (Harding et al., 2006). Data mining is the technique applied to sort massive datasets for patterns and build relations to resolve issues by analyzing data, and is applied to assist decision-making (Padmavaty et al., 2020).

Data mining is known as the intermediate layer of computer science and statistics. From the data mining process, the valuable insights obtained can be used to raise profits and computational complexities. A variety of techniques exist in the data mining process, including artificial intelligence, machine learning and statistics (Kumar et al., 2014).

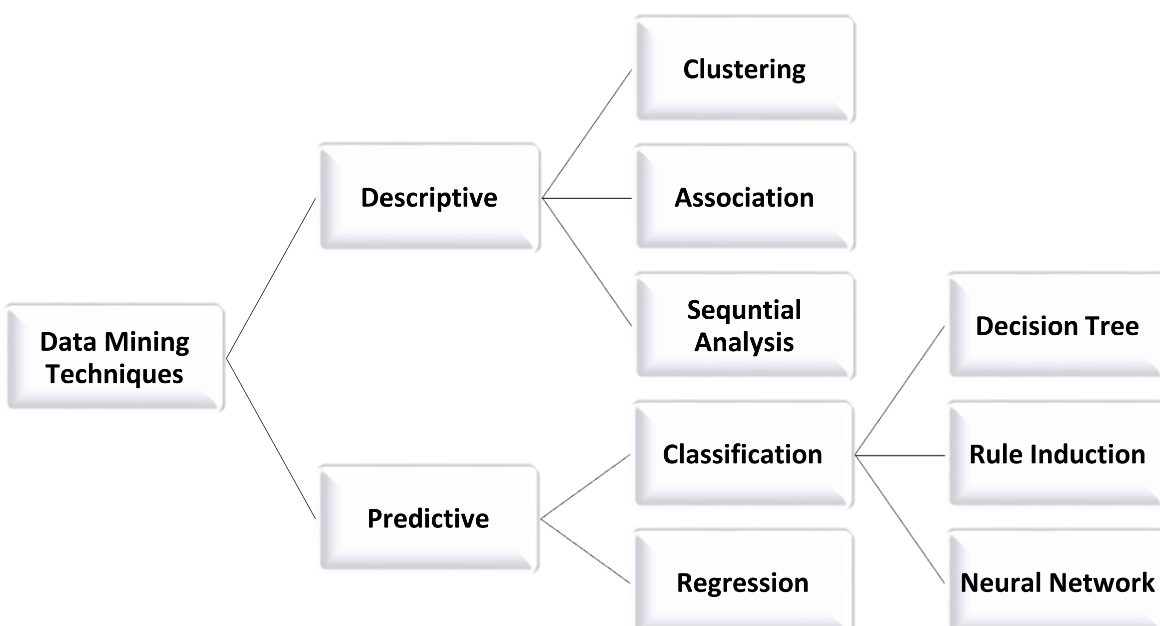

Figure 4: Data Mining Techniques (Kaur and Dhiman, 2016)

Figure 5 depicts the techniques of data mining. The purpose of the figure is to demonstrate the many data mining techniques depicting the type of mining and data recovery operations, including prediction, classification, clustering, sequential patterns, association, and decision trees. In association, the relationship of a specific item in a data transaction on other items in the same transaction is used to predict patterns (Kaur and Dhiman, 2016). In classification, the methods are considered for learning different functions that represent each element of the chosen data in one of the predefined group of classes (Kaur and Dhiman, 2016). Prediction is associated with regression techniques (Kaur and Dhiman, 2016). The request for sequential pattern analysis is to uncover alike patterns in data transaction over a commercial period (Olson and Delen, 2008). In addition, clustering is considered to be the process of grouping data into similar properties (Sajana et al., 2016).

Siguenza-Guzman et al. (2015) classify the functions, or models, of data mining depending on the task carried out: clustering, association, classification, and regression. Statistics, artificial intelligence, and machine learning are enumerated data mining algorithms. Machine learning is a blend of advanced statistical techniques and artificial intelligence heuristics applied for data analytics and knowledge discovery. In contrast to other techniques, data mining advantages also embrace extracting patterns, describing trends, and behavior prediction. The authors depict the mining process as an interactive sequence of stages. The first stage consists in integrating data from various data sources and formats. In the second step, the cleansing process is carried out in order to eliminate noise, duplication and inconsistencies in the data. Additionally, data transformation in an appropriate format is applied, filtering and aggregation techniques are implemented to retrieve the summary data. In the third step, knowledge is taken from the transformed data and the information is analyzed to identify the really interesting patterns. Lastly, knowledge is also displayed to the user.

Sharma et al. (2012) elaborate Knowledge Discovery Process (KDD) as key to the application of data mining projects. The KDD process includes several phases, beginning of an understanding of the business, then understand and prepare the data, modeling, evaluation, and deployment. Business understanding is described as the initial stage focused on comprehending the goals and requests from a business viewpoint, then convert that into a data mining problem definition and a draft plan to meet the targets. Data understanding is described as a phase that begins with a first data collection process and carries out data familiarization events, to notice quality difficulties, to uncover first glimpses of the data or to create assumptions about concealed information. In addition, preparing the data is described as a stage that encompasses every activity to build the final data set based on the primary raw data. During the modelling (data mining) phase, different techniques (e.g., decision tree, regression, clustering) are chosen and implemented, and the elements are calibrated. The evaluation phase consists of systematically evaluating the model and reviewing the stages completed to build the model to ensure that it appropriately attains the operational objectives. Lastly, the deployment model is described as a creation of the model and is not usually the end of the action. Although the aim of the model is to improve the knowledge of the data, the knowledge acquired should be organized and presented for use by customers.

2.4.1.1 Data mining tools as modern solutions

Gergin et al. (2019) note that data science and tools are the modern state-of-the-art solutions that the businesses are applying in this digital transformation era. Businesses increase the profitability of quality control processes by applying data mining techniques. Data is handled using data mining tools to understand and inspect patterns and relationships. Table 4 presents the results on the identification of potential data mining tools and their advantages.

Table 4: Data Mining Tools

Data Mining Tool: Type	Advantages
Weka (Java): Machine Learning	Easy user interface, handles many data mining tasks
Rapid Miner (Java): Statistical Analysis; Data mining; Predictive Analysis	Visualization (user-friendly), highly flexible, provides procedures (selecting attributes and identifying outliers)
R (C, Fortran, R): Statistical Computing	Strong selection of data mining tasks, rapid application of many machine learning algorithms, improved graphics, has particular data types
Orange (C++, Python, C): Machine Learning; Data mining; Data visualization	Best debugger, shortest scripts, poor statistics

Jovic et al. (2014) outline and compare seven data mining applications, such as RapidMiner, R, Weka, Orange, KNIME, and scikit-learn. The authors emphasize that RapidMiner, R, Weka, and KNIME have numerous features required to provide a comprehensive operational data mining platform and as such their use can be recommended for almost all data mining tasks. Moreover, Dušanka et al. (2017) outlined as important and compared data mining technologies and technologies with free algorithm implementations, such as Weka, RapidMiner, H2O and Apache Spark.

Weka, a release based on Java, is a highly advanced tool and is applied to various applications comprising visualization and techniques for data analytics and prediction modelling. Weka handles numerous data mining tasks, including data pre-processing, clustering, classification, regression, visualization, and feature selection. Weka contains various machine learning techniques for data mining tasks (Kaur and Dhiman, 2016). Weka is applied to make predictions in real time in very challenging real-world applications (Janošcová, 2016).

Janošcová (2016) asserts that, in the case of dealing with large datasets, command-line interface is proposed to be used or to write code directly in Java, Groovy or Jython. Weka has few graphical user interfaces that provide easy accessibility to the respective functions. The key graphical user interface, the "Explorer", provides an interface based on panel, where the different panels address various data mining tasks. Weka is applied to make predictions in real time in very challenging real-world applications.

Dušanka et al. (2017) and Jovic et al. (2014) introduce RapidMiner as a data mining tool that provides an integrated environment with an attractive and easy-to-use visual interface. All in RapidMiner is concentrated on processes. The processes involve operators, as visual components. RapidMiner also provides the application helpers option to build the process automatically. While RapidMiner is very powerful with its base operator package, extensions are what makes it even more advantageous. Popular add-ons contain operator packages for text and web mining, time series analysis, etc. Moreover, the provision of deep learning methodologies and certain of the most sophisticated specific machine learning algorithms are limited at present. Nevertheless, big data analysis via the Hadoop cluster (Radoop) is handled.

Jovic et al. (2014) introduce the R data mining tool as an open-source tool and as a great option for data mining tasks. The source code for R is drawn up in C++, Fortran, and even R. The R tool is a language of interpretation and mainly optimized for matricial computations. The data mining tool turns on only an easy graphical user interface with a shell for input, and that is not a user friendly. Moreover, from the point of view of the data mining user, R enables the implementation of numerous machine learning algorithms very rapidly. The R data mining tool has specific data types for handling big data, supports parallelization, data streams, web mining, graph mining, spatial mining, and numerous other advanced tasks, comprising a limited support for deep learning methods.

2.4.2 Machine Learning

Machine learning, as a subsection of artificial intelligence, is scientific research on algorithms and statistics patterns used by computing machines to complete tasks with no explicit directives, for instance the use of pattern recognition and inference (OECD, 2019). The application of data-intensive machine learning methods can be noted across science, technology, and business, contributing to stronger evidence-based decision-making used across multiple disciplines (Jordan and Mitchell, 2015).

Simon et al. (2016) present supervised and unsupervised machine learning. Supervised machine learning is explained as a program 'trained' using a predetermined set of 'training examples', which then eases its capacity to achieve a precise outcome when it supplies new information. Unsupervised machine learning is explained as a specific program in a stack of data and has to find models and relations. Hence, the authors consider deep learning and big data to be two priority domains of data science. Deep learning algorithms retrieve composite data patterns through a hierarchical learning process of analysis and learning of unsupervised data (big data) pools.

Baştanlar and Ozuysal (2014), in their paper term, machine learning as a computer capable of making effective predictions based on previous experiences has recently experienced an impressive evolution recently due to the rapid growth in the storage capacity and computing power. Furthermore, the authors present the classification of machine learning tasks according to the intended performance of a machine-learned system. Unsupervised learning techniques include only the input feature values in the training data and the learning algorithm includes hidden structure in the training data based on them. Clustering techniques that endeavor to divide data into coherent groups fall into this grouping. Supervised learning methods require that the output variable for each training sample to be known. Subsequently, each training sample is a pair of input and output values. The algorithm then trains a model that predicts the value of the output variables from the input variables utilizing the defined features in the process. If the output variables are continuously evaluated, then the prediction model is referred to as a “regression function”. If the output variables are discrete values, then the predictive model is known as a “classifier”.

2.4.3 Natural language processing

Natural language processing is stated as a domain of artificial intelligence, computer science and linguistics concerned about the way computers relate with human languages (Fabian and Alexandru-Nicolae, 2009).

Natural language processing is essential because is helping us build models and processes the pieces of information that enter the voice or text or both and handles them according to the technique in the computer. In this way, the output of a natural language processing system processes speech and written text (Jain et al., 2018). Natural language processing appliances compose numerous areas of studies, like natural language text processing and summarization, multilingual and cross language information retrieval, speech recognition, artificial intelligence, etc. (Baştanlar and Ozuysal, 2014).

Friedman et al. (2013) outline the way in which natural language processing systems are being developed to facilitate decision-making, in addition to supporting the retrieval of information and relationships. Natural language processing implementations empower decision-making.

2.4.3.1 Chatbot - A Software to Revolutionize the Business

Smutny and Schreiberova (2020) consider chatbots as a software tool that works with users on a given subject or within a particular area in a natural and conversational manner using text and voice. In addition, Bhosale et al. (2020) present the notion of chatbots as a system that tries to replicate typed discussion, intended to, at least temporarily, “cheating” the human being into believing that he/she was communicating with another person. Accordingly, chatbots are referred to as conversational agents who work with operators on a particular subject.

In addition, Lacerda and Aguiar (2019) elaborate chatbots as conversational software agents that need natural language processing to understand the intentions of users and have become continually popular. Chatbots therefore converse with people, integrating services, users, and communication channels. Pal and Singh (2019) describe chatbots as digital tools, meaning hardware or software practices machine learning and artificial intelligence approaches emulate human behaviors and deliver a task-oriented framework and a progressing dialogue to engage in the conversation.

According to Kaczorowska-Spychalska (2019), chatbots are among the symptoms of digital transformation. Chatbots can empower a nicer and gentler manner of interacting, as a result, making it easier for humans to admit machines as cognitive cohorts. As a result, the role of chatbots is to better understand customer requests and behaviors and make the customer experience a significant step towards a positive digital transformation.

Lastly, chatbots play an important role in a business's digital journey and in achieving next-generation intelligent customer service (Deloitte, 2018).

2.4.3.2 Benefits of Chatbots in Driving Digital transformation

Palanica et al. (2019) identify profitability as a potential advantage for chatbots in the conduct of digital transformation. The authors explain that chatbots are profitable and can run 24 hours a day and can communicate in a wide variety of languages to better meet the needs of customers. Additionally, chatbots can automate repeated administration duties. Ukpabi et al. (2019), claim chatbots lower costs for customers and businesses. Customers lower their communication expenses by not having to call, and businesses will not need to hire customer service representatives or contract out call center response services.

Ivanov and Webster (2017) consider labor cost savings as the most obvious financial gain from adopting chatbots. The authors prove that chatbots work 24/7, which is much more than the typical 40-hour workweek for human labor. Moreover, chatbots can serve more than one customer at a time, which is not standard for human workers. Consequently, if chatbots could be applied, productivity and financial benefits would be evident. In addition, chatbots have a positive impact on sales. Businesses will not massively substitute their human workers with chatbots but use chatbots to boost productivity growth and increase customer service with equal or slightly less workers. Furthermore, chatbots can be applied to enhance communication, reduce service and support costs, and reduce time spent in routine processes (Mohan, 2019; Zumstein and Hundertmark, 2017).

Richad et al. (2019) argue that the customer experience is a basic requirement for chatbots when driving a digital transformation. Through chatbots, businesses can offer 24-hour customer services in one week; chatbots are accessible everywhere and offer efficient customer service operations. Additionally, chatbots provide a quick response to customer queries, improving the delivery of a great customer experience. Nuruzzaman and Hussain (2018) explain how the combination of a new wave of thinking with newly developed artificial intelligence technology, as chatbot technology, has the potential to completely transform customer experience to offer excellent service in a way that resonates with modern customers.

Kaczorowska-Spychalska (2019) highlights that attaining a better understanding of the client needs and behaviors is a very important step towards a successful digital transformation that sustains and changes the customer experience. In addition, chatbots let businesses or brands to enhance the customer experience while enhancing their engagement and satisfaction. As a result, chatbots spare as much as 30% of client support expenses and assist companies' lower client service expenses by accelerating answer times and meeting 80% of common requests.

2.5 Data Mining Technologies, Algorithms and Tools

2.5.1 Clustering Algorithms

Clustering is introduced as the process of grouping of similar objects together. The object set is referred to as a cluster and has similar objects in relation to the objects of the other cluster. A variety of clustering techniques can be applied depending on how the data behaves (Kumar, 2013). Clustering techniques are regarded to generate high-quality of clusters by grouping unlabeled data together (Sajana et al., 2016). Clustering is known as unsupervised classification and (unsupervised) learning (Saxena et al., 2017).

2.5.1.1 K-means algorithm

K-means clustering is presented as the method for classification or grouping elements to k groups (where k represents the number of preselected groups). Clustering is achieved by reducing the sum of the (Euclidean) distances per square between the elements and the respective centroids. In terms of advantages, K-means can be computationally faster and scalable, leading to clusters that are narrower than hierarchical clusters. One of the drawbacks of the K-means is the challenge in comparison to the quality of the clusters generated. The following drawbacks of the K-means is the challenge of K prediction and operation with non-globular clusters. K-means runs in well-formed clusters and is responsive to outlier values (noise). The value of k must be pre-determined (Zafar and Ilyas, 2015).

2.5.1.2 Farthest-first algorithm

The farthest-first algorithm is presented in detail. The farthest-first traversal k-center is a fast algorithm. Here, k points are initially chosen as cluster centers. The first center is chosen randomly and the second center is avidly chosen as the most distant point from the first. Each sustainable center was designated by the avid selection of the most distant point of all the centers already selected, and the remaining points, are attached to the nearest cluster whose center (Kumar, 2013). In terms of the benefits presented, the method using the farthest-point heuristic is quick and suitable for large-scale data mining implementations. The farthest-first has the similar drawbacks as the K-means clustering (Zafar and Ilyas, 2015).

2.5.1.3 Canopy clustering algorithm

Canopy clustering is an unsupervised pre-clustering technique, and it is widely applied as a pre-processing venture; it is intended to accelerate clustering assignments on large datasets (Ravichandran, 2017). In terms of advantages, the canopy clustering does not decrease the precision of grouping but increases the efficiency of the computation. As a result, canopy clustering accelerates the clustering actions on large datasets, and is a very easy, rapid, and precise technique. In addition, canopy clustering can be utilized in the design of MapReduce applying the Hadoop cluster and to accelerate prototype-based clustering techniques like K- means and expectation-maximization (EM) (Kumar et al., 2014).

2.5.1.4 Density-based group of clustering algorithm

The group based on the density of the clustering algorithm is depicted as a dataset in the same manner as partitioning methods and prevails over other methods with noise management capabilities and identifying clusters with arbitrary shapes (Breitkreutz and Casey, 2008). In terms of advantages, the group based on the density of the clustering technique does not require a priori knowledge of the number of clusters in the data; it can find clusters of random shape; it can even find clusters fully surrounded by a different cluster and has a sense of noise (Zafar and Ilyas, 2015).

2.5.1.5 Filtered clusterer

The Filter is depicted as a subset or a partly ordered set and that adds a new nominal attribute that indicates the clusters chosen for each instance. The algorithm is constructed in two manners; it is either constructed with the first dataset or has a serial cluster model to use (Shrivastava and Arya, 2012; Gnanapriya et al, 2017).

2.5.2 Classification Algorithms

Ragab et al. (2014) describe the classification as one of the more common data mining techniques which is primarily used in a separate dataset for each proceeding and assigns that proceeding to a separate class. Classification is implemented to obtain models that correctly describe important data classes. To begin with, the pattern is constructed by performing a classification algorithm within the training dataset. At the end-stage, the acquired model is tested versus a predefined test data set to assess performance. Accordingly, the classification technique identifies the class label for the dataset for which the class label is not recognized.

Classification techniques, as elaborated in Figure 6, consist of decision trees, k-nearest neighbors (KNN), Naïve Bayes, support vector machines, artificial neural networks (ANN), and more.

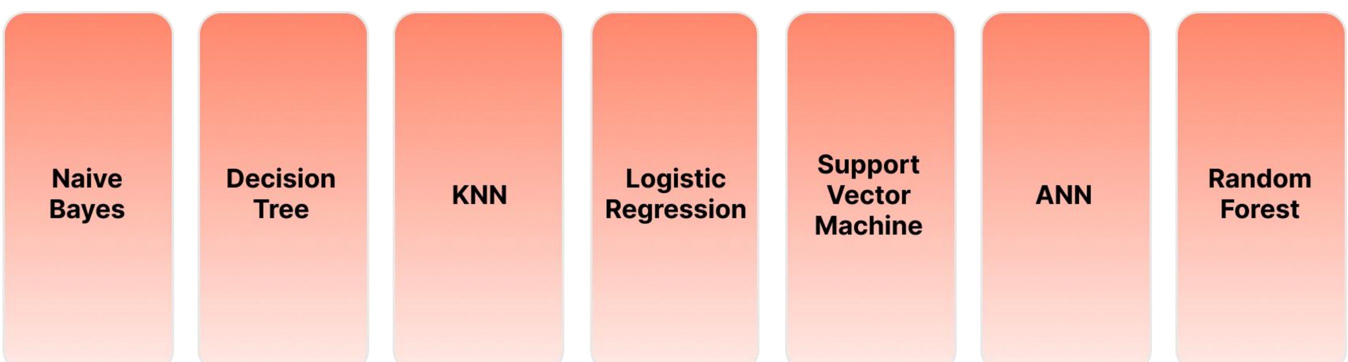

Figure 5: Classification Algorithms used with Data Mining Technology

Classification techniques involve a broad scope of implementations in various perspectives like social network analysis, credit card rating, artificial intelligence, and so on. (Jadhav and Channe, 2016).

2.5.2.1 Decision Tree Algorithm (J48)

Kaur and Chhabra (2014) describe the decision tree algorithm as an opportunity to uncover the function of the attribute vector for some instances. Newly created instances are identified in accordance with class training instances. The algorithm generates the rules for predicting the target variable's value class, and makes the distribution of critical data easy to understand. The J48 results in the rules based on which the separate identity of the data is generated. Discrete and continuous attributes are handled through J48.

Jadhav and Channe (2016) investigate the pros and cons of the algorithm. The authors cite the benefits of the technique as being mere and rapid, providing great accuracy, the depiction can be easily understood, fostering progressive learning and less memory, and managing noisy data as well. The detected drawbacks of the technique involve a lengthy training period, a much more difficult representation for certain conceptions because of a replication issue and an over the fitting issue.

2.5.2.2 Random Forest Algorithm

Avand et al. (2019) state that the technique is a common, tree-based approach that involves regression and classification trees. Furthermore, random forest is a non-parametric technique for the construction of decision tree techniques data. Changes in the results of each tree are the main problem encountered with this method. A random forest algorithm is proposed to reduce the estimate of variation and variance. Moreover, it is a blend of a multiple decision trees that incorporate more than one bootstrap sample of the data, and multiple input parameters participating at random in the tree model.

Furthermore, Ali et al. (2012) present the benefits of the random forest technique, for example addresses the issue of overfitting, have less sensitivity to outliers, can readily establish the parameters and removes the necessity to prune the trees, and stands for a variable significance and instantly produces accuracy.

2.5.2.3 Naïve Bayes Algorithm

Saputra et al. (2018) regard the Naïve Bayes algorithm as a common probabilistic classification method, even though this technique argues all attributes are independent of each other. Therefore, Naïve Bayes works to predict the probability of a particular data case relating to a specific class. The Naïve Bayes algorithm can work fine in reality. Furthermore, the Naïve Bayes algorithm requires little training data to attribute for estimating the parameters of the classification practice.

Jadhav and Channe (2016) examine the pros and cons of the algorithm. The study points to the benefits of the technique, for example, require little time to calculate the training, improves the performance of Naïve Bayes by removing non-significant characteristics, and provides highly accurate and valuable outcomes. The noted drawbacks of the algorithm are that it implies a huge quantity of data (tuples) to get high performance and less accuracy when compared with other classifiers on certain datasets.

2.5.2.4 Logistic Regression Algorithm

Zhu et al. (2019) argue that the logistic regression algorithm is useful where the objective is to categorize data elements. When using logistics regression, the target variable is two-classes, meaning that the target variable may be categorized as 1 or 0. The implementation of the logistic regression model has showcased the dominance in several fields.

Ray (2019) introduces logistic regression pros and cons. The pros of the technique include easy application, computational efficiency, and ease of regulation. At the same time, the drawbacks of the technique include the inability to comprehend a nonlinear issue because the decisional area is linear, prone to overfitting, does not perform correctly if all independent parameters are not uncovered. Logistic Regression can be applied in various areas, like a prediction of the threat of growing a particular disease, prediction of the probability of the process or system defects (in engineering), etc.

2.5.2.5 K-Nearest Neighbor (KNN) Algorithm

Furthermore, Avand et al. (2019) claim that the K-Nearest Neighbor (KNN) is a classification algorithm of an unrecognized set if we have data with specific properties (X) and the value of the relation (Y). A nonparametric learning method is a KNN classifier. Within the scope of classification, the technique assesses the distance between the target point and nearby points based on the value provided for K and obtain compliance with the maximum number of votes from these close points.

Jadhav and Channe (2016) draw up the pros and cons of the algorithm. The authors point out the pros of the technique, indicating that applying and understanding is easy, training is very quick, it is consistent with noise training data, and provides great performance and accuracy. The noted drawbacks of the technique are lazy learners, high computational costs, local data structure sensitivity, limited memory and lag of execution.

2.6 An Overview of Data Mining Applications in Different Business Areas

The primary focus of the overview of data mining applications is to comprehend the views of researchers on the application of data mining in various fields and digital transformation. Data mining is applied to various types of businesses and impacts the digital transformation of various businesses (Reddy, 2011). The aim of studying of the implementation of data mining is to facilitate the understanding of businesses as they adapt their strategies to a new digital age. As a result, several studies are selected to assist in understanding and explaining data mining in the retail trade, e-commerce, banking services and manufacturing sectors.

2.6.1 Data mining in retail

Chen et al. (2012) provide a study aimed at helping businesses to understand more about their customers and, consequently, to conduct more successful customer-oriented marketing. A customer-focused business intelligence study on an online retailer is introduced. The research contributes to the process of the consumer's understanding in terms of profit, and the implementation of appropriate marketing and decision-making strategies, speeding up digital transformation efforts. It is noted that numerous eminent online retail brands embrace data mining technologies as an essential tool for competition in the marketplace. Association analyses supported the establishment of consumer purchasing habits related to products that have been purchased together.

Castelo-Branco et al. (2020) develop an approach to enable a plan to use the data mining tools and technologies at the retail level in an appropriate manner, and introduce conceptions such as market basket, association rules and cross-sell and up-sell. Market basket analysis in marketing aims to provide the retailer with evidence to comprehend the purchasing behavior of the buyer that can support the retailer improve decision-making. The authors highlight that retail data mining can contribute to find out customer purchasing behaviors and patterns and trends, identify ways in which customer service quality can be advanced, increase customer loyalty and satisfaction, improve consumer ratios of goods, develop more impactful goods transportation and distribution policies, and lower business costs. The Cross

Industry Standard Process for data mining (CRISP-DM) model is presented. In addition, examples of successful future of retail and data mining applications are highlighted, such as: 1) Hewlett-Packard's analytics team conducted a manual pilot project in the small and medium-sized business online shop and call center. 2) Sweden's indoor gigantic designer IKEA has focused on image recognition as well as augmented reality, leading to greater customer satisfaction and reduced returns. 3) Macy, the world-class department store, uses big data for a smarter customer experience. 4) Amazon Go is the retail tech that is supposed to drive the approach to the future of artificial intelligence in retail. The key concept behind Amazon Go is the one store that prospers on the concept of no pay claim. 5) The Starbucks CTO suggested combining transactional information with other data, such as weather conditions, promotions, etc., for a responsive and customized service.

Kaur and Jagdev (2017) examine how changes in the retail sector are influenced by big data. It is noted that retail is totally dependent upon big data analytics. The mining of customer analytics aims to drive profits, growth, and competitiveness, both in-store or online. With digital technologies, as the key driver of digital transformation, retailers are making smart decisions based on online data. Then, there is a demonstration of how machine learning uses big data to collect, use and predict a meaningful perception of customer spending habits and how these habits change. Big data is smart for having a direct impact on sales by gathering data on accurate consumer consumption patterns. For instance, Amazon is fast reacting to the competition through an analytics system that delivers dynamic pricing and, in comparison with other retail dealers, carries out this modification approximately every three months. Moreover, Metro Group applies retail analytics to ascertain how goods are moving through stores in order to provide real-time information to the relevant staff and customers in order to use it easily. In addition, Staples uses Hadoop and big data techniques for sales prediction through the handling of approximately ten million data transfers a week at 1,100 locations across the U.S.

2.6.2 Data mining in E-Commerce

Ismail et al. (2015) introduce data mining for e-commerce with the following techniques: association, clustering and prediction. The addressed advantages of data mining in electronic commerce are linked to market segmentation, sales forecasting, basket analysis, merchandise planning and customer relationship management. The tackled e-commerce data mining issues relate to spider identification, data transformation and make the data model commercially user-friendly, support enabling data transformation and model constructing. The notion of data mining in electronic commerce is incorporated as an incorporation of statistics, databases, and artificial intelligence with certain topics with the aim of enhancing decision-making. Cloud computing, an important technology in the digital transformation era, in electronic commerce is seen as an efficient way to use resources and reduce costs to business that embrace efficient data mining. Association rules, clustering and prediction, as key techniques, are explained. Additionally, some common data mining tools are described in detail, such as Weka, NLTK and Spider Miner. The final product of data mining provides a leadership opportunity that is capable of monitoring the buying habits of its customers, demand trends and locations, effectively implement the strategic decision for business growth and the income.

Zhao (2018) presents several big data algorithms as well as their implementation in e-commerce to deliver input to e-commerce businesses as part of the deployment of big data mining. In the age of digital transformation, big data mining a key player in developing e-commerce and represents the future of global e-commerce. The drawbacks of electronic commerce applications are outlined. Mobile e-commerce is presented as a rapid change in consumer behaviors and habits, such as multi-scenario consumption and diverse buying patterns. In the same way, e-commerce linked to social networks is presented as an important business model. The introduced e-commerce models employ big data technology, increase the effectiveness of online marketing, provide accurate recommendations by predicting consumer buying behavior, follow the behavior of competitors, prevent risks, do strategies and increase profitability, etc. In the paper, eBay is featured as a template for the biggest web-based corporate site and data mining is the focus. The eBay big data business model is made up of three layers such as the data platform layer, the data integration layer, and the data access layer. Ultimately, the objectives of the big data mining techniques in electronic commerce are primarily to a) optimize the e-commerce site or system to improve the customers' online shopping experience; b) enhance the ability of customer relation management to enhance decision-making; c) provide personalized services to increase the online sales; d) provide a value-added service.

2.6.3 Data mining in banking

Preethi and Vijayalakshmi (2017) analyze the data mining algorithms and notions needed in the banking industry to positively influence their performance. The authors regard the processing of huge transactional data and decision-making as a vital task; second, manual and conservative handling of decision-making is time-lagging, tedious and subject to errors. In this case, data mining techniques provide an efficient way of decision-making. In the study, two treated fields of banking applications are customer relationship management and fraud detections and the main techniques applied are clustering and association methods. In addition, data mining, using processes, enables information to be used in applications as fraud detection, market analysis, production control, etc. Most importantly, the banks use a customer relationship system to create brand value, identify and understand the needs of their customers. In these cases, data mining techniques enable the discovery of new customers by using clustering techniques and the retention of customers by using Apriori algorithms. Algorithms are also covered. Fraud detection is the primary area of focus, as anti-fraud measures are a major concern for banks, and most banks implement a hybrid approach to detect fraud trends. In conclusion, as regards digital transformation, by employing data mining techniques in data handling and decision-making, banks significantly increase their profits.

Hasheminejad and Khorrami (2018) examine the data mining algorithms used to analyze bank clients and reveal their clients and devise better marketing strategies. Introduction is carried out for some data mining techniques that are used to classify customers and help businesses make decisions. Customer relationship management is regarded to have the objective to improve the business relations with clients and to maximize the lifetime value of a customer to a business. Data mining will make it easier to study client analysis. Focus is placed on identifying customers and their needs as a pillar of digital transformation. By clustering customers, banks can identify critical customer attributes. Customer clustering is emphasized as an enabler to help guide the implementation of design marketing strategies for each customer group. A study on the datasets of recent studies of various researchers is conducted. Data mining techniques are studied and synthesized to identify customer behaviors, providing the banking industry with a competitive advantage.

2.6.4 Data mining in manufacturing

Harding et al. (2006) provide an overview of the uses of data mining in manufacturing. The purpose of this paper is to present the relationship between data mining and manufacturing industry and how data mining becomes an important component of digital transformation and decision-making. The main focus of this paper was on algorithmic applications. Comprehensive research has been carried out on applications in fields such as engineering design, manufacturing systems, decision support systems, failure detection and quality improvement, data mining in maintenance and customer relationship management.

Oliff and Liu (2017) discuss the use of data mining techniques and their advanced use in Industry 4.0 ready works. As a result, the methodology presented incorporates the principles of data mining and uses them to facilitate quality decision-making. Advanced data analytics and machine learning, which are part and parcel of digital transformation, are critical technologies and data mining and knowledge discovery for the use of extensive data to understand manufacturing processes are outlined in this study. Algorithms are a next topic. A case study is carried out to demonstrate how to utilize data mining to drive assembly and quality control processes forward and to apply them to existing systems without difficulties. Most algorithms are based on rule-based learning in which they operate through mathematical relationships. Intelligent systems and processes in Industry 4.0 are demonstrated as becoming feasible by the utilization of data and the investigation of ways in the utilization of existing records.

Dai et al. (2020) provide insight and discussion on the needs and challenges of big data analytics within manufacturing internet of things (MIoT). MIoT analytics, as a key driver of digital transformation, offers advantages, such as improvements in operation and production, improved supply chain efficiency, lowered costs, and improved customer experience. Consequently, data analytics are important in extracting informative values, estimating the coming events, and predicting the increase or decrease of products. In addition, predicting consumer behavior plays an important role in the manufacturing sector, e.g. in improving purchasing decision-making. Data analytics challenges addressed are: data temporal and spatial correlation, effective data mining systems and privacy and security. In addition, this research provides a prototype system to demonstrate the feasibility of distributed computing models in MIoT.

2.7 Benefits of Applying Data Mining

The adoption of data mining is advantageous from a process digitalization perspective, as data mining empowers predictive analytics, enhances decision-making, increases cost-effectiveness, lessens risks, and reflects customer behaviors (Saeed, 2020). Digital transformation enables new business models that rely on intelligent analytics-based digital decision-making that are used to gain new ideas and to improve decision-making (Roedder, 2016). Investigating the benefit of data mining in decision-making aims to facilitate the process of inducing businesses to adapt their strategies to the new digital era.

2.7.1 Improvement of Decision-Making

Milovic and Milovic (2012) address the context of predicting in two stages such as the learning and the decision-making stage. The decision-making stage for individuals is generally not regarded as qualitative when there is a lot of data to classify. Therefore, information gained through data mining techniques can be helpful for effective decisions to enhance an organization's success. Pulakkazhy and Balan (2013) speculate that valuable information is concealed in the amount of operational and historical data that can be used for critical decision-making processes if they are uncovered. The authors assert that data mining improves decision-making and make the decision-making process easier and productive. Businesses that use data mining techniques benefit greatly and gain an advantage over other businesses that do not, because interesting trends and insights can be gleaned from the sheer volume of data that can then be applied to decision-making practice. Ltifi et al. (2016) suggest a model with visual data mining techniques to support dynamic data mining. Furthermore, the authors point out the significance of trends like visualization, data mining and dynamic decision-making. Incorporation of visual data mining methods into real-time decision-making is essential for dealing with complex and time-based data. The application of data mining techniques enables to digitize the processes of extraction of valuable knowledge and regularity of data, and to provide guidance for decision-making. Consequently, decision makers are capable of visually predicting trends in the time-based data and their behavior. Manita et al. (2020) highlight the growing importance of digitization and the impact of digital technologies on businesses. Furthermore, the study highlights the importance of implementation of digital strategies to bring needed changes to regulators as part of the digital transformation. Novel digital technologies, such as data analytics and data mining, improve decision-making efficiency and help predict the performance of businesses with greater confidence.

2.8 Challenges and Limitations of applying data mining

Some of the research articles reviewed will facilitate the process of uncovering the challenges and limitations in the application of business data mining. The research is intended to facilitate data mining application approach, considering the limitations encountered.

Zain and Rahman (2017) identify the five common drawbacks of the application of data mining, including technology, expertise, incoherent or missing data, and privacy and data security. Wrong technology used in data mining promotes improper data manipulation, differently powerful technology helps drive efficient data mining processes and can prevent technology issues. Skills, in terms of controlling the data, are essential for manipulating and managing the large volume of data, and the data mining tool. Inconsistent or missing data could negatively impact businesses, in many ways. Data mining is a breach of privacy and a threat to data security as it poses a threat to statistical security and privacy.

Sharma et al. (2013), enumerate three limitations to data mining: privacy issues, security issues and misuse of information/inaccurate information. Businesses gather customer information in different manners, and it is possible to infringe upon the privacy of users. Security is regarded as a high-risk challenge as data are gathered and the potential for data to be hacked is a significant concern for businesses. Finally, improper or inaccurate use of information is regarded as highly harmful and has severe consequences if it is used by non-ethical individuals.

Ikenna et al. (2014) addressed the following challenges related to the implementation of mining in a business: problem of poor data quality, employee empowerment, data integrity and security issues, and integration complexity. Data mining enables business decision-making processes. The authors imply that the application of data mining techniques to a business issue implies the elimination of high-quality data that are generally primary data generated in operational functions. Additionally, finding an expert on a data mining project within a company can be challenging. Moreover, integrity and security are identified as critical challenges for any data gathering that is shared and used by a business. Lastly, the complexity of integration is seen as a crucial question because it is difficult to efficiently integrate data mining projects into operational processes.

2.9 Evaluation of Data Mining Model Outcomes

Mohd Selamat et al. (2018) reckon CRISP-DM as the most prominent data mining model and is primarily applied for predicting and facilitating decision-making processes. The reputation of CRISP-DM model is that it applies across all industries and is considered as the most commonly used model. In contrast to the KDD and SEMMA models, the CRISP-DM process model is an ongoing life cycle operating mode.

Martínez-Plumed et al. (2019) have studied CRISP-DM methodology that was created to catalogue and lead the steps in data mining projects. CRISP-DM is considered as the canonical approach and explains and expands the phases in six stages: business understanding, data understanding, data preparation, modelling, evaluation, and deployment. In the evaluation stage, the generated models, as an outcome of the modelling stage, are then assessed in a relation to a given quality standard, and the best model is then applied for making predictions. At this stage, the results are evaluated, and the process is reviewed to ascertain whether it achieves the objectives or not.

Dåderman and Rosander (2018) highlight that the work performed must be assessed and the result must meet operational requirements. In addition, upon completion of this evaluation stage, a decision on the outcome of the data mining would have been achieved.

Markusoski et al. (2019) consider KDD as an exploratory analysis and modelling of large databases, and develop KDD in nine stages: develop an understanding of application areas, select and create a dataset, preprocess and clean, transform data, choose the correct data mining task, choose the data mining technique, employ the data mining technique, evaluate and use the discovered knowledge. Moreover, Tariq et al. (2019) emphasize the steps in the SEMMA methodology, such as sample, explore, modify, model, and assess.

Table 5 presents results on identifying data mining process models and comparison of data mining process models.

Table 5: Data Mining Process Models

Data Mining Models		**CRISP-DM**	**SEMMA**	**Knowledge Discovery Process (KDD)**
Year		1996	1999	1996
Phases	A	Business Understanding		Developing and understanding of the application domain
	B	Data Understanding	Sample Explore	Selecting and building a target data set Preprocessing, cleansing
	C	Data Preparation	Modify	Data Transformation
	D	Modelling	Model	Selecting the proper data mining task Selecting the data mining algorithm Employing the data mining algorithm
	E	Evaluation	Asses	Evaluate and interpret the mined patterns
	F	Deployment		Using the discovered knowledge

2.9.1 Evaluation Metrics - Clustering

Singh and Surya (2015) highlight that clustering is tasked with obtaining a structure in an unlabeled data set. Clusters are measured according to the resemblance between objects, which are present within the same cluster and simultaneously, or dissimilarity, which is seen among objects belonging to various clusters as the distance between cluster points. If the clustering algorithm divides different and similar observations, then it has worked well. The cluster count, the cluster instances, square error, and the time needed to construct a/the model and the log likelihood are the components for comparing clustering algorithms. Accordingly, assessing the performance of a clustering algorithm differs from the case classification algorithms.

Renjith et al. (2020), in their study present an approach in the assessment of performance of clustering algorithms in three steps. The approach takes the first step of determining the optimal number of clusters. Therefore, at this step, it is necessary to specify the optimum number of clusters as an entry for the clustering process. Finally, the algorithm's performance follows by entering the mean processing time per algorithm against a varying number of records (cardinality) with a constant number of attributes (dimensionality).

Patel and Patel (2019) analyze the performance and evaluate clustering algorithms according to specific parameters, including the number of clusters created, improperly clustered instances, how long it takes to construct the model and other metrics, including the number of instances within the dataset, as well as the type of the data set.

Valarmathy and Krishnaveni (2019) regard that the clustering technique can be applied for classes that are unknown and do not look at the instances labeled by class are used in the classification. The attribute that provides the correct similarity should be identified to increase the measurement of similarity between clusters. Cluster properties can be parsed to identify the patterns that differ between clusters. The efficiency of good clustering technique is evaluated by its capability to identify the models that are concealed and produces maximal intraclass resemblance and reduce interclass resemblance among other objects in the clusters.

2.9.2 Evaluation Metrics - Classification

Hossin and Sulaiman (2015) argue that measurement of assessment is important in reaching the optimum classifier in classification training. Selecting appropriate assessment measures is an important part of the discrimination and acquisition of the optimum classifier.

Çığşar and Ünal (2019) describe the classifiers' performance criteria. Classification accuracy is called the ability of the model to properly predict the class label as indicated in percentage terms. Speed is the time needed to build the model. Robustness is defined as the model’s ability to properly predict despite data with noise and missing values. Scalability means a model's ability to be precise and productive while handling a growing volume of data. Interpretability is referred to as the level of comprehension and insight that model brings.

