UNIVERSITETI I EVROPËS JUGLINDORE
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА
SOUTH EAST EUROPEAN UNIVERSITY

FAKULTETI I SHKENCAVE DHE TEKNOLOGJIVE BASHKËKOHORE
ФАКУЛТЕТ ЗА СОВРЕМЕНИ НАУКИ И ТЕХНОЛОГИИ
FACULTY OF CONTEMPORARY SCIENCES AND TECHNOLOGIES

POSTGRADUATE STUDIES-SECOND CYCLE

THESIS:

"Data-Driven Intelligent C2C Mobile Solution"

CANDIDATE:
Imer Ademi

MENTOR:
Assoc. Prof. Dr. Visar Shehu

Tetovo, 2021

UNIVERSITETI I EVROPËS JUGLINDORE
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА
SOUTH EAST EUROPEAN UNIVERSITY

FAKULTETI I SHKENCAVE DHE TEKNOLOGJIVE BASHKËKOHORE
ФАКУЛТЕТ ЗА СОВРЕМЕНИ НАУКИ И ТЕХНОЛОГИИ.
FACULTY OF CONTEMPORARY SCIENCES AND TECHNOLOGIES

POSTGRADUATE STUDIES-SECOND CYCLE

THESIS:

"Data-Driven Intelligent C2C Mobile Solution"

CANDIDATE:
Imer Ademi

MENTOR:
Assoc. Prof. Dr. Visar Shehu

Tetovo, 2021

Personal Statement

I, Imer Ademi, a master's student in the Software Engineering study program offered by the Faculty of Contemporary Sciences and Technologies, state that the master thesis entitled 'Data-Driven Intelligent C2C Mobile Solution' is originally and uniquely written by me. The sources used are cited regularly.

Acknowledgment

Sincere thanks to my mentor, professor Visar Shehu, to commission members prof. Marika Apostolova and prof. Xhemal Zenuni, to all my professors and friends for the help and the great support through the process of writing this master thesis.

Also, special thanks to my family for being there for me and supporting me at every stage of my life.

I, Nedmiran Beqiri, professor of English language and literature, on date 01.07.2021 with full responsibility give this:

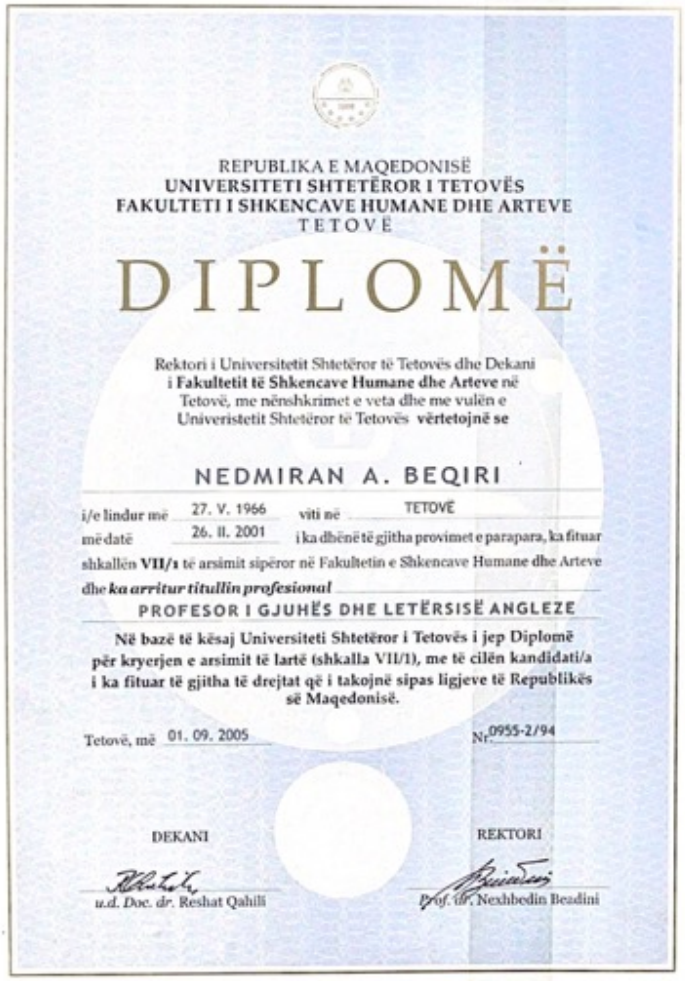
STATEMENT

By which I state that I have revised and confirm the correct form of writing in the English language of the Master Thesis titled:

“Data-Driven Intelligent C2C Mobile Solution”

written by the candidate: Imer Ademi

Tetovë, 01.07.2021 Nedmiran Beqiri



REPUBLIKA E MAQEDONISË
UNIVERSITETI SHTETËROR I TETOVËS
FAKULTETI I SHKENCAVE HUMANE DHE ARTEVE
TETOVË

DIPLOMË

Rektori i Universitetit Shtetëror të Tetovës dhe Dekani i **Fakultetit të Shkencave Humane dhe Arteve** në Tetovë, me nënshkrimet e veta dhe me vulën e Univeristetit Shtetëror të Tetovës **vërtetojnë se**

NEDMIRAN A. BEQIRI

i/e lindur më 27. V. 1966 viti në TETOVË
më datë 26. II. 2001 i ka dhënë të gjitha provimet e parapara, ka fituar shkallën **VII/1** të arsimit sipëror në Fakultetin e Shkencave Humane dhe Arteve dhe ***ka arritur titullin profesional***

PROFESOR I GJUHËS DHE LETËRSISË ANGLEZE

Në bazë të kësaj Universiteti Shtetëror i Tetovës i jep Diplomë për kryerjen e arsimit të lartë (shkalla VII/1), me të cilën kandidati/a i ka fituar të gjitha të drejtat që i takojnë sipas ligjeve të Republikës së Maqedonisë.

Tetovë, më 01. 09. 2005 Nr. 0955-2/94

DEKANI
u.d. Doc. dr. Reshat Qahili

REKTORI
Prof. dr. Nexhbedin Beadini

РЕПУБЛИКА МАКЕДОНИЈА
ДРЖАВЕН УНИВЕРЗИТЕТ ВО ТЕТОВО
ФАКУЛТЕТ ЗА ХУМАНИ НАУКИ И УМЕТНОСТ
ТЕТОВО

ДИПЛОМА

Ректорот на Државниот Универзитет во Тетово и Деканот на **Факултет За Хумани Науки и Уметност** во Тетово, со своите потписи и со печатот на Државниот Универзитет во Тетово **потврдуваат дека**

НЕДМИРАН А. БЕ́КИРИ

роден/а на 27. V. 1966 год. во ТЕТОВО
на ден 26. II. 2001 год. го завршил полагањето на пропишаните испити и се здобил со високо образование **VII/1** на Факултет за Хумани Науки и Уметност и ***стекнал право на стручен назив***

ПРОФЕСОР ПО АНГЛИСКИ ЈАЗИК И КНИЖЕВНОСТ

Врз основа на тоа Државниот Универзитет во Тетово му/ѝ издаде Диплома за завршено високо образование (VII/1) степен, со која кандидатот/ка ги стекнува сите права што му/ѝ припаѓаат според законите на Република Македонија

Во Тетово, 01. 09. 2005 Бр. 0955-2/94

ДЕКАН
в.д. Доц. д-р. Решат Кахили

РЕКТОР
Проф. д-р. Недбедин Беадини

Table of Contents

Table of Figures

Abstract

Nowadays, smartphones have become one of the keyways people search for any kind of help for their problems. One of the problems of the new generation these days is to find a service provider for a specific issue they have. Also, service providers struggle to find new customers to continue and grow their business.

So, a data-driven mobile platform that would allow service providers to connect with customers is much needed in our country. Data generated from this application will be used to analyze the market needs in North Macedonia and decide how to proceed with the development according to the recommendations.

This master thesis uses methods for quantitative data gathering to analyze the needs of the market in North Macedonia (in the form of surveys and interviews) and build a mobile application that will help people find service providers more easily and in a qualitative way. On the other hand, to help service providers to find new customers, grow their business, organize their job requests, etc.

Also, data generated from this application will be modeled for developing predictive methods. First, we will analyze the needs of the market in North Macedonia. And then, the data gathered will be designated to support classification techniques that will be used