The author, Verma (2019), provides an overview of evaluation metrics, which are regarded significant when assessing the classification algorithm for a dataset. “TPR is the measure of exactness; more specifically, it is the percentage of tuples in a test dataset that are assigned as positive. Recall is the measure of completeness; more specifically, it is the percentage of positive tuples in a test dataset that the classifier is assigned as positive. F measure is the harmonic mean of precision and recall. ROC curves are a visual tool that helps us in comparing two classification models. An ROC curve for a certain model brings out the trade-off between the TPR and the FPR. For a certain test dataset and a model, TPR is the ratio of positive tuples that the models label correctly; FPR is the ratio of negative tuples that are mislabeled as positive. PRC area is the area under the precision-recall curve that measures the accuracy of the classification model on a class dataset" (Verma, 2019).

Table 6 outlines the measures used to evaluate classification.

Table 6: Metrics for classification evaluations

Metrics	**Formula**
Accuracy	$\frac{(tp+tn)}{(tp+fp+tn+fn)}$
Error Rate	$\frac{(fp+fn)}{(tp+fp+tn+fn)}$
Precision (p)	$\frac{tp}{(tp+fp)}$
Specificity	$\frac{tn}{(fp+tn)}$
Recall (r)	$\frac{tp}{(tp+fn)}$
Sensitivity	$\frac{tp}{(tp+fn)}$
F Measure	$\frac{(2*p*r)}{(p+r)}$
MCC	$\frac{(tp*tn)-(fp*fn)}{[(tp+fn)*(tp+fp)*(tn+fp)*(tn+fn)]^{1/2}}$
Note: **tp** - true positive, **fp** - false positive, **fn** - false negative, **tn** - true negative	

Specificity is used as a measure of how many negative patterns are classified with precision. Sensitivity is used as a measurement of the proportion of positive trends that are classified with precision. Precision is used to measure positive outcomes that are accurately predicted based on the total predicted outcomes in a positive class. Moreover, accuracy measures the ratio between accurate predictions and the total number of cases. MCC factor is a measure of the quality (ratio) in binary classifications (Hossin and Sulaiman, 2015; Halimu et al., 2019).

2.10 Home Appliances After-sales Services

Home appliance customers are demanding and require comprehensive after-sales services support. Home appliances are considered durable products and are expected to be operational for some time (Murali et al., 2016).

Home appliance after-sales services are identified as high-margin businesses, which make up a larger portion of businesses and corporate profits and are the most sustainable source of income and require the most modest investments. Delivering the after-sales services is difficult and quite distinct from production and distribution supply chains (Altekin et al., 2017).

The after-sales services is introduced as an important concept in the home appliance industry to create good customer relations that enhance performance for consistent results. The authors, through the research from three different perspectives, underline the impact of after-sales services impact on customer satisfaction (Wickramasinghe and Mathusinghe, 2016).

Considering it is important for home appliances after-sales services businesses to implement digital technologies, in particular digital technologies that impact customer experience.

Durugbo (2020) systematically conducts a literature review of after-sales services and aftermarket support. In general, after sales is the period when the vendor or manufacturer guarantees its support to the purchaser, maintaining or repairing the purchased goods. The objective of after-sales service is to maintain extended warranty options that ensure product reliability and reduce servicing costs – along with the basic guarantee as the repair defects or failures of the product. The objective of the after-sales services business is to meet the warranty (guarantee) i.e., a time-limited contract assignment by original equipment manufacturers (OEM) to service the equipment.

Figure 7 illustrates the most important activities, like providing technical support in the field, distributing spare parts, and selling accessories and customer service (Durugbo, 2020). Moreover, secondary markets are mutually supportive product markets which are often bought from other related products. The purpose of the figure is to introduce the typical operations of a home appliances after-sales services business.

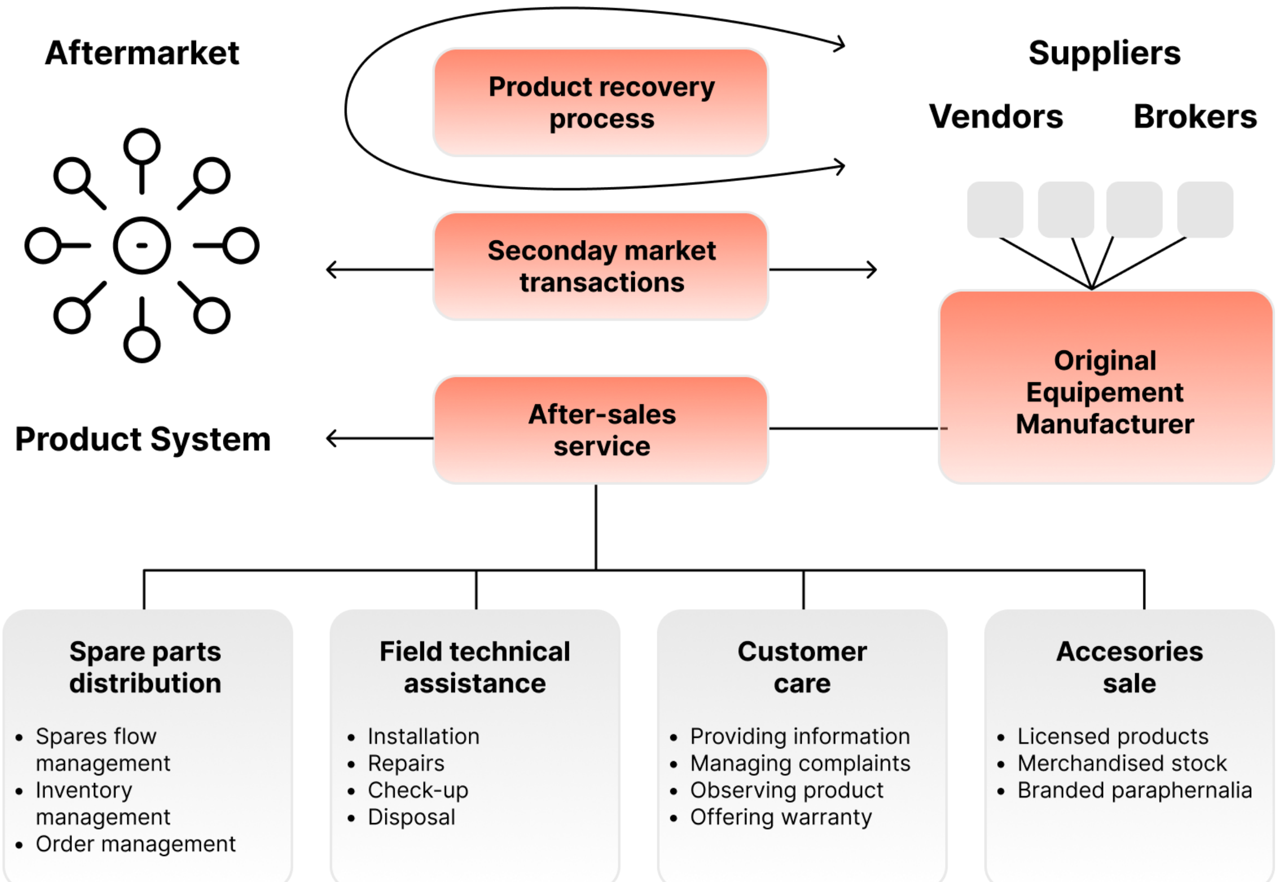

Figure 6: After-sales Service Business Processes (Durugbo, 2020)

Murali et al. (2016) term after-sales services as activities which take place after clients have purchased and are committed to assisting clients to use and dispose of goods. The home appliance sector is among the largest in the world and is among the most consistent categories in terms of customer satisfaction. Home appliances are relatively costly goods with short and long-term economic impacts which consumers can struggle to manage from a household budget perspective.

Esmaeilpour (2016) singles out home appliances after-sales service as a factor in business success. After-sales services address all the problems of companies that sell goods to their clients such as service warranty, delivery of products, installation services, parts supply, and repair services. One of the issues with purchasers is the inability to provide a proper the after-sales service guarantee for sustainable domestic appliances. Undesirable quality of home appliances after-sales services provided by the business results in a threatening business relationship.

2.11 Performance measurement of after-sales service business activity: nonfinancial aspects

Nowadays, fact-based decision-making is applied. Analytics and data mining approach is used to tackle and facilitate strategic business decision-making, predict diverse situations and to measure performance of the business.

After-sales service presents a significant role in profit generation and customer satisfaction (Saccani et al., 2006). After-sales service business is a division and its managers must obtain acceptable financial outcomes (revenue growth, profitability) and competitive performance (market share, customer satisfaction and loyalty, competitors' outcomes) (Saccani et al., 2006).

Financial performance measures are not sufficient to describe a business' trends (Nagy et al., 2011). Nonfinancial performance measures complement financial performance measures and are the drivers of financial performance measures (Nagy et al., 2011).

Both management and stakeholders consider non-financial indicators very important (Laura et al., 2019). The presence of non-financial indicators in a business' performance measurement supports business' strategic alignment and influences on organizational effectiveness (Laura et al., 2019).

The balanced scorecard is among the most commonly used performance metrics models and strategy management instruments and is used to explain, communicate and implement strategies (Oliveira et al., 2021).

Performance measures are categorized in nonfinancial performance measures, such as customer satisfaction, business internal process and innovation and improvement activities, and financial performance measures (Fowke, 2010).

The performance measures for managerial decision-making (Balanced Scorecard Model, Fowke, 2010) are presented in Figure 8.

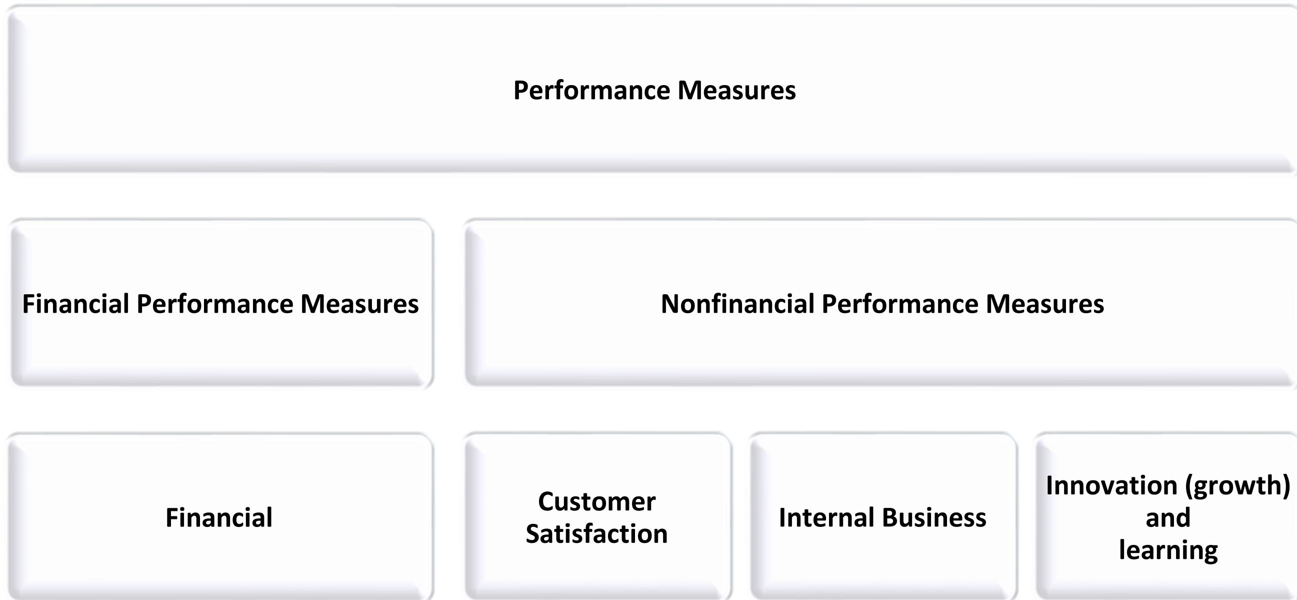

Figure 7: Performance measures for managerial decision-making

The customer satisfaction measures are focused on finding the existing gap between the customer expectations and the performance of the business, relating the characteristics of the output (service) provided (Saccani et al., 2006).

Saccani et al. (2006), in their paper, proposed performance measurement system -framework, which is divided into four stages such as business area, process level, activity level and development and innovation level. Hereinafter, process performance can be measured by customer satisfaction, agility, and efficiency. Furthermore, the authors have performed four case studies with after-sales service businesses. Based on the first case study, the after sales business is more concerned about response times and customer satisfaction than with financial performance. In the second case study, the strategy of the business focuses on customer satisfaction, loyalty and retention in the long run. Moreover, in the third case study, the home appliances after sales considers customer satisfaction as the main objective of the strategy. Lastly, in the fourth case study of the paper, the goals of after-sales service concentrate on customer retention and on the progress of the brand across customer satisfaction.

Service quality is an assessment of service performance if it conforms to the customer's expectations (Ramya et al., 2019). Parasuraman et al. (1988) developed a model with principal dimensions of service quality and a model that measures service quality (SERVQUAL). Service quality denotes a gap standing between customer expectations and service performance.

Murali et al. (2016) consider after-sales services contribute significantly to customer satisfaction. The authors consider that the main objective of any business is to satisfy its customers. After-sales service quality is studied and evaluated by measuring customer satisfaction through SERVQUAL dimensions covering after-sales service attributes. The work presented, examined the relation between customer satisfaction and the quality of after-sales service.

Golrizgashti et al. (2020), the authors elaborate and study how after-sales service activities provide more quality than what customers expect. High service quality results in a competitive advantage and has a significant impact on customer decision-making. The study is performed through SERVQUAL dimensions.

Alireza et al. (2011), in their research, explores the relation of after-sales services and customer satisfaction. The author concludes that customer satisfaction is highly affected by the quality of the after-sales service provided.

The SERVQUAL dimensions and attributes (Murali et al., 2016; Golrizgashti et al., 2020; Parasuraman et al., 1988; Ramya et al., 2019) are presented in Table 7.

Table 7: SERVQUAL dimensions (service quality) based on after-sales service attributes

Dimensions	**Description of Attributes**
Reliability (RL)	Consistency of service quality
	Options and variety of services
	Supply of required spare parts
	Service delivery as promised
Responsiveness (RS)	Reasonable warranty policy
	Responsiveness to customer complaints
	Time taken for resolving the complaint
	Affordable cost of service
Assurance (AS)	Customer handling
	Professionalism of personnel
	Technical skills of personnel
	Interpersonal behavior of personnel
Empathy (EM)	Accessibility of personnel
	Ease of contacting personnel
	Comprehending customers' needs
Tangible (TA)	Supply of service tools/equipment
	Accessibility of service center
	Complaint reporting facilities
	Quality & availability of technical service documentation
	Info and guidance available at service center

Murali (2016) and Parasuraman et al. (1988), in their papers, tackle the SERVQUAL, a multi-dimensional research instrument, which is used to measure five dimensions of service quality in different areas including home appliances. Service quality and customer satisfaction affect financial performance. The main aspects of service quality include (Parasuraman et al., 1988):

1. **Reliability** which is referred to as the ability to perform as promised.

2. **Responsiveness** which is termed as the willingness to assist customers and to provide prompt service.
3. **Assurance** which is denoted to the ability of a business' personnel to convey confidence and trust to customers.
4. **Tangibles** which are physical things about the business such as personnel appearance, equipment, and facilities.
5. **Empathy** which is the degree of care provided by the business to its customers.

Finally, customer satisfaction and service quality are two important nonfinancial performance measures for managerial decision-making. Accordingly, our research will focus on studying and experimenting mainly with customer satisfaction and service quality as key points in after-sales decision-making and after-sales service business strategy.

2.12 Summary

In this section, we presented all the findings obtained from the related work research, gaps and contributions.

The objective of the related work was to explore different views and areas of intelligent applications and advanced analytics implementation, important aspects of data mining in driving digital transformation, and benefits and challenges of data mining along the business digital transformation journey.

The findings on digital transformation dimensions, technologies, tools and techniques and other results obtained from the related work are presented in Figure 9.

Dimensions of digital transformation	• Customers • Strategies • Technologies • Operations • Organization and culture
Digital transformation technologies	• Big data analytics • Iot • Cloud • Mobile • Social networks
Advanced and big data analytics	• Machine learning • Text mining (NLP) • Data mining
Data mining techniques	• Classification • Clustering • Regression • Association • Sequential Analysis
Data mining tools	• Weka (Java) • Rapid miner (Java) • R (C, Fortran, R) • Orange (C++, Python, C)
Clustering algorithms	• K-means • Farthest-first • Canopy • Density-based group • Filtered clusterers
Classification algorithms	• Decision tree • Random forest • Naive bayes • Logistic regression • K-nearest neighbour (KNN)

Figure 8: Scheme on digital transformation dimensions, technologies, tools and techniques

The findings on the business areas that have used and benefited from data mining techniques aiming to improve decision-making along the business digital transformation and the results obtained from the related work are presented in Figure 10. The data mining technology is embraced in many industries for its benefits. The major areas where data mining is taking place are retail, e-commerce, banking, and manufacturing.

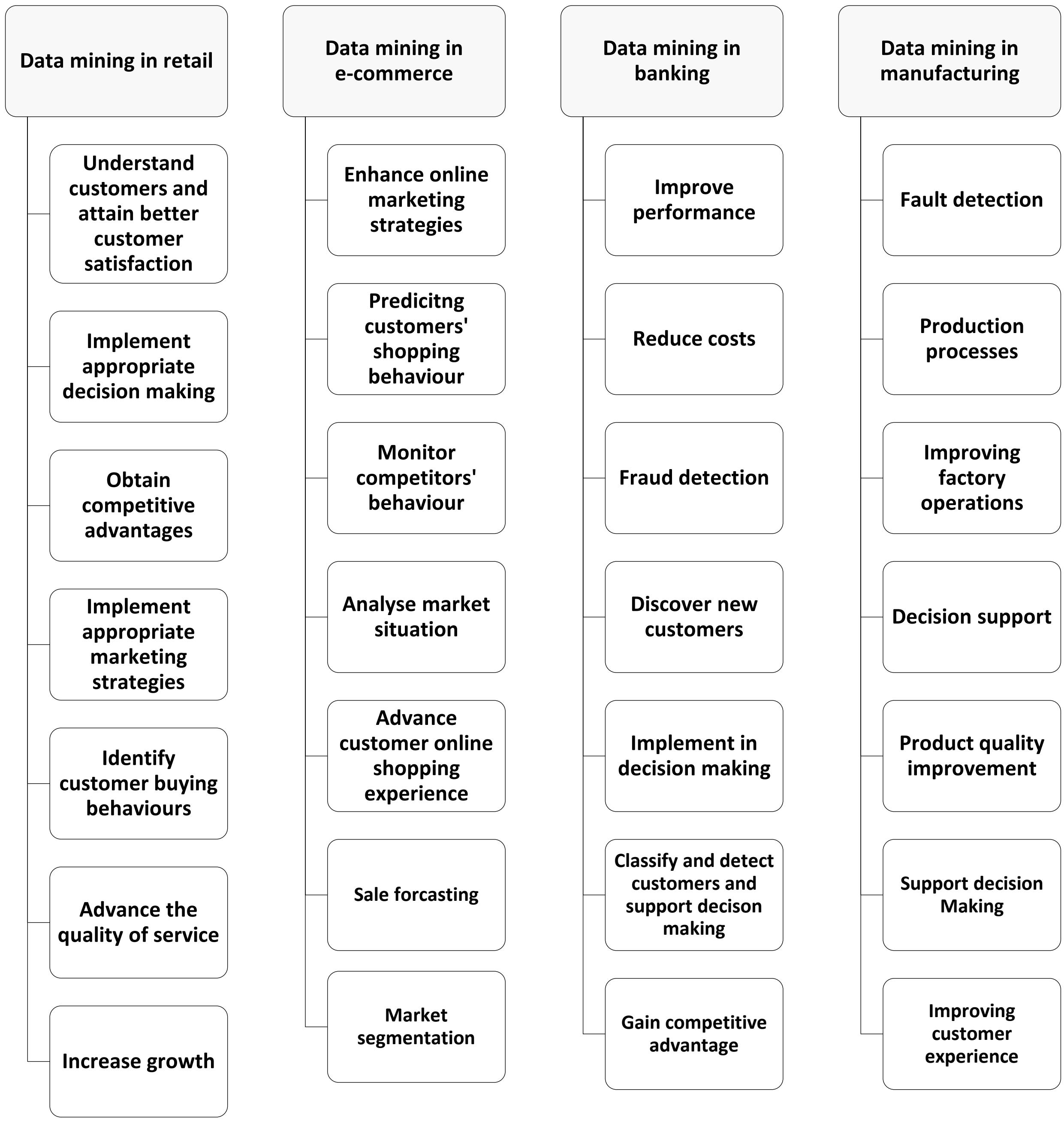

Figure 9: Scheme on business areas benefited from data mining techniques

The findings on the challenges and benefits of applying data mining along businesses' digital transformation journey and the results obtained from the related work are presented in Figure 11.

Figure 10: Findings on the challenges and benefits of applying data mining

The findings on performance measurement for decision-making are presented in Figure 12.

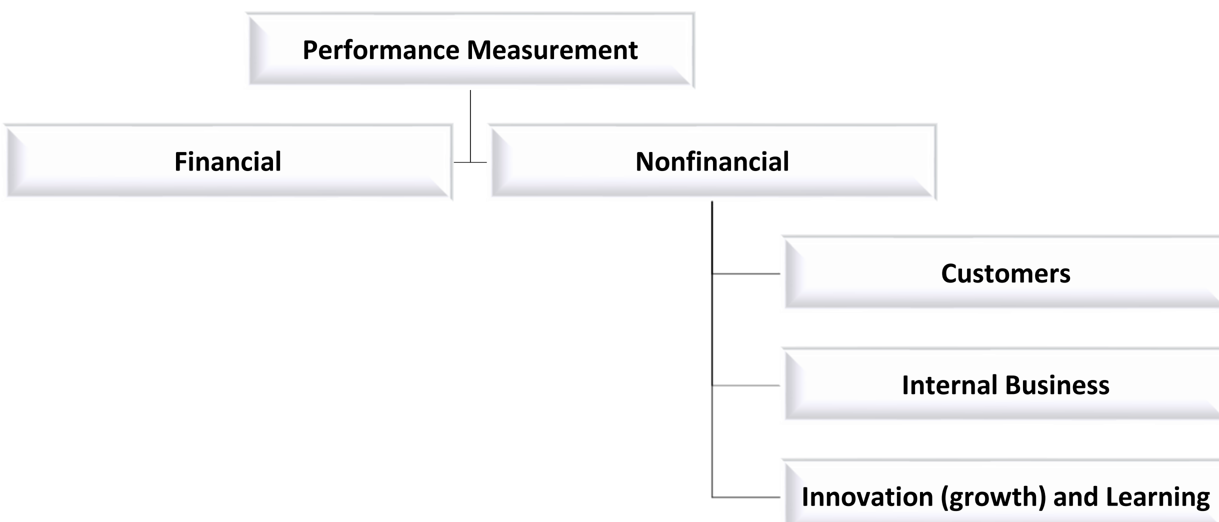

Figure 11: Findings on performance measurement for managerial decision-making

The conduction of related work results by identifying research gaps. Our research study addresses two main gaps, such as:

1) The exploration of understanding after-sales service businesses and their challenges in after-sales service businesses in the Kosovo market is the gap that we try to fill with the qualitative research - case study. The results obtained by the research will identify areas and aspects of improvement and the strategic, effective solutions offered to the management of after-sales services businesses are beneficial and impact the decision-making processes.
2) Moreover, decision-making must be based on empirical evidence and consequently lead to better outcomes and impact strategies and non-financial decisions. In this context, our research study identifies gaps that we try to fill with the data mining approach through the quantitative research study.

The main contribution of this research study is that:

- Our research proposes a model and an outcome - a recommendation system - as a part of the features of the data mining system, to streamline decision-making throughout the business digital transformation journey in after-sales service businesses that help to make the right decisions.

3 METHODOLOGY

This section contains the methodology of the research study and includes the combination of different tasks. Reflections to complement this research were included in the presentation of the steps of implementing a case study approach in the area of after-sales services home appliances.

The Figure 13 presents methodology of analysis, which comprises two types of research, qualitative and quantitative study, and alleviates the provision of conclusions in this context.

Figure 12: Methodology analysis with two research types of study

In the context of *business understanding*, as a case study, investigated the home appliances after-sales service businesses on the Kosovar market. The selected business is a business that provides after-sales service for home appliances, electronics repairs, including a broad range of appliances. In addition, the business provides appliance installation services including air conditioning, washing machine, dishwasher, etc.

Figure 14 shows the key after sales actions carried out by the business, including field technical support, distribution of spare parts, customer service and logistics.

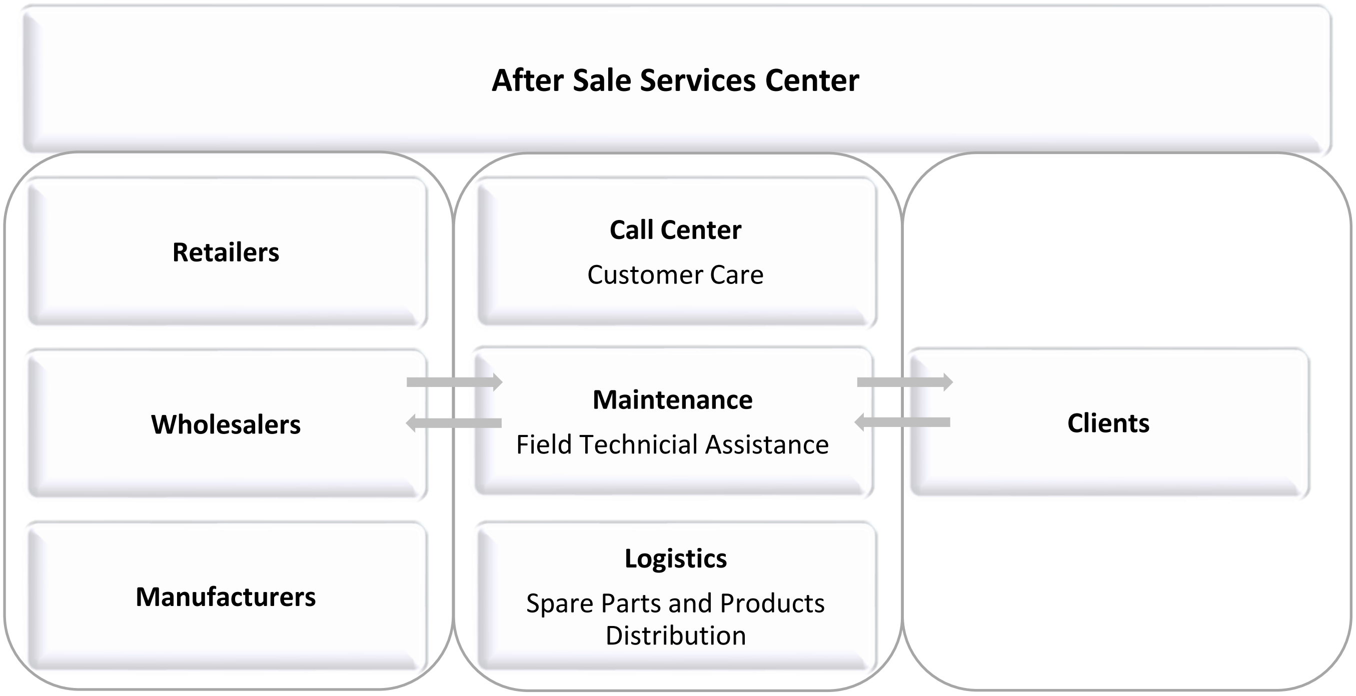

Figure 13: Company's After-sales Service Business Process

The business does not rely on data mining tools for advanced analytics. Thus, data mining has been lacking.

Considering how important analytics results are enhancing decision-making and efficiency, implementing analytical platforms and smart applications is a challenge to after-sales service activities. The next issue is to retrieve and enter valid, dependable, and useful information and insights from huge databases with minimum human interference. Lastly, the other challenge is to integrate and the link delivery with business logic, influencing decision-making.

3.1 Qualitative Research

In this study, a qualitative research approach is selected to understand and identify the weak areas of performance and challenges of the business. The purpose of this exploratory research on home appliances after-sales service businesses is to discover and seek new insights and as a result help the decision makers to create proper strategies to improve the business performance and enhance customer satisfaction. Based on the important performance metrics derived from the related work study, this research will examine the relationship between after-sales services and customer satisfaction.

Moreover, it includes descriptive research to provide a study on the actual situation of the after-sales service business in the Kosovo market.

Based on the information gathered through the interview, we are able to generate productive conclusions and answer to the research question two RQ2:

What challenges are faced by the businesses? What are some of the challenges we find solvent or hard to overcome, and how do we come up with effective solutions?

Figure 15 provides a graphical presentation of the qualitative case study research methodology.

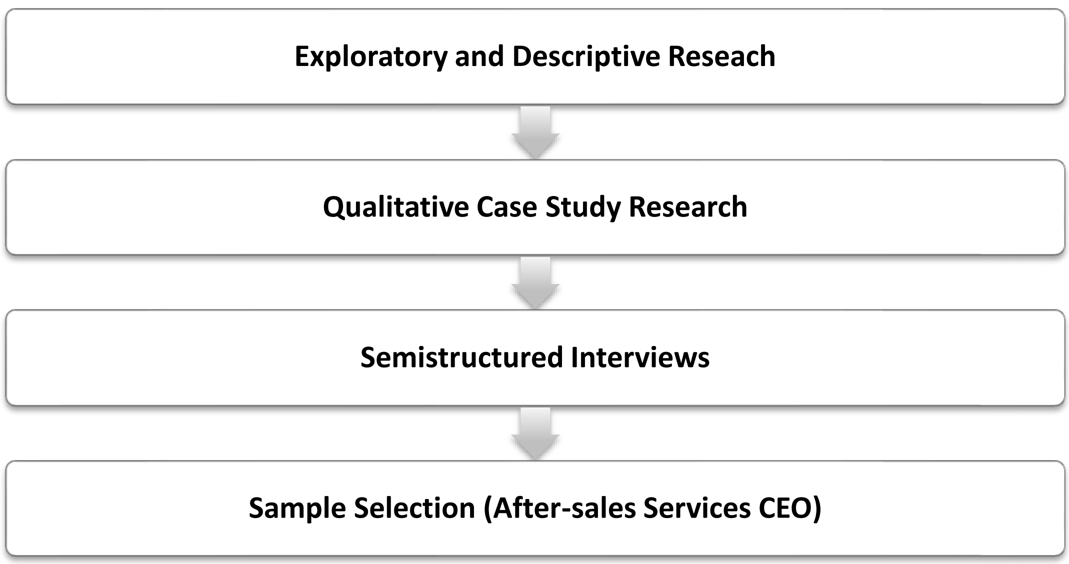

Figure 14: Qualitative Case Study Research Workflow

Qualitative research study workflow is comprised of four steps of research as exploratory and descriptive research, qualitative case study research, semi-structured interviews, and sample selection.

3.1.1 Research design, sampling, and instrument

Empirical methods were used to carry out this study. Semi-structured interview was conducted to benefit from a very potent sample of respondents, regardless of its region, which are from home appliances after-sales service businesses on the Kosovo market and is the subject of this study. This sampling design has included two sizes of businesses, micro-sized and small sized businesses. The interview was held with 19 respondents over the period of late June to early October 2021. The interviewed top management officers were selected based on their accurate and wide knowledge on business, and their appropriateness on getting answers to questions. Through the data collected from the interview several types of findings are obtained. Table 8 below shows the sample selection.

Table 8: Sample Selection for Qualitative Research

Number of **Respondents** (CEO)	19
Number of respondents (R) based on **Region of Business**	1 (R) - All Around Kosovo
	2 (R) - Ferizaj Region
	3 (R) - Peja Region
	1 (R) - Gjakova Region
	2 (R) - Prishtina Region
	3 (R) - Mitrovica Region
	3 (R) - Prizren Region
	4 (R) - Gjilan Region
Number of respondents (R) based on **Business' Expertise**	1 (R) - Home appliances after-sales service business for repairing all brands and categories
	18 (R) - Home appliances after-sales service business for repairing distinct categories
Number of respondents (R) based on **Business' Size**	18 (R) - A micro-sized enterprise (<10 employees)
	1 (R) - A medium-sized enterprise (~50 employees)

The interview, with the information gathered by the exploratory research, enabled us to discover and seek new insights to facilitate the design of the strategy of after-sales service business. Moreover, the research was descriptive research which studied the actual situation, mainly the challenges, of the after-sales service business in the Kosovo market. The dimensions of service quality of the interview were mainly related to nonfinancial performance measure – customer satisfaction.

3.1.2 Validity and Reliability

Through validity, which is an important aspect, we make sure the ability of the interview is to measure what is planned. Through reliability, we indicate the trustiness of the interview.

The validity of the research is based on:

- The interview which is designed to respond to the research questions
- The findings of the related work study on management decision-making measures that contribute to the formulation of the interview to evaluate and measure the service quality impacting customer satisfaction, a nonfinancial performance measure, through SERVQUAL dimensions
- The SERVQUAL dimensions which are key for the preparation of interview questions
- The interview of top management officers of home appliances after sales-service businesses in the Kosovo market
- The interview that is developed through meetings and phone calls with the management
- The interview questions that are sent before the meeting in order to have time to review the question and prepare answers preliminarily
- The respondents that are very well-informed of the after-sales services business activities
- Other documentations that are used as a source of evidence
- The documents produced by the research that is delivered to the chief executive officer (CEO) for approval of the results

Reliability, which is also an important aspect of the research, is based on:

- Objectivity that is ensured through the interview process
- Application of Albanian language applied during the interview

3.2 Quantitative Research

The quantitative research study provides analysis and practical comparative study of the application of data mining techniques and the implementation of recommendation system, as a portion of data mining system functionalities.

Based on the studies by implementing data mining approach, we are able to generate productive conclusions and answer to the research question three RQ3:

How does a model support an after-sales service business that is considering using data mining to improve decision-making along the business digital transformation? How to analyze and design the platforms that can improve the analytics of the data?

Quantitative research study comprises three steps of research as data mining algorithms comparison implementation, study and analysis on service quality attributes and customer satisfaction, and recommendation system implementation (Figure 16).

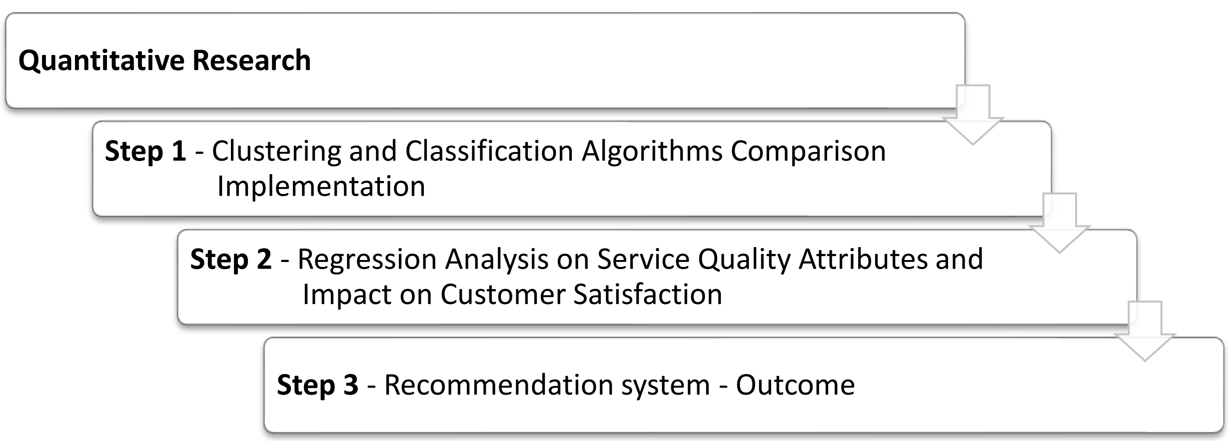

Figure 15: Quantitative Methodology Workflow

3.2.1 Clustering and Classification Algorithms Comparison Implementation

The methodology is followed by comparing various types of classification and clustering methods. The purpose of comparing clustering and classification methods is to come up with the right technique among the various methods that would help in this case study.

The classification algorithms such as Naive Bayes, J48, random forest, K-Nearest Neighbor, and logistic regression are implemented in the dataset.

The clustering algorithms such as SimpleKMeans, Canopy, MakeDensity BasedClusterer, FarthestFirst and FilteredClusterer are implemented in the dataset.

Weka 3.8.4 was implemented for clustering and classification algorithms in this case. Attwal and Dhiman (2020) delve into Weka, which is an open-source platform based on Java and is an integration "of machine learning algorithms and data pre-processing tools". The Weka platform applies techniques to implement of main data mining, supervised and unsupervised tasks. The author, Janošcová (2016), emphasizes that Weka provides allowance to new packages for distributed data mining. It is not sufficient to simply obtain information, but to pick the most significant information, properly interpret the information obtained and then act effectively.

Moreover, the dataset applied here is of a business providing home appliances after-sales service. The methodology began by preparing data for the company's dataset. The dataset received is managed by the customer service unit. Figure 20 presents the steps of the data preparation workflow.

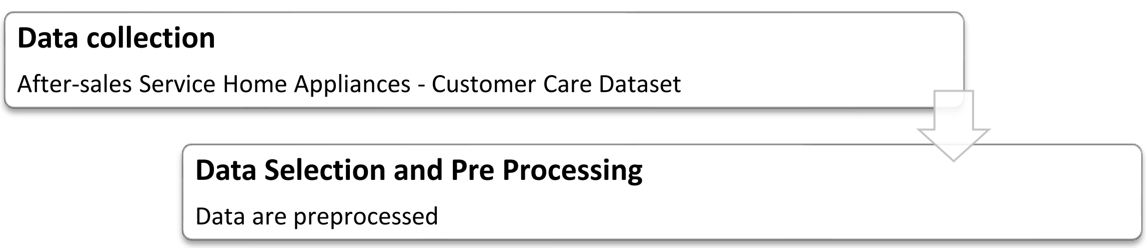

Figure 16: The preparation of data workflow

Data collection - Secondary data have been used to understand the types of after-sales services. Within the existing business workflow, customer inquiries are gathered from customer service. The customer service representative will proceed with the customer's inquiry. Afterwards, the technical service meets with the customer to understand their needs. Under certain circumstances, the technical service should contact the customers for a detailed understanding of their needs.

Data selection - The CSV type dataset consists of approximately 62,000 observations and 8 variables. The attributes used are authorized service, brand, category, repaired by age, guarantee, action, sum, guarantee and days to repair.

Data pre-processing - All the data are pre-processed. As a data preprocessing is a major in the data mining process, the chosen data are eliminated of inconsistent, incorrect, and missing data. Using the data mining platform with algorithms, is the goal of successful analysis and comprehending the models and trends concealed within the business data.

In the case study, the clustering comparison implementation, *feature selection techniques*, such as correlation-based attribute evaluation, information gain attribute evaluation and wrapper attribute evaluation, are used to reduce class variable attributes by eliminating the excessive, non-predictive information characteristics the dataset. As a result, the high impact attributes to predict repairs are guarantee, authorized service, brand, action, category, repaired by age, days to repair and sum. The guarantee is the response and class attribute. And, the experimental section provides a comparison of clustering algorithms to denote the best clustering algorithm that helps in the area of after-sales services to a business.

Table 9 outlines the feature selection methods carried out in our dataset.

Table 9: Feature Selection Methods Performed

Feature Selection Method	High Impact Attributes
Correlation-based attribute evaluation	authorizedservice, sum, category, brand, repairedbyage, action, daystorepair
Information gain attribute evaluation	authorizedservice, sum, repairedbyage, category, daystorepair, brand, action
Wrapper attribute evaluation	authorizedservice, category, action, repairedbyage, sum

Velmurugan and Anuradha (2016) consider feature selection mainly as a part of data pre-processing and the feature selection techniques are carried out to reduce class variable attributes by eliminating the redundant and unpredictable data attributes. Correlation-based feature selection is a simple filtering method that classifies feature subsets in relation to a correlated heuristics evaluation feature. Information gain attribute evaluation evaluates an attribute through the measurement of information gain. The Wrapper technique relies on the learning algorithm and uses it as a black box for estimating the accuracy of subsets of variables within the prediction task.

Figure 21 presents the visualization results of the pre-processing phase and provides in particular the distribution of after-sales services according to the "repairedbyage" attribute.

The attribute “repairedbyage” is a numeric value; therefore, you can see the statistics describing the distribution of values in the data - minimum, maximum, mean and standard deviation. The attribute is specified for all instances (no missing values); it has twelve distinct values. Minimum = 1 is the lowest “repairedbyage”, maximum = 12 is the highest “repairedbyage”. The blue represents ‘yes’ of products with guarantee through the distribution of values of ‘repairedbyage’ attribute. The red represents ‘no’ of products without guarantee through the distribution of values of ‘repairedbyage’ attribute.

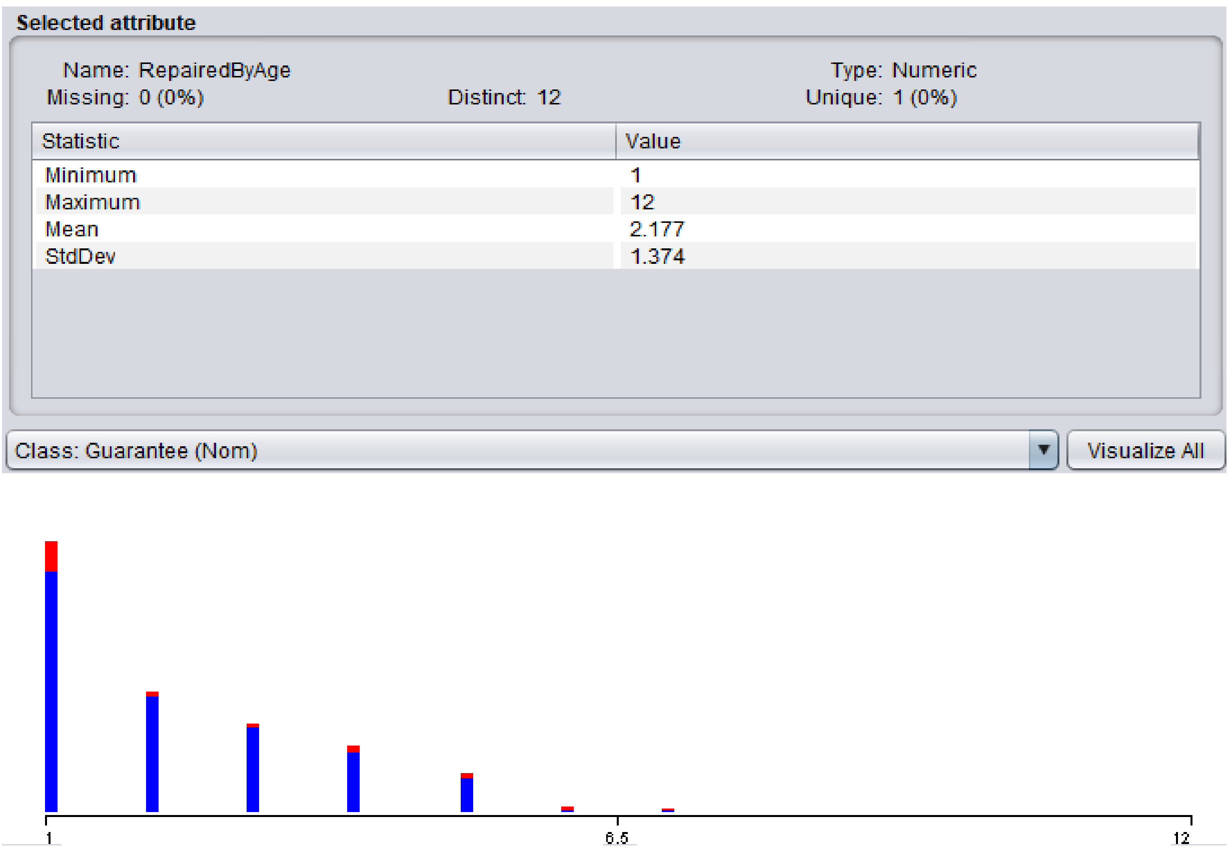

Figure 17: Visualization of pre-processing stage

The chosen data in the dataset is cleared of inconsistent and incorrect data, missing data and outliers. Table 10 presents the attributes in the dataset.

Table 10: The structure of dataset for data mining comparison implementation

Attribute	Description	Values
Guarantee	If the product is guaranteed or not guaranteed	Yes or No
Authorized Service	If the product is serviced through its own service center or a subcontractor	Service center, Subcontractor
Brand	Brand is the manufacturer of the product with 10 different values	Brand 1 - 9, Other Brands
Action	The kind of service supplied to the product	Servicing, Installation, Change
Category	The category includes small domestic appliances, personal care, TV, mobile, air conditioning, etc., with nine distinct values	Category 1 – 9
Repaired By Age	The product's age and is a numeric value	Numeric Value
Days to Repair	The number of days during which the service is performed from the time requested	Numeric Value
Sum	The amount (€) to serve and is a numerical value - if the product service is in guarantee, the amount is 0	Numeric Value

Only three row data is presented below in XML format.

```
<?xml?>
..
   <Row>
    <Cell><Data ss:Type="String">AuthorizedService</Data></Cell>
    <Cell><Data ss:Type="String">Category</Data></Cell>
    <Cell><Data ss:Type="String">Brand</Data></Cell>
    <Cell><Data ss:Type="String">Action</Data></Cell>
    <Cell><Data ss:Type="String">RepairedByAge</Data></Cell>
    <Cell><Data ss:Type="String">DaysToRepair</Data></Cell>
    <Cell><Data ss:Type="String">Sum</Data></Cell>
    <Cell><Data ss:Type="String">Guarantee</Data></Cell>
   </Row>
   <Row>
    <Cell><Data ss:Type="String">Subcontractors</Data></Cell>
    <Cell><Data ss:Type="String">AirConditioning</Data></Cell>
    <Cell><Data ss:Type="String">LG</Data></Cell>
    <Cell><Data ss:Type="String">Installation</Data></Cell>
    <Cell><Data ss:Type="Number">1</Data></Cell>
    <Cell><Data ss:Type="Number">2</Data></Cell>
    <Cell><Data ss:Type="Number">0</Data></Cell>
    <Cell><Data ss:Type="String">Yes-guarantee</Data></Cell>
   </Row>
   <Row>
    <Cell><Data ss:Type="String">ServiceCenter</Data></Cell>
    <Cell><Data ss:Type="String">AirConditioning</Data></Cell>
    <Cell><Data ss:Type="String">BRAND 6</Data></Cell>
    <Cell><Data ss:Type="String">Servicing</Data></Cell>
    <Cell><Data ss:Type="Number">1</Data></Cell>
    <Cell><Data ss:Type="Number">1</Data></Cell>
    <Cell><Data ss:Type="Number">0</Data></Cell>
    <Cell><Data ss:Type="String">Yes-guarantee</Data></Cell>
   </Row>
...
</Workbook>
```


The goal of this project is to explore how implementing data mining algorithm results to better decision-making and improve productivity through business prediction analysis.

3.2.2 Regression Analysis on Service Quality Attributes and Impact on Customer Satisfaction

Quantitative research provides regression analysis of service quality attributes and examine the impact of service quality on customer satisfaction, a nonfinancial performance measure for managerial decision-making. The main phases of this step are data collection and selection, pre-processing, reliability test and correlation analysis, simple regression analysis and multiple regression analysis.

Figure 17 provides a graphical presentation of the steps of analysis workflow.

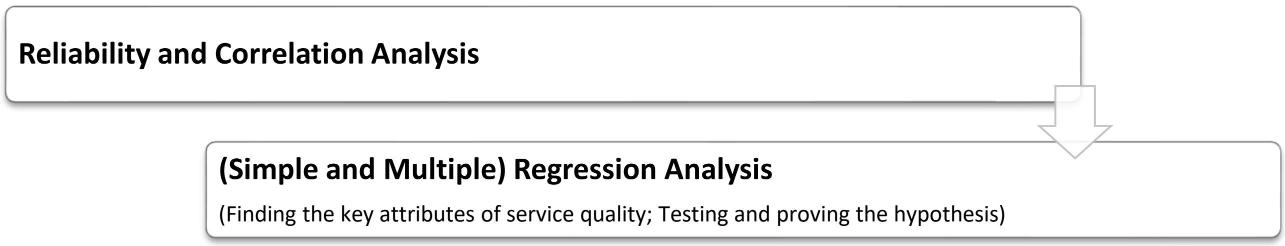

Figure 18: The steps of statistical and regression analysis

We conducted analyses and tests with Stata, a statistical data science program.

Data collection - Secondary data have been used to identify the service quality factors impacting customer satisfaction of after-sales services business. Customer inquiries are gathered through customer service. For this study, approximately 52K data were extracted from the previous dataset and the available variables were grouped for the study purposes.

Data selection - The CSV type dataset consists of approximately 52,000 observations and 5 variables. The used attributes include brand with spare parts (nominal), authorized service (nominal), payment (nominal), time to repair (nominal) and customer satisfaction (nominal). The selected data are eliminated from inconsistent, incorrect, and missing data.

Data pre-processing - All the data are pre-processed. Table 11 presents the needed attributes in the dataset.

Table 11: The structure of dataset for performance measurement

Attributes	Description	Values
Authorized Service	If the product is serviced through its own service center or a subcontractor	Service center, Subcontractor (1 or 0)
Brand with spare parts	Brand with spare parts means the manufacturer enables (initially) pre-provision for spare parts to the authorized service. It means the after-sales service business has the option in replacing immediately the spare parts at the time of repairing and problem identification without delay to deliver spare parts by the manufacturer	Yes or No (1 or 0)
Time to Repair	The number of weeks during which the service is performed from the time requested	1, 2, 3 and 4
Customer Satisfaction	The received service is without any complaint – if the service of the product is without complaint, it is considered an expected service, otherwise it is received service	Expected or Received Service (1 or 0)
Payment	Whether the product serviced is paid. The product is under warranty, but the service type is out of warranty	Yes or No (1 or 0)

The objective of the study is to measure the dimensions of service quality with an impact on customer satisfaction across the SERVQUAL dimensions by ensuring a few after-sales service attributes according to the dataset provided. The attributes such as 1) affordable cost of service, 2) supply of required spare parts, 3) time taken for resolving the complaint, and 4) technical skills and professionalism of ASEE personnel are a few attributes that are considered in the evaluation if the customer is receiving qualitative service who possess products in guarantee.

Factors – dimensions available of service quality are as follows: responsiveness, reliability, and assurance. Table 12 presents the available attributes in accordance service quality dimensions and attributes in the dataset.

Table 12: Dimensions and attributes of service quality

Dimension	Service Quality Attributes	Variable
Reliability (RL)	Supply of required spare parts	BrandwithSpareparts
Responsiveness (RS)	Affordable cost of service	Payment
	Time taken for resolving the complaint	TimeToRepair
Assurance (AS)	Technical skills and professionalism of personnel	AuthorizedService

The objective of this project is to analyze how the service quality attributes impact customer satisfaction and find the most important attribute impacting customer satisfaction.

3.2.3 Recommendation System

The design of the recommendation system rests on several principles, such as data mining approach and business-to-business observation, usability for the end-user friendliness and interactive interface through the implementation of a streamlit API design – recommendation system with machine learning concepts. The web-based recommendation system facilitates the process in presenting, with data and graphs, the prediction probabilities of the product repairs that fulfil their requirements.

Python has been selected for this implementation. This recommendation system implements streamlit library which is python library. The streamlit was applied by providing the python libraries for simplicity and flexibility. A visual studio code was used to design the graphical user interface. While developing the recommendation system, many meetings were held with the management of business in order to incorporate most important business elements and requirements in the model. The recommendation system has been designed to meet the requirements of the users (decision makers) of after-sales service business.

Figure 18 presents the concept design of the implementation and module functionalities. The module delivers prediction probabilities, capabilities, and full privileges to a user of the system. The objective of the system is to provide a web-based, the easy-to-use recommendation system that meets operational expectations. The schema of the web-based recommendation system is easy, understandable and accessible, useful and user-friendly.

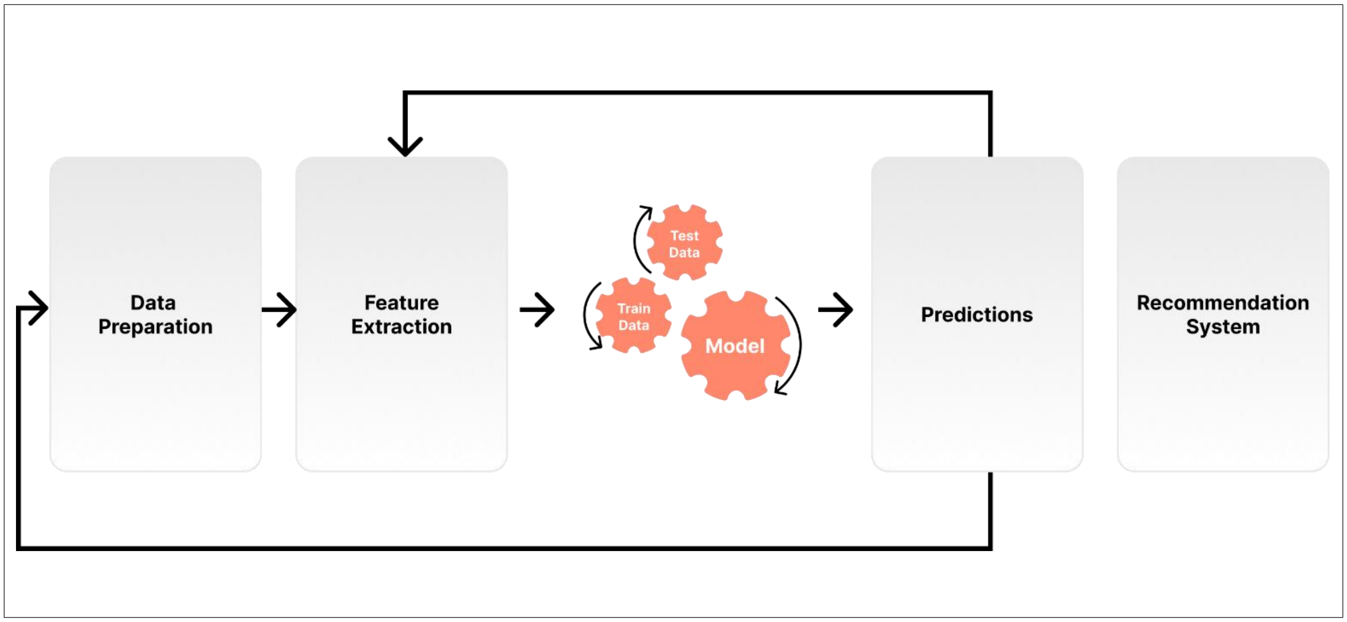

Figure 19: Functionalities of the recommendation system - as an outcome - implementation module

The studies of different researchers that contributed to conceptually develop our recommendation system are different. E.g., Lundkvsit (2014), in his/her study gives an overview of the implementation of a front end (interface), which then processes the flow of information into the classification algorithm – decision tree. Capozzoli, A. et al. (2016) presented a general architecture to collect and analyze data.

Moreover, the authors present the informative dashboards in the application layer, and the latter layer in Figure 19 that generate valuable information for various users and propose actions or strategies ready for implementation. Different types of graphs can be utilized to demonstrate extracted knowledge to end users in an explanatory and user-friendly manner.

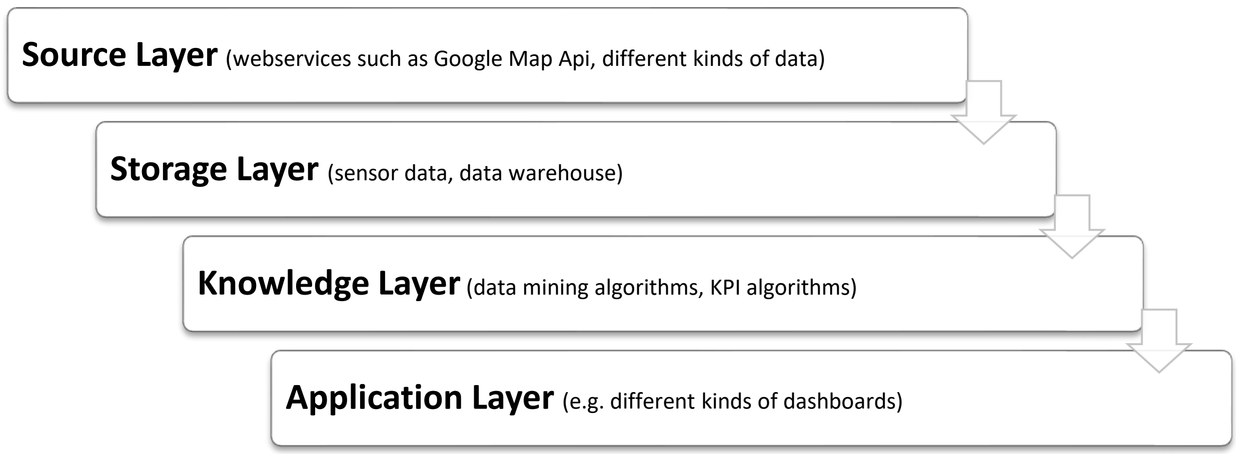

Figure 20: A general architecture to collect, store and analyze data (Source: Capozzoli, A. et al., 2016)

3.3 Summary

The proposed model is a two-phase research study, the qualitative and quantitative research study. The qualitative research study is one step study where quantitative research study is composed of three step study.

The proposed model is graphically presented as below in Figure 22.

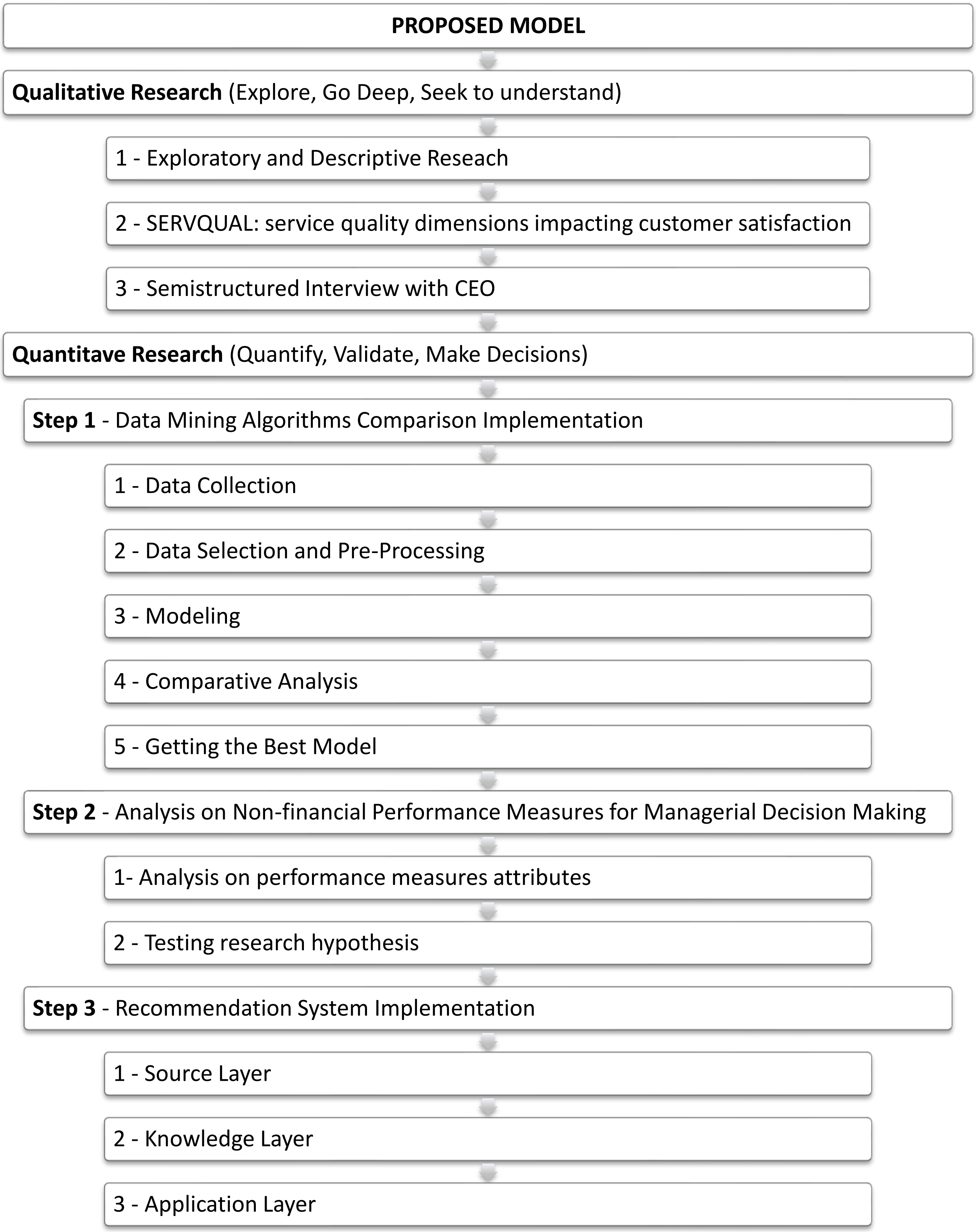

Figure 21: Proposed Model

Qualitative case study research, as a research methodology, by semi-structured interview conducting, has the aim to understand and identify the weak areas of performance and challenges of the home appliances after-sales service business in Kosovo market, and the intention of the study is to evaluate and measure service quality impacting customer satisfaction, a nonfinancial performance measure for managerial decision-making, through SERVQUAL dimensions.

Quantitative research study comprises three steps.

Step 1 of quantitative research provides an analysis and practical comparative study of the application of different classification and clustering algorithms in a home appliance after-sales service business. The main phases of this step are data collection, pre-processing, modelling, comparative analysis and getting the best model.

Step 2 of quantitative research provides regression analysis on service quality and customer satisfaction, nonfinancial performance measures for managerial decision-making, and testing research hypotheses. The main phases of this step are data collection and selection, pre-processing, reliability test and correlation analysis, simple and multiple regression analysis.

Step 3 of quantitative research provides analysis and implementation of recommendation system, as a portion of data mining system functionalities, which will be used to predict and digitize manual processes of data presentation for the after-sales service business. The main layers of the recommendation system architecture include application layer, knowledge layer and source layer.

4 EXPERIMENTS AND RESULTS

The performed experiments and results provide the analysis and results of comparative research on the application of clustering and classification algorithms, analysis on service quality attributes and examine impact of service quality on customer satisfaction, a nonfinancial performance measure for managerial decision-making, and the implementation of recommendation system in the home appliance after-sales services field, and provide guidance for businesses aiming to improve decision-making along the digital transformation journey.

Step 1 provides experiments and results from the qualitative case study approach, semi-structured interview conducting, that evaluates and measures service quality impacting customer satisfaction, a nonfinancial performance measure for managerial decision-making, through SERVQUAL dimensions by covering after-sales service attributes. As a result, this study aims to outcome models that can help top management and decision makers to formulate proper strategies to improve the business performance and enhance customer satisfaction.

Step 2 provides experiments and results from the comparative research on the implementation of the clustering and classification techniques in a home appliance after-sales service business. The results can enhance decision-making and increase efficiency due to business prediction analysis. Furthermore, quantitative research provides analysis on service quality attributes and examine impact of service quality on customer satisfaction, a nonfinancial performance measure for managerial decision-making. Furthermore, Step 2 provides implementation in recommendation system, as a portion of data mining system functionalities, which is used to predict and digitize manual processes of data presentation for the after-sales service business.

4.1 The challenges and effective solutions to the after-sales service businesses

As a result, based on the studies and qualitative research performed, we produce recommendations that can help top management and decision makers solve different situations they cope with to improve business performance and financial performance measures. The recommendations comprise effective solutions to home appliances after-sales service businesses.

4.1.1 The challenges faced by the after-sales service businesses

The intention of the study is to evaluate and measure service quality impacting customer satisfaction, a nonfinancial performance measure for managerial decision-making, through SERVQUAL dimensions by covering after-sales service attributes.

Respondents state few challenges that affect customer satisfaction. The respondents assert that 1) responsiveness to customer complaints, 2) time taken for resolving the complaint, 3) info and guidance available at service center, 4) affordable cost of service, 5) complaint reporting facilities, 6) supply of required spare parts and service delivery as promised, 7) reasonable warranty policy, 8) comprehending customers' needs, 9) options and variety of services, 10) accessibility of service center, 11) technical skills of personnel, and 12) consistency of service quality are attributes that are considered challenging to qualitatively serve customers.

Table 13 shows in tabular format the challenges faced by the home appliances after sales businesses to serve customers.

Table 13: What are some of the challenges we find solvent or hard to overcome?

SERVQUAL Dimension	**Challenges**
Responsiveness to customer complaints	Digital responsivity to complaints
Time taken for resolving the complaint	- Long period of time to repair the product - Peak periods in which the repair load is enormous or unexpected as a result of various situations and conditions
Info and guidance available at the service center	Lack of online resources increases number of complaints
Affordable cost of service	- Long term guarantee of products causes customers to buy new products instead of costly repairs of low-quality products with short life cycles or high-quality products with long life cycles - Differences on servicing cost/price between the authorized services and informality in the market
Complaint reporting facilities	Practice of traditional registration processing of complaints
Supply of required spare parts, Service delivery as promised	Provision of missing spare parts and end of life of spare parts
Reasonable warranty policy	Reasonable guarantee policy that affects guarantee period only for some brands and categories
Comprehending customers' needs	Lack of understanding the preferences and expectations
Accessibility of service center	In case the products need to be repaired in a service center, logistics or transportation from deep zones is a challenge
Technical skills of personnel	Kosovo economic and market instability cause employee leave, in particular technicians leave
Options and variety of services	Lack of information by customers on options and variety of services for the products with expired guarantee offered by an authorized after-sales service business causes reduction or lack of customers
Consistency of service quality	Informality in the (black) market causes the reduction or lack of customers

4.1.2 Strategizing effective solutions to increase customer satisfaction and service quality

The contributions are the outcomes of qualitative research which enables strategizing effective solutions for the after-sales services businesses.

A related work as an approach is performed and used in order to propose effective solutions in accordance with the issues and challenges stated that service quality impacts customer satisfaction.

Challenge - Digital responsiveness to customer complaints.

Parida et al. (2019) and Rudnick et al. (2020) consider digitalization, real-time applications and adoption of new technologies promotes linkages between customers and business. Moreover, Silva and Lima (2018), Meiryani (2019) and Rudnick et al. (2020) elaborate how the sophisticated any real-time case-management systems can improve operation and facilitate the accuracy of data and statistical reporting. Thereupon, the main advantage of the system is the digitization of retrieving day-to-day tasks data and of database population with case status.

Challenge - Long period of time to repair the product.

Saccani et al. (2006) empowers the role and influence of information systems in creating an effective service organization. Furthermore, Pagalday et al. (2018) state that, due to technology, machines can predict, monitor, and extract information for analysis, decision-making, and prompt action. Lele (1997) propose to use framework to predict product and service strategies in response to evolving customer needs.

Lastly, Cascio and Montealegre (2016) assert that the relationships among counterparts are important to the success of industrial societies. Moreover, new infrastructure technologies make it possible control, coordinate, and collaborate on activities more easily among the counterparts.

Challenge - Lack of online manuals increases the number of complaints and as a result affects customer satisfaction.

Murali et al. (2015) consider that the after-sales service business delivers high performance by providing manuals with concise directives and giving proper information and guidance about the features and functions of the product to the customer.

Challenge - Long term guarantee of products causes customers to buy new products instead of costly repairs of low-quality products with short life cycles or high-quality products with long life cycles.

Laitala et al. (2020) state that the phenomena of buying new products instead of repairing broken appliances are impacted by intensive advertisement of new products and better economic situations, GDP. In this context, Domazet et al. (2018) propose design efficient after-sales service marketing strategies by representing the length of the warranty term and the geographic coverage of the service network, as well as the price of the new item.

Challenge - Practice of traditional registration processing of complaints by using relationships or power of their position in society/work in order to accelerate the process.

Murali (2016) and Camilleri (2017a) point out that designing efficient customer-centric marketing strategies is essential to promote modern services and accessibility over traditional services and accessibility; Frey et al. (2017) refer that information technology and digital platforms are essential to drive change and innovation.

Challenge - Provision of missing spare parts and end of life of spare parts.

Chari et al. (2012), in their paper, emphasize a business must provide spare parts throughout the guarantee period in order to stay competitive, and a business must provide customers with spare parts for the full length of the warranty period after the product has been outdated. Under these conditions, the authors consider negotiating and carefully determine the guarantee period to offer with products at the time of setting agreements with manufacturers and retailers. Accordingly, replacing the spare parts at the time of repairing without delay is a sign delivering good performance.

Challenge - Reasonable guarantee policy that affects guarantee period only for some brands and categories.

The authors, Chari et al. (2012), consider to carefully determine the guarantee period to offer with products and negotiate the warranty policy in the time of setting the agreements with manufacturers and retailers.

Challenge - Lack of understanding the preferences, and expectations.

Camilleri (2017a, 2017b) highlights that in order to remain successful in competitive marketplace businesses, one needs to analyze and identify customers' behavior to gain competitive advantage by understanding customer needs, preferences and expectations.

Challenge - Lack of information from customers on options and variety of services for the products offered by an authorized after-sales service business causes a reduction or lack of customers.

Camilleri (2017b) highlights the need to analyze customer needs and preferences in order to remain successful in competitive businesses. Camilleri (2017a) considers designing efficient customer-centric marketing is essential for a business in the context where businesses understand their customers' needs and preferences and inform customers on options and variety of services for the products they offer. Moreover, Kang and Hustvedt (2013), the authors consider it is very important to build trustable linkage between the business and customers. Thereupon, trust is a customer's consideration that a business will perform in the best interests of its own customers.

Challenge - In case the products need to be repaired in a service center, logistics or transportation from deep zones is a challenge.

Altekin et al. (2017) propose better planning or increased capacities for faster transport of products in the field of after-sales service. Advanced logistics systems improve fast delivery of service parts, maintenance, and product repair services.

Challenge - Peak periods in which the repair load is enormous or unexpected as a result of various situations and conditions.

Enshassi (2008), in his/her paper, elaborates how main contractor use subcontractor to execute specific operation throughout different situations and to execute according to the required specific activity and complexity of the work in order to respond to customer complaints and reduce the work pressure on the main contractor. Moreover, Hammermann and Mohnen (2012) tackle the way how treats overtime work and study how the monetary compensation and incentives help to overcome problems, build effective relationships between employees and employers and effect employees' satisfaction.

Challenge - Kosovo economic and market instability cause employee leave, in particular technicians leave.

Murali et al. (2016) affirm that technical skills and interpersonal behavior of those in service have a significant impact on customer satisfaction. Moreover, ILO (2010) recommends connecting with training providers who offer qualified/skilled technicians with skills trained and creating strategies to improve the technical and professional skills of their service staff. Another solution is demonstrated by Mydyti and Kadriu, (2020) to offer internship which will facilitate the process of finding talent that matches the internal culture.

Challenge - Informality in the (black) market causes the reduction or lack of customers.

The author, Ulyssea (2020), describes non-formal businesses as those which are non-compliant with regulations e.g., by not registered in the tax administration. The most productive policy to decrease informality is considered to be intensifying enforcement of laws and regulations, and as a result the government is increasing the formality for official small businesses. Moreover, Oviedo et al. (2009) elaborate how governments are performing many activities to ease the alteration of businesses and individuals from informality to formality.

Challenge - Differences in servicing cost/price between the authorized services and informality in the market.

In their paper, Hinchliffe et al. (2020), recommend steps to take on building partnerships such as through mechanisms that give the financial incentives to the informal market. Moreover, the authors recommend involving joint workshops with the informal sector and as a result agree on getting fair prices and a pricing mechanism which is accepted by the informal sector.

4.2 Data Mining Comparison Implementation

4.2.1 Clustering and Classification Algorithms Comparison Implementation

The objective of this research is to analyze the comparative study of how data clustering and classification algorithms are applied in a home appliance after-sales service business.

Figure 25 provides a graphical presentation of the steps of the clustering algorithms workflow.

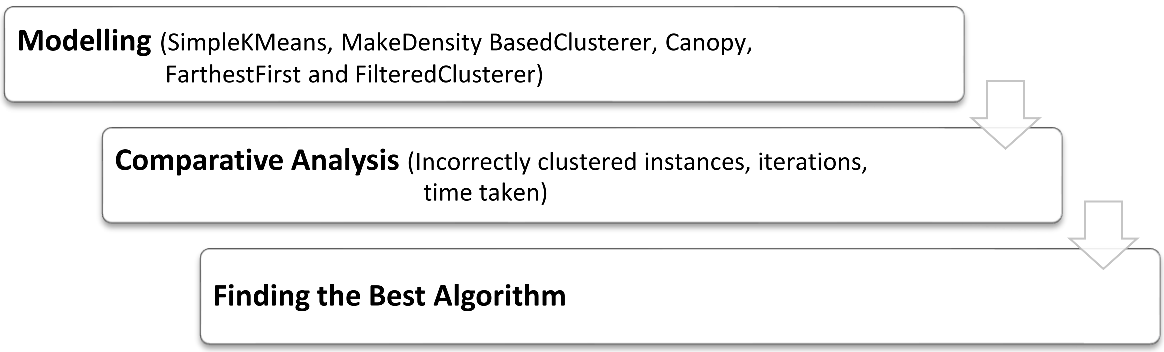

Figure 22: The steps of clustering algorithms comparison implementation

Figure 26 provides a graphical presentation of the classification algorithms comparison workflow.

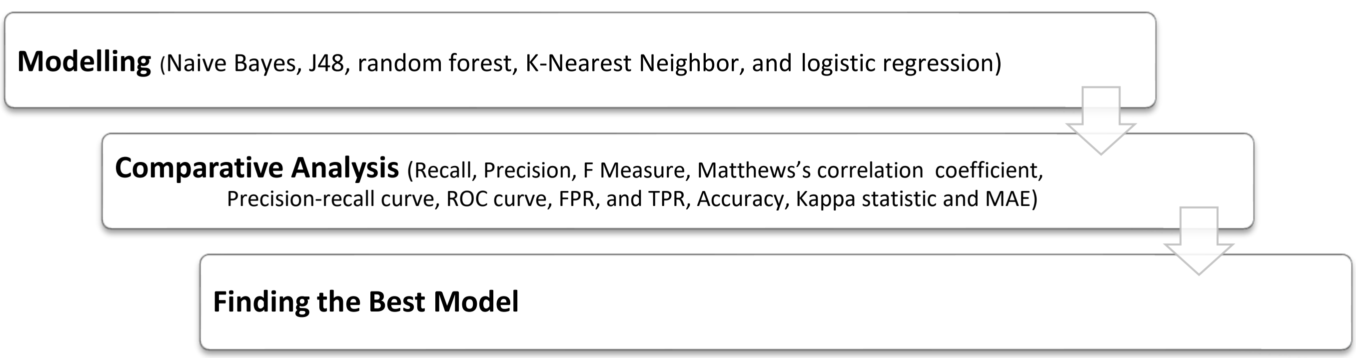

Figure 23: The steps of classification algorithms comparison implementation

Moreover, we will provide one model and the results of this study guide businesses aiming to deliver better decision-making and optimized strategic objectives.

4.2.1.1 Clustering - Results

We conducted an analysis with five clustering algorithms, i.e., SimpleKMeans, MakeDensity BasedClusterer, Canopy, FarthestFirst and FilteredClusterer. In the five algorithms, the output is generated according to similar objects and the time needed to generate these clusters. The chosen classification option is the classes to cluster evaluation. Furthermore, the chosen option is the guarantee attribute for the algorithms to be evaluated. Table 14 presents the findings of various clustering techniques.

Table 14: Results from Different Clustering

Algorithm	No. of Clusters	No. of Iteration	Log Likelihood	Sum of squared errors
Canopy	25	-		
SimpleKMeans	2	2		30809.5
MakeDensity BasedClusterer	2	2	-12.3	30809.5
FarthestFirst	2	-		
FilteredClusterer	2	2		30809.5

SimpleKMeans, MakeDensityBasedClusterer, FarthestFirst and FilteredClusterer algorithm, different from the Canopy algorithm, have the same clustered instances. In terms of likelihood, only MakeDensityBasedClusterer denotes a value that provides more reliable clusters.

Table 15: Performance Comparison - Clustering

Algorithm	Incorrectly clustered instances (%)	Accuracy (%)	Time taken to build (seconds)
Canopy	53.18	46.82	0.28
SimpleKMeans	47.30	52.70	0.38
MakeDensity BasedClusterer	47.29	52.71	0.58
FarthestFirst	11.54	88.46	0.22
FilteredClusterer	47.30	52.70	0.23

Table 15 shows the incorrectly clustered instances, the percentage accuracy and the duration of each experiment. Canopy, SimpleKMeans, FilteredCluterer, MakeDensity BasedClusterer algorithms takes more time to build a cluster and have high incorrectly clustered instances; that is why the FarthestFirst algorithm outperforms other algorithms. If the value of incorrectly clustered instances is high, it does not make a good cluster. As a result, FarthestFirst is the best algorithm as it takes considerably less time to construct a model and has a lower level of incorrectly clustered instances. The Figure 27 graphically presents the results of the comparison of clustering algorithms.

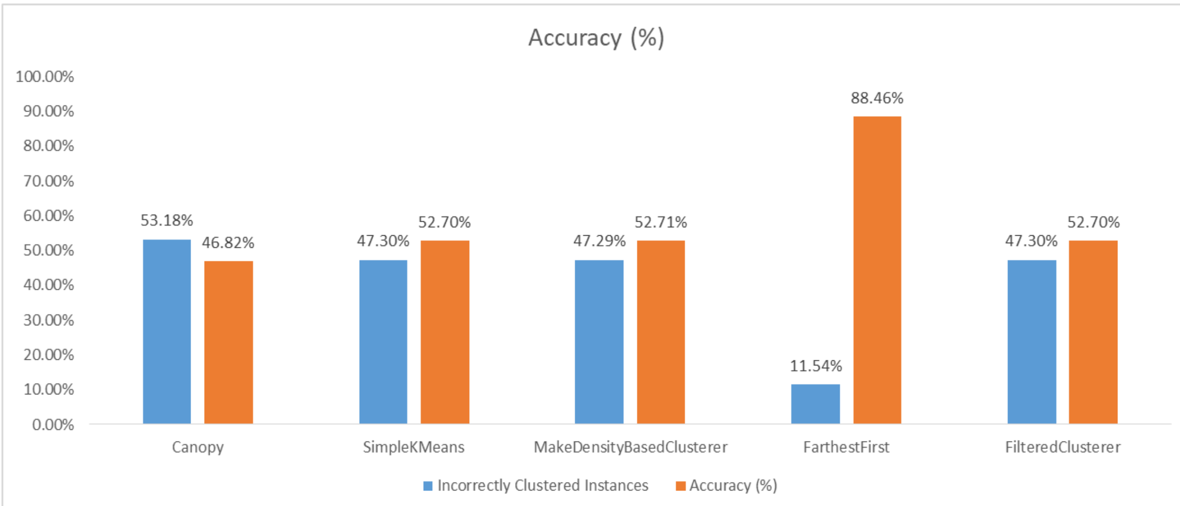

Figure 24: Findings on Clustering Algorithms

The best clustering results show that the prediction is categorized as binary because of the class attribute has two output values, within which product guarantee can be differentiated with or without guarantee (Figure 28).

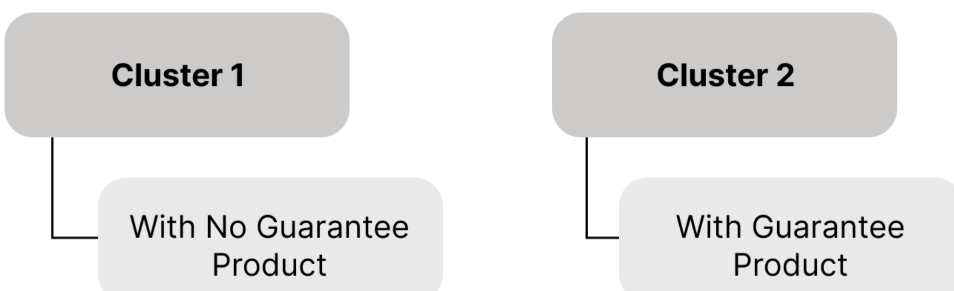

Figure 25: Key generated clusters

Lastly, the best algorithm identified is FarthestFirst clustering. It takes less time and is the most accurate, lower level of poorly clustered instances than other clustering algorithms to uncover similar clusters for the customer care service dataset. The result achieved helps to make the decision as to which algorithm is suitable for the company. As a consequence, the FarthestFirst clustering algorithm delivers useful insights for the company and the information generated by the clustering is suitable for use in all operational functions specially to bring a cost-effective consumer response.

4.2.1.2 Classification - Results

The growing demand of classification-based data mining in the area of home appliances after-sales service has demonstrated the demand to practice data mining techniques for enhanced decision-making. The findings and discussions related to the comparison of the classification techniques are detailed below.

The primary goal for comparing algorithms is to identify the right algorithm to address or solve business issues. Five classification algorithms have been compared to find the most suitable classification algorithm. The dataset includes 62k instances, and seven attributes and one class attribute (Guarantee attribute). The selected test option is 10-fold cross-validation, which means the dataset is split into 10 sections. The first nine sections are used for algorithm training, and the tenth section is used for algorithm evaluation. Cross validation is a valuable feature for fine-tuning the data mining model. Comparison analysis of classification measures involving recall, precision, F Measure, Matthews's correlation coefficient (MCC), precision-recall curve (PRC), ROC curve, FPR, and TPR have been extracted.

Table 16: Performance measurement results (comparison of classification algorithms)

Algorithm	Performance Measure							
	TPR	FPR	Precision	Recall	F-measure	MCC	Area	
							ROC	PRC
J48	0.953	0.288	0.950	0.953	0.951	0.709	0.859	0.940
RandomForest	0.949	0.328	0.946	0.949	0.946	0.680	0.911	0.957
IBk	0.946	0.359	0.943	0.946	0.943	0.660	0.882	0.948
Logistic	0.914	0.773	0.901	0.914	0.887	0.305	0.833	0.924
NaïveBayes	0.905	0.785	0.876	0.905	0.879	0.223	0.797	0.905

Comparing the classification techniques presented in Table 16 summarizes the simulation results from five classifiers applied and their related performance measurements. It shows that J48, RandomForest and K-Nearest Neighbour (IBk at Weka) achieve the best results. The other useful statistical features, receiver operating characteristic (ROC) area, where the value must be greater than 0.5 of all the algorithms, proves the presence of a statistical dependency. The two most common indicators are the overall accuracy and the Kappa Coefficient. The implementation of the J48 technique is suitable for the business, accordingly by adding value to business in streamlining decision-making.

Table 17: Classifier algorithm results

Algorithm	Accuracy (%)	Kappa Statistic	MAE
J48	95.2604	0.7071	0.0844
RandomForest	94.8711	0.6755	0.0787
IBk	94.6373	0.6536	0.0789
Logistic	91.3985	0.2207	0.1334
Naïve Bayes	90.4935	0.178	0.1253

The findings of the classification techniques shown in Table 17 summarize the findings of the five classifiers regarding accuracy, Kappa statistic and mean absolute error (MAE). It indicates J48, RandomForest and IBk as the most accurate algorithms amongst others. Logistic and Naïve Bayes algorithms provide the weakest accuracy. In addition, it indicates that J48, RandomForest and IBk yield the smallest error amongst the other techniques investigated. The Naïve Bayes yields the greatest error. The J48 Kappa statistic has a value of 0.7071, indicating the presence of a statistical dependency and is elevated. The accuracy value of J48 is 95.2604%, demonstrating the high percentage of correctly classified cases.

Table 18: MAE, RMSE, RAE and RRSE Metrics

Algorithm	MAE	RMSE	RAE	RRSE
J48	0.0844	0.2058	49.0169 %	70.1451 %
RF	0.0787	0.208	45.7368 %	70.8913 %
IBk	0.0789	0.2187	45.8332 %	74.5325 %
Logistic	0.1333	0.2536	77.4472 %	86.4468 %
Naïve Bayes	0.1253	0.2819	72.7974 %	96.0723 %

Table 18 shows that the J48 algorithm is less prone to errors than other algorithms. The indicated errors are mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE).

The model is according to its predictive yield. J48 algorithm is the finest using the MAE, RMSE, RAE and RRSE, receiver operating characteristic area, accuracy, Matthews's correlation coefficient, precision-recall curve, precision, F-measure, recall and statistical criteria. As a result, the J48 is proven to be the most effective, and can be applied to create the prediction model and predict repairs according to the product guarantee period. In addition, all the algorithm details are developed visually using outputs from the application of various binary classifiers. The Figure 29 graphically presents the results of the comparison of classification algorithms.

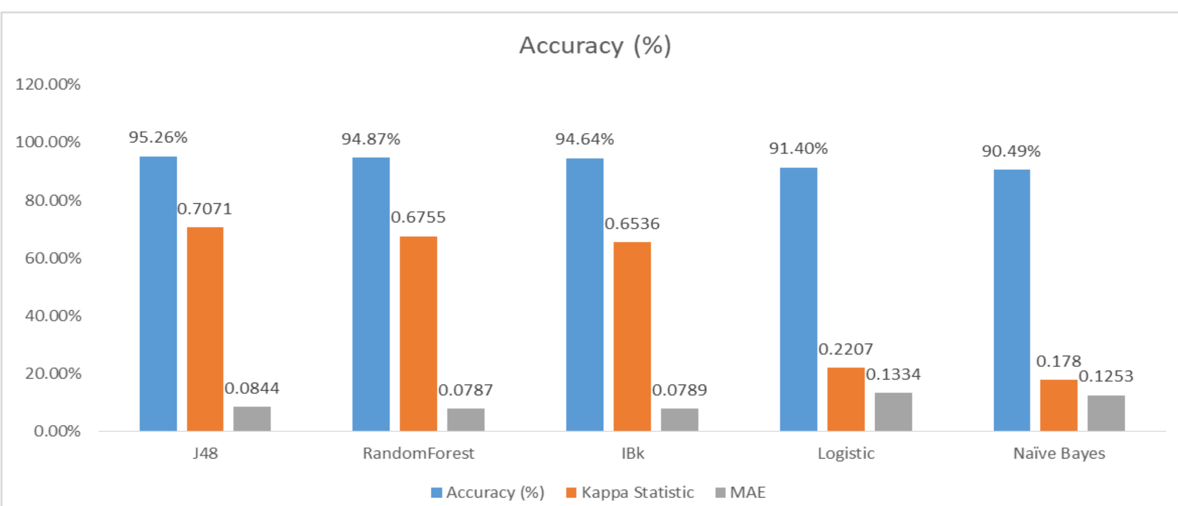

Figure 26: Findings on Classification Algorithms

In Figure 30, the findings demonstrate effectiveness of the Random Forest algorithm in predicting the repairs according to the guarantee period of the product. The model was 94.8791% accurate. The training dataset confusion matrix demonstrates the accuracy of the predicted class. Training data is composed of 62,431 cases - with guarantee 56,492 and without guarantee 5,939. Consequently, the produced tree classified 59,234 cases properly over 62,431 cases.

```
Correctly Classified Instances       59234               94.8791 %
Incorrectly Classified Instances      3197                5.1209 %
Kappa statistic                          0.676
Mean absolute error                      0.0787
Root mean squared error                  0.208
Relative absolute error                 45.7368 %
Root relative squared error             70.8913 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.981    0.361    0.963      0.981    0.972      0.680    0.911     0.984     Yes-guarantee
                 0.639    0.019    0.783      0.639    0.704      0.680    0.911     0.696     No-guarantee
Weighted Avg.    0.949    0.328    0.946      0.949    0.946      0.680    0.911     0.957

=== Confusion Matrix ===

     a     b   <-- classified as
 55438  1054 |     a = Yes-guarantee
  2143  3796 |     b = No-guarantee
```

Figure 27: Results generated by the Random Forest binary classifier application

In figure 31, the findings demonstrate effectiveness of the IBk algorithm in predicting the repairs according to the guarantee period of the product. The model was 94.6373% accurate. The training dataset confusion matrix demonstrates the accuracy of the predicted class. Training data is composed of 62,431 cases - with guarantee 56,492 and without guarantee 5,939. Consequently, the produced tree classified 59,083 cases properly over 62,431 cases.

```
Correctly Classified Instances       59083               94.6373 %
Incorrectly Classified Instances      3348                5.3627 %
Kappa statistic                          0.6536
Mean absolute error                      0.0789
Root mean squared error                  0.2187
Relative absolute error                 45.8332 %
Root relative squared error             74.5325 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.982    0.395    0.959      0.982    0.971      0.660    0.882     0.981     Yes-guarantee
                 0.605    0.018    0.782      0.605    0.682      0.660    0.882     0.640     No-guarantee
Weighted Avg.    0.946    0.359    0.943      0.946    0.943      0.660    0.882     0.948

=== Confusion Matrix ===

     a     b   <-- classified as
 55487  1005 |     a = Yes-guarantee
  2343  3596 |     b = No-guarantee
```

Figure 28: Results generated by the IBk binary classifier application

In Figure 32, the findings demonstrate effectiveness of the Logistic algorithm in predicting the repairs according to the guarantee period of the product. The model was 91.4001% accurate. The training dataset confusion matrix demonstrates the accuracy of the predicted class. Training data is composed of 62,431 cases - with guarantee 56,492 and without guarantee 5,939. Consequently, the produced tree classified 57,062 cases properly over 62,431 cases.

```
Correctly Classified Instances       57062               91.4001 %
Incorrectly Classified Instances      5369                8.5999 %
Kappa statistic                          0.2209
Mean absolute error                      0.1333
Root mean squared error                  0.2536
Relative absolute error                 77.4472 %
Root relative squared error             86.4468 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.995    0.853    0.917      0.995    0.954      0.306    0.833     0.964     Yes-guarantee
                 0.147    0.005    0.742      0.147    0.245      0.306    0.833     0.547     No-guarantee
Weighted Avg.    0.914    0.772    0.901      0.914    0.887      0.306    0.833     0.925

=== Confusion Matrix ===

     a     b   <-- classified as
 56189   303 |     a = Yes-guarantee
  5066   873 |     b = No-guarantee
```

Figure 29: Results generated by the Logistic binary classifier application

In figure 33, the findings demonstrate effectiveness of the Naïve Bayes algorithm in predicting the repairs according to the guarantee period of the product. The model was 90.4935% accurate. The training dataset confusion matrix demonstrates the accuracy of the predicted class. Training data is composed of 62,431 cases - with guarantee 56,492 and without guarantee 5,939. Consequently, the produced tree classified 56,496 cases properly over 62,431 cases.

```
Correctly Classified Instances       56496               90.4935 %
Incorrectly Classified Instances      5935                9.5065 %
Kappa statistic                          0.178
Mean absolute error                      0.1253
Root mean squared error                  0.2819
Relative absolute error                 72.7974 %
Root relative squared error             96.0723 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.986    0.866    0.915      0.986    0.949      0.223    0.797     0.966     Yes-guarantee
                 0.134    0.014    0.501      0.134    0.211      0.223    0.797     0.328     No-guarantee
Weighted Avg.    0.905    0.785    0.876      0.905    0.879      0.223    0.797     0.905

=== Confusion Matrix ===

     a     b   <-- classified as
 55703   789 |     a = Yes-guarantee
  5146   793 |     b = No-guarantee
```

Figure 30: Results generated by the Naïve Bayes binary classifier application

In Figure 34, the findings demonstrate effectiveness of the J48 algorithm in predicting the repairs according to the guarantee period of the product. The model was 95.2604% accurate. The training dataset confusion matrix demonstrates the accuracy of the predicted class. Training data is composed of 62,431 cases - with guarantee 56,492 and without guarantee 5,939. Consequently, the produced tree classified 59,472 cases properly over 62,431 cases. These findings suggest that of the five algorithms tested, the J48 classifier can greatly improve the classification methods for application in home appliances after-sales services field.

```
Correctly Classified Instances       59472               95.2604 %
Incorrectly Classified Instances      2959                4.7396 %
Kappa statistic                          0.7071
Mean absolute error                      0.0844
Root mean squared error                  0.2058
Relative absolute error                 49.0169 %
Root relative squared error             70.1451 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.981    0.316    0.967      0.981    0.974      0.709    0.859     0.973     Yes-guarantee
                 0.684    0.019    0.790      0.684    0.733      0.709    0.859     0.620     No-guarantee
Weighted Avg.    0.953    0.288    0.950      0.953    0.951      0.709    0.859     0.940

=== Confusion Matrix ===

     a     b   <-- classified as
 55412  1080 |     a = Yes-guarantee
  1879  4060 |     b = No-guarantee
```

Figure 31: Results generated by the J48 binary classifier application

Figure 35 presents the J48 visualization. The resulting decision tree comprises repaired by age, authorized service, and sum, as root, and class for a leaf node, which is the product guarantee. The leaf node was introduced by rectangle and root nodes were introduced by an oval. The prediction is categorized as binary (two classes) classification due to the class attribute has two output values, within which product guarantee can be differentiated with or without guarantee. This case study addressed a gap and built a binary classification model (decision tree) to predict the potential repairs based on the product guarantee period.

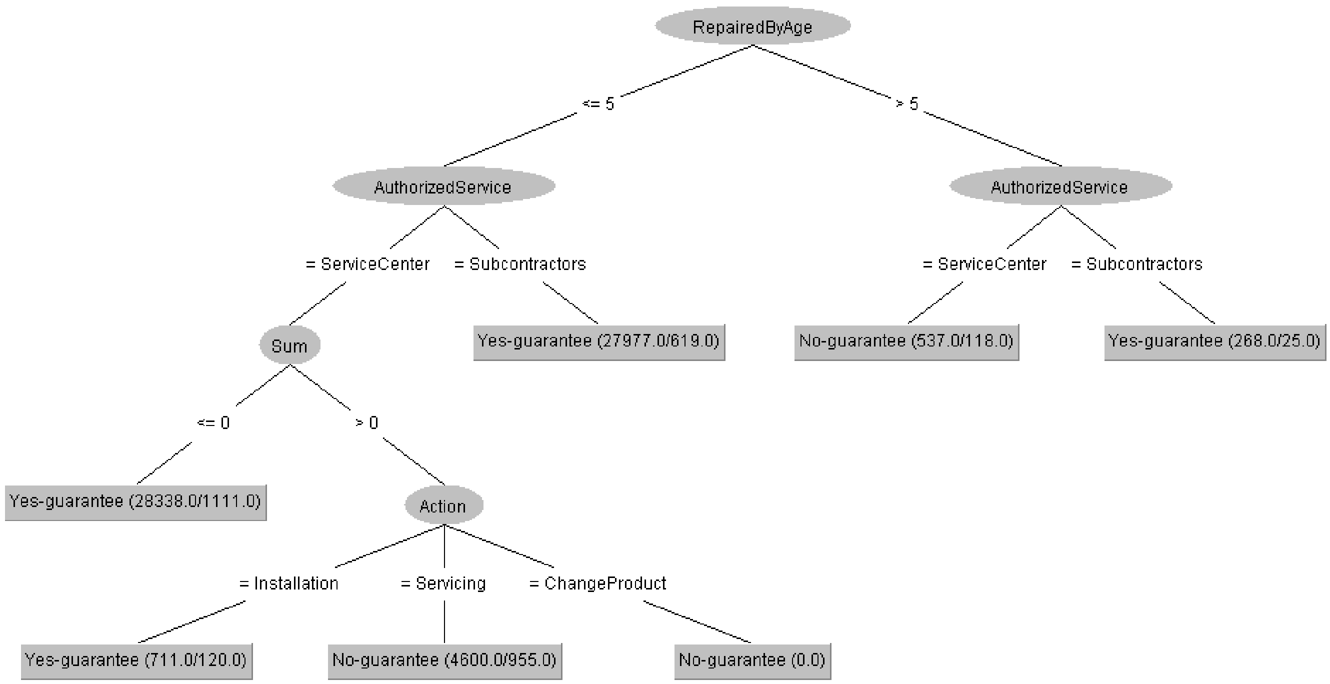

Figure 32: Decision Tree (J48) visualization

The J48 predicts that the subcontractors provide services and repair the products under guarantee. The authorized company requires the services of subcontractors for the products with a guarantee when the load of repairs is enormous or unforeseen because of various circumstances and conditions. The authorized service is required to provide services for the products under guarantee, and therefore requires the assistance of subcontractors.

In addition, if the value of the age of the products is equal or less than five years and with sum (payment) zero, the decision tree predicts the repairs of products with a guarantee. If the value of the age of the products is greater than five years, the decision tree predicts the repairs of products without guarantee.

Moreover, if the value of the age of the products is equal or less than five years and the value of the sum (payment) is greater than zero, the decision tree predicts the repairs of products without guarantee and predicts the installation of products under guarantee. Installing a product is an extra service, which is paid for, and typically the product is under guarantee.

The model anticipates the repairs according to the guarantee period. The longer the product guarantee period, the greater the need for servicing. The high-predicted value of the age of the repairs under guarantee is not an indicator of benefit. As a general rule, the service activities of home appliances are independent of the sales activities. Therefore, the interest of after-sales services businesses is not to have a longer guarantee period as this affects profit and customer experience. The first recommendation is that agreements with manufacturers and dealers should be based on the results of the predictive model.

In addition, the business can provide a service to the customers who own the products with expired guarantee. The second results of the prediction model indicate low servicing for expired guarantee products. Herein, the model's second result points to the need to improve customer experience to improve services by an authorized after-sales services business. The second recommendation is that the business can notify its own customers with expired-guarantee products that can obtain the service of an authorized after-sales services business.

As a result, the binary prediction model can be used to predict product repairs in home appliances after-sales service business. The model depicts the relationships between the variables and the result. This binary prediction model helps decision-makers in identifying gaps and in making the right decisions. Moreover, the recommendations aim to improve business services, decision-making processes and customer experience of the home appliances after-sales service business along the business digital transformation journey.

4.3 Regression Analysis on Service Quality Attributes and Impact on Customer Satisfaction

Through the regression analysis, hypothesis two "H2: Service quality impacts customer satisfaction in after-sales service business" have been tested and proved.

We studied and implemented data mining comparison techniques in a case study and the conclusions extracted will help to extend the study with the intention of understanding and identifying important aspects in offering qualitative services in after-sales services businesses.

Accordingly, the binary predictive model, based on the results obtained from the data mining, comparison implementation, foresees the repairs in the relation to the guarantee period and consequently the study extends and focuses only with services of products in guarantee.

4.3.1 Reliability test

Reliability test is performed by measuring Cronbach's alpha coefficient. Cronbach (1951) estimates that the range of the alpha coefficient as a minimum of 0.7. The formula below presents the Cronbach's alpha formula (Eq. 1).

$$\alpha = \frac{Nc}{v + (N-1)c} \tag{1}$$

N = number of items, c = average inter-item covariance among the items; c= average variance

We conducted this test to measure how closely related a set of variables are as a group. This indicates the reliability of the scale. The reliability is used and assessed by the method of Cronbatch's alpha. The Cronbach's alpha reliability of the five items (variables) is 0.7054, suggesting that the items (variables) are relatively consistent internally.

Table 19: Cronbach's Alpha value (1)

Research variables	**Cronbach's Alpha value**
BrandwithSpareparts, Payment, TimeToRepair, AuthorizedService and CustomerSatisfaction	**0.7054**

Table 19 presents the Cronbach Alpha value between customer satisfaction and other four independent variables such as time to repair variable, brand with spare parts, payment, and authorized service. The reliability test presented by measuring Cronbach's Alpha is acceptable and indicates reliability.

4.3.2 Correlation analysis

We conducted this analysis to examine the correlation between a dependent variable (customer satisfaction) and independent variables (brand with spare parts, time to repair, authorized service and payment variables).

The following figure presents the correlation coefficients for the independent variables to the dependent variable.

```
. pwcorr customersatisfaction authorizedservice brandwithspareparts weekstorepair payment, star(0.05) sig

             | custom~n author~e brandw~s weekst~r  payment
-------------+---------------------------------------------
customersa~n |   1.0000
             |
             |
authorized~e |  -0.2990*  1.0000
             |   0.0000
             |
brandwiths~s |   0.1932* -0.3619*  1.0000
             |   0.0000   0.0000
             |
weekstorep~r |  -0.8512*  0.1625* -0.1449*  1.0000
             |   0.0000   0.0000   0.0000
             |
     payment |   0.2404* -0.7632*  0.3061* -0.1378*  1.0000
             |   0.0000   0.0000   0.0000   0.0000
             |
```

Figure 33: Results of the correlation analysis between variables

In all cases, there is a positive correlation between customer satisfaction and other independent variables. Correlations are based on all 52K observations. Correlations estimate the strength and direction of the linear relation among the five variables.

The Pearson correlation coefficient (-0.8512) indicates an inverse relationship between time to repair variable and customer satisfaction variable, when time to repair values increase the customer satisfaction variable decreases. These two variables, time to repair attribute and customer satisfaction, indicate strong and highest correlation.

The next Pearson correlation coefficients are (0.1932), (-0.2990) and (0.2404) and indicates a relationship between other variables and customer satisfaction variable. These three variables and customer satisfaction, have the lowest correlation compared to time to repair variable and customer satisfaction variable.

4.3.3 Multiple Regression Analysis

The equation (Eq. 2) presents the customer satisfaction as a function of time to repair the product, brand with spare parts, authorized service and payment:

$$\text{Customer satisfaction} = f(\text{TimeToRepair, Brand with Spare Parts, Payment, Authorized Service}) \tag{2}$$

Customer satisfaction according to equation (Eq. 2) is presented as an outcome variable. Time to repair the product, brand with spare parts, authorized service, payment are predictor variables.

Figure 37 presents the multiple regression analysis between customer satisfaction and other four independent variables such as time to repair variable, brand with spare parts, payment and authorized service.

. reg customersatisfaction authorizedservice brandwithspareparts weekstorepair payment, beta

Source	SS	df	MS
Model	5331.08467	4	1332.77117
Residual	1764.45355	52165	.033824471
Total	7095.53822	52169	.136010624

Number of obs	=	52170
F(4, 52165)	=	39402.57
Prob > F	=	0.0000
R-squared	=	0.7513
Adj R-squared	=	0.7513
Root MSE	=	.18391

customersatisfact~n	Coef.	Std. Err.	t	P>|t|	Beta
authorizedservice	-.1172033	.0025552	-45.87	0.000	-.1588345
brandwithspareparts	.0145283	.0021165	6.86	0.000	.0161664
weekstorepair	-.4632308	.0012512	-370.22	0.000	-.8229492
payment	.000626	.0025493	0.25	0.806	.0008307
_cons	1.480191	.0034164	433.26	0.000	.

Figure 34: Results of the regression analysis between five variables

Based on the results of the multiple regression analysis, an equation (Eq. 3) with four independent variables is formulated as follows:

$$\text{Customer satisfaction} = 1.480191 - 0.1172033 * \text{AuthorizedService} - 0.4632308 * \text{TimeToRepair} + 0.0145283 * \text{BrandwithSpareparts} + 0.000626 * \text{Payment} \tag{3}$$

T statistic, for the TimeToRepair variable is (-370.22), for the BrandWithSpareparts variable is (6.86), for the AuthorizedService variable is (-45.87), are very highly significant. The t-values present the importance of TimeToRepair, BrandWithSpareparts and AuthorizedService variables in the model. The high t-values reject the null hypothesis according to which the coefficient differs from 0.

F statistic, 39,402.57, is highly significant and F-test confirms the conclusion of the t-test.

P-value, 0.000, of time to repair and brand with spare parts have a statistically significant relationship with customer satisfaction variable and statistically significant in explaining customer satisfaction. Moreover, we are 95% confident that predictor variables have an effect or an impact on customer satisfaction.

β_5: The customer satisfaction of a case increases by (0.000626) value if the customer pays the service for the product under warranty.

β_4: The customer satisfaction of a case increases by (0.0145283) value if the product belongs to the brand with spare parts.

β_3: The customer satisfaction of a case decreases by (0.4632308) value for an additional week increase in weeks repair the product.

β_2: The customer satisfaction of a case decreases by (0.1172033) value if the product repairs in service center against the subcontractor.

β_1: The intercept, 1.480191, means that each product repair within the first week of the repair and belongs to the brand without spare parts will have high possibility to be considered acceptable and satisfying by the customer.

$\mathbf{R^2}$, coefficient of determination, 0.7513, meaning TimeToRepair, BrandWithSpareparts, Payment and AuthorizedService variables explain only 75.13 % of the variance in customer satisfaction.

Root MSE (root mean squared error), 0.18391, is closer to zero and as a result is better the fit.

We have sufficient statistical evidence to reject the null hypothesis at 1%, 5%, 10% level of significance.

In a different way, the value calculated of the F statistics and R^2 through *ESS, TSS* and *RSS* obtained in the customer satisfaction equations (Eq. 4) and (Eq. 5) is as follows.

$$\text{F Statistics} = \frac{ESS/2}{RSS/(T-3)} = \frac{\Sigma\left(\hat{Y}_i - \bar{Y}\right)^2/4}{(\Sigma\left(Y_i - \hat{Y}\right)^2)/(\mathrm{T}-5)} = \frac{5{,}331.08467/4}{1{,}764.45355\ /52{,}165} = 3{,}9402.57 \tag{4}$$

$$R^2 = \frac{ESS}{TSS} = \frac{\Sigma\left(\hat{Y}_i - \bar{Y}\right)^2}{\Sigma(Y_i - \bar{Y})^2} = \frac{4{,}392.75487}{6{,}049.39786} = 0.7513 = 75.13\% \tag{5}$$

In a different way, the calculated value of the adjusted coefficient of determination (AdjR2) and the results obtained are presented in the equation Eq. 6 as below.

$$\mathrm{Adj}R^2 = 1 - \frac{n-1}{n-k}(1-R^2) = 1 - \frac{52169}{52165}(1-0.7513) = 1 - 0.2487 = 0.7513 \tag{6}$$

In a different way, the calculated value of the Root Mean Square Error (RootMSE) and results obtained are presented in the equation Eq. 7 as below.

$$\mathrm{RootMSE} = \sqrt{\frac{RSS}{(n-k)}} = \sqrt{\frac{5331.08467}{52165}} = 0.18391 \tag{7}$$

MSS - Model Sum of Squares; RSS - Residual Sum of Squares; TSS - Total Sum of Squares; k - #predictors; n - # of observations;

β, the standardized coefficient beta, for the TimeToRepair variable is (-0.8229492) and for the AuthorizedService variable is (-0.1588345), are very highly significant. The standardized coefficient beta is significant and is used in the regression equations to determine the presence and power of relations between customer satisfaction and time to repair variables.

Based on results obtained, the strong correlation and strong linear relationship between customer satisfaction and four independent variables is gained. The best predictor is the time taken for resolving the complaint.

Time taken for resolving the complaint (responsiveness) is one of the key attributes of service quality impacting customer satisfaction.

Consequently, we extract the following main hypothesis:

Hypothesis H2: Service quality impacts customer satisfaction in after-sales service business.

Figure 38 provides a graphical presentation for the hypothesis H2 extracted.

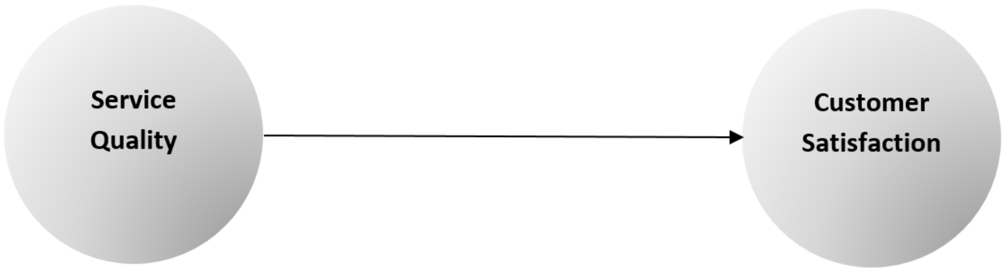

Figure 35: Time taken for resolving the complaint impact on customer satisfaction

4.3.4 Testing research hypothesis

We conducted this analysis to predict the customer satisfaction variable based on time taken for resolving the complaint variable. The linear regression model is used to test the hypothesis.

The basic regression equation Eq. 8 stands as:

$$\gamma = \beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + \xi \tag{8}$$

Y is an explained variable, X is an explanatory variable, ξ is a residual; β_1/β_2 assesses the impact of X_1/ X_2 upon *Y*. The constant term (β_0) measures Y when X_1 and X_2 are 0.

The relationship between customer satisfaction and the time to repair variable can be regarded as a mathematical function as presented in equation (Eq. 9) as follows:

$$\text{Customer satisfaction} = f(\text{TimeToRepair}) \tag{9}$$

Customer satisfaction is an outcome variable (*Y)*. Time to repair is a predictor variable (*X*).

Figure 39 presents the simple regression analysis between customer satisfaction and time to repair variable.

```
. reg customersatisfaction weekstorepair, beta

      Source |       SS       df       MS              Number of obs =   52170
-------------+------------------------------           F(  1, 52168) =       .
       Model |  5141.21163     1  5141.21163           Prob > F      =  0.0000
    Residual |  1954.32659 52168  .037462172           R-squared     =  0.7246
-------------+------------------------------           Adj R-squared =  0.7246
       Total |  7095.53822 52169  .136010624           Root MSE      =  .19355

------------------------------------------------------------------------------
customersat~n |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
weekstorepair |  -.4791423   .0012934  -370.46   0.000                -.8512166
        _cons |   1.451952   .0018623   779.67   0.000                        .
------------------------------------------------------------------------------
```

Figure 36: The regression between customer satisfaction and time to repair

From the linear regression analysis, the simple regression equation (Eq. 10) can be expressed in the following manner:

$$\text{Customer satisfaction} = 1.451952 - 0.4791423 * \text{TimeToRepair} \tag{10}$$

T statistic, - 370.46, is highly significant. The t-values presents the importance of a time to repair variable in the model.

P-value, 0.000, we have a statistically significant relationship, and we are 95% confident that time to repair have an effect on customer satisfaction. P-value of the model tests whether $\mathbf{R^2}$ is different from 0. A P-value below 0.05 denotes a significant relationship of *X* to *Y*.

β_2: Customer satisfaction of a case decrease by (- 0. 4791423) value for an additional week increase in time to repair the product.

β_1: The intercept, 1.451952, means that each product repair and service within the first week of the repair will have 97.28% possibility to be considered acceptable and satisfying by the customer.

$\mathbf{R^2}$, 0.7246, meaning time to repair variable explains only 72.46 % of the variance in customer satisfaction. $\mathbf{R^2}$ displays the amount of variance of *Y* explained by *X*.

α, 0.8424, meaning measured Cronbach's Alpha is acceptable and indicates reliability.

Table 20: Cronbach's Alpha value between customer satisfaction and time to repair

Research variables	**Cronbach's Alpha value**
Customer satisfaction and time to repair	**0.8424**

Table 20 presents the Cronbach Alpha value between customer satisfaction variable and time to repair variable. The reliability test presented by measuring Cronbach's Alpha is acceptable and indicates reliability.

Root MSE (root mean squared error), 0.19355, is closer to zero and as a result is better the fit. Root MSE presents the mean distance from the estimator to the mean.

The t statistic for the coefficient of time to repair, customer satisfaction is (- 370.46), highly significant. Thus, we have sufficient statistical evidence to reject the null hypothesis at 1%, 5%, 10% level of significance. It is a point performing a t test on the intercept, means that each product repair within the first week of the repair will have 98.86% possibility to be considered acceptable and satisfying by the customer.

In a different way, the value calculated of the F statistics and R^2 through *ESS, TSS* and *RSS* obtained in the customer satisfaction equations Eq. 11 and Eq. 12 is as follows.

$$\text{F Statistics} = \frac{ESS}{RSS/(T-2)} = \frac{\Sigma\left(\hat{Y}_i - \bar{Y}\right)^2}{\left(\Sigma\left(Y_i - \hat{Y}\right)^2\right)/(\mathrm{T}-2)} = \frac{5141.21163}{1954.32659/(52170-2)} = 137{,}237.414 \tag{11}$$

$$R^2 = \frac{ESS}{TSS} = \frac{\Sigma\left(\hat{Y}_i - \bar{Y}\right)^2}{\Sigma(Y_i - \bar{Y})^2} = \frac{5141.21163}{7095.53822} = 0.7246 = 72.46\% \tag{12}$$

The results indicate that the mean sum of squares of the model is about 137,237.414 times greater than that of the residual. Moreover, 72.46% of the variation in the dependent variable are explainable by the model.

β, (-0.8512), the standardized coefficient beta is significant.

Accordingly, the main hypothesis, H2: Service quality impacts customer satisfaction in after-sales service business, is accepted.

4.4 Recommendation System Implementation

Through quantitative research, the hypothesis three "*H3: Business strategic decision-making in after-sales service impacts after-sales service quality*" has been proved.

Data mining applications and systems are used as a means to improve the managerial decision-making in home appliances after-sales service business. We will provide a recommendation system, with the best classification algorithm implementation, a portion of data mining system functionalities which will serve as an opportunity in influencing decision-making.

One of the important requirements of the home appliances after-sales service business is to digitize manual processes of data presentation where it presents the prediction results in a manner that is easy to be interpreted. This recommendation system, as an outcome, will contribute to decision makers and management of after-sales service businesses and reflect positively by increasing efficiencies and because of the analysis of business prediction.

This section extends the study based on the conclusions of previous qualitative and quantitative research studies as follows:

1. Visually analyze and predict trends of services to ease decision-making
2. Practical comparative study of the implementation of the clustering algorithms: The prediction is categorized as binary (two classes) classification, where the guarantee of the product can be differentiated with or without guarantee
3. Practical comparative study of the implementation of the classification algorithms: The decision tree algorithm (J48) leads to the best results, and can be used to produce the prediction model and predict repairs according to guarantee period of the product
4. Study on service quality and customer satisfaction: Time taken for resolving the complaint (responsiveness) is one of the key attributes of service quality impacting customer satisfaction, a nonfinancial performance measure for managerial decision-making.

4.4.1 Source Layer

The chosen data is cleared of inconsistencies, inaccurate and missing data, and outlier values. The used attributes include brand (nominal), category (nominal), and weeks to repair (nominal). The weeks to repair is the response and class attribute. The dataset is utilized for exploring and studying with 55k cases with three attributes and a class attribute (weeks to repair product). The model is assessed according to its predictive accuracy performance. The model achieved 80.85% accuracy.

Table 21 presents the needed grouped attributes in the dataset.

Table 21: The structure of dataset for recommendation system implementation

Attributes	Description	Values
Weeks to Repair	The number of weeks during which the service is performed from the time requested	1, 2, 3 and 4
Brand	Brand refers to the manufacturer of the product with six different values.	Brand One, Brand Two, Brand Three, Brand Four, Brand Five, Brand Six
Category	The product category includes domestic appliances (MDA), small domestic appliances and personal care (SDAP), TV, heating, air conditioning, with five different values	Air Conditioning, Heating, MDA, SDAP, TV

4.4.2 Knowledge Layer

For the purpose of our prediction study, the decision tree algorithm (J48) has been identified as leading to the finest results and can be used to produce the prediction model. Moreover, the python has been selected as best suited implementation for the recommendation system, as a portion of data mining system functionalities.

Rokach et al. (2005) describe a decision tree as a classifier rendered as a recursive partition from characteristic space to subspaces which provide a predictive basis. Decision tree algorithms create a decision tree with a particular dataset and their aim is to obtain an optimal decision tree by reducing the generalization error, the nodes or the average depth.

Since Python (scikit-learn library) will be implemented, the criterion parameter is used to obtain the optimal decision tree. The criterion parameter is the function applied to the quality of a division in order to select *gini* or *entropy*.

The authors in their study define the **Gini Index** as a criterion on impurities which calculates the differences among the probability distributions above the target attribute values (Rokach et al., 2005). The Gini index is defined in equation (Eq. 13) as follows:

$$Gini(y,S) = 1 - \sum_{c_j \in dom(y)} \left(\frac{|\delta_{y=c_j} S|}{|S|} \right)^2 \tag{13}$$

Furthermore, the evaluation criterion for selecting the attribute a_i is denoted in GiniGain equation (Eq. 14) as follows:

$$\text{GiniGain}(a_i\,, \text{S}) \;=\; \text{Gini}(\text{y}, \text{S}) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\delta_{a_i=v_{i,j}} S|}{|S|} Gini(y, \delta_{a_i=v_{i,j}} S) \tag{14}$$

S presents the training set; y presents the target feature; a_i presents the input feature set; (v_1, ... , v_n) presents the outcomes of discrete functions.

Information Gain is defined as an impurity-based criterion, as presented in the InformationGain equation Eq. 15, that applies the entropy measure as the impurity measure (Rokach et al., 2005).

$$\begin{aligned} &\text{InformationGain}(a_i\,, \text{S}) \\ &\quad = \text{Entropy}(\text{y}, \text{S}) - \sum_{v_{i,j} \in dom(a_i)} \frac{|\delta_{a_i=v_{i,j}} S|}{|S|} Entropy(y, \delta_{a_i=v_{i,j}} S) \end{aligned} \tag{15}$$

Information gain as in equation Eq. 16 is based on entropy where:

$$Entropy(y,S) = \sum_{c_j \in dom(y)} - \left(\frac{|\delta_{y=c_j} S|}{|S|} \right) \log_2 \left(\frac{|\delta_{y=c_j} S|}{|S|} \right) \tag{16}$$

If the subset observations in a dataset are identical, so there is neither impurity nor randomness, or belong to a class, then the entropy of that dataset would be minimum value, which is 0. Entropy equals the sum of the probability for every label multiplied by the logarithmic probability of the label itself (Tangirala, S., 2020).

Tangirala (2020), the author studied and concluded empirically that both Information gain and gini index generate the same accuracy for classification problems. The findings gained demonstrate that there is no substantial variance in model performance using gini index and Information gain, regardless of if the dataset is balanced or imbalanced.

The model is built with Python code by using Jupyter Notebook for data science projects by following the steps as below in Figure 42. The steps of building the model for the recommendation system are: 1) importing required libraries, 2) loading dataset, 3) Feature Selection - Splitting Dataset in Features and Target Variable, 4) Preparing the Data for Model Building - Splitting Data, 5) Building Decision Tree - Classification Model, 6) Evaluating Model and 7) Visualizing the Decision Tree.

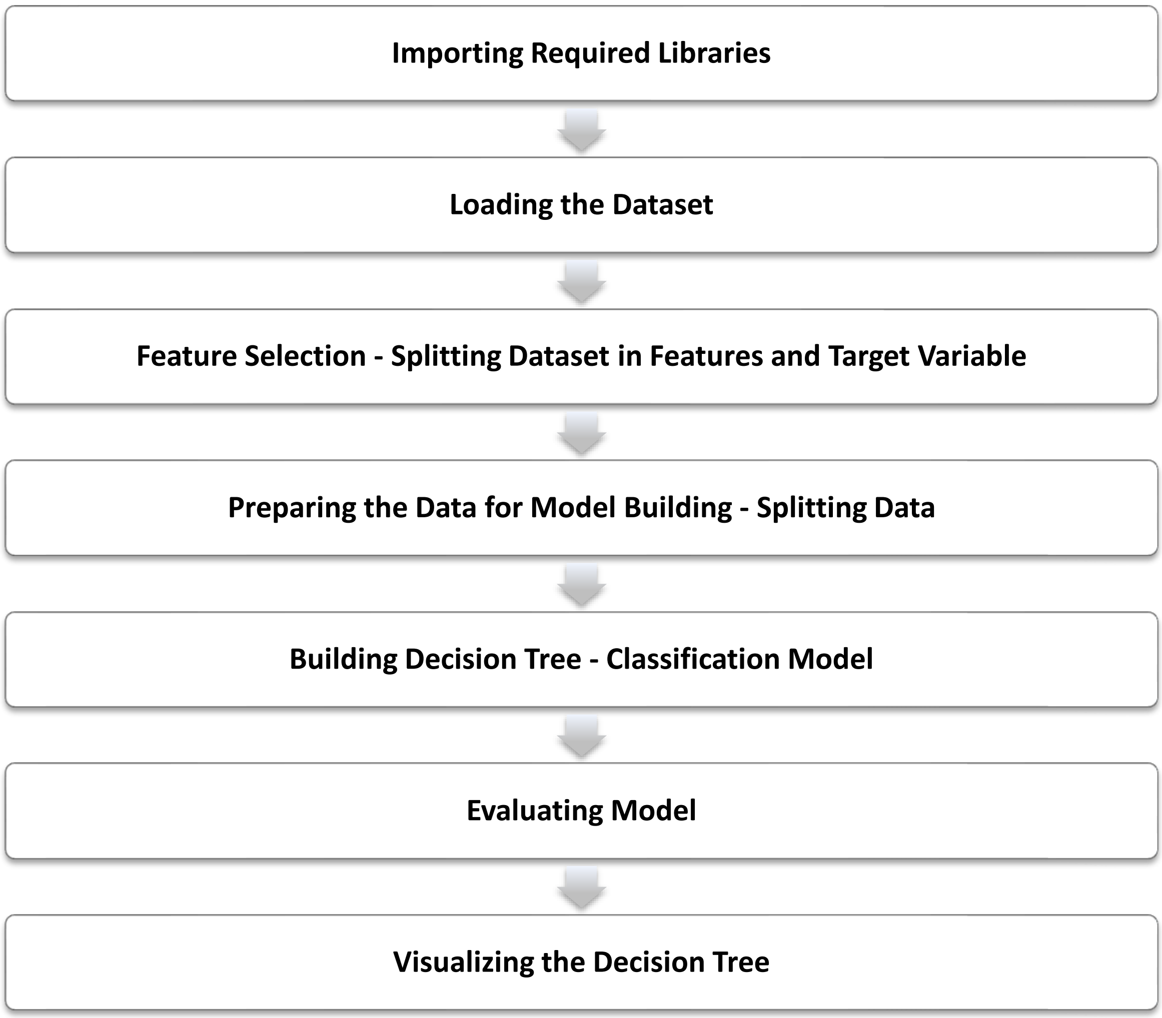

Figure 37: Steps to Build the Model

4.4.3 Application Layer

The purpose of this recommendation system is to facilitate the process of decision-making due to reliable product repairing predictions and increase the business' productivity and efficiency. This contribution goes to the decision makers of after-sales service business and the recommendation system is to improve the service quality towards customers, and as a result enhance customer experience. The recommendation system will serve as a way of increasing the service quality and business communication between after-sales service business and subcontractors, retailers, wholesalers, and manufacturers. Service quality can be achieved by using the data mining approach and appliances.

Consequently, the main objective of the recommendation system is to improve the decision-making and to enhance the customer experience in home appliances after-sales service.

The users interact with the recommendation system, without having to log in, which in turn handles the information generated by the classification algorithm – the decision tree. The users benefit by obtaining the probability prediction - data and graphs.

The built model is a binary classification model to predict the service quality of home appliances after-sales services as a function of the time taken to resolve the complaint. The solution of our model follows the machine learning concepts and implements a decision tree algorithm for its possibility to handle categorical predictors and interpret the results.

This scope is composed by one module: The Decision Maker Module.

The recommendation system is built with Python code by using Streamlit package for data science projects by following the steps as below in Figure 43 for the prediction of time taken for resolving the complaint.

Figure 43 presents the steps implemented to build the recommendation system. The steps of building the recommendation system are: 1) Collecting user input features into dataframe, 2) Combining user input features with after-sales service dataset, 3) Encoding of ordinal features, 4) Reading from saved classification model, 5) Applying model to make predictions and 6) Visualizing the data in bar plot by using "Plotly Chart".

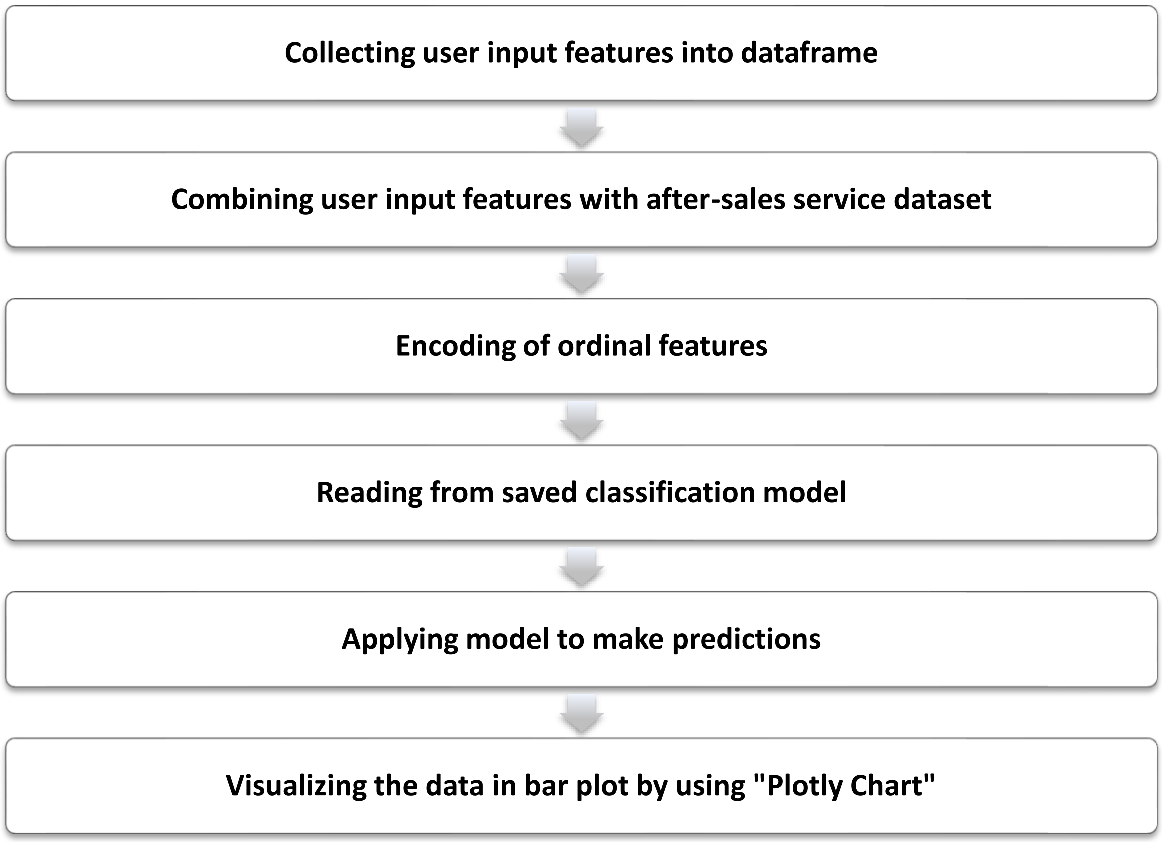

Figure 38: Steps to Build the Recommendation System

This module delivers all the capabilities for the users who have all the privileges. The prediction data and relevant graphs have been presented in a suitable way. The first part of recommendation system displays the prediction probability of 'time taken for resolving the complaint' numerically in percentage. The second part of recommendation system displays graphically the comparison of the prediction results that handles different classification inputs. The user will be able to show repair prediction probability by category and brand of the products. The users can seek predictions for products and brands of the products. Moreover, the prediction probability of service quality (time taken for resolving the complaint) demonstration by seeking option is the key feature of this recommendation system. The target is 'weeks to repair' as a categorical predictor and the model predicts service quality with four groups of the response, such as week one, week two, week three and week four. The target presents the number of weeks the repair is completed. The left sidebar of the module offers the options to select which brand or category the user wants to predict the time to repair the product. Figure 44 shows the options to select on the left sidebar in order to obtain prediction probability of weeks to repair, such as:

- Six main product brands such as: 1) Brand One, 2) Brand Two, 3) Brand Three, 4) Brand Four, 5) Brand Five And 6) Brand Six
- Five main product categories such as: 1) Air Conditioning, 2) Heating, 3) MDA, 4) SDAP and 5) TV.

Figure 44 presents the first part of recommendation system and the menu with the selection of option of brands and categories.

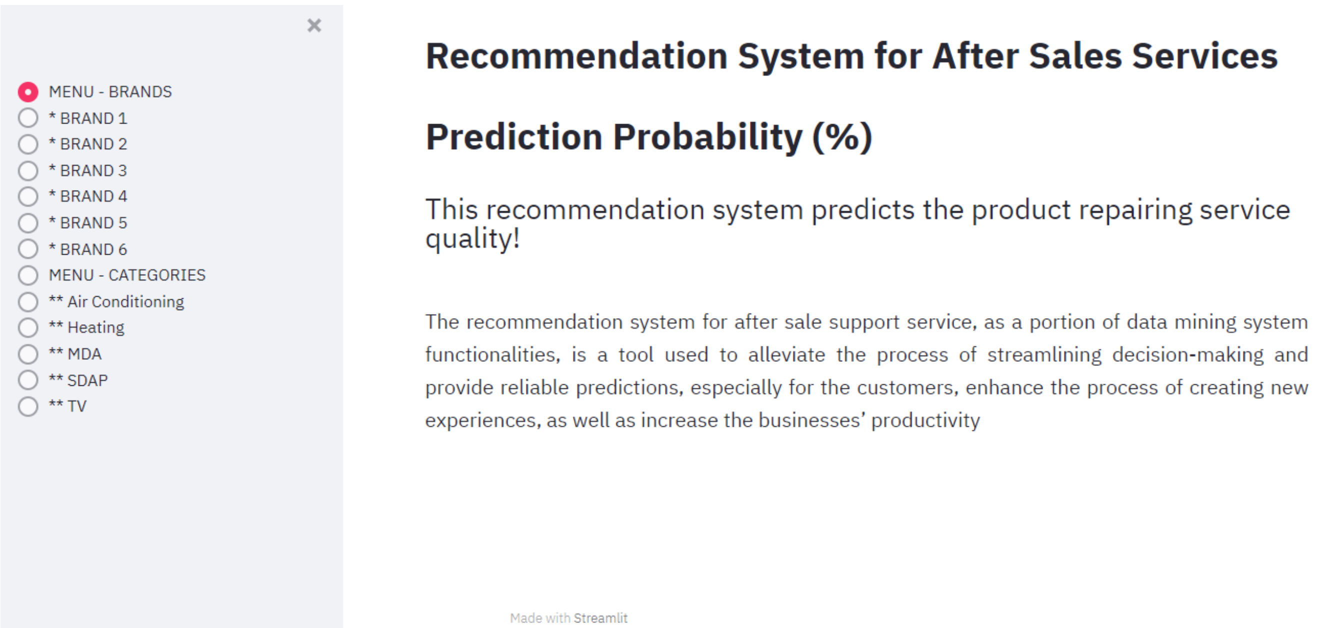

Figure 39: Recommendation System for a decision-maker

The advantages expected and provided by the recommendation system are security, speed, ease of accessibility and reduction of loading time.

4.4.4 Brand Prediction

The brand prediction of the recommendation system has the links (contents on the left sidebar) for the predictions of time to repair the product by brand. After selecting e.g., the first option – Brand 1 on the left sidebar, the obtained prediction probability results are demonstrated in Figure 45.

Figure 45 shows the Brand One prediction in recommendation system. The brand prediction in recommendation system presents the prediction probability of service quality (time taken for resolving the complaint) for the five categories. The recommendation system presents the quality of service, more specifically the prediction time to complete the repair. The percentage in green presents the prediction probability to complete the repair for week one for each category of Brand 1. The percentage in blue presents the prediction probability to complete the repair for two weeks each category of Brand 1. The percentage in red presents the prediction probability to complete the repair for three weeks for each category of Brand 1. The percentage in orange presents the prediction probability to complete the repair for four weeks for each category of Brand 1. The prediction probability for categories heating and TV is better, particularly in the sense that the probability of the completion of product repair within one week is higher. The prediction probability for category SDAP is worse, particularly in the sense that the probability of the completion of product repair for more than one week is higher.

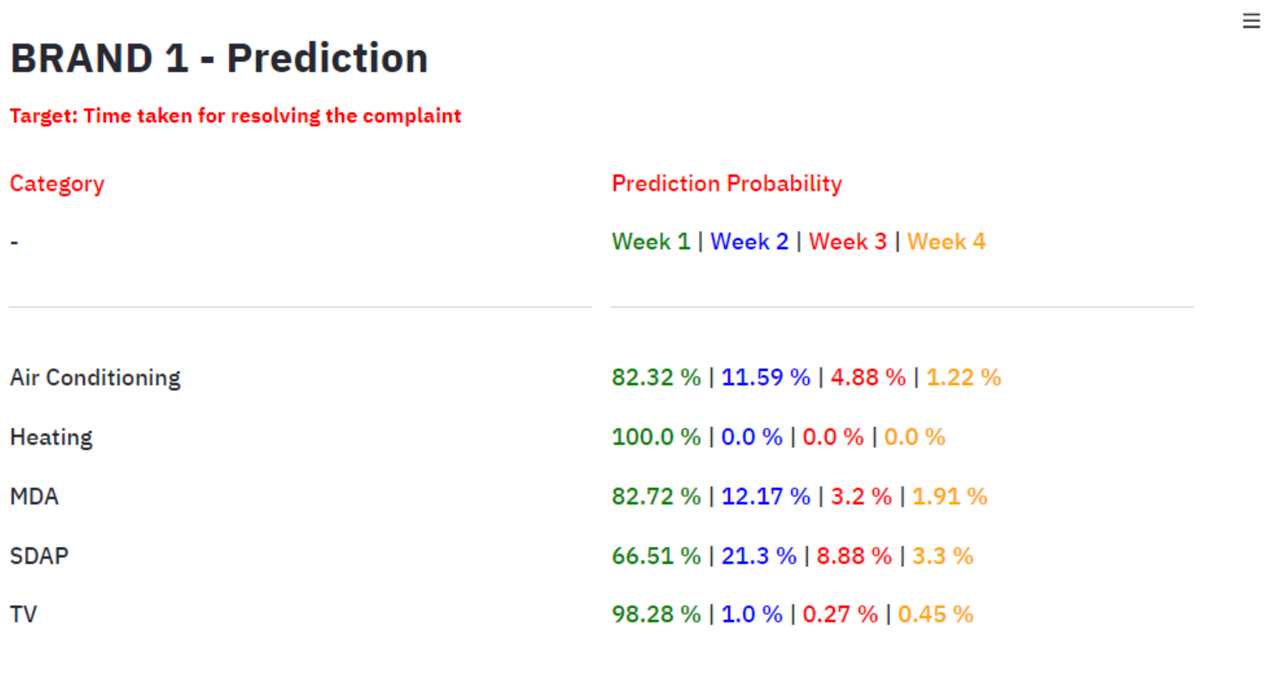

Figure 40: Brand 1 Prediction of the Recommendation System

The chart in Figure 45 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint. The comparison demonstrations refer to each category of Brand 1.

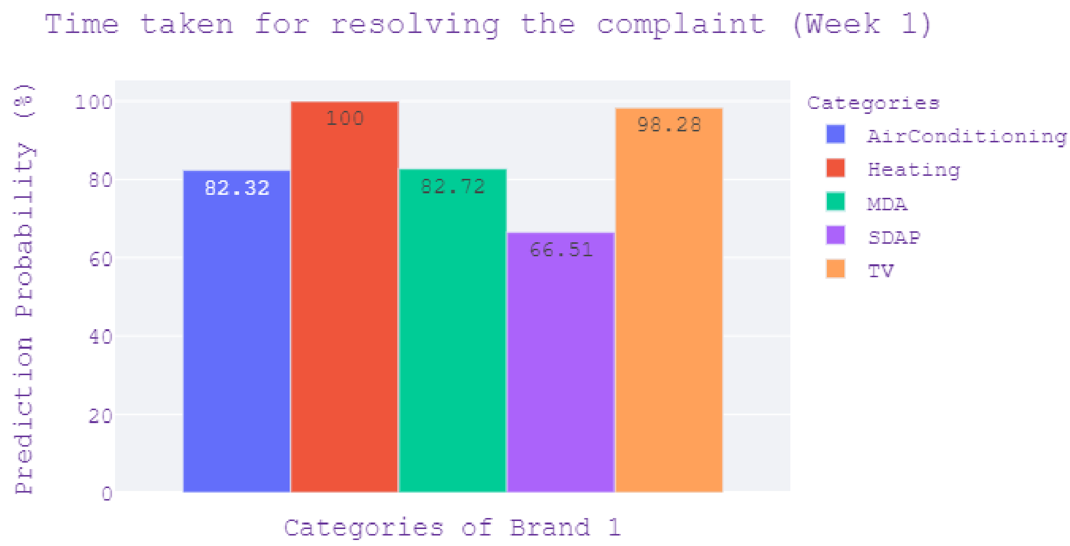

Figure 41: Brand 1 App – Chart of Repair Prediction for One Week

The chart in Figure 46 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint within one week. The comparison demonstrations belong to each category of Brand 1.

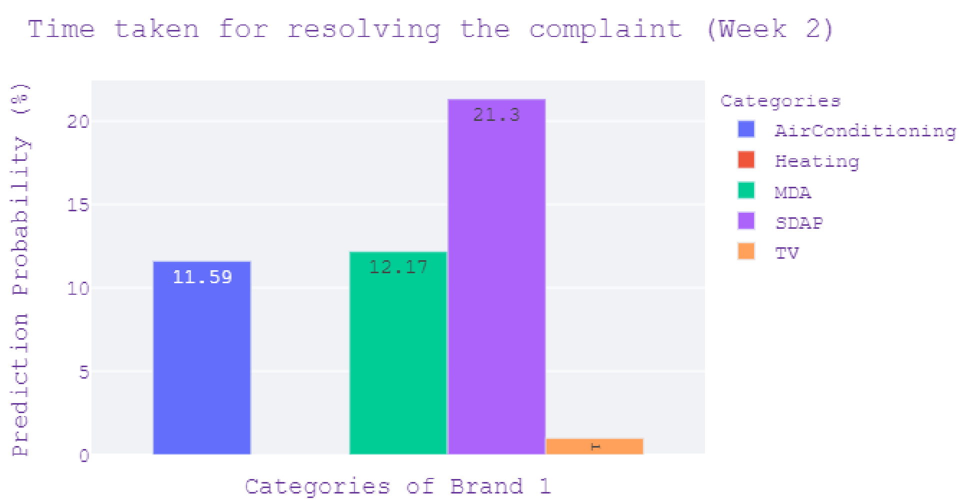

Figure 42: Brand 1 App – Chart of Repair Prediction for Two Weeks

The chart in Figure 47 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint within two weeks. The comparison demonstrations belong to each category of Brand 1.

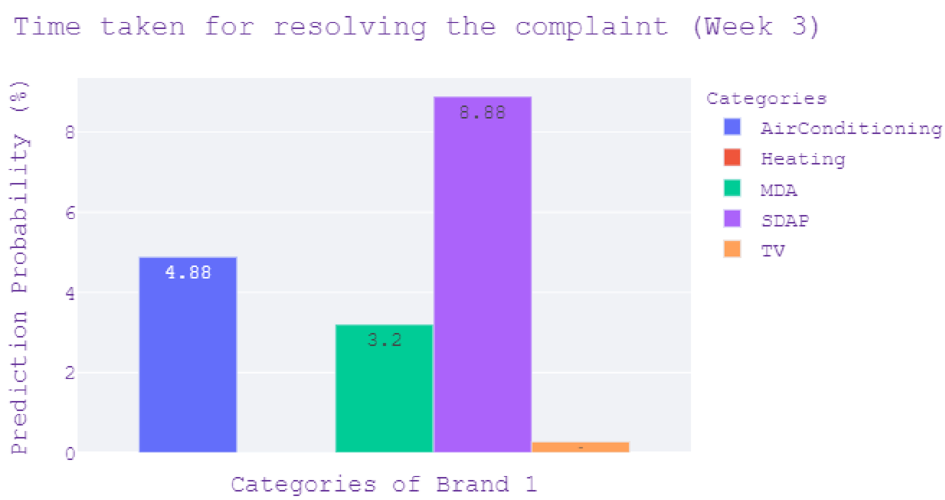

Figure 43: Brand 1 App – Chart of Repair Prediction for Three Weeks

The chart in Figure 48 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint within three weeks. The comparison demonstrations belong to each category of Brand 1.

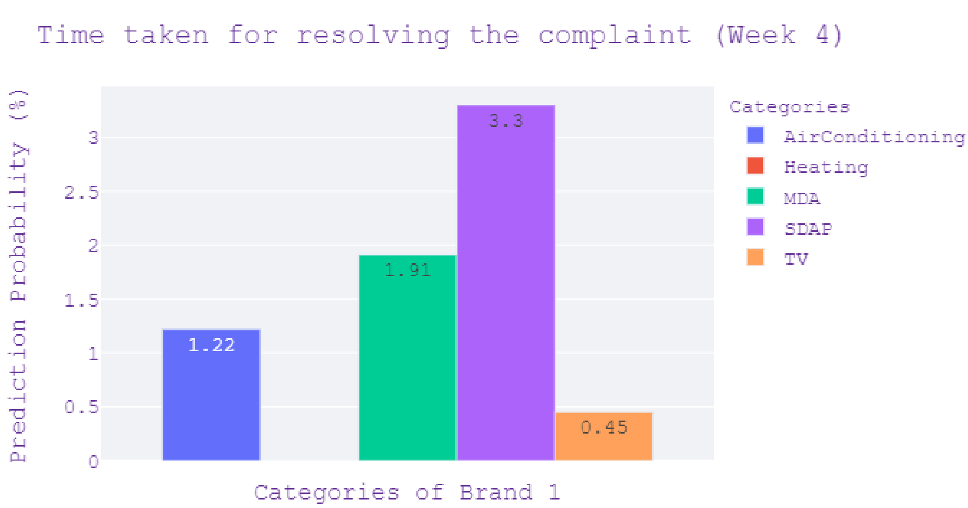

Figure 44: Brand 1 App – Chart of Repair Prediction for Four Weeks

The chart in Figure 49 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint within four weeks. The comparison demonstrations belong for each category of Brand 1.

On the left sidebar, identical parts in recommendation system are provided for all the brands and present prediction probability of time taken for resolving the complaint for each category of brands.

4.4.5 Category Prediction

The category prediction of the recommendation system has the links (contents on the left sidebar) for the predictions by category. After selecting e.g., the first option of categories - MDA on the left sidebar, we obtained results for the prediction probability and are demonstrated in Figure 50. The category part of the recommendation system presents prediction probability of time taken for resolving the complaint for six brands. The prediction probability for Brand 1 is better, particularly in the sense that the probability of the completion of product repair within one week is higher. The prediction probability for Brand three is worse, particularly in the sense that the probability of the completion of product repair for more than one week is higher.

MDA - Prediction

Target: Time taken for resolving the complaint

Brand	Prediction Probability
-	Week 1 | Week 2 | Week 3 | Week 4
BRAND 1	82.72 % | 12.17 % | 3.2 % | 1.91 %
BRAND 2	69.76 % | 20.09 % | 6.48 % | 3.67 %
BRAND 3	61.93 % | 26.45 % | 7.05 % | 4.57 %
BRAND 4	76.09 % | 15.76 % | 3.26 % | 4.89 %
BRAND 5	72.03 % | 19.21 % | 5.08 % | 3.67 %
BRAND 6	67.31 % | 21.93 % | 5.79 % | 4.97 %

Figure 45: MDA Prediction of the Recommendation System

The chart in Figure 50 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint. The comparison demonstrations refer to each brand of MDA category.

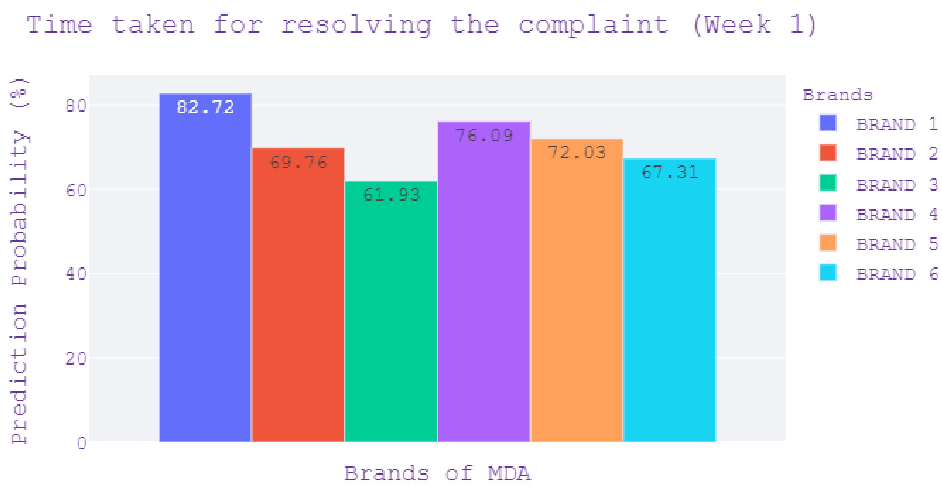

Figure 46: MDA Category App – Chart of Repair Prediction for One Week

The chart in Figure 51 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint within one week. The comparison demonstrations refer to each brand of MDA category.

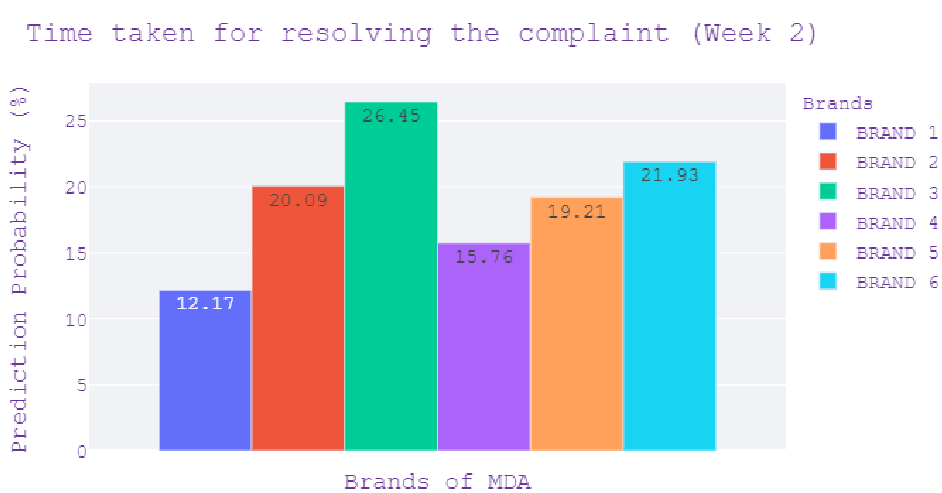

Figure 47: MDA Category App – Chart of Repair Prediction for Two Weeks

The chart in Figure 52 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint the complaint within two weeks. The comparison demonstrations refer to each brand of MDA category.

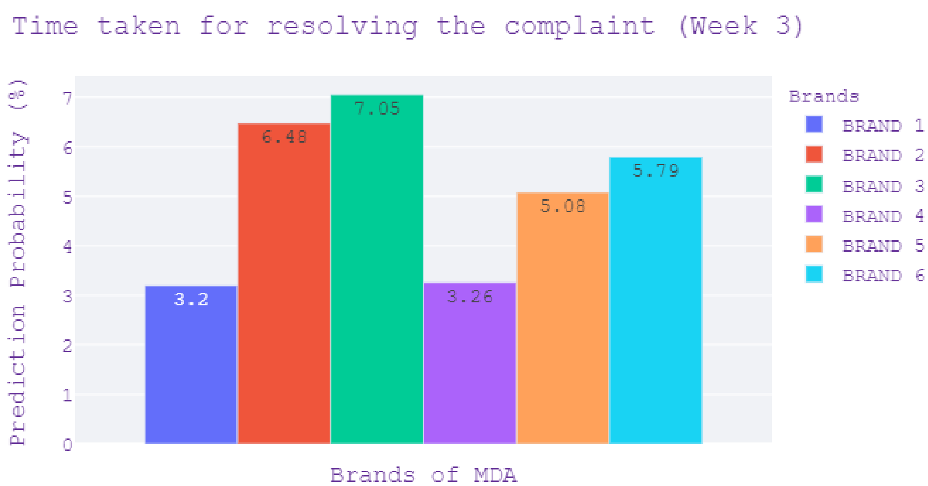

Figure 48: MDA Category App – Chart of Repair Prediction for Three Weeks

The chart in Figure 53 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint the complaint within three weeks. The comparison demonstrations refer to each brand of MDA category.

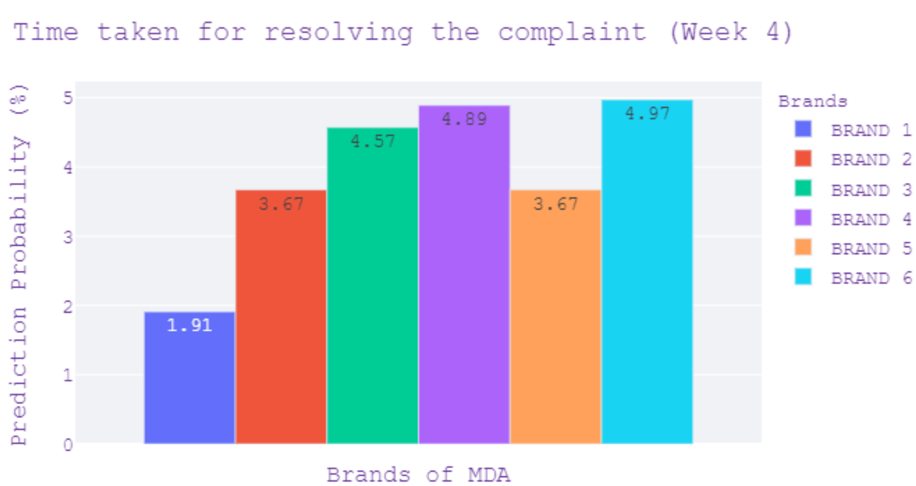

Figure 49: MDA Category App – Chart of Repair Prediction for Four Weeks

The chart in Figure 54 demonstrates graphically the comparison of percentage probability of time taken for resolving the complaint the complaint within four weeks. The comparison demonstrations refer to each brand of MDA category.

On the left sidebar, identically parts of the recommendation system are provided for the categories, as air conditioning, heating, SDAP and TV, and present prediction probability of time taken for resolving the complaint for each brand categories.

4.4.6 Quality of the Recommendation System

The quality of the prediction by the recommendation system implementation is based on the studies performed, which is the best classification algorithm and the speed to respond to the users. The recommendation system ensures that it adapts the needs of the users, it results to be effective and efficient, and it has no negative outcome by its usage. Moreover, the recommendation system is tested by identifying bugs and errors in the system in order to assure high quality of the implementation and meets business and technical requirements.

4.4.7 Results - System Usability Testing

Usability testing as a technique is performed to identify problems. Usability testing benefits both the users (decision makers) and the after-sales service business in general by predicting the quality of service. Moreover, users were satisfied interacting with the recommendation system. A simple test was performed to test the implementation with only one user, and it is important that we got the user's feedback during the process of testing.

The testing was performed on recommendation system to find incorrect classification, errors, and test behavior and performance. The following activities were tested:

- Flexibility/Responsiveness of recommendation system in different browsers
- Accessibility/Easiness
- Incorrect/Correct classification by decision tree algorithm - prediction probability

The recommendation system fulfils the user's expectations. The interface, usability and performance are important aspects of a recommendation system.

Table 22 displays the performed test steps and the expected results and outcomes.

Table 22: Testing Results of the Recommendation System Performance

Steps Taken	**Expected Results**	**Outcome**
Flexibility/Responsiveness of recommendation system in different browsers to be tested. Three web browsers such as Chrome, Firefox and Internet Explorer were selected.	Condition to be met, all elements of recommendation system to be loaded without over-flow and be responsive.	The process of testing flexibility/responsiveness of recommendation system is completed successfully with no overflow.
Accessibility/Easiness to be tested.	Testing on accessibility/easiness is achieved by not entering credentials and no error catching.	The phase of testing accessibility - easiness is completed successfully.
Incorrect/correct prediction probability percentage to be tested.	According to the user, the prediction probability percentage for days to repair – target was mostly correct.	The phase of testing incorrectness/correctness of the prediction probability percentage is completed successfully.
User Satisfaction during testing process.	The business reported their satisfaction during the testing process.	Passed by reaching a good performance.

4.4.8 Evaluation of the impact of the outcome of the recommendation system

Through evaluation of impact, hypothesis one “H1: The data mining approach improves decision-making competency and decision-making processes in after-sales service business” has been tested and proved.

In this section, we present experiments and evaluation of time taken for resolving the complaint and the positive results made impact through the outcome of the recommendation system. Time taken for resolving the complaint is one of the key attributes of service quality impacting customer satisfaction, a nonfinancial performance measure for managerial decision-making.

The comparison and the positive results are displayed for two years, 2020 and 2021. The reduction of the yearly average of time taken for resolving the complaint is a proof of positive impact. The yearly average of time taken for resolving the complaint has reduced by 7.78% (0.54 days) in 2021 in comparison to the year of 2020. The evaluation of the impact is presented in Figure 55.

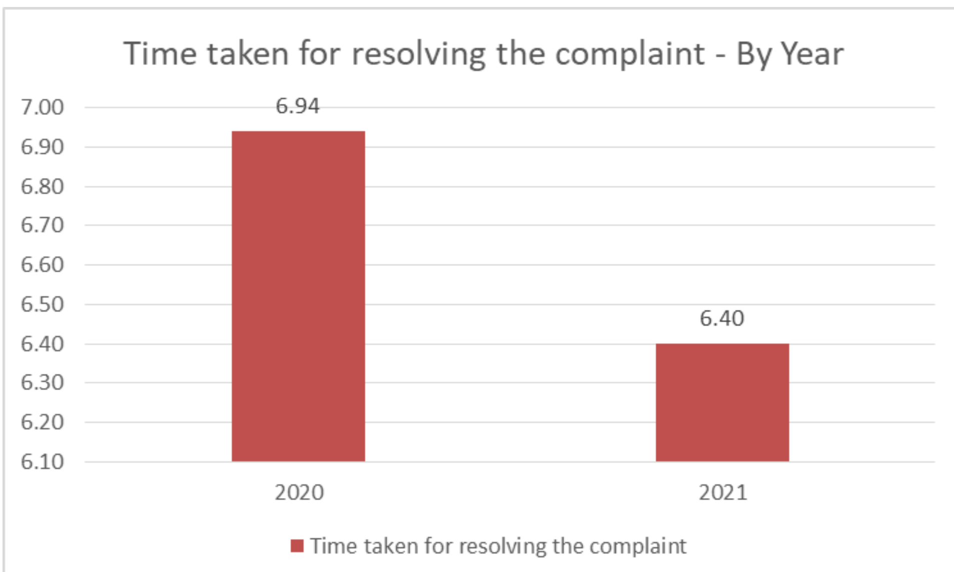

Figure 50: Yearly Comparison – Time taken for resolving the complaint

In Figure 56, the quarterly comparison results of decrease are presented for the average time taken for resolving the complaint. The quarterly comparison and the positive results are displayed for two years, 2020 and 2021. The reduction in quarters of the average of time taken for resolving the complaint is a proof of positive impact. The average of time taken for resolving the complaint has reduced in three quarters in 2021 in comparison to the year of 2020.

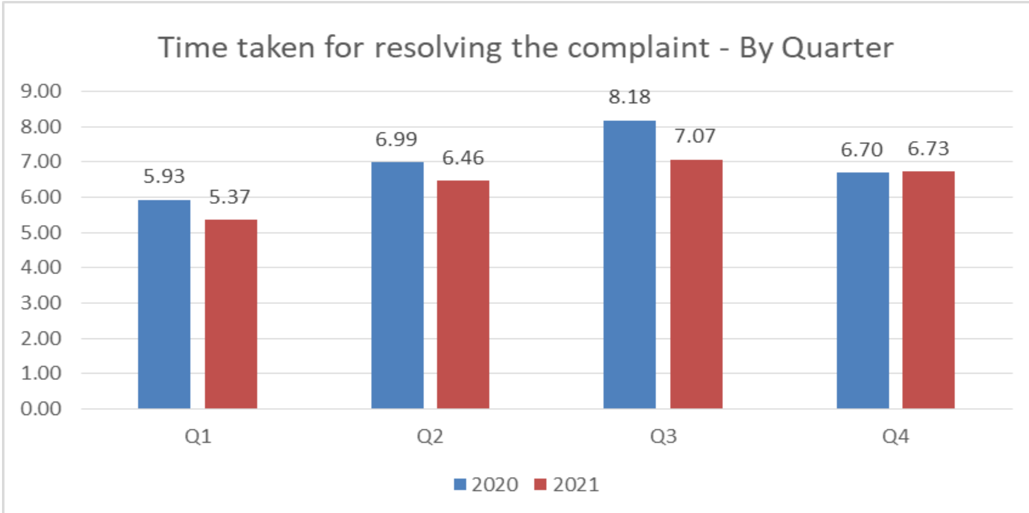

Figure 51: Quarterly Comparison - Time taken for resolving the complaint

For the quarter 1, the decrease of the target – average time taken for resolving the complaint is about 0.56 days. For the quarter 2, the decrease of the target – average time taken for resolving the complaint is about 0.53 days. For the quarter 3, the decrease of the target – average time taken for resolving the complaint is about 1.11 days. For the quarter 4, the target – average time taken for resolving the complaint is almost the same.

In Figure 57, the yearly comparison for all brands results of decrease of the average time taken for resolving the complaint is presented.

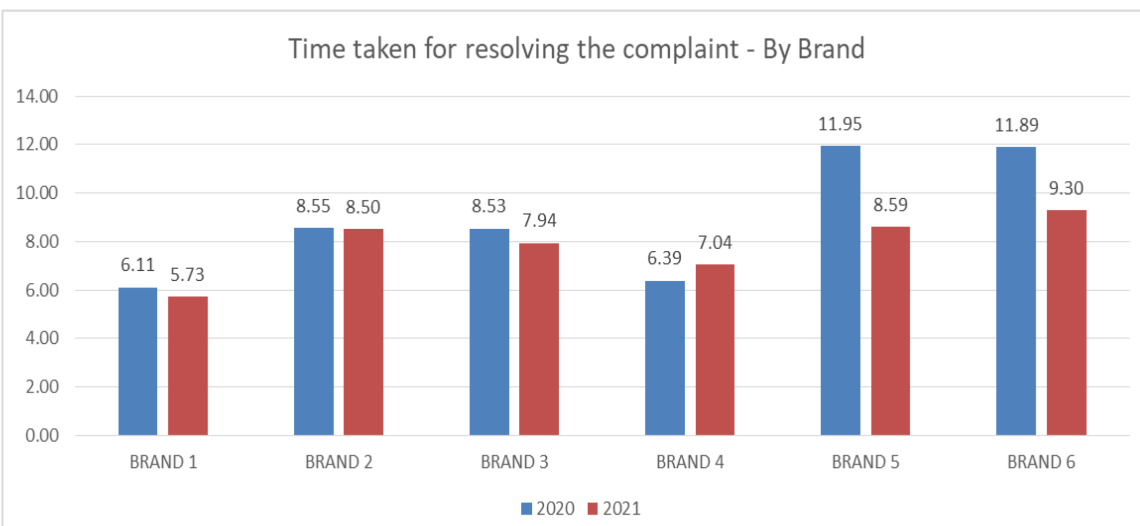

Figure 52: Yearly Comparison, By Brand - Time taken for resolving the complaint

The yearly comparison for the brands and the positive results are displayed for two years, 2020 and 2021. The reduction in years for the brands of the average of time taken for resolving the complaint is a proof of positive impact. The average of time taken for resolving the complaint for each brand has reduced in 2021 in comparison to the year of 2020.

For Brand 1, the decrease of the target – average time taken for resolving the complaint is about 0.38 days. For Brand 2, the decrease of the target – average time taken for resolving the complaint is slight about 0.05 days. For Brand 3, the decrease of the target – average time taken for resolving the complaint is about 0.59 days. For Brand 4, the increase of the target – average time taken for resolving the complaint is about 0.65 days. For Brand 5, the decrease of the target – average time taken for resolving the complaint is about 3.36 days. For Brand 6, the decrease of the target – average time taken for resolving the complaint is about 2.59 days.

4.5 Summary

Based on the findings of the qualitative research, the following potential key effective strategic solutions as a recommendation for increasing customer satisfaction and improving after-sales service quality and have an impact on decision-making. Figure 24 presents grouped recommendations to be applied by after-sales services business based the findings of the research.

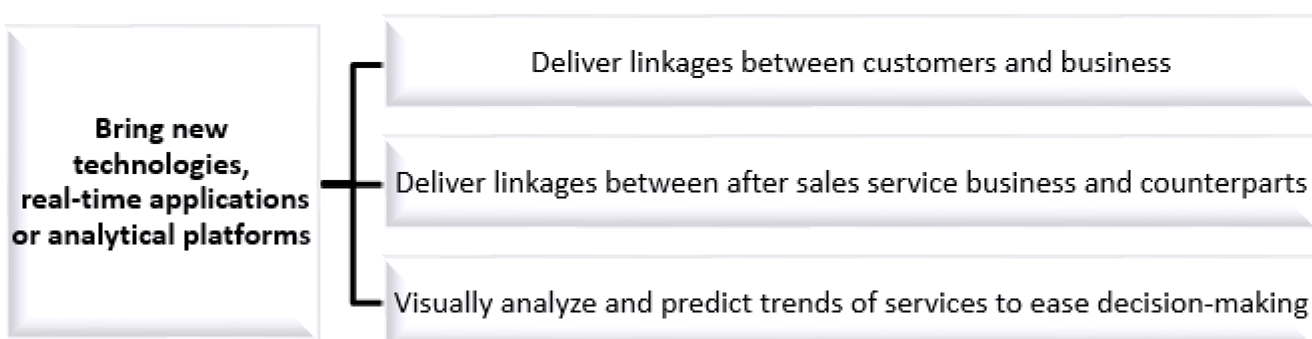

Figure 53: Potential Key Findings as solutions for the challenges to effectively serve customers

The contributions of qualitative research in terms of results are:

- Qualitative research, semi-structured interviews conducted, through the empirical methodologies, provides an understanding of the challenges and issues of home appliances after-sales service businesses in Kosovo market.
- Effective strategic solutions are proposed, and research recommendations have helped analyze business performance to design strategies by decision makers to improve the customer satisfaction. The solutions proposed are such as: a) bring new technologies, real-time applications or analytical platforms, b) design efficient marketing strategies, c) negotiate when dealing guarantee policy and period terms with manufacturers and retailers, d) understand customer needs, preferences and expectations, e) improve operational activities, f) connect with training providers, and g) offer internships.
- The results obtained identify areas and aspects of improvement and the recommendations offered to management of after-sales services businesses are beneficial to the 1) decision-making processes and the 2) research proves the importance of analysis in studying the impact of service quality on customer satisfaction, a nonfinancial performance measure for managerial decision-making.

Moreover, a case study was conducted by applying classification and clustering techniques to the after-sales services for home appliances throughout the business digital transformation journey. We gained a practical sample of the application of the data mining approach. The main goal of the case data mining comparison implementation was to identify the best clustering and classification techniques by comparative study of a variety of clustering and classification algorithms.

Furthermore, the purpose of the regression analysis was to predict the value of the customer satisfaction variable and estimate the effect of four independent variables on the dependent customer satisfaction variable. Figure 40 presents the analysis which are reliable to identify the effects and influence of these four independent - explanatory variables, especially the time to repair variable, to customer satisfaction variable.

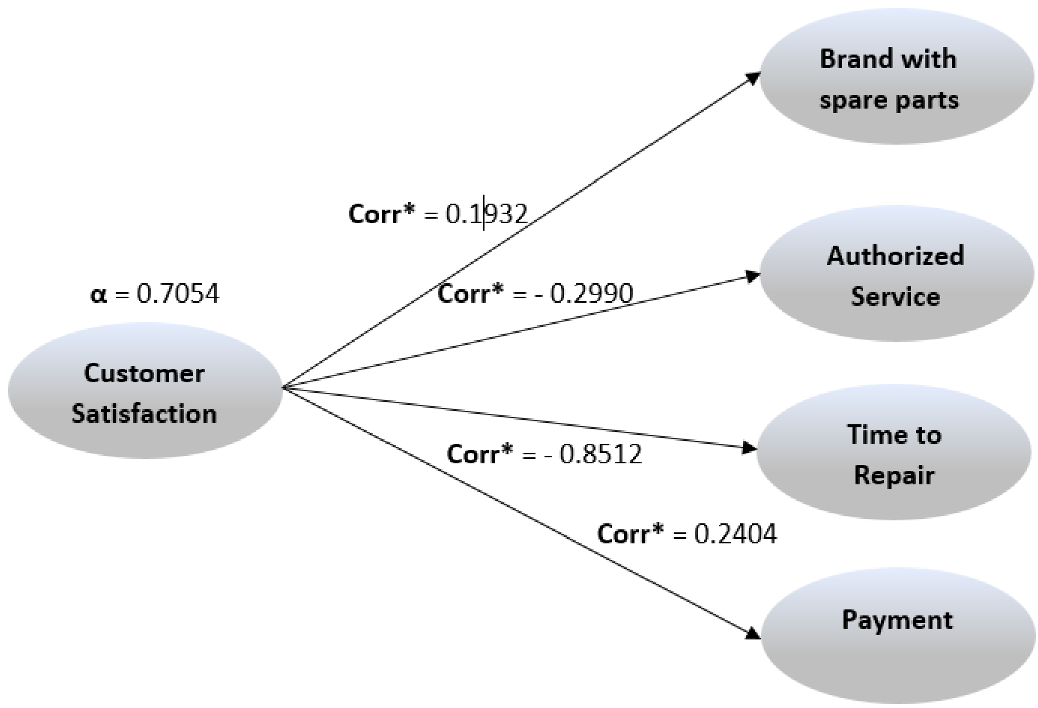

Figure 54: Findings on Correlation and Reliability Results

The empirical conclusion achieved, as in Figure 41, tells that the time taken for resolving the complaint is a good predictor and influences customer satisfaction. Consequently, service quality plays an important role in after-sales service businesses. Finally, the study found that time taken for resolving the complaint impacts customer satisfaction, a nonfinancial performance measure for managerial decision-making.

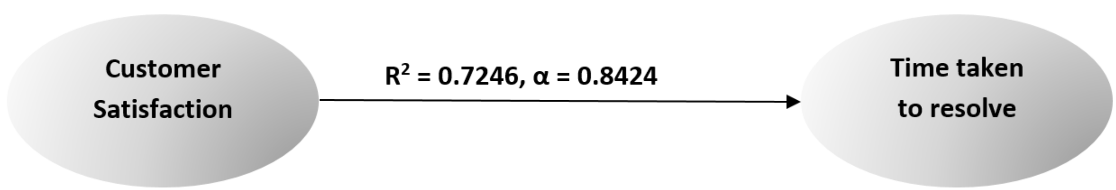

Figure 55: Findings on Reliability and Regression Results

Finally, the recommendation system implementation section contributes to elaborating all parts of the implementing of the data mining approach to the recommendation system and making a model for other businesses with the similar concerns. In this case study, we gained a concrete sample of the application of the data mining approach for analyzing information and predicting processes in business.

The model of decision tree, as the best classification algorithm, is chosen in the implementation of recommendation system, as a portion of data mining system functionalities. The recommendation system will facilitate processes between the counterparts and the after-sales services businesses. The implemented recommendation system presents data that mainly predicts the probability of time taken for resolving the complaint - time to repair the product and digitizes manual processes of data presentation. Python has been selected as a trend in system development as best suited implementation for the recommendation system, as a portion of data mining system functionalities, as an outcome of this study. The recommendation system ensures that it adapts the needs of the users, it results to be effective and efficient. Moreover, the recommendation system is of high quality of the implementation and meets business and technical requirements.

More importantly, the evaluation results prove the impact of the implementation of data mining approach by facilitating the process of decision-making due to reliable product repairing predictions and increase the business' productivity and efficiency.

5 DISCUSSION OF FINDINGS

The identification of important factors towards the expectation of the analytical applications, intelligent applications and the descriptive, predictive, and prescriptive models in the field of after-sales services were an important point of study. As a result, our experiments addressed questions such as: 1) Which techniques, trends and tools can influence a data mining system? Which business areas have used and benefited from data mining techniques aiming to improve decision-making along the business digital transformation? 2) What challenges are faced by the businesses? What are some of the challenges we find solvent or hard to overcome, and how do we come up with effective solutions? 3) How does a model support an after-sales service business that is considering using data mining to improve decision-making along the business digital transformation? How to analyze and design the platforms that can improve the analytics of the data?

The thesis broadens the understanding and experience of key perspectives on data mining as an opportunity for businesses to transform digitally. The thesis presents the innovative approach by proving that the data mining approach is a revolutionary solution and the latest trend in digital transformation by providing businesses with research on the advantages and challenges according to their needs.

5.1 Hypothesis Testing

According to the empirical findings from the statistical analysis, three hypotheses were accepted, as indicated in the following table.

Table 23: Hypotheses

Hypotheses	Accepted
H1: The data mining approach improves decision-making competency and decision-making processes in after-sales service business	√
H2: Service quality impacts customer satisfaction in after-sales service business	√
H3: Business strategic decision-making in after-sales service impacts after-sales service quality	√

H1: The data mining approach improves decision-making competency and decision-making processes in after-sales service business.

Through evaluation of impact, hypothesis one H1 has been tested, proved, and accepted. The implicit conclusion is that the decision makers of businesses need to consider the implementation of a data mining techniques as a chance to enhance decision-making competency throughout the business digital transformation.

H2: Service quality impacts customer satisfaction in after-sales service business.

Through the regression analysis, hypothesis two H2 has been tested, proved, and accepted. The implicit conclusion is that service quality impacts customer satisfaction in after-sales service business.

H3: Business strategic decision-making in after-sales service impacts after-sales service quality.

Through quantitative research, the hypothesis three H3 has been tested, proved, and accepted. The implicit conclusion is that decision-making competency impacts after-sales service quality.

5.2 Proposed Model

The research in question presents advanced analytics techniques and tools by demonstrating their importance and the need for digital transformation, influencing businesses. The results show that big data analytics is one of accelerators, and that data mining, machine learning, and natural language processing are the main techniques of big data and advanced analytics of digital transformation of businesses. Data mining is one of the main techniques of advanced analytics of digital transformation of the businesses considered. The most common data mining techniques highlighted are clustering, classification, regression, sequential analysis, and association.

Furthermore, the main objective of the aspects addressed is to attract businesses to use chatbots technology, built by using natural language processing technologies, as a way to boost digital transformation. An important benefit of digital technologies is to be emphasized, the implementation of cloud-based services, as a key of digital technology, free consumers from maintenance of hardware and software and gain benefits. The results show that the CRISP-DM - data mining model is applicable across all industries and is considered as the most commonly used model.

Herewith, the thesis demonstrates the benefits and impact of the business digital transformation by evaluating progress, and the most common dimensions of digital transformation. The digital maturity model is an effective means to provide directives for a clear path along the digital transformation journey. The findings indicate that the digital maturity model is to evaluate the capabilities mainly on certain common business dimensions, such as customers, strategies, technologies, operations, organization, and culture. Agile, the modern approach, results to become a driving framework in the digital transformation of businesses.

The following are the challenges and limitations of applying data mining: technology, skills, problem of poor data quality, misuse of information/inaccurate information, complexity of integration as well as data security, and privacy. The benefits of data mining include simplifying decision-making, increasing efficiency and business productivity, and improving customer experience throughout their digital transformation journey.

The potential key findings derived from the related work study that impacted in proposing the model (See Figure 1; Figure 22) are schematically presented below in Figure 58.

Digital Transformation Dimensions	•Technologies; Customers; Strategies
Digital Transformation Technologies	•Big Data Analytics; Advanced Analytics
Advanced Analytics	•Data Mining; Machine Learning
Data Mining Techniques	•Classification; Clustering; Regression
Benefits of Applying Data Mining	•Improvement of Decision Making; Enhancing Customer Experience
Data Mining Model Process	•Business Understanding; Data Understanding and Preparation; Modelling; Evaluation; Deployment;
Performance Measurement: Nonfinancial Aspects	•Customer Satisfaction; Service Quality

Figure 56: The potential key findings derived from the related work study

Finally, the aims were to build and implement a recommendation system and determine the sustainability and reliability of the recommendation system, by selecting the best trends and data mining technologies and tools. The related work findings influenced building our recommendation system for increasing customer satisfaction, improving after-sales service quality and have an impact on decision-making.

5.3 Recommendation System

The thesis aimed to perform qualitative and quantitative research and bring answers to research questions. In line with our hypotheses, the outcome showed that the proposed model, and recommendation system as an outcome, alleviate the process of decision-making along the business digital transformation.

Qualitative research provides an understanding of the challenges and issues of home appliances after-sales service businesses in the Kosovo market. Strategic solutions are recommended and elaborate how the recommendations offered have aided decision makers in analyzing the performance of businesses. Our study follows the potential recommendation as in visually analyzing and predicting trends of services to ease decision-making and delivering linkages between after-sales service business and counterparts (Saccani et al., 2006; Pagalday et al., 2018; Lele, 1997; Cascio and Montealegre, 2016;).

The potential key findings derived from the qualitative research are schematically presented below. Figure 59 elaborates key effective strategic solutions as recommendations.

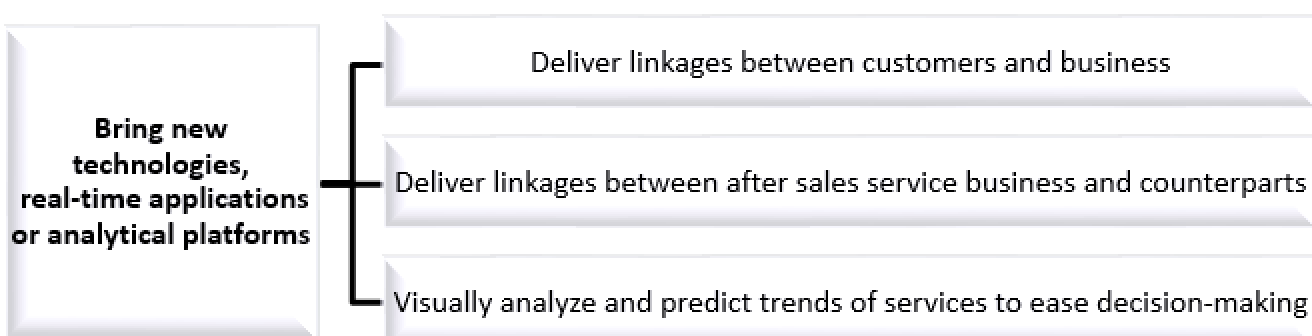

Figure 57: Potential Key Findings as solutions for the challenges to effectively serve customers

The quantitative research study applied in home appliances after-sales service, the practical comparative study of the implementation of the clustering and classification algorithms and the regression analysis study on nonfinancial performance measures, demonstrates that the implementation of recommendation system implementation is sustainable and reliable as an outcome, and proves that the data mining approach has a positive impact on improvements in decision-making processes and decision-making competency along the business digital transformation journey.

The main findings are achieved by applying the comparison of clustering and classification algorithms in a case study of after-sales services business. The best clustering and classification techniques are found out by comparative analysis of different clustering and classification algorithms. The dataset is updated by the company's client service. Regarding the quantitative study, about 62K cases were retrieved from the database. The studies of different researchers that contributed on application, selection and comparison of clustering and classification algorithms, and their evaluation metrics, are such as Singh and Surya (2015), Renjith et al. (2020), Patel and Patel (2019), Valarmathy and Krishnaveni (2019), Hossin and Sulaiman (2015), Çığşar and Ünal (2019), Verma (2019), Hossin and Sulaiman, 2015, Chen et al. (2012), Halimu et al., 2019, etc.

The practical comparative study of the implementation of the clustering algorithms obtains the finding that the prediction is categorized as binary (two classes) classification due to the class attribute has two output values, within which product guarantee can be differentiated with or without guarantee (See Figure 28). And, FarthestFirst is the most effective algorithm because it takes much less time to construct a model and has a lower level of poorly clustered instances. The comparison of algorithms is carried out on three parameters, such as accuracy, incorrectly clustered instance and time needed to construct a repair prediction model (See Figure 27). The practical comparative study of the implementation of the classification algorithms obtains the finding that the decision tree (J48) data mining algorithm gives the best results and can predict repairs according to guarantee period of the product. According to the results, the constructed decision tree (classification algorithm) will be used as a sample model for the recommendation system designing. The comparison of the five-classifier algorithms is carried out on three parameters, such as accuracy, Kappa statistics and mean absolute error (MAE) (See Figure 29).

Moreover, the studies of different researchers that contributed on application of statistical, correlation, and regression analysis are such as Golrizgashti et al. (2020), Murali (2016), Murali et al. (2016), etc. Our case study examines the relationship between service quality variables and customer satisfaction. The analysis predicts the value of the customer satisfaction variable and estimate the effect of four independent variables on the dependent customer satisfaction variable. Consequently, service quality plays an important role in after-sales service businesses. The potential key finding is derived from the regression analysis study on nonfinancial performance measures (See Figure 41). The time taken for resolving the complaint is a good predictor and influences customer satisfaction. Finally, the empirical conclusion achieved, and the study found that time taken for resolving the complaint impacts customer satisfaction, a nonfinancial performance measure for managerial decision-making.

Furthermore, the quantitative and qualitative research study proves that a strong relationship can be built between customers and after-sales service businesses by offering good service quality. Responsiveness is one of five main dimensions of service quality impacting customer satisfaction, a nonfinancial performance measure for managerial decision-making. Customer satisfaction is an important factor affecting business strategic decision-making processes in after-sales service.

The previous results of the studies are valuable in application of recommendation system implementation (See Figure1; Figure 18; Figure 42; Figure 43) and are presented in the chronological manner as in Figure 60.

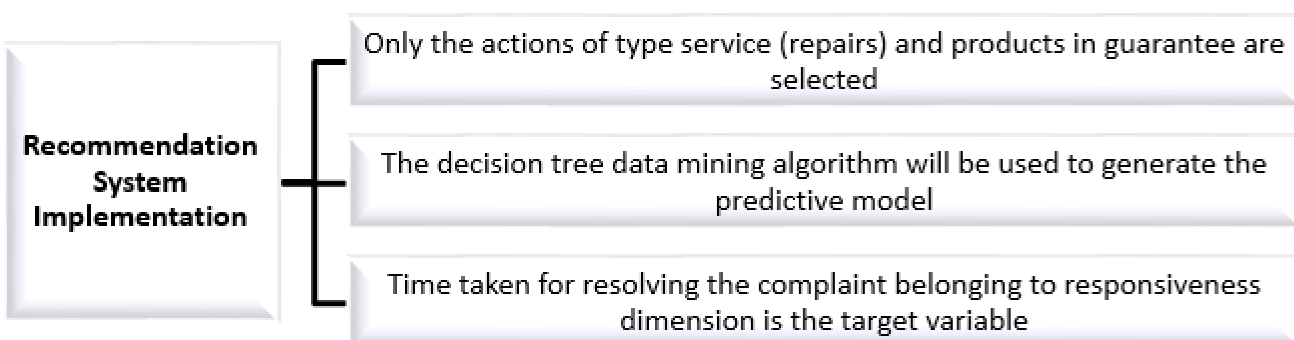

Figure 58: Results of quantitative study to be applied in building recommendation system

5.4 Data Mining Approach Impact in Decision-making Processes and Competency

The implication of the findings of the quantitative research study and the qualitative research study in building the recommendation system implementation in home appliances after-sales service demonstrate that the data mining approach has a positive impact on improving decision-making processes and business productivity along the business digital transformation journey. Lastly and more importantly, the data mining approach impacting the optimization of businesses' strategic goals made the most valuable part of research and as a result reinforcement of linkages between businesses and customers.

The positive impact in the main attribute of service quality - time taken for resolving the complaint - is evaluated and presented (See Figure 55; Figure 56; Figure 57;) and the results of decrease of the target - time taken for resolving the complaint in each quarter, prove the impact of the implementation of data mining approach by facilitating the process of decision-making due to reliable product repairing predictions, and increase the business' productivity and efficiency.

5.5 Contributions

Our study has six contributions as research outcomes such as:

Contribution 1:

We carried out interdisciplinary research involving the spectrum of the data mining, and the digital business transformation, the decision-making competency and processes, where the research outcome is the result of this approach.

Contribution 2:

We carried out qualitative research that provides an understanding of the challenges and issues facing home appliances after-sales service businesses in the Kosovo market. Effective strategic solutions are proposed. The solutions proposed are such as: a) bring new technologies, real-time applications or analytical platforms, b) improve operational activities, etc.

Contribution 3:

We carried out quantitative research, comparative research on the implementation of the clustering and classification techniques in a home appliance after-sales service business. FarthestFirst clustering indicates in being the finest algorithm. The prediction is categorized as binary (two classes) classification due to the class attribute has two output values. The decision tree (J48) data mining algorithm gives the best results.

Contribution 4:

We conducted a quantitative research, regression analysis on service quality attributes and examine the impact of service quality on customer satisfaction, a nonfinancial performance measure for managerial decision-making. The study found that time taken for resolving the complaint impacts customer satisfaction, a nonfinancial performance measure for managerial decision-making.

Contribution 5:

The main contribution of our research is bringing a proposed model with an implementation of a recommendation system, as a portion of data mining system functionalities - as an outcome - to alleviate the process of streamlining decision-making along the business digital transformation journey in after-sales service businesses which help take appropriate decisions. The decision tree, as the best classification algorithm, is chosen in the implementation of recommendation system. The recommendation system will facilitate processes between the counterparts the after-sales services businesses. The implemented recommendation system presents data that mainly predicts the probability of time taken for resolving the complaint and digitizes manual processes of data presentation.

Contribution 6:

The evaluation results prove the impact of the implementation of data mining approach by facilitating the process of decision-making due to reliable product repairing predictions and increase the business' productivity and efficiency.

6 CONCLUSIONS

This section covers the conclusions of the research study. This thesis proposes data mining by demonstrating its aim and digital transformation as the modern trend impacting the entire business.

The IoT, cloud and big data analytics, as the most appropriate technologies, have been introduced as accelerators of business digital transformation. Additionally, Weka, RapidMiner, Orange and R have been identified as key data mining tools. Moreover, data mining, machine learning, and natural language processing are the primary techniques big data and advanced analytics. Furthermore, businesses will be more customer-focused by moving their services forward and saving time for their customers by reinforcing their processes. Data mining technology is embraced across many industries because of its benefits. The primary lines of business where data mining is applied are retail, e-commerce, banking and manufacturing.

The objective of the case study was to identify the best clustering and classification algorithms by conducting a practical comparative study of a variety of algorithms. The clustering algorithms are compared with different parameters, such as incorrectly clustered instances, iterations, time taken, etc., for prediction of repairs and installations. FarthestFirst clustering is the best algorithm amongst other algorithms considered, including Canopy, SimpleKMeans, FilteredClusterer and MakeDensityBasedClusterer. The FarthestFirst algorithm took a minimal amount of time to construct a cluster with an accuracy of 88.46%. The FarthestFirst clustering consists of two clusters, where the class attribute - guarantee of products - can be identified with or with no guarantee. The FarthestFirst algorithm demonstrates that it is a good choice for the after-sales services business. The results indicate that the data mining approach demonstrates the effect on improved decision-making.

Moreover, the implementation of the data mining approach involves discovering the best classification technique by comparing several classification algorithms in practice. The eight attributes applied include: authorized service, brand, category, repaired by age, guarantee, action, sum, guarantee and days to repair. Naive Bayes, J48, random forest, K-Nearest Neighbor, and logistic regression classification algorithms have been applied in the dataset to predict repairs based on guarantee period. Among the other algorithms, J48 has proved the most accurate and give the lowest error. The benefit of the built binary prediction model is the effect on improving the customer experience and increasing efficiencies as a result of the business prediction analysis.

Based on regression analysis of service quality attributes and examine the impact of service quality on customer satisfaction, time taken for resolving the complaint impacts customer satisfaction, a nonfinancial performance measure for managerial decision-making.

Moreover, the recommendations given make the businesses are aware of the possible solutions for the issues they are facing with which are indicated and found by the qualitative research study. Finally, the proposed solutions make businesses implement right strategic decision-making which positively impact customer satisfaction.

In this stage, the study contributes to the implementation of the data mining approach and the creation of a model for other businesses with similar concerns. With the case implementation, we obtain a practical sample of the implementation of the recommendation system, as a portion of data mining system functionalities, on analyzing information, predicting, and monitoring processes in business.

More importantly, the positive impact of the main attributes of service quality - time taken for resolving the complaint - is evaluated. The evaluation proves that the implementation data mining approach improves and impacts decision-making and increases efficiency due to business prediction analysis.

6.1 Model Validation and Managerial Implication

After completing the research study, we interviewed the CEO of the largest home appliance after-sales services business in the region where our model and recommendation system is presented and applied. The interview and discussions resulted positively. The business provided us with a recommendation letter to prove that the recommendation system has a positive significant impact on improving decision-making and expresses interest in participation of the next phases of the research. More importantly, the evaluation results prove the impact of the implementation of data mining approach by facilitating the process of decision-making due to reliable product repairing predictions and increase the business' productivity and efficiency.

The data mining approach provides an opportunity for managers make strategic decisions and data mining technologies are important for improving business performance. This thesis helps businesses, managers and researchers, as key beneficiaries, figure out the positive side of the data mining approach, as a modern approach, along the business digital transformation. The recommendations provided will ensure that home appliances after-sales service managers are aware of the advantages of implementing the data mining approach. The benefit of the data mining approach implementation in case study deals with analyzing information and predicting repairs and installations in the business.

The goal of the research in the implementation of data mining techniques is to attract and foster businesses and managers in applying the data mining approach along the business digital transformation journey.

6.2 Limitations of the Research

Our study has five limitations such as:

- The implementation of the analytics systems and intelligent applications, a proposed model, endures a challenge to be implemented in the home appliances after-sales service businesses in the Kosovo market
- There is no prior study, implementation, or application on the Kosovo market, not even in the region, which would help even for comparative studies
- The evaluation of the impact of data mining approach is limited only on the impact of nonfinancial performance measures, such as service quality and customer satisfaction. Furthermore, financial performance studies were not authorized by businesses
- The examination or quantitative research was conducted for just five service quality impact on customer satisfaction.
- Digitizing and automating more processes in the data mining system can be enhanced.

6.3 Recommendations for Future Research

Here are some recommendations for future studies:

- Conduct further studies to examine additional factors and attributes. Thus, it is recommended that further studies be conducted with more independent attributes and additional factors in order to improve the quality of the findings in the context of nonfinancial and financial performance measures for managerial decision-making
- Collect and analyze data using a larger sample size and gain feedback from different businesses to increase the chance of findings of data mining approach improving decision-making processes of a business and decision-making competency
- Consider more comprehensive sampling design with a much larger sample size of different business types in order to gather complementary information
- The new research aim leads to a practical implementation of the other data mining techniques, not limited to clustering, classification, and regression techniques
- In the long term, the aim is to focus on new perspectives and the intentions of data mining in pushing digital transformation by businesses, especially the small ones, and to be based on empirical studies.

7 PUBLICATIONS AND CONFERENCES

Mydyti, H., & Kadriu, A. (2021, August). Data Mining Approach Improving Decision-Making Competency Along the Business Digital Transformation Journey: A Literature Review. In International Conference for Emerging Technologies in Computing (pp. 129-146). Springer, Cham.

Mydyti, H. (2021). Data Mining Approach Improving Decision-Making Competency along the Business Digital Transformation Journey: A Case Study–Home Appliances after Sales Service. *SEEU Review, 16*(1), 45-65.

Mydyti, H., & Kadriu, A. (2021, June). The Data Mining Approach: A Case Study-Clustering Algorithms for After Sales Service. In *2021 10th Mediterranean Conference on Embedded Computing (MECO*) (pp. 1-6). IEEE.

Mydyti, H., & Kadriu, A. (2021). The Impact of Chatbots in Driving Digital Transformation. *International Journal of E-Services and Mobile Applications (IJESMA), 13*(4), 88-104.

8 BIBLIOGRAPHY

Abdelaal M., Khater, M. and Zaki, M. (2018). Digital Business Transformation and Strategy: What Do We Know So Far? *Cambridge Service Alliance*, 10, pp.1-35.

Ali, J., Khan, R., Ahmad, N. and Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI), 9*(5), p.272.

Alonso, J., Escalante, M. and Orue-Echevarria Arrieta L. (2016). Transformational Cloud Government (TCG): Transforming Public Administrations with a Cloud of Public Services. *Procedia Computer Science*, 97, pp.43-52.

Altekin, F. T., Aylı, E. and Şahin, G. (2017). After-sales services network design of a household appliances manufacturer. *Journal of the Operational Research Society, 68*(9), pp.1056-1067.

Attwal, K. P. S. and Dhiman, A. S. (2020). Exploring Data Mining Tool-Weka And Using Weka To Build And Evaluate Predictive Models. *Advances and Applications in Mathematical Sciences, 19*(6), pp.451-469.

Avand, M., Janizadeh, S., Naghibi, S. A., Pourghasemi, H. R., Khosrobeigi Bozchaloei, S. and Blaschke, T. (2019). A comparative assessment of random forest and k-nearest neighbor classifiers for gully erosion susceptibility mapping. *Water, 11*(10), p.2076.

Avinash B.M., Akarsha B.M. (2017). Big data analytics For E-Commerce – Its Impact on Value Creation. *International Journal of Advanced Research in Computer and Communication Engineering, 6*(12), pp.181-188.

Balashova, E. and Gromova, E. (2018). Russian industrial sector in the conditions of the Fourth Industrial Revolution. In *IOP Conference Series: Materials Science and Engineering, 404*(1), p.012014. IOP Publishing.

Baştanlar, Y. and Özuysal, M. (2014). Introduction to machine learning. In *miRNomics: MicroRNA Biology and Computational Analysis*, 1107, pp.105-128. Humana Press, Totowa, NJ.

Benlian, A., Kettinger, W. J., Sunyaev, A. and Winkler, T. J. (2018). Special Section Introduction: The Transformative Value of Cloud Computing: A Decoupling, Platformization, and Recombination Theoretical Framework. *Journal of Management Information Systems. 35*(3), pp.719-739.

Betz, C., Olagunju, A. O. and Paulson, P. (2016). The impacts of digital transformation, agile, and DevOps on future IT curricula. In *Proceedings of the 17th Annual Conference on Information Technology Education* (pp. 106-106).

Bhosale, S., Patil, R. & Khupe, Hasham. (2020). Chatbot Application Using Artificial Intelligence. *International Journal of Advance and Innovative Research. 7*(1), pp.2394 – 7780.

Boström, E. and Celik, O. C. (2017). Towards a Maturity Model for Digital Strategizing: A qualitative study of how an organization can analyze and assess their digital business strategy. Academic Press. IT Management Master Thesis submitted to the Department of informatics, UMEA Universitet Umea.

Breitkreutz, D. and Casey, K. (2008). Clusterers: a comparison of partitioning and density-based algorithms and a discussion of optimisations. Technical Report 11999, James Cook University, Townsville, Australia, 14 p.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N. and Trench, M. (2017). Artificial intelligence: The next digital frontier. *McKinsey Global Institute*, pp.1-80.

Bumblauskas, D., Nold, H., Bumblauskas, P., and Igou, A. (2017). Big data analytics: transforming data to action. *Business Process Management Journal*, pp.703-720.

Camilleri, M. A. (2017a). Market Segmentation, Targeting and Positioning. In *book: Travel Marketing, Tourism Economics and the Airline Product: An Introduction to Theory and Practice*, pp.69-83. Springer, Cham, Switzerland.

Camilleri, M. A. (2017b). Understanding Customer Needs and Wants. In *book: Travel Marketing, Tourism Economics and the Airline Product: An Introduction to Theory and Practice*. Springer, Milan, Italy.

Capozzoli, A. and Cerquitelli, T. and Piscitelli, M. (2016). Enhancing energy efficiency in buildings through innovative data analytics technologies. *Next Generation Platforms for Intelligent Data Collection*; Dobre, C., Xhafa, F., Eds, pp.353-389.

Cascio, W. and Montealegre, R. (2016). How Technology is Changing Work and Organizations. *Annual Review of Organizational Psychology and Organizational Behavior, 3*(1), pp.349-375.

Castelo-Branco, F., Reis, J. L., Vieira, J. C. and Cayolla, R. (2020). Business Intelligence and Data Mining to Support Sales in Retail. In *Marketing and Smart Technologies* (pp. 406-419). Springer, Singapore.

Chari, N., Diallo, C., Venkatadri, U. and Ait-Kadi, D. (2012). EOL Warranty and Spare Parts Manufacturing Strategy using New and Reconditioned Parts. In *54th Annual Conference of the Canadian Operational Research Society and the 10th International Conference on Multiple Objective Programming and Goal Programming CORS/MOPGP, Niagara Falls, Canada.*

Chen, D., Sain, S. L. and Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management, 19*(3), pp.197-208.

Chen, W. J., Kamath, R., Kelly, A., Lopez, H. H. D., Roberts, M. and Yheng, Y. P. (2015). Systems of insight for digital transformation*: Using IBM operational decision manager advanced and predictive analytics*. IBM Redbooks.

Çığşar, B. and Ünal, D. (2019). Comparison of data mining classification algorithms determining the default risk. *Scientific Programming*.

Coetzee, L. and Eksteen, J. (2011). The Internet of Things-promise for the future? An introduction. In *2011 IST-Africa Conference Proceedings* (pp. 1-9). IEEE.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, pp.297–334.

Dåderman, A. and Rosander, S. (2018). Evaluating frameworks for implementing machine learning in signal processing: A comparative study of CRISP-DM, SEMMA and KDD. Skolan För Elektroteknik Och Datavetenskap.

Dai, H. N., Wang, H., Xu, G., Wan, J. and Imran, M. (2020). Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. *Enterprise Information Systems, 14*(9-10), pp.1279-1303.

Davenport, T., Guha, A., Grewal, D. and Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), pp.24-42.

Deloitte Digital Maturity Model (2018). Achieving digital maturity to drive growth.

Deloitte Artificial Intelligence. (2018). *Chatbots Point of View*. . https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/deloitte-analytics/deloitte-nl-chatbots-moving-beyond-the-hype.pdf.

Dias, J., Khanna, S., Paquette, C., Rohr, M., Seitz, B., Singla, A., Sood, R. van Ouwerkerk, J. (2017). Introducing the next-generation operating model. *McKinsey & Company, New York, 41*.

Domazet, I.S., Stošić, I., & Lazić, M. (2018). Competitive relations in the aftersales market of major appliances in Serbia. *Economic Analysis, 51*(1-2), p.47.

Dremel, C., Wulf, J., Herterich, M. M., Waizmann, J. C. and Brenner, W. (2017). How AUDI AG Established Big Data Analytics in Its Digital Transformation. *MIS Quarterly Executive, 16*(2).

Durugbo, C. M. (2020). After-sales services and aftermarket support: a systematic review, theory and future research directions. *International Journal of Production Research, 58*(6), pp.1857-1892.

Dušanka, D., Darko, S., Srdjan, S., Marko, A. and Teodora, L. (2017). A comparison of contemporary data mining tools. In XVII International Scientific Conference on Industrial Systems (IS'17), Novi Sad, Serbia, October, 4(6).

Elgendy, N. and Elragal, A. (2014). Big data analytics: a literature review paper. In *Industrial conference on data mining* (pp. 214-227). Springer, Cham.

Elgendy, N. and Elragal, A. (2016). Big data analytics in support of the decision-making process. *Procedia Computer Science, 100*, pp.1071-1084.

Enshassi, A. (2008). The contractor – subcontractor relationship: the general contractor's view. In *Proceedings from International Conference on Building Education and Research (BEAR)*, pp.1520 – 1527.

Eremina, Y., Lace, N. and Bistrova, J. (2019). Digital Maturity and Corporate Performance: The Case of the Baltic States. *Journal of Open Innovation: Technology, Market, and Complexity, 5*(3), p.54.

Esmaeilpour, M. (2016). Analyzing after Sales Services in House Appliances Products and Measuring Customers Satisfaction: A Survey in Bushehr, Iran. *Journal of Harmonized Research in Management, 2*(2), pp.204-215.

Fabian, R. and Alexandru-Nicolae, M. (2009). Natural language processing implementation on Romanian ChatBot. In *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering* (No. 5). WSEAS.

Felch, V., Asdecker, B. and Sucky, E. (2019). Maturity models in the age of Industry 4.0–Do the available models correspond to the needs of business practice? *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 5165–5174).

Fowke, R. A. (2010). Performance Measures for Managerial Decision Making: Performance Measurement Synergies in Multi-Attribute Performance Measurement Systems. *Dissertations and Theses*. Portland State University. Paper 164. https://doi.org/10.15760/etd.164.

Frey, A., Trenz, M. and Veit, D. (2017). The Role of Technology for Service Innovation in Sharing Economy Organizations – a Service-Dominant Logic Perspective. In *Proceedings of the 25th European Conference on Information Systems (ECIS)*, pp.1885-1901.

Friedman, C., Rindflesch, T. C. and Corn, M. (2013). Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. *Journal of biomedical informatics, 46*(5), pp.765-773.

Galetsi, P., Katsaliaki, K. and Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques, and prospects. *International Journal of Information Management, 50,* pp.206-216.

Genzorová, T., Corejova, T. and Stalmašeková, N. (2019). How Digital Transformation can influence business model, Case study for transport industry. *Transportation Research Procedia*, 40, pp.1053-1058. DOI:10.1016/j.trpro.2019.07.147.

Gergin, Z., Üney-Yüksektepe, F., Buldanlı, B. B., Şahin, T., Elçi, L. and Ekinci, M. (2019). Data Mining Approach for Quality Control Process Improvement.

Gimpel, H., Hosseini, S., Huber, R. X. R., Probst, L., Röglinger, M. and Faisst, U. (2018). Structuring Digital Transformation: A Framework of Action Fields and its Application at ZEISS. *Journal of Information Technology, Theory and Application, 19*(1), p.3.

Gnanapriya, S., Freeda, R. and Sowmiya, M. (2017). Evaluation of Clustering Capability Using Weka Tool. *IJIET, 1*(8).

Golrizgashti, S. and Hejaz, A. & F., Kimia. (2020). Assessing after-sales services quality: integrated SERVQUAL and fuzzy Kano's model. *International Journal of Services Economics and Management, 11* (2), pp.137-166.

Gunasekaran, A., Yusuf, Y. Y., Adeleye, E. O., Papadopoulos, T., Kovvuri, D. and Geyi, D. A. G. (2019). Agile manufacturing: an evolutionary review of practices. *International Journal of Production Research, 57*(15-16), pp.5154-5174.

Hagberg, J., Sundström, M. and Egels-Zandén, N. (2016). The digitalization of retailing: an exploratory framework. *International Journal of Retail & Distribution Management*. 44. 694-712. DOI:10.1108/IJRDM-09-2015-0140.

Halimu, C., Kasem, A. and Newaz, S. S. (2019). Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification. In *Proceedings of the 3rd international conference on machine learning and soft computing* (pp. 1-6).

Hammermann, A. and Mohnen, A. (2012). Who Benefits from Benefits? Empirical Research on Tangible Incentives. *Review of Managerial Science, 8*(3), pp.327-350.

Harding, J. A., Shahbaz, M. and Kusiak, A. (2006). Data mining in manufacturing: a review. *Journal of Manufacturing Science and Engineering-transactions of The Asme - J MANUF SCI ENG*, 128.

Hasheminejad, S. M. H. and Khorrami, M. (2018). Data mining techniques for analyzing bank customers: A survey. *Intelligent Decision Technologies, 12*(3), pp.303-321.

Hausberg, J. P., Liere-Netheler, K., Packmohr, S., Pakura, S. and Vogelsang, K. (2019). Research streams on digital transformation from a holistic business perspective: a systematic literature review and citation network analysis. *Journal of Business Economics, 89*(8-9), pp.931-963.

Hie, B. P. (2019). Impact of Transforming Organizational Culture and Digital Transformation Governance toward Digital Maturity in Indonesian Banks. *International Review of Management and Marketing, 9*(6), pp.51-57.

Hinchliffe, D. et al. (2020). Case studies and approaches to building partnerships between the informal and the formal sector for sustainable e-waste management. https://www.step-initiative.org/files/_documents/publications/Partnerships-between-the-informal-and-the-formal-sector-for-sustainable-e-waste-management.pdf

Hossin, M. and Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*(2), p.1.

Ikenna, O. et al. (2014). Review of Data Mining as a Tool for Organization's Growth and Productivity. *IJCSMC, 9*(3), pp.284-290.

International Labour Office (ILO). (2010). A Skilled Workforce for Strong, Sustainable and Balanced Growth. A G20 Training Strategy. *Geneva: International Labor Office.*

Ismail, M., Ibrahim, M. M., Sanusi, Z. M. and Nat, M. (2015). Data mining in electronic commerce: benefits and challenges. *International Journal of Communications, Network and System Sciences, 8*(12), p.501.

Ivanov, S. H. and Webster, C. (2017). Adoption of robots, artificial intelligence and service automation by travel, tourism and hospitality companies–a cost-benefit analysis. In *International Scientific Conference Contemporary Tourism – Traditions and Innovations* (pp. 190-203). Cambridge Scholars Publishing.

Jadhav, S. D. and Channe, H. P. (2016). Comparative study of K-NN, naive Bayes and decision tree classification techniques. *International Journal of Science and Research (IJSR), 5*(1), pp.1842-1845.

Jain, A., Kulkarni, G. and Shah, V. (2018). Natural language processing. *International Journal of Computer Sciences and Engineering, 6*(1), pp.161-167.

Janošcová, R. (2016). Mining Big Data in WEKA. *11th International Workshop on Knowledge Management - IWKM*, Bratislava, Slovakia.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), pp.255-260.

Jovic, A., Brkic, K. and Bogunovic, N. (2014). An overview of free software tools for general data mining. In 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 1112-1117). IEEE.

Kaczorowska-Spychalska, D. (2019). How chatbots influence marketing. *Management. 23*(1), pp.251-270.

Kang, J. and Hustvedt, G. (2013). Building Trust between Consumers and Corporations: The Role of Consumer Perceptions of Transparency and Social Responsibility. *Journal of* Business Ethics, *125*(2), pp.253-265.

Kaur, G. and Chhabra, A. (2014). Improved J48 classification algorithm for the prediction of diabetes. *International journal of computer applications, 98*(22).

Kaur, K. and Dhiman, S. (2016). Review of data mining with Weka tool. *International Journal of Computer Sciences and Engineering (IJCSE), 4*(8), pp.41-44.

Kaur, R. and Jagdev, G. (2017). Big Data in retail sector-an evolution that turned to a revolution. *International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 4*(4), pp.43-52.

Kumar, A., Ingle, Y. S., Pande, A. and Dhule, P. (2014). Canopy clustering: a review on pre-clustering approach to K-means clustering. *International Journal of Innovations & Advancement in Computer Science (IJIACS), 3*(5), pp.22-29.

Kumar, M. (2013). An optimized farthest first clustering algorithm. In *2013 Nirma University International Conference on Engineering (NUiCONE)* (pp. 1-5). IEEE.

Lacerda, A. R., & Aguiar, C. S. (2019). FLOSS FAQ chatbot project reuse: how to allow nonexperts to develop a chatbot. In *Proceedings of the 15th International Symposium on Open Collaboration* (pp. 1-8).

Laitala, K., Klepp, I., Haugrønning, V., Throne-Holst, H. and Strandbakken, P. (2020). Increasing repair of household appliances, mobile phones and clothing: Experiences from consumers and the repair industry. *Journal of Cleaner Production, 282*, 125349.

Laura, B., Dainelli, F., Giunta, F. and Simoni, L. (2019). Are non-financial KPIs in annual reports really "key”? An investigation of company disclosure and analyst reports in the UK. *ICAS.*

Lele, M.M. (1997). After-sales service - necessary evil or strategic opportunity? *Managing Service Quality: An International Journal, 7*(3), pp.141-145.

Ltifi, H., Benmohamed, E., Kolski, C. and Ayed, M. B. (2016). Enhanced visual data mining process for dynamic decision-making. *Knowledge-Based Systems*, 112, pp.166-181.

Lundkvsit, E. (2014). Decision Tree Classification and Forecasting of Pricing Time Series Data. Master’s Degree Project. KTH, School of Electrical Engineering (EES), Automatic Control.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, pp.46-60.

Manita, R., Elommal, N., Baudier, P. and Hikkerova, L. (2020). The digital transformation of external audit and its impact on corporate governance. *Technological Forecasting and Social Change, 150*, 119751.

Markusoski, L., Zdravkoski, I., Andonovski, M., Jovanoska, A. (2019). Knowledge Discovery Da-tabases (KDD) Process in Data Mining. In *International Conference Proceeding* (Pp. 529-539). Faculty of Economic Prilep.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Orallo, J. H., Kull, M., Lachiche, N., Ramírez-Quintana, M. J. and Flach, P. A. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering.*

Mazumdar, A. and Alharahsheh, H. (2019). Insights of Trends and Developments in Cloud Computing. *South Asian Research Journal of Engineering and Technology 1*(3), pp.98-107. South Asian Research Publication.

McCormick, J., Doty, C. A., Sridharan, S., Curran, R., Evelson, B., Hopkins, B., Little, C., Leganza, G., Purcell, B. and Miller, E. (2016). Predictions 2017: artificial intelligence will drive the insights revolution. *Forrester Research for Customer Insights Professionals*.

Memon, M. A., Soomro, S., Jumani, A. K. and Kartio, M. A. (2017). Big data analytics and its applications. *arXiv* preprint arXiv:1710.04135.

Meiryani, A. S. (2019). Database Management System. *International Journal of Scientific & Technology Research, 8*(6).

Milovic, B. and Milovic, M. (2012). Prediction and decision making in health care using data mining. *International Journal of Public Health Science (IJPHS), 1*(2), pp.69-78.

Moe, N. B., Aurum, A. and Dybå, T. (2012). Challenges of shared decision-making: A multiple case study of agile software development. *Information and Software Technology, 54*(8), pp.853-865.

Mohan, R. (2019). The Chat bot revolution and the Indian HR Professionals. *International journal of information and computer science. 6*(3), pp.489-499.

Mohd Selamat, S. A., Prakoonwit, S., Sahandi, R., Khan, W. and Ramachandran, M. (2018). Big data analytics - A review of data-mining models for small and medium enterprises in the transportation sector. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(3), e1238.

Morakanyane, R., Grace, A. A. and O'Reilly, P. (2017). Conceptualizing Digital Transformation in Business Organizations: A Systematic Review of Literature. In *30th Bled eConference: Digital Transformation – from Connecting Things to Transforming our Lives* (p. 427-444).

Murali, S. (2016). Assessment of impact of after sales service attributes on customer satisfaction for a home-appliances manufacturing firm. *International Journal of Enterprise Network Management*, 7, pp.27-41.

Murali, S., Pugazhendhi, S. and Muralidharan, C. (2015). Evaluation of performance of after sales service - a comparative study involving home appliances manufacturing firms. *ARPN Journal of Engineering and Applied Sciences, 20*(13), pp.5614-5619.

Murali, S., Pugazhendhi, S. and Muralidharan, C. (2016). Modelling and investigating the relationship of after sales service quality with customer satisfaction, retention and loyalty–a case study of home appliances business. *Journal of retailing and consumer services*, *30*, pp.67-83.

Mydyti, H. and Kadriu, A. (2020). Using Internship Management System to Improve the Relationship between Internship Seekers, Employers and Educational Institutions. *ENTRENOVA-ENTerprise REsearch InNOVAtion, 6*(1), pp.97-104.

Nagy, A., Băbăiłă, C. and Schneider (Ispas), A. (2011). Non-financial performance measurement of hotel activity. *Recent Researches in Tourism and Economic Development,* pp.225-228.

Neware, R. and Khan, A. (2018). Cloud Computing Digital Forensic challenges. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1090-1092). IEEE.

Nuruzzaman, M. and Hussain, O. K. (2018). A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)* (pp. 54-61). IEEE.

O'Donovan, P. and Leahy, K. and Bruton, K. and O'Sullivan, D. (2015). Big data In Manufacturing: A Systematic Mapping Study. *Journal of Big Data, 2*(1), pp.1-22.

O’Regan G. (2017). Agile Methodology. In: Concise Guide to Software Engineering. Undergraduate Topics in Computer Science. Springer, Cham.

OECD (2019). Artificial Intelligence in Society. OECD Publishing, Paris.

OECD and the United Kingdom Department for Business, Energy and Industrial Strategy (BEIS). (2018). *Implications of the Digital Transformation for the Business*. http://www.oecd.org/sti/ind/digital-transformation-business-sector-summary.pdf.

Oliveira, C., Martins, A., Camilleri, M. A. and Jayantilal, S. (2021). Using the Balanced Scorecard for Strategic Communication and Performance Management. In *Camilleri, M.A. (Ed.) Strategic Corporate Communication in the Digital Age*, Emerald, Bingley, UK.

Oliff, H. and Liu, Y. (2017). Towards industry 4.0 utilizing data-mining techniques: a case study on quality improvement. *Procedia CIRP, 63*, pp.167-172.

Olson, D. L. and Delen, D. (2008). Advanced data mining techniques. *Springer Science & Business Media*.

Orfanidis, P. (2018). Prominence of Big Data in the Digital Transformation Era. [Unpublished master thesis dissertation]. School of Economics, Business Administration & Legal Studies, Thessaloniki – Greece.

Othman, B. A., Harun, A., De Almeida, N. M. and Sadq, Z. M. (2020). The effects on customer satisfaction and customer loyalty by integrating marketing communication and after sale service into the traditional marketing mix model of Umrah travel services in Malaysia. *Journal of islamic marketing*.

Oviedo, A. M., Thomas, M. R. and Karakurum-Özdemir, K. (2009). Economic Informality Causes, Costs, and Policies—A Literature Survey. *World Bank Working Paper*. No. 167.

Ozdogan, B., Gacar, A. and Aktaş, H. (2017). Digital Agriculture Practices in the Context of Agriculture 4.0. *Journal of Economics, Finance and Accounting, 4*(2), pp.186-193.

Padmavaty, V., Geetha, C. and Priya, N. (2020). Analysis of Data Mining Tool Orange. *International Journal of Modern Agriculture, 9*(4), pp.1146-1150.

Pal, S. N., & Singh, D. (2019). Chatbots and virtual assistant in Indian banks. *Industrija*. *47*(4), pp.75-101.

Palanica, A., Flaschner, P., Thommandram, A., Li, M. and Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of medical Internet research. 21(*4), e12887.

Pagalday, G., Zubizarreta, P., Uribetxebarria, J., Erguido, A. and Castellano, E. (2018). Efficient development and management of after sale services. *Procedia Manufacturing, 19*, pp.18-25.

Palmer, A., Jiménez, R. and Gervilla, E. (2011). Data mining: Machine learning and statistical techniques. *Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.),* pp.373-396.

Parasuraman, A., Berry, L. L., and Zeithaml, V. A. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), pp.12–40.

Parida, V., Sjödin, D. and Reim, W. (2019). Reviewing Literature on Digitalization, Business Model Innovation, and Sustainable Industry: Past Achievements and Future Promises. *Sustainability, 11*(2), p.391.

Paschou, T., Rapaccini, M., Peters, C., Adrodegari, F. and Saccani, N. (2019). Developing a Maturity Model for Digital Servitization in Manufacturing Firms. In *International Joint conference on Industrial Engineering and Operations Management* (pp. 413-425). Springer, Cham.

Patel, S. and Patel, A. (2019). Performance Analysis and Evaluation of Clustering Algorithms. *International Journal of Innovative Technology and Exploring Engineering, 6S2*(8), pp.179-183.

Alireza, F., Fatemeh, B. and Pegah, M. (2011). How after-sales service quality dimensions affect customer satisfaction. *African Journal of Business Management, 5*(17), pp.7658-7664.

Pflaum, A. A. and Gölzer, P. (2018). The IoT and digital transformation: toward the data-driven enterprise. *IEEE pervasive computing, 17*(1), pp.87-91.

Prabhu, C. S. R., Chivukula, A. S., Mogadala, A., Ghosh, R. and Livingston, L. J. (2019). Big Data Analytics. In *Big Data Analytics: Systems, Algorithms, Applications* (pp. 1-23). Springer, Singapore.

Preethi, M. and Vijayalakshmi, M. (2017). Data Mining in Banking Sector. *International Journal of Advanced Networking & Applications (IJANA), 8*(5), pp.1-4.

Pulakkazhy, S. and Balan, R. S. (2013). Data mining in banking and its applications-a review. *Journal of Computer Science, 9*(10), pp.1252-1259.

Quinn, M., Strauss, E. and Kristandl, G. (2014). The effects of cloud technology on management accounting and business decision-making. *Financial Management, 10*(6), pp.1-12.

Rachinger, M., Rauter, R., Müller, C., Vorraber, W. and Schirgi, E. (2019). Digitalization and its influence on business model innovation*. Journal of Manufacturing Technology Management* (pp. 1143-1160). Emerald Publishing Limited.

Ragab, A. H. M., Noaman, A. Y., Al-Ghamdi, A. S. and Madbouly, A. I. (2014). A comparative analysis of classification algorithms for students college enrollment approval using data mining. In *Proceedings of the 2014 Workshop on Interaction Design in Educational Environments* (pp. 106-113).

Ravichandran, S. (2017). Performance Comparison of Clustering Techniques In Gain Ratio Features Reduced Mcdr. *International Research Journal of Computer Science (IRJCS), 4*(8), pp.56-61.

Ramya, N., Kowsalya, A. And Dharanipriya, K. (2019). Service Quality and its Dimensions. *EPRA International Journal of Research and Development (IJRD), 4* (2), pp.38-41.

Ray, S. (2019). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.

Reddy, D. L. C. (2011). A review on data mining from past to the future. *International Journal of Computer Applications*, 975, p.8887.

Rehman, M. H., Yaqoob, I., Salah, K., Imran, M., Jayaraman, P. P. and Perera, C. (2019). The role of big data analytics in industrial Internet of Things. *Future Generation Computer Systems, 99*, pp.247-259.

Reis, J., Amorim, M., Melão, N. and Matos, P. (2018). Digital transformation: a literature review and guidelines for future research. In *World conference on information systems and technologies* (pp. 411-421). Springer, Cham.

Renjith, S., Sreekumar, A. and Jathavedan, M. (2020). Performance evaluation of clustering algorithms for varying cardinality and dimensionality of data sets. *Materials Today: Proceedings*, 27, pp.627-633.

Richad, R., Vivensius, G., Sfenrianto, S. and Kaburuan, E. R. (2019). Analysis of factors influencing millennial's technology acceptance of chatbot in banking industry in Indonesia. *International Journal of Civil Engineering and Technology, 10(*4), pp.1270-1281.

Ridge M., Johnston, K. A., O'Donovan, B. (2015). The use of big data analytics in the retail industries in South Africa. *African Journal of Business Management*, *9*(19), pp.688 - 703.

Roedder, N., Dauer, D., Laubis, K., Karaenke, P., Weinhardt, C. (2016). The digital transformation and smart data analytics: An overview of enabling developments and application areas. In *2016 IEEE International Conference on Big Data* (pp. 2795-2802). IEEE.

Rohanizadeh, S. S. and Bameni, M. M. (2009). A proposed data mining methodology and its application to industrial procedures. *Journal of Industrial Engineering 4*(2009), pp.37-50.

Rokach L., Maimon O. (2005). Decision Trees. In: Maimon O., Rokach L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_9

Rudnick, M., Riezebos, J., Powell, D. J. and Hauptvogel, A. (2020). Effective after-sales services through the lean servitization canvas. *International Journal of Lean Six Sigma, 5*(11), pp.943-956.

Saccani, N., Songini, L. and Gaiardelli, P. (2006). The role and performance measurement of after-sales in the durable consumer goods industries: An empirical study. *International Journal of Productivity and Performance Management*, *55*(3/4), pp.259 – 283.

Sadiku, M. N., Adebo, P. O. and Musa, S. M. (2018). Big Data in Business. *International Journal of Advanced Research in Computer Science and Software Engineering, 8*(1), pp.160-162.

Saeed, T. (2020). Data Mining for Small and Medium Enterprises: A Conceptual Model for Adaptation. *Intelligent Information Management, 12*(5), pp.183-197.

Sajana, T., Rani, C. S. and Narayana, K. V. (2016). A survey on clustering techniques for big data mining. *Indian journal of Science and Technology*, 9(3), pp.1-12.

Saputra, M. F. A., Widiyaningtyas, T. and Wibawa, A. P. (2018). Illiteracy classification using K means-Naïve Bayes algorithm. JOIV: *International Journal on Informatics Visualization, 2*(3), pp.153-158.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W and Lin, C. T. (2017). A review of clustering techniques and developments. *Neurocomputing, 267*, pp.664-681.

Schwertner, K. (2017). Digital transformation of business. *Trakia Journal of Sciences, 15*(1), pp.388-393.

Serran, W. (2018). Digital Systems in Smart City and Infrastructure: Digital as a Service. *Smart Cities, 1*(1), pp. 134-153.

Sharma, B. R., Kaur, D. and Manju, A. (2013). Review on Data Mining: Its Challenges, Issues and Applications. *International Journal of Current Engineering and Technology, 3*(2), pp.695-700.

Sharma, S., Osei-Bryson, K. M. and Kasper, G. M. (2012). Evaluation of an integrated Knowledge Discovery and Data Mining process model. *Expert Systems with Applications, 39*(13), pp.11335-11348.

Shrivastava, V. and Arya, N. (2012). A study of various clustering algorithms on retail sales data. *IJCCN, 1*(2).

Silva, M. and Lima, C. (2018). The Role of Information Systems in Human Resource Management. In *book: Management of Information Systems, 16*, pp.113-126.

Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J. and Cattrysse, D. (2015). Literature review of data mining applications in academic libraries. *The Journal of Academic Librarianship, 41*(4), pp.499-510.

Sima, V., Gheorghe, I., Subić, J. and Nancu, D. (2020). Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review. *Sustainability, 12*(10), p.4035.

Simon, A., Deo, M. S., Venkatesan, S. and Babu, R. (2016). An overview of machine learning and its applications. *International Journal of Electrical Sciences & Engineering, 1*(1), pp.22-24.

Singh, P. and Surya, A. (2015). Performance analysis of clustering algorithms in data mining in Weka. *International journal of advances in engineering & technology, 7*(6), p.1866.

Smutny, P. and Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education, 151*, p.103862.

Srivastava, U. and Gopalkrishnan, S. (2015). Impact of big data analytics on banking sector: Learning for Indian banks. *Procedia Computer Science, 50*, pp.643-652.

Sultan, J. and Bechter, C. (2019). Big data analytics In Islamic Banking. *International Academic Journal of Business Management, 6*(01), pp.21-31.

Tangirala, S. (2020). Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*. *International Journal of Advanced Computer Science and Applications (IJACSA), 11*(2), pp.612-619.

Tariq, H. I., Sohail, A., Aslam, U., Batcha, N. K. (2019). Loan Default Prediction Model Using Sample, Explore, Modify, Model, and Assess (SEMMA). *JCTN, 16*(8), pp.3489-3503.

Teichert, R. (2019). Digital transformation maturity: A systematic review of literature. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 67*(6), pp.1673-1687.

Telegescu, T. (2018). IT in the workspace–The need for digital transformation. In *Proceedings of the International Conference on Business Excellence 12*(1), pp.952-965. Sciendo.

Ukpabi, D. C., Aslam, B. and Karjaluoto, H. (2019). Chatbot adoption in tourism services: A conceptual exploration. *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and* Hospitality (pp. 105-121). Emerald Publishing Limited.

Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics, 12(*1), pp.525-546.

Uyoyo, E. (2014). Big data analytics And Its Application in E-Commerce. *E-Commerce Technologies, 1*. University of Bradford.

Valarmathy, N. and Krishnaveni, S. (2019). Performance Evaluation and Comparison of Clustering Algorithms used in Educational Data Mining. *International Journal of Recent Technology and Engineering (IJRTE), ISSN,* pp.2277-3878.

Valdez-de-Leon, O. (2016). A digital maturity model for telecommunications service providers. *Technology innovation management review, 6*(8).

Velmurugan, T. and Anuradha, C. (2016). Performance evaluation of feature selection algorithms in educational data mining. *Performance Evaluation, 5*(02), pp.131-139.

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research, 122*, pp.889-901.

Verma, A. (2019). Evaluation of classification algorithms with solutions to class imbalance problem on bank marketing dataset using WEKA. *International Research Journal of Engineering and Technology, 5*(13), pp.54-60.

Vivekananth, P. and Baptist, A. (2015). An Analysis of Big Data Analytics Techniques. *International Journal of Engineering and Management Research (IJEMR), 5*(5), pp.17-19.

Warner, K. S. and Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning, 52*(3), pp.326-349.

West, D. M. and Allen, J. R. (2018). *How artificial intelligence is transforming the world*. Report, Brookings Institution. April 24, 2018.

Wickramasinghe, V. and Mathusinghe, K. (2016). After-sales services of home appliances: evidence from Sri Lanka*. International Journal of Consumer Studies, 40*(1), pp.115-124.

Wiesböck, F. and Hess, T. (2018). Understanding the capabilities for digital innovations from a digital technology perspective (No. 1/2018). Arbeitsbericht.

Williams, C., Schallmo, D., Lang, K. and Boardman, L. (2019). Digital Maturity Models for Small and Medium-sized Enterprises: A Systematic Literature Review. In *ISPIM Conference Proceedings* (pp. 1-15). The International Society for Professional Innovation Management (ISPIM).

Zafar, M. H. and Ilyas, M. (2015). A clustering-based study of classification algorithms. *International journal of database theory and application, 8*(1), pp.11-22.

Zeinab, K. A. M. and Elmustafa, S. A. A. (2017). Internet of things applications, challenges and related future technologies. *World Scientific News, 2*(67), pp.126-148.

Zhao, X. (2018). A Study on the Application of Big Data Mining in E-commerce. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)* (pp. 1867-1871). IEEE.

Zhu, C., Idemudia, C. U. and Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. *Informatics in Medicine Unlocked, 17*, 100179.

Ziyadin, S., Suieubayeva, S. and Utegenova, A. (2019). Digital transformation in business. In *International Scientific Conference "Digital Transformation of the Economy: Challenges, Trends, New Opportunities"* (pp. 408-415). Springer, Cham.

Zain, M. S. I. M. and Rahman, S. A. (2017). Challenges of Applying Data Mining in Knowledge Management towards Organization. *International Journal of Academic Research in Business and Social Sciences (IJ-ARBSS), 7*(12), pp.405-412.

Zumstein, D. and Hundertmark, S. (2017). Chatbots--An Interactive Technology for Personalized Communication, Transactions and Services. *IADIS International Journal on WWW/Internet. 15*(1), pp.96-109.

9 APPENDIX

9.1 Qualitative Research

9.1.1 SERVQUAL dimensions

Dimension	Attributes of Service Quality
Reliability (RL)	1. Consistency of service quality
	2. Options and variety of services
	3. Supply of required spare parts
	4. Service delivery as promised
Responsiveness (RS)	1. Reasonable warranty policy
	2. Responsiveness to customer complaints
	3. Time taken for resolving the complaint
	4. Affordable cost of service
Assurance (AS)	1. Customer handling
	2. Professionalism of personnel
	3. Technical skills of personnel
	4. Interpersonal behavior of personnel
Empathy (EM)	1. Accessibility of personnel
	2. Ease of contacting personnel
	3. Comprehending customers’ needs
Tangible (TA)	1. Supply of service tools/equipment
	2. Accessibility of service center
	3. Complaint reporting facilities
	4. Quality & availability of technical service documentation
	5. Info and guidance available at service center

9.1.2 The challenges faced by the after-sales service businesses

SERVQUAL Dimension	**Challenges**
Responsiveness to customer complaints	Digital responsiveness to customer complaints
Time taken for resolving the complaint	- Long period of time to repair the product - Peak periods in which the repair load is enormous or unexpected as a result of various situations and conditions
Info and guidance available at service center	Lack of online resources increases number of complaints
Affordable cost of service	- Long term guarantee of products causes customers to buy new products instead of costly repairs of low-quality products with short life cycles or high-quality products with long life cycles - Differences on servicing cost/price between the authorized services and informality in the market
Complaint reporting facilities	Practice of traditional registration processing of complaints
Supply of required spare parts, Service delivery as promised	Provision of missing spare parts and end of life of spare parts
Reasonable warranty policy	Reasonable guarantee policy that affects guarantee period only for some brands and categories
Comprehending customers' needs	Lack of understanding the preferences and expectations
Accessibility of service center	In case the products need to be repaired in a service center, logistics or transportation from deep zones is a challenge
Technical skills of personnel	Kosovo economic and market instability cause employee leave, in particular technicians leave
Options and variety of services	Lack of information by customers on options and variety of services for the products with expired guarantee offered by an authorized after-sales service business causes reduction or lack of customers
Consistency of service quality	Informality in the (black) market causes the reduction or lack of customers

9.2 Data Mining Algorithms Results

9.2.1 Clustering Algorithms

9.2.1.1 Filtered Clusterer

```
=== Run information ===

Scheme:       Weka.clusterers.FilteredClusterer -F "Weka.filters.AllFilter " -W
Weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000
-min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "Weka.core.EuclideanDistance -R first-
last" -I 500 -num-slots 1 -S 10
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
Ignored:
              Guarantee
Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

FilteredClusterer using Weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -
periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A
"Weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10 on data
filtered through Weka.filters.AllFilter

Filtered Header
@relation database-Weka.filters.unsupervised.attribute.Remove-R8-
Weka.filters.AllFilter

@attribute AuthorizedService {ServiceCenter,Subcontractors}
@attribute Category
{TV+AV,AirConditioning,MDA,Other,SDAP,Mobile,IT,Accesories,Heating}
@attribute Brand {Other,BRAND 1,BRAND 2, BRAND 3,BRAND 4,BRAND 5,BRAND 6,BRAND
7,BRAND 8}
@attribute Action {Installation,Servicing,ChangeProduct}
@attribute RepairedByAge numeric
@attribute DaysToRepair numeric
@attribute Sum numeric

@data
```

```
Clusterer Model

kMeans
======

Number of iterations: 2
Within cluster sum of squared errors: 30809.502482559958

Initial starting points (random):

Cluster 0: ServiceCenter,MDA,BRAND 1,Servicing,3,3,0
Cluster 1: Subcontractors,MDA,BRAND 1,Servicing,7,6,0

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute                Full Data               0              1
                         (62431.0)       (34186.0)      (28245.0)
================================================================
AuthorizedService    ServiceCenter   ServiceCenter Subcontractors
Category                       MDA             MDA            MDA
Brand                        BRAND 1           BRAND 1           BRAND 1
Action                   Servicing       Servicing      Servicing
RepairedByAge               2.1766          1.9427         2.4597
DaysToRepair                7.0186          7.6616         6.2404
Sum                         8.2631          4.4437        12.8858



Time taken to build model (full training data) : 0.45 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      34186 ( 55%)
1      28245 ( 45%)


Class attribute: Guarantee
Classes to Clusters:

     0     1  <-- assigned to cluster
 28891 27601 | Yes-guarantee
  5295   644 | No-guarantee

Cluster 0 <-- No-guarantee
Cluster 1 <-- Yes-guarantee

Incorrectly clustered instances :   29535.0   47.3082 %
```

9.2.1.2 MakeDensityBasedClusterer

```
=== Run information ===

Scheme:       Weka.clusterers.MakeDensityBasedClusterer -M 1.0E-6 -W
Weka.clusterers.SimpleKMeans -- -init 0 -max-candidates 100 -periodic-pruning 10000
-min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "Weka.core.EuclideanDistance -R first-
last" -I 500 -num-slots 1 -S 10
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
Ignored:
              Guarantee
Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===

MakeDensityBasedClusterer:

Wrapped clusterer:
kMeans
======

Number of iterations: 2
Within cluster sum of squared errors: 30809.502482559958

Initial starting points (random):

Cluster 0: ServiceCenter,MDA,BRAND 1,Servicing,3,3,0
Cluster 1: Subcontractors,MDA,BRAND 1,Servicing,7,6,0

Missing values globally replaced with mean/mode

Final cluster centroids:
                                       Cluster#
Attribute                Full Data            0              1
                         (62431.0)    (34186.0)      (28245.0)
================================================================
AuthorizedService    ServiceCenter  ServiceCenter Subcontractors
Category                       MDA            MDA            MDA
Brand                           BRAND 1            BRAND 1            BRAND 1
Action                   Servicing      Servicing      Servicing
RepairedByAge               2.1766         1.9427         2.4597
DaysToRepair                7.0186         7.6616         6.2404
Sum                         8.2631         4.4437        12.8858

Fitted estimators (with ML estimates of variance):
Cluster: 0 Prior probability: 0.5476
Attribute: AuthorizedService
```

```
Discrete Estimator. Counts =  34187 1  (Total = 34188)
Attribute: Category
Discrete Estimator. Counts =  4248 536 23470 36 5318 421 6 2 158  (Total = 34195)
Attribute: Brand
Discrete Estimator. Counts =  893 20567 941 808 3322 1728 1623 3080 1233  (Total =
34195)
Attribute: Action
Discrete Estimator. Counts =  950 33196 43  (Total = 34189)
Attribute: RepairedByAge
Normal Distribution. Mean = 1.9427 StdDev = 1.324
Attribute: DaysToRepair
Normal Distribution. Mean = 7.6616 StdDev = 12.1212
Attribute: Sum
Normal Distribution. Mean = 4.4437 StdDev = 17.8334

Cluster: 1 Prior probability: 0.4524

Attribute: AuthorizedService
Discrete Estimator. Counts =  1 28246  (Total = 28247)
Attribute: Category
Discrete Estimator. Counts =  29 1200 26770 69 166 6 1 1 12  (Total = 28254)
Attribute: Brand
Discrete Estimator. Counts =  268 26282 215 453 248 32 209 306 241  (Total = 28254)
Attribute: Action
Discrete Estimator. Counts =  1024 27211 13  (Total = 28248)
Attribute: RepairedByAge
Normal Distribution. Mean = 2.4597 StdDev = 1.3808
Attribute: DaysToRepair
Normal Distribution. Mean = 6.2404 StdDev = 13.0551
Attribute: Sum
Normal Distribution. Mean = 12.8858 StdDev = 14.7311


Time taken to build model (full training data) : 1.06 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      34178 ( 55%)
1      28253 ( 45%)


Log likelihood: -12.30436


Class attribute: Guarantee
Classes to Clusters:

     0     1  <-- assigned to cluster
 28884 27608 | Yes-guarantee
  5294   645 | No-guarantee

Cluster 0 <-- No-guarantee
Cluster 1 <-- Yes-guarantee

Incorrectly clustered instances :   29529.0  47.2986 %
```

9.2.1.3 SimpleKMeans

```
=== Run information ===

Scheme:       Weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-
pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A
"Weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
Ignored:
              Guarantee
Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===


kMeans
======

Number of iterations: 2
Within cluster sum of squared errors: 30809.502482559958

Initial starting points (random):

Cluster 0: ServiceCenter,MDA,BRAND 1,Servicing,3,3,0
Cluster 1: Subcontractors,MDA,BRAND 1,Servicing,7,6,0

Missing values globally replaced with mean/mode

Final cluster centroids:
                                         Cluster#
Attribute                Full Data              0              1
                         (62431.0)      (34186.0)      (28245.0)
================================================================
AuthorizedService    ServiceCenter  ServiceCenter Subcontractors
Category                       MDA            MDA            MDA
Brand                            BRAND 1          BRAND 1          BRAND 1
Action                   Servicing      Servicing      Servicing
RepairedByAge               2.1766         1.9427         2.4597
DaysToRepair                7.0186         7.6616         6.2404
Sum                         8.2631         4.4437        12.8858


Time taken to build model (full training data) : 0.75 seconds

=== Model and evaluation on training set ===
```

```
Clustered Instances

0      34186 ( 55%)
1      28245 ( 45%)


Class attribute: Guarantee
Classes to Clusters:

     0     1  <-- assigned to cluster
 28891 27601 | Yes-guarantee
  5295   644 | No-guarantee

Cluster 0 <-- No-guarantee
Cluster 1 <-- Yes-guarantee

Incorrectly clustered instances :   29535.0  47.3082 %
```

9.2.1.4 FarthestFirst

```
=== Run information ===

Scheme:       Weka.clusterers.FarthestFirst -N 2 -S 1
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
Ignored:
              Guarantee
Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===


FarthestFirst
==============

Cluster centroids:

Cluster 0
     Subcontractors MDA BRAND 1 Servicing 1.0 3.0 10.0
Cluster 1
     ServiceCenter AirConditioning Brand2 Installation 5.0 3.0 0.0



Time taken to build model (full training data) : 0.25 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      60931 ( 98%)
1       1500 (  2%)


Class attribute: Guarantee
Classes to Clusters:

     0     1  <-- assigned to cluster
 55109  1383 | Yes-guarantee
  5822   117 | No-guarantee

Cluster 0 <-- Yes-guarantee
Cluster 1 <-- No-guarantee

Incorrectly clustered instances :   7205.0   11.5407 %
```

9.2.1.5 Canopy

```
=== Run information ===

Scheme:       Weka.clusterers.Canopy -N -1 -max-candidates 100 -periodic-pruning
10000 -min-density 2.0 -t2 -1.0 -t1 -1.25 -S 1
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
Ignored:
              Guarantee
Test mode:    Classes to clusters evaluation on training data

=== Clustering model (full training set) ===


Canopy clustering
=================

Number of canopies (cluster centers) found: 26
T2 radius: 1.042
T1 radius: 1.302

Cluster 0: Subcontractors,MDA,BRAND 1,Servicing,2.415415,6.531469,8.41393,{45203}
<0,7,21>
.....
Cluster 25: Subcontractors,Mobile,Other,Servicing,1,3.5,20.06,{4} <21,25>

Time taken to build model (full training data) : 0.99 seconds

=== Model and evaluation on training set ===

Clustered Instances

 0      26034 ( 42%)
 1       2663 (  4%)
 2       3533 (  6%)
 3       1882 (  3%)
 4       1189 (  2%)
 5        110 (  0%)
 6        550 (  1%)
 7      18733 ( 30%)
 8        980 (  2%)
 9        225 (  0%)
10        118 (  0%)
11         15 (  0%)
12       1063 (  2%)
13        155 (  0%)
14         89 (  0%)
```

```
15          32 (  0%)
16          10 (  0%)
17         172 (  0%)
18        1205 (  2%)
19        2119 (  3%)
20         128 (  0%)
21         306 (  0%)
22         558 (  1%)
23         155 (  0%)
24         109 (  0%)
25         298 (  0%)


Class attribute: Guarantee
Classes to Clusters:

     0     1     2     3     4     5     6     7     8     9    10    11    12
13    14    15    16    17    18    19    20    21    22    23    24    25  <--
assigned to cluster
 25435  2574  2807  1785  1183   104   443 14936   898   220   110    15   998
151    86    30    10   169  1130  2022   125   273   446   149   104   289 | Yes-
guarantee
   599    89   726    97     6     6   107  3797    82     5     8     0    65
4     3     2     0     3    75    97     3    33   112     6     5     9 | No-
guarantee

Cluster  0 <-- Yes-guarantee
Cluster  1 <-- No class
Cluster  2 <-- No class
Cluster  3 <-- No class
Cluster  4 <-- No class
Cluster  5 <-- No class
Cluster  6 <-- No class
Cluster  7 <-- No-guarantee
Cluster  8 <-- No class
Cluster  9 <-- No class
Cluster 10 <-- No class
Cluster 11 <-- No class
Cluster 12 <-- No class
Cluster 13 <-- No class
Cluster 14 <-- No class
Cluster 15 <-- No class
Cluster 16 <-- No class
Cluster 17 <-- No class
Cluster 18 <-- No class
Cluster 19 <-- No class
Cluster 20 <-- No class
Cluster 21 <-- No class
Cluster 22 <-- No class
Cluster 23 <-- No class
Cluster 24 <-- No class
Cluster 25 <-- No class

Incorrectly clustered instances :   33199.0   53.1771 %
```

9.2.2 Classification Algorithms

9.2.2.1 J48

```
=== Run information ===

Scheme:       Weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
              Guarantee
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

RepairedByAge <= 5
|   AuthorizedService = ServiceCenter
|   |   Sum <= 0: Yes-guarantee (28338.0/1111.0)
|   |   Sum > 0
|   |   |   Action = Installation: Yes-guarantee (711.0/120.0)
|   |   |   Action = Servicing: No-guarantee (4600.0/955.0)
|   |   |   Action = ChangeProduct: No-guarantee (0.0)
|   AuthorizedService = Subcontractors: Yes-guarantee (27977.0/619.0)
RepairedByAge > 5
|   AuthorizedService = ServiceCenter: No-guarantee (537.0/118.0)
|   AuthorizedService = Subcontractors: Yes-guarantee (268.0/25.0)

Number of Leaves  :     7

Size of the tree :  12


Time taken to build model: 4.39 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       59472               95.2604 %
Incorrectly Classified Instances      2959                4.7396 %
Kappa statistic                          0.7071
Mean absolute error                      0.0844
Root mean squared error                  0.2058
Relative absolute error                 49.0169 %
Root relative squared error             70.1451 %
Total Number of Instances            62431
```

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  
PRC Area  Class
                 0.981    0.316    0.967      0.981    0.974      0.709    0.859     
0.973     Yes-guarantee
                 0.684    0.019    0.790      0.684    0.733      0.709    0.859     
0.620     No-guarantee
Weighted Avg.    0.953    0.288    0.950      0.953    0.951      0.709    0.859     
0.940     

=== Confusion Matrix ===

     a     b   <-- classified as
 55412  1080 |     a = Yes-guarantee
  1879  4060 |     b = No-guarantee
```

9.2.2.2 RandomForest

```
=== Run information ===

Scheme:       Weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -
M 1.0 -V 0.001 -S 1
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
              Guarantee
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

Weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-
capabilities

Time taken to build model: 48.6 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       59234               94.8791 %
Incorrectly Classified Instances      3197                5.1209 %
Kappa statistic                          0.676
Mean absolute error                      0.0787
Root mean squared error                  0.208
Relative absolute error                 45.7368 %
Root relative squared error             70.8913 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area
PRC Area  Class
                 0.981    0.361    0.963      0.981    0.972      0.680    0.911
0.984     Yes-guarantee
                 0.639    0.019    0.783      0.639    0.704      0.680    0.911
0.696     No-guarantee
Weighted Avg.    0.949    0.328    0.946      0.949    0.946      0.680    0.911
0.957
```

```
=== Confusion Matrix ===

     a     b   <-- classified as
 55438  1054 |     a = Yes-guarantee
  2143  3796 |     b = No-guarantee
```

9.2.2.3 Naïve Bayes

```
=== Run information ===

Scheme:       Weka.classifiers.bayes.NaiveBayes
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
              Guarantee
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

                               Class
Attribute           Yes-guarantee  No-guarantee
                           (0.9)         (0.1)
================================================
AuthorizedService
  ServiceCenter             28892.0        5296.0
  Subcontractors            27602.0         645.0
  [total]                   56494.0        5941.0

Category
....

Brand
...

Action
  Installation               1848.0         126.0
  Servicing                 54592.0        5815.0
  ChangeProduct                55.0           1.0
  [total]                   56495.0        5942.0

RepairedByAge
  mean                       2.1545        2.3868
  std. dev.                  1.3154        1.8295
  weight sum                  56492          5939
  precision                       1             1

DaysToRepair
  mean                       7.0284        7.0869
  std. dev.                 12.6446       11.8262
  weight sum                  56492          5939
  precision                  2.6906        2.6906
```

```
Sum
  mean                      7.2538      17.6394
  std. dev.                14.1528      32.3323
  weight sum                 56492         5939
  precision                 3.5319       3.5319


Time taken to build model: 0.3 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       56496               90.4935 %
Incorrectly Classified Instances      5935                9.5065 %
Kappa statistic                          0.178
Mean absolute error                      0.1253
Root mean squared error                  0.2819
Relative absolute error                 72.7974 %
Root relative squared error             96.0723 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area
PRC Area  Class
                 0.986    0.866    0.915      0.986    0.949      0.223    0.797
0.966     Yes-guarantee
                 0.134    0.014    0.501      0.134    0.211      0.223    0.797
0.328     No-guarantee
Weighted Avg.    0.905    0.785    0.876      0.905    0.879      0.223    0.797
0.905

=== Confusion Matrix ===

     a     b   <-- classified as
 55703   789 |     a = Yes-guarantee
  5146   793 |     b = No-guarantee
```

9.2.2.4 Logistic

```
=== Run information ===

Scheme:       Weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-
places 4
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
              Guarantee
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...
                                           Class
Variable                           Yes-guarantee
================================================
AuthorizedService=Subcontractors          3.0484
C...
Category=Heating                          1.6565
...
Action=Installation                     -14.6868
Action=Servicing                        -15.7033
Action=ChangeProduct                   1090.3848
RepairedByAge                            -0.1007
DaysToRepair                              0.0035
Sum                                      -0.0572
Intercept                                19.0173


Odds Ratios...
                                           Class
Variable                           Yes-guarantee
================================================
AuthorizedService=Subcontractors         21.0809
...
Action=ChangeProduct                    Infinity
RepairedByAge                             0.9042
DaysToRepair                              1.0035
Sum                                       0.9444


Time taken to build model: 7.26 seconds

=== Stratified cross-validation ===
=== Summary ===
```

```
Correctly Classified Instances       57062               91.4001 %
Incorrectly Classified Instances      5369                8.5999 %
Kappa statistic                          0.2209
Mean absolute error                      0.1333
Root mean squared error                  0.2536
Relative absolute error                 77.4472 %
Root relative squared error             86.4468 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area
PRC Area  Class
                 0.995    0.853    0.917      0.995    0.954      0.306    0.833
0.964     Yes-guarantee
                 0.147    0.005    0.742      0.147    0.245      0.306    0.833
0.547     No-guarantee
Weighted Avg.    0.914    0.772    0.901      0.914    0.887      0.306    0.833
0.925

=== Confusion Matrix ===

     a     b   <-- classified as
 56189   303 |     a = Yes-guarantee
  5066   873 |     b = No-guarantee
```

9.2.2.5 IBk

```
=== Run information ===

Scheme:       Weka.classifiers.lazy.IBk -K 1 -W 0 -A
"Weka.core.neighboursearch.LinearNNSearch -A \"Weka.core.EuclideanDistance -R
first-last\""
Relation:     database
Instances:    62431
Attributes:   8
              AuthorizedService
              Category
              Brand
              Action
              RepairedByAge
              DaysToRepair
              Sum
              Guarantee
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification


Time taken to build model: 0.11 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       59083               94.6373 %
Incorrectly Classified Instances      3348                5.3627 %
Kappa statistic                          0.6536
Mean absolute error                      0.0789
Root mean squared error                  0.2187
Relative absolute error                 45.8332 %
Root relative squared error             74.5325 %
Total Number of Instances            62431

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area
PRC Area  Class
                 0.982    0.395    0.959      0.982    0.971      0.660    0.882
0.981     Yes-guarantee
                 0.605    0.018    0.782      0.605    0.682      0.660    0.882
0.640     No-guarantee
Weighted Avg.    0.946    0.359    0.943      0.946    0.943      0.660    0.882
0.948

=== Confusion Matrix ===

     a     b   <-- classified as
 55487  1005 |     a = Yes-guarantee
  2343  3596 |     b = No-guarantee
```

9.2.3 Reliability and Regression Analysis Results

9.2.3.1 Correlation results

```
 (5 vars, 52170 obs)

. pwcorr customersatisfaction authorizedservice brandwithspareparts weekstorepair payment


             | custom~n author~e brandw~s weekst~r  payment
-------------+---------------------------------------------
customersa~n |   1.0000
authorized~e |  -0.2990   1.0000
brandwiths~s |   0.1932  -0.3619   1.0000
weekstorep~r |  -0.8512   0.1625  -0.1449   1.0000
     payment |   0.2404  -0.7632   0.3061  -0.1378   1.0000
```

9.2.3.2 Cronbach Alpha (1)

```
. alpha customersatisfaction weekstorepair


Test scale = mean(unstandardized items)
Reversed item:  customersatisfaction
Average interitem covariance:     .2056783
Number of items in the scale:            2
Scale reliability coefficient:      0.8424
```

9.2.3.3 Cronbach Alpha (2)

```
. alpha customersatisfaction authorizedservice brandwithspareparts weekstorepair payment


Test scale = mean(unstandardized items)
Reversed items:  authorizedservice weekstorepair
Average interitem covariance:      .079217
Number of items in the scale:            5
Scale reliability coefficient:      0.7054
```

9.2.3.4 Mutliple regression results

```
. reg customersatisfaction authorizedservice brandwithspareparts weekstorepair payment, beta

      Source |       SS       df       MS              Number of obs =   52170
-------------+------------------------------           F(  4, 52165) =39402.57
       Model |  5331.08467     4  1332.77117           Prob > F      =  0.0000
    Residual |  1764.45355 52165  .033824471           R-squared     =  0.7513
-------------+------------------------------           Adj R-squared =  0.7513
       Total |  7095.53822 52169  .136010624           Root MSE      =  .18391
------------------------------------------------------------------------------------
customersatisfact~n |      Coef.   Std. Err.      t    P>|t|                  Beta
--------------------+---------------------------------------------------------------
  authorizedservice |  -.1172033   .0025552   -45.87   0.000             -.1588345
brandwithspareparts |   .0145283   .0021165     6.86   0.000              .0161664
      weekstorepair |  -.4632308   .0012512  -370.22   0.000             -.8229492
            payment |    .000626   .0025493     0.25   0.806              .0008307
              _cons |   1.480191   .0034164   433.26   0.000                     .
------------------------------------------------------------------------------------
```

9.2.3.5 Simple Regression results

```
. reg customersatisfaction weekstorepair

      Source |       SS       df       MS              Number of obs =   52170
-------------+------------------------------           F(  1, 52168) =       .
       Model |  5141.21163     1  5141.21163           Prob > F      =  0.0000
    Residual |  1954.32659 52168  .037462172           R-squared     =  0.7246
-------------+------------------------------           Adj R-squared =  0.7246
       Total |  7095.53822 52169  .136010624           Root MSE      =  .19355
------------------------------------------------------------------------------
customersat~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
weekstorepair |  -.4791423   .0012934  -370.46   0.000    -.4816773   -.4766072
        _cons |   1.451952   .0018623   779.67   0.000     1.448302    1.455602
```

9.3 Recommendation System Coding (Partially)

9.3.1 Model Building

```
# Importing libraries in Python
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
import seaborn as sns
# Loading the dataset
mng_servicing = pd.read_csv('dataset.csv')
mng_servicing.head()

df = mng_servicing.copy()
target = 'weekstorepair'
encode = ['brand','category']
for col in encode:
    dummy = pd.get_dummies(df[col], prefix=col)
    df = pd.concat([df,dummy], axis=1)
    del df[col]
target_mapper = {'1-to-1':0, '2-to-2':1, '3-to-3':2, '4-to-4':3 }
def target_encode(val):
    return target_mapper[val]
df['weekstorepair'] = df['weekstorepair'].apply(target_encode)
# Separating X and y
X = df.drop('weekstorepair', axis=1)
Y = df['weekstorepair']

#Preparing the data for Model Building
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test=train_test_split(X, Y, test_size=0.3,random_state=1)

#Importing Decision Tree from Sklearn to build a classification model
from sklearn.tree import DecisionTreeClassifier
clf=DecisionTreeClassifier()
clf.fit(X_train,y_train)
y_pred = clf.predict(X_test)

#Check accuracy of the model build
clf.score(X_test, y_test)
accuracy = clf.score(X_test, y_test)
print(f"Accuracy with train data: {accuracy:0.2f}")
#prediction
y_pred = clf.predict(X_test)
```

9.3.2 Application 1

```
import streamlit as st
import pandas as pd
import numpy as np

def app():
    st.title('Recommendation System for After-sales services')
    st.write("""# Prediction Probability (%) """)
    st.header("""This recommendation system predicts the product repairing service 
quality!  """)

    st.markdown('<p style="text-align: justify;font-size:20px;">The recommendation 
system for after sale support service, as a portion of data mining system 
functionalities, is a tool used to alleviate the process of streamlining decision-
making and provide reliable predictions, especially for the customers, enhance the 
process of creating new experiences, as well as increase the businesses’ 
productivity</p>', unsafe_allow_html=True)
```

9.3.3 Application 2

```
import streamlit as st
import pandas as pd
import numpy as np
import pickle
import plotly.express as px

def app():
    st.title('BRAND 1 - Prediction')
    load_clf = pickle.load(open('dataset.pkl', 'rb'))
    def brand1_ac():
            brand = 'BRAND1'
            category = 'AirConditioning'
            data10 = {'brand':brand,
                    'category':category}
            features10 = pd.DataFrame(data10, index=[0])
            return features10
    input_arc_ac = brand1_ac()
    servicing_raw10 = pd.read_csv('dataset.csv')
    servicing10 = servicing_raw10.drop(columns=['weekstorepair'])
    df10 = pd.concat([input_arc_ac,servicing10],axis=0)
    encode = ['brand','category']
    for col10 in encode:
        dummy10 = pd.get_dummies(df10[col10], prefix=col10)
        df10 = pd.concat([df10,dummy10], axis=1)
        del df10[col10]
    df10 = df10[:1]
    prediction_proba10 = load_clf.predict_proba(df10)

    def brand1_heat():
            brand = ‘BRAND 1’
            category = 'Heating'
            data11 = {'brand':brand,
                    'category':category}
            features11 = pd.DataFrame(data11, index=[0])
            return features11
    input_arc_heat = brand1_heat()
    servicing_raw11 = pd.read_csv('dataset.csv')
    servicing11 = servicing_raw11.drop(columns=['weekstorepair'])
    df11 = pd.concat([input_arc_heat,servicing11],axis=0)
    for col11 in encode:
        dummy11 = pd.get_dummies(df11[col11], prefix=col11)
        df11 = pd.concat([df11,dummy11], axis=1)
        del df11[col11]
```

```
df11 = df11[:1]
prediction_proba11 = load_clf.predict_proba(df11)

def brand1_mda():
        brand = ‘BRAND 1’
        category = 'MDA'
        data12 = {'brand':brand,
                'category':category}
        features12 = pd.DataFrame(data12, index=[0])
        return features12
input_arc_mda = brand1_mda()
servicing_raw12 = pd.read_csv('dataset.csv')
servicing12 = servicing_raw12.drop(columns=['weekstorepair'])
df12 = pd.concat([input_arc_mda,servicing12],axis=0)
for col12 in encode:
    dummy12 = pd.get_dummies(df12[col12], prefix=col12)
    df12 = pd.concat([df12,dummy12], axis=1)
    del df12[col12]
df12 = df12[:1]
prediction_proba12 = load_clf.predict_proba(df12)

def brand1_sdap():
        brand = ‘BRAND 1’
        category = 'SDAP'
        data14 = {'brand':brand,
                'category':category}
        features14 = pd.DataFrame(data14, index=[0])
        return features14
input_arc_sdap = brand1_sdap()
servicing_raw14 = pd.read_csv('dataset.csv')
servicing14 = servicing_raw14.drop(columns=['weekstorepair'])
df14 = pd.concat([input_arc_sdap,servicing14],axis=0)
for col14 in encode:
    dummy14 = pd.get_dummies(df14[col14], prefix=col14)
    df14 = pd.concat([df14,dummy14], axis=1)
    del df14[col14]
df14 = df14[:1]
prediction_proba14 = load_clf.predict_proba(df14)

def brand1_tv():
        brand = ‘BRAND 1’
        category = 'TV'
        data15 = {'brand':brand,
                'category':category}
        features15 = pd.DataFrame(data15, index=[0])
```

```
            return features15
    input_arc_tv = brand1_tv()
    servicing_raw15 = pd.read_csv('dataset.csv')
    servicing15 = servicing_raw15.drop(columns=['weekstorepair'])
    df15 = pd.concat([input_arc_tv,servicing15],axis=0)
    for col15 in encode:
        dummy15 = pd.get_dummies(df15[col15], prefix=col15)
        df15 = pd.concat([df15,dummy15], axis=1)
        del df15[col15]
    df15 = df15[:1]
    prediction_proba15 = load_clf.predict_proba(df15)

    t_prediction = '<p style="font-size: 18px; color: red;"><b>Target: Time taken for
resolving the complaint</b> </p>'
    st.markdown(t_prediction, unsafe_allow_html=True)

    categ, valpre = st.columns(2)
    with categ:
        st.markdown(f"<h3 style='text-align: left;color: red;'>Category</h3>",
unsafe_allow_html=True)
        st.markdown(f"<h3>-</h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
        text1 = 'Air Conditioning'
        st.markdown(f"<h3 style='text-align: left;'>{text1}</h3>",
unsafe_allow_html=True)
        text2 = 'Heating'
        st.markdown(f"<h3 style='text-align: left;'>{text2}</h3>",
unsafe_allow_html=True)
        text3 = 'MDA'
        st.markdown(f"<h3 style='text-align: left;'>{text3}</h3>",
unsafe_allow_html=True)
        text5 = 'SDAP'
        st.markdown(f"<h3 style='text-align: left;'>{text5}</h3>",
unsafe_allow_html=True)
        text6 = 'TV'
        st.markdown(f"<h3 style='text-align: left;'>{text6}</h3>",
unsafe_allow_html=True)

    with valpre:
        st.markdown(f"<h3 style='text-align: left; color: red;'>Prediction
Probability</h3>", unsafe_allow_html=True)
        st.markdown(f"<h3 style='text-align: left;'><label style='color:green'>Week
1</label> | <label style='color:blue'>Week 2</label> | <label style='color:red'>Week
3</label> | <label style='color:orange'>Week 4</label> </h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
```

```
        number10 = round(prediction_proba10[0][0]*100 , 2)
        number101= round(prediction_proba10[0][1]*100 , 2)
        number102= round(prediction_proba10[0][2]*100 , 2)
        number103= round(prediction_proba10[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number10} %</label> | <label style='color:blue'>{number101}
%</label> | <label style='color:red'>{number102} %</label> | <label
style='color:orange'>{number103} %</label> </h3>", unsafe_allow_html=True)

        number11 = round(prediction_proba11[0][0]*100 , 2)
        number111= round(prediction_proba11[0][1]*100 , 2)
        number112 = round(prediction_proba11[0][2]*100 , 2)
        number113= round(prediction_proba11[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number11} %</label> | <label style='color:blue'>{number111}
%</label> | <label style='color:red'>{number112} %</label> | <label
style='color:orange'>{number113} %</label> </h3>", unsafe_allow_html=True)
        number12 = round(prediction_proba12[0][0]*100 , 2)
        number121= round(prediction_proba12[0][1]*100 , 2)
        number122 = round(prediction_proba12[0][2]*100 , 2)
        number123= round(prediction_proba12[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number12} %</label> | <label style='color:blue'>{number121}
%</label> | <label style='color:red'>{number122} %</label> | <label
style='color:orange'>{number123} %</label> </h3>", unsafe_allow_html=True)
        number14 = round(prediction_proba14[0][0]*100 , 2)
        number141= round(prediction_proba14[0][1]*100 , 2)
        number142 = round(prediction_proba14[0][2]*100 , 2)
        number143= round(prediction_proba14[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number14} %</label> | <label style='color:blue'>{number141}
%</label> | <label style='color:red'>{number142} %</label> | <label
style='color:orange'>{number143} %</label> </h3>", unsafe_allow_html=True)
        number15 = round(prediction_proba15[0][0]*100 , 2)
        number151= round(prediction_proba15[0][1]*100 , 2)
        number152 = round(prediction_proba15[0][2]*100 , 2)
        number153 = round(prediction_proba15[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number15} %</label> | <label style='color:blue'>{number151}
%</label> | <label style='color:red'>{number152} %</label> | <label
style='color:orange'>{number153} %</label> </h3>", unsafe_allow_html=True)

    st.markdown("<hr/>",unsafe_allow_html=True)
    st.title('Comparison of categories')
```

```
    df = pd.DataFrame([["", round(prediction_proba10[0][0]*100 , 2),
round(prediction_proba11[0][0]*100 , 2), round(prediction_proba12[0][0]*100 , 2),
round(prediction_proba14[0][0]*100 , 2), round(prediction_proba15[0][0]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 1)",
    xaxis_title="Categories of Brand 1",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][1]*100 , 2),
round(prediction_proba11[0][1]*100 , 2), round(prediction_proba12[0][1]*100 , 2),
round(prediction_proba14[0][1]*100 , 2), round(prediction_proba15[0][1]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 2)",
    xaxis_title="Categories of Brand 1",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][2]*100 , 2),
round(prediction_proba11[0][2]*100 , 2), round(prediction_proba12[0][2]*100 , 2),
round(prediction_proba14[0][2]*100 , 2), round(prediction_proba15[0][2]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 3)",
    xaxis_title="Categories of Brand 1",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][3]*100 , 2),
round(prediction_proba11[0][3]*100 , 2), round(prediction_proba12[0][3]*100 , 2),
```

```
round(prediction_proba14[0][3]*100 , 2), round(prediction_proba15[0][3]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 4)",
    xaxis_title="Categories of Brand 1",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
```

9.3.4 Application 3

```
import streamlit as st
import pandas as pd
import numpy as np
import pickle
import plotly.express as px
import plotly.graph_objs as go

def app():
    st.title('Air Conditioning - Prediction')
    load_clf = pickle.load(open('dataset.pkl', 'rb'))
    def samg_ac():
            brand = ‘BRAND 1’
            category = 'AirConditioning'
            data10 = {'brand':brand,
                    'category':category}
            features10 = pd.DataFrame(data10, index=[0])
            return features10
    input_sam_ac = samg_ac()
    servicing_raw10 = pd.read_csv('dataset.csv')
    servicing10 = servicing_raw10.drop(columns=['weekstorepair'])
    df10 = pd.concat([input_sam_ac,servicing10],axis=0)
    encode = ['brand','category']
    for col10 in encode:
        dummy10 = pd.get_dummies(df10[col10], prefix=col10)
        df10 = pd.concat([df10,dummy10], axis=1)
        del df10[col10]
    df10 = df10[:1]
    prediction_proba10 = load_clf.predict_proba(df10)

    def samg_choc():
            brand = ‘BRAND 2’
            category = 'AirConditioning'
            data11 = {'brand':brand,
                    'category':category}
            features11 = pd.DataFrame(data11, index=[0])
            return features11
    input_sam_choc = samg_choc()
    servicing_raw11 = pd.read_csv('dataset.csv')
    servicing11 = servicing_raw11.drop(columns=['weekstorepair'])
    df11 = pd.concat([input_sam_choc,servicing11],axis=0)
    for col11 in encode:
        dummy11 = pd.get_dummies(df11[col11], prefix=col11)
        df11 = pd.concat([df11,dummy11], axis=1)
```

```
    del df11[col11]
df11 = df11[:1]
prediction_proba11 = load_clf.predict_proba(df11)

def samg_dw():
        brand = 'BRAND 3'
        category = 'AirConditioning'
        data12 = {'brand':brand,
                'category':category}
        features12 = pd.DataFrame(data12, index=[0])
        return features12
input_sam_dw = samg_dw()
servicing_raw12 = pd.read_csv('dataset.csv')
servicing12 = servicing_raw12.drop(columns=['weekstorepair'])
df12 = pd.concat([input_sam_dw,servicing12],axis=0)
for col12 in encode:
    dummy12 = pd.get_dummies(df12[col12], prefix=col12)
    df12 = pd.concat([df12,dummy12], axis=1)
    del df12[col12]
df12 = df12[:1]
prediction_proba12 = load_clf.predict_proba(df12)

def samg_ht():
        brand = 'BRAND 4'
        category = 'AirConditioning'
        data13 = {'brand':brand,
                'category':category}
        features13 = pd.DataFrame(data13, index=[0])
        return features13
input_sam_ht = samg_ht()
servicing_raw13 = pd.read_csv('dataset.csv')
servicing13 = servicing_raw13.drop(columns=['weekstorepair'])
df13 = pd.concat([input_sam_ht,servicing13],axis=0)
for col13 in encode:
    dummy13 = pd.get_dummies(df13[col13], prefix=col13)
    df13 = pd.concat([df13,dummy13], axis=1)
    del df13[col13]
df13 = df13[:1]
prediction_proba13 = load_clf.predict_proba(df13)

def samg_mh():
        brand = 'BRAND 5'
        category = 'AirConditioning'
        data14 = {'brand':brand,
                'category':category}
```

```
            features14 = pd.DataFrame(data14, index=[0])
            return features14
    input_sam_mh = samg_mh()
    servicing_raw14 = pd.read_csv('dataset.csv')
    servicing14 = servicing_raw14.drop(columns=['weekstorepair'])
    df14 = pd.concat([input_sam_mh,servicing14],axis=0)
    for col14 in encode:
        dummy14 = pd.get_dummies(df14[col14], prefix=col14)
        df14 = pd.concat([df14,dummy14], axis=1)
        del df14[col14]
    df14 = df14[:1]
    prediction_proba14 = load_clf.predict_proba(df14)

    def samg_rf():
            brand = ‘BRAND 6’
            category = 'AirConditioning'
            data15 = {'brand':brand,
                    'category':category}
            features15 = pd.DataFrame(data15, index=[0])
            return features15
    input_sam_rf = samg_rf()
    servicing_raw15 = pd.read_csv('dataset.csv')
    servicing15 = servicing_raw15.drop(columns=['weekstorepair'])
    df15 = pd.concat([input_sam_rf,servicing15],axis=0)
    for col15 in encode:
        dummy15 = pd.get_dummies(df15[col15], prefix=col15)
        df15 = pd.concat([df15,dummy15], axis=1)
        del df15[col15]
    df15 = df15[:1]
    prediction_proba15 = load_clf.predict_proba(df15)

    t_prediction = '<p style=" font-size: 18px;"><label><b>Target: Time taken for
resolving the complaint</b></label></p>'
    st.markdown(t_prediction, unsafe_allow_html=True)

    categ, valpre = st.columns(2)
    with categ:
        st.markdown(f"<h3 style='text-align: left;color:red;'>Brand</h3>",
unsafe_allow_html=True)
        st.markdown(f"<h3>-</h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
        text1 = ‘BRAND 1’
        st.markdown(f"<h3 style='text-align: left;'>{text1}</h3>",
unsafe_allow_html=True)
        text2 = ‘BRAND 2’
```

```
        st.markdown(f"<h3 style='text-align: left;'>{text2}</h3>",
unsafe_allow_html=True)
        text3 = ‘BRAND 3’
        st.markdown(f"<h3 style='text-align: left;'>{text3}</h3>",
unsafe_allow_html=True)
        text4 = ‘BRAND 4’
        st.markdown(f"<h3 style='text-align: left;'>{text4}</h3>",
unsafe_allow_html=True)
        text6 = ‘BRAND 5’
        st.markdown(f"<h3 style='text-align: left;'>{text6}</h3>",
unsafe_allow_html=True)
        text5 = ‘BRAND 6’
        st.markdown(f"<h3 style='text-align: left;'>{text5}</h3>",
unsafe_allow_html=True)

    with valpre:
        st.markdown(f"<h3 style='text-align: left; color: red;'>Prediction
Probability</h3>", unsafe_allow_html=True)
        st.markdown(f"<h3 style='text-align: left;'><label style='color:green'>Week
1</label> | <label style='color:blue'>Week 2</label> | <label style='color:red'>Week
3</label> | <label style='color:orange'>Week 4</label> </h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
        number10 = round(prediction_proba10[0][0]*100 , 2)
        number101= round(prediction_proba10[0][1]*100 , 2)
        number102= round(prediction_proba10[0][2]*100 , 2)
        number103= round(prediction_proba10[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number10} %</label> | <label style='color:blue'>{number101}
%</label>  | <label style='color:red'>{number102} %</label> | <label
style='color:orange'>{number103} %</label>  </h3>", unsafe_allow_html=True)
        number11 = round(prediction_proba11[0][0]*100 , 2)
        number111= round(prediction_proba11[0][1]*100 , 2)
        number112 = round(prediction_proba11[0][2]*100 , 2)
        number113= round(prediction_proba11[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number11} %</label> | <label style='color:blue'>{number111}
%</label> | <label style='color:red'>{number112} %</label> | <label
style='color:orange'>{number113} %</label> </h3>", unsafe_allow_html=True)
        number12 = round(prediction_proba12[0][0]*100 , 2)
        number121= round(prediction_proba12[0][1]*100 , 2)
        number122 = round(prediction_proba12[0][2]*100 , 2)
        number123= round(prediction_proba12[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number12} %</label> | <label style='color:blue'>{number121}
```

```
%</label> | <label style='color:red'>{number122} %</label> | <label
style='color:orange'>{number123} %</label>  </h3>", unsafe_allow_html=True)
        number13 = round(prediction_proba13[0][0]*100 , 2)
        number131= round(prediction_proba13[0][1]*100 , 2)
        number132 = round(prediction_proba13[0][2]*100 , 2)
        number133= round(prediction_proba13[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number13} %</label> | <label style='color:blue'>{number131}
%</label> | <label style='color:red'>{number132} %</label> | <label
style='color:orange'>{number133} %</label>  </h3>", unsafe_allow_html=True)
        number14 = round(prediction_proba14[0][0]*100 , 2)
        number141= round(prediction_proba14[0][1]*100 , 2)
        number142 = round(prediction_proba14[0][2]*100 , 2)
        number143= round(prediction_proba14[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number14} %</label> | <label style='color:blue'>{number141}
%</label>  | <label style='color:red'>{number142} %</label> | <label
style='color:orange'>{number143} %</label> </h3>", unsafe_allow_html=True)
        number15 = round(prediction_proba15[0][0]*100 , 2)
        number151= round(prediction_proba15[0][1]*100 , 2)
        number152 = round(prediction_proba15[0][2]*100 , 2)
        number153 = round(prediction_proba15[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number15} %</label> | <label style='color:blue'>{number151}
%</label> | <label style='color:red'>{number152} %</label> | <label
style='color:orange'>{number153} %</label> </h3>", unsafe_allow_html=True)
    st.markdown("<hr/>",unsafe_allow_html=True)
    st.title('Comparison of Brands')
    df = pd.DataFrame([["", round(prediction_proba10[0][0]*100 , 2),
round(prediction_proba11[0][0]*100 , 2), round(prediction_proba12[0][0]*100 ,
2),round(prediction_proba13[0][0]*100 , 2), round(prediction_proba14[0][0]*100 , 2),
round(prediction_proba15[0][0]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 1)",
    xaxis_title="Brands of Air Conditioning",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][1]*100 , 2),
round(prediction_proba11[0][1]*100 , 2), round(prediction_proba12[0][1]*100 ,
```

```
2),round(prediction_proba13[0][1]*100 , 2), round(prediction_proba14[0][1]*100 , 2),
round(prediction_proba15[0][1]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 2)",
    xaxis_title="Brands of Air Conditioning",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][2]*100 , 2),
round(prediction_proba11[0][2]*100 , 2), round(prediction_proba12[0][2]*100 ,
2),round(prediction_proba13[0][2]*100 , 2), round(prediction_proba14[0][2]*100 , 2),
round(prediction_proba15[0][2]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 3)",
    xaxis_title="Brands of Air Conditioning",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][3]*100 , 2),
round(prediction_proba11[0][3]*100 , 2), round(prediction_proba12[0][3]*100 ,
2),round(prediction_proba13[0][3]*100 , 2), round(prediction_proba14[0][3]*100 , 2),
round(prediction_proba15[0][3]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 4)",
    xaxis_title="Brands of Air Conditioning",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
```

9.3.5 Application 4

```
import streamlit as st
import pandas as pd
import numpy as np
import pickle
import plotly.express as px

def app():
    st.title('BRAND 2 - Prediction')
    load_clf = pickle.load(open('dataset.pkl', 'rb'))
    def brand2_ac():
            brand = ‘BRAND 2’
            category = 'AirConditioning'
            data10 = {'brand':brand,
                    'category':category}
            features10 = pd.DataFrame(data10, index=[0])
            return features10
    input_bsc_ac = brand2_ac()
    servicing_raw10 = pd.read_csv('dataset.csv')
    servicing10 = servicing_raw10.drop(columns=['weekstorepair'])
    df10 = pd.concat([input_bsc_ac,servicing10],axis=0)
    encode = ['brand','category']
    for col10 in encode:
        dummy10 = pd.get_dummies(df10[col10], prefix=col10)
        df10 = pd.concat([df10,dummy10], axis=1)
        del df10[col10]
    df10 = df10[:1]
    prediction_proba10 = load_clf.predict_proba(df10)

    def brand2_heat():
            brand = ‘BRAND 2’
            category = 'Heating'
            data11 = {'brand':brand,
                    'category':category}
            features11 = pd.DataFrame(data11, index=[0])
            return features11
    input_bsc_heat = brand2_heat()
    servicing_raw11 = pd.read_csv('dataset.csv')
    servicing11 = servicing_raw11.drop(columns=['weekstorepair'])
    df11 = pd.concat([input_bsc_heat,servicing11],axis=0)
    for col11 in encode:
        dummy11 = pd.get_dummies(df11[col11], prefix=col11)
        df11 = pd.concat([df11,dummy11], axis=1)
        del df11[col11]
```

```
df11 = df11[:1]
prediction_proba11 = load_clf.predict_proba(df11)

def brand2_mda():
        brand = 'BRAND 2'
        category = 'MDA'
        data12 = {'brand':brand,
                'category':category}
        features12 = pd.DataFrame(data12, index=[0])
        return features12
input_bsc_mda = brand2_mda()
servicing_raw12 = pd.read_csv('dataset.csv')
servicing12 = servicing_raw12.drop(columns=['weekstorepair'])
df12 = pd.concat([input_bsc_mda,servicing12],axis=0)
for col12 in encode:
    dummy12 = pd.get_dummies(df12[col12], prefix=col12)
    df12 = pd.concat([df12,dummy12], axis=1)
    del df12[col12]
df12 = df12[:1]
prediction_proba12 = load_clf.predict_proba(df12)

def brand2_sdap():
        brand = 'BRAND 2'
        category = 'SDAP'
        data14 = {'brand':brand,
                'category':category}
        features14 = pd.DataFrame(data14, index=[0])
        return features14
input_bsc_sdap = brand2_sdap()
servicing_raw14 = pd.read_csv('dataset.csv')
servicing14 = servicing_raw14.drop(columns=['weekstorepair'])
df14 = pd.concat([input_bsc_sdap,servicing14],axis=0)
for col14 in encode:
    dummy14 = pd.get_dummies(df14[col14], prefix=col14)
    df14 = pd.concat([df14,dummy14], axis=1)
    del df14[col14]
df14 = df14[:1]
prediction_proba14 = load_clf.predict_proba(df14)

def brand2_tv():
        brand = 'BRAND 2'
        category = 'TV'
        data15 = {'brand':brand,
                'category':category}
        features15 = pd.DataFrame(data15, index=[0])
```

```
            return features15
    input_bsc_tv = brand2_tv()
    servicing_raw15 = pd.read_csv('dataset.csv')
    servicing15 = servicing_raw15.drop(columns=['weekstorepair'])
    df15 = pd.concat([input_bsc_tv,servicing15],axis=0)
    for col15 in encode:
        dummy15 = pd.get_dummies(df15[col15], prefix=col15)
        df15 = pd.concat([df15,dummy15], axis=1)
        del df15[col15]
    df15 = df15[:1]
    prediction_proba15 = load_clf.predict_proba(df15)

    t_prediction = '<p style="font-size: 18px; color: red;"><b>Target: Time taken for
resolving the complaint</b> </p>'
    st.markdown(t_prediction, unsafe_allow_html=True)

    categ, valpre = st.columns(2)
    with categ:
        st.markdown(f"<h3 style='text-align: left;color: red;'>Category</h3>",
unsafe_allow_html=True)
        st.markdown(f"<h3>-</h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
        text1 = 'Air Conditioning'
        st.markdown(f"<h3 style='text-align: left;'>{text1}</h3>",
unsafe_allow_html=True)
        text2 = 'Heating'
        st.markdown(f"<h3 style='text-align: left;'>{text2}</h3>",
unsafe_allow_html=True)
        text3 = 'MDA'
        st.markdown(f"<h3 style='text-align: left;'>{text3}</h3>",
unsafe_allow_html=True)
        text5 = 'SDAP'
        st.markdown(f"<h3 style='text-align: left;'>{text5}</h3>",
unsafe_allow_html=True)
        text6 = 'TV'
        st.markdown(f"<h3 style='text-align: left;'>{text6}</h3>",
unsafe_allow_html=True)

    with valpre:
        st.markdown(f"<h3 style='text-align: left; color: red;'>Prediction
Probability</h3>", unsafe_allow_html=True)
        st.markdown(f"<h3 style='text-align: left;'><label style='color:green'>Week
1</label> | <label style='color:blue'>Week 2</label> | <label style='color:red'>Week
3</label> | <label style='color:orange'>Week 4</label> </h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
```

```
        number10 = round(prediction_proba10[0][0]*100 , 2)
        number101= round(prediction_proba10[0][1]*100 , 2)
        number102= round(prediction_proba10[0][2]*100 , 2)
        number103= round(prediction_proba10[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number10} %</label> | <label style='color:blue'>{number101}
%</label> | <label style='color:red'>{number102} %</label> | <label
style='color:orange'>{number103} %</label> </h3>", unsafe_allow_html=True)

        number11 = round(prediction_proba11[0][0]*100 , 2)
        number111= round(prediction_proba11[0][1]*100 , 2)
        number112 = round(prediction_proba11[0][2]*100 , 2)
        number113= round(prediction_proba11[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number11} %</label> | <label style='color:blue'>{number111}
%</label> | <label style='color:red'>{number112} %</label> | <label
style='color:orange'>{number113} %</label> </h3>", unsafe_allow_html=True)
        number12 = round(prediction_proba12[0][0]*100 , 2)
        number121= round(prediction_proba12[0][1]*100 , 2)
        number122 = round(prediction_proba12[0][2]*100 , 2)
        number123= round(prediction_proba12[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number12} %</label> | <label style='color:blue'>{number121}
%</label> | <label style='color:red'>{number122} %</label> | <label
style='color:orange'>{number123} %</label> </h3>", unsafe_allow_html=True)
        number14 = round(prediction_proba14[0][0]*100 , 2)
        number141= round(prediction_proba14[0][1]*100 , 2)
        number142 = round(prediction_proba14[0][2]*100 , 2)
        number143= round(prediction_proba14[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number14} %</label> | <label style='color:blue'>{number141}
%</label> | <label style='color:red'>{number142} %</label> | <label
style='color:orange'>{number143} %</label> </h3>", unsafe_allow_html=True)
        number15 = round(prediction_proba15[0][0]*100 , 2)
        number151= round(prediction_proba15[0][1]*100 , 2)
        number152 = round(prediction_proba15[0][2]*100 , 2)
        number153 = round(prediction_proba15[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number15} %</label> | <label style='color:blue'>{number151}
%</label> | <label style='color:red'>{number152} %</label> | <label
style='color:orange'>{number153} %</label> </h3>", unsafe_allow_html=True)

    st.markdown("<hr/>",unsafe_allow_html=True)
    st.title('Comparison of categories')
```

```
    df = pd.DataFrame([["", round(prediction_proba10[0][0]*100 , 2),
round(prediction_proba11[0][0]*100 , 2), round(prediction_proba12[0][0]*100 , 2),
round(prediction_proba14[0][0]*100 , 2), round(prediction_proba15[0][0]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 1)",
    xaxis_title="Categories of Brand 2",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][1]*100 , 2),
round(prediction_proba11[0][1]*100 , 2), round(prediction_proba12[0][1]*100 , 2),
round(prediction_proba14[0][1]*100 , 2), round(prediction_proba15[0][1]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 2)",
    xaxis_title="Categories of Brand 2",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][2]*100 , 2),
round(prediction_proba11[0][2]*100 , 2), round(prediction_proba12[0][2]*100 , 2),
round(prediction_proba14[0][2]*100 , 2), round(prediction_proba15[0][2]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 3)",
    xaxis_title="Categories of Brand 2",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][3]*100 , 2),
round(prediction_proba11[0][3]*100 , 2), round(prediction_proba12[0][3]*100 , 2),
```

```
round(prediction_proba14[0][3]*100 , 2), round(prediction_proba15[0][3]*100 ,
2)]],columns=["Product", "AirConditioning", "Heating", "MDA", "SDAP", "TV"])
    fig = px.bar(df, x="Product", y=["AirConditioning", "Heating", "MDA", "SDAP",
"TV"],barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 4)",
    xaxis_title="Categories of Brand 2",
    yaxis_title="Prediction Probability (%)",
    legend_title="Categories",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
```

9.3.6 Application 5

```
import streamlit as st
import pandas as pd
import numpy as np
import pickle
import plotly.express as px
import plotly.graph_objs as go

def app():
    st.title('Heating - Prediction')
    load_clf = pickle.load(open('dataset.pkl', 'rb'))
    def samg_ac():
            brand = ‘BRAND 1’
            category = 'Heating'
            data10 = {'brand':brand,
                    'category':category}
            features10 = pd.DataFrame(data10, index=[0])
            return features10
    input_sam_ac = samg_ac()
    servicing_raw10 = pd.read_csv('dataset.csv')
    servicing10 = servicing_raw10.drop(columns=['weekstorepair'])
    df10 = pd.concat([input_sam_ac,servicing10],axis=0)
    encode = ['brand','category']
    for col10 in encode:
        dummy10 = pd.get_dummies(df10[col10], prefix=col10)
        df10 = pd.concat([df10,dummy10], axis=1)
        del df10[col10]
    df10 = df10[:1]
    prediction_proba10 = load_clf.predict_proba(df10)

    def samg_choc():
            brand = ‘BRAND 2’
            category = 'Heating'
            data11 = {'brand':brand,
                    'category':category}
            features11 = pd.DataFrame(data11, index=[0])
            return features11
    input_sam_choc = samg_choc()
    servicing_raw11 = pd.read_csv('dataset.csv')
    servicing11 = servicing_raw11.drop(columns=['weekstorepair'])
    df11 = pd.concat([input_sam_choc,servicing11],axis=0)
    for col11 in encode:
        dummy11 = pd.get_dummies(df11[col11], prefix=col11)
        df11 = pd.concat([df11,dummy11], axis=1)
```

```
    del df11[col11]
df11 = df11[:1]
prediction_proba11 = load_clf.predict_proba(df11)

def samg_dw():
        brand = ‘BRAND 3’
        category = 'Heating'
        data12 = {'brand':brand,
                'category':category}
        features12 = pd.DataFrame(data12, index=[0])
        return features12
input_sam_dw = samg_dw()
servicing_raw12 = pd.read_csv('dataset.csv')
servicing12 = servicing_raw12.drop(columns=['weekstorepair'])
df12 = pd.concat([input_sam_dw,servicing12],axis=0)
for col12 in encode:
    dummy12 = pd.get_dummies(df12[col12], prefix=col12)
    df12 = pd.concat([df12,dummy12], axis=1)
    del df12[col12]
df12 = df12[:1]
prediction_proba12 = load_clf.predict_proba(df12)

def samg_ht():
        brand = ‘BRAND 4’
        category = 'Heating'
        data13 = {'brand':brand,
                'category':category}
        features13 = pd.DataFrame(data13, index=[0])
        return features13
input_sam_ht = samg_ht()
servicing_raw13 = pd.read_csv('dataset.csv')
servicing13 = servicing_raw13.drop(columns=['weekstorepair'])
df13 = pd.concat([input_sam_ht,servicing13],axis=0)
for col13 in encode:
    dummy13 = pd.get_dummies(df13[col13], prefix=col13)
    df13 = pd.concat([df13,dummy13], axis=1)
    del df13[col13]
df13 = df13[:1]
prediction_proba13 = load_clf.predict_proba(df13)

def samg_mh():
        brand = ‘BRAND 5’
        category = 'Heating'
        data14 = {'brand':brand,
                'category':category}
```

```
            features14 = pd.DataFrame(data14, index=[0])
            return features14
    input_sam_mh = samg_mh()
    servicing_raw14 = pd.read_csv('dataset.csv')
    servicing14 = servicing_raw14.drop(columns=['weekstorepair'])
    df14 = pd.concat([input_sam_mh,servicing14],axis=0)
    for col14 in encode:
        dummy14 = pd.get_dummies(df14[col14], prefix=col14)
        df14 = pd.concat([df14,dummy14], axis=1)
        del df14[col14]
    df14 = df14[:1]
    prediction_proba14 = load_clf.predict_proba(df14)

    def samg_rf():
            brand = ‘BRAND 6’
            category = 'Heating'
            data15 = {'brand':brand,
                    'category':category}
            features15 = pd.DataFrame(data15, index=[0])
            return features15
    input_sam_rf = samg_rf()
    servicing_raw15 = pd.read_csv('dataset.csv')
    servicing15 = servicing_raw15.drop(columns=['weekstorepair'])
    df15 = pd.concat([input_sam_rf,servicing15],axis=0)
    for col15 in encode:
        dummy15 = pd.get_dummies(df15[col15], prefix=col15)
        df15 = pd.concat([df15,dummy15], axis=1)
        del df15[col15]
    df15 = df15[:1]
    prediction_proba15 = load_clf.predict_proba(df15)

    t_prediction = '<p style=" font-size: 18px;"><label><b>Target: Time taken for
resolving the complaint</b></label></p>'
    st.markdown(t_prediction, unsafe_allow_html=True)

    categ, valpre = st.columns(2)
    with categ:
        st.markdown(f"<h3 style='text-align: left;color:red;'>Brand</h3>",
unsafe_allow_html=True)
        st.markdown(f"<h3>-</h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
        text1 = ‘BRAND 1’
        st.markdown(f"<h3 style='text-align: left;'>{text1}</h3>",
unsafe_allow_html=True)
        text2 = ‘BRAND 2’
```

```
        st.markdown(f"<h3 style='text-align: left;'>{text2}</h3>",
unsafe_allow_html=True)
        text3 = ‘BRAND 3’
        st.markdown(f"<h3 style='text-align: left;'>{text3}</h3>",
unsafe_allow_html=True)
        text4 = ‘BRAND 4’
        st.markdown(f"<h3 style='text-align: left;'>{text4}</h3>",
unsafe_allow_html=True)
        text6 = ‘BRAND 5’
        st.markdown(f"<h3 style='text-align: left;'>{text6}</h3>",
unsafe_allow_html=True)
        text5 = ‘BRAND 6’
        st.markdown(f"<h3 style='text-align: left;'>{text5}</h3>",
unsafe_allow_html=True)

    with valpre:
        st.markdown(f"<h3 style='text-align: left; color: red;'>Prediction
Probability</h3>", unsafe_allow_html=True)
        st.markdown(f"<h3 style='text-align: left;'><label style='color:green'>Week
1</label> | <label style='color:blue'>Week 2</label> | <label style='color:red'>Week
3</label> | <label style='color:orange'>Week 4</label> </h3>", unsafe_allow_html=True)
        st.markdown("<hr/>",unsafe_allow_html=True)
        number10 = round(prediction_proba10[0][0]*100 , 2)
        number101= round(prediction_proba10[0][1]*100 , 2)
        number102= round(prediction_proba10[0][2]*100 , 2)
        number103= round(prediction_proba10[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number10} %</label> | <label style='color:blue'>{number101}
%</label>  | <label style='color:red'>{number102} %</label> | <label
style='color:orange'>{number103} %</label>  </h3>", unsafe_allow_html=True)
        number11 = round(prediction_proba11[0][0]*100 , 2)
        number111= round(prediction_proba11[0][1]*100 , 2)
        number112 = round(prediction_proba11[0][2]*100 , 2)
        number113= round(prediction_proba11[0][3]*100 , 2)
        st.markdown(f
"<h3 style='text-align: left;'><label style='color:green'>{number11} %</label> | <label
style='color:blue'>{number111} %</label> | <label style='color:red'>{number112}
%</label> | <label style='color:orange'>{number113} %</label> </h3>",
unsafe_allow_html=True)
        number12 = round(prediction_proba12[0][0]*100 , 2)
        number121= round(prediction_proba12[0][1]*100 , 2)
        number122 = round(prediction_proba12[0][2]*100 , 2)
        number123= round(prediction_proba12[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number12} %</label> | <label style='color:blue'>{number121}
```

```
%</label> | <label style='color:red'>{number122} %</label> | <label
style='color:orange'>{number123} %</label>  </h3>", unsafe_allow_html=True)
        number13 = round(prediction_proba13[0][0]*100 , 2)
        number131= round(prediction_proba13[0][1]*100 , 2)
        number132 = round(prediction_proba13[0][2]*100 , 2)
        number133= round(prediction_proba13[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number13} %</label> | <label style='color:blue'>{number131}
%</label> | <label style='color:red'>{number132} %</label> | <label
style='color:orange'>{number133} %</label>  </h3>", unsafe_allow_html=True)
        number14 = round(prediction_proba14[0][0]*100 , 2)
        number141= round(prediction_proba14[0][1]*100 , 2)
        number142 = round(prediction_proba14[0][2]*100 , 2)
        number143= round(prediction_proba14[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number14} %</label> | <label style='color:blue'>{number141}
%</label>  | <label style='color:red'>{number142} %</label> | <label
style='color:orange'>{number143} %</label> </h3>", unsafe_allow_html=True)
        number15 = round(prediction_proba15[0][0]*100 , 2)
        number151= round(prediction_proba15[0][1]*100 , 2)
        number152 = round(prediction_proba15[0][2]*100 , 2)
        number153 = round(prediction_proba15[0][3]*100 , 2)
        st.markdown(f"<h3 style='text-align: left;'><label
style='color:green'>{number15} %</label> | <label style='color:blue'>{number151}
%</label> | <label style='color:red'>{number152} %</label> | <label
style='color:orange'>{number153} %</label> </h3>", unsafe_allow_html=True)
    st.markdown("<hr/>",unsafe_allow_html=True)
    st.title('Comparison of Brands')
    df = pd.DataFrame([["", round(prediction_proba10[0][0]*100 , 2),
round(prediction_proba11[0][0]*100 , 2), round(prediction_proba12[0][0]*100 ,
2),round(prediction_proba13[0][0]*100 , 2), round(prediction_proba14[0][0]*100 , 2),
round(prediction_proba15[0][0]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 1)",
    xaxis_title="Brands of Heating",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][1]*100 , 2),
round(prediction_proba11[0][1]*100 , 2), round(prediction_proba12[0][1]*100 ,
```

```
2),round(prediction_proba13[0][1]*100 , 2), round(prediction_proba14[0][1]*100 , 2),
round(prediction_proba15[0][1]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 2)",
    xaxis_title="Brands of Heating",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][2]*100 , 2),
round(prediction_proba11[0][2]*100 , 2), round(prediction_proba12[0][2]*100 ,
2),round(prediction_proba13[0][2]*100 , 2), round(prediction_proba14[0][2]*100 , 2),
round(prediction_proba15[0][2]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 3)",
    xaxis_title="Brands of Heating",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
    df = pd.DataFrame([["", round(prediction_proba10[0][3]*100 , 2),
round(prediction_proba11[0][3]*100 , 2), round(prediction_proba12[0][3]*100 ,
2),round(prediction_proba13[0][3]*100 , 2), round(prediction_proba14[0][3]*100 , 2),
round(prediction_proba15[0][3]*100 , 2)]],columns=["Product", “BRAND 1”, “BRAND 2”,
“BRAND 3”, “BRAND 4”, “BRAND 5”, “BRAND 6”])
    fig = px.bar(df, x="Product", y=[“BRAND 1”, “BRAND 2”, “BRAND 3”, “BRAND 4”, “BRAND
5”, “BRAND 6”], barmode='group', height=400)
    fig.update_layout(
    title="Time taken for resolving the complaint (Week 4)",
    xaxis_title="Brands of Heating",
    yaxis_title="Prediction Probability (%)",
    legend_title="Brands",
    font=dict(family="Courier New, monospace", size=14, color="RebeccaPurple"))
    fig.update_traces(texttemplate='%{y}',textposition='inside')
    st.plotly_chart(fig)
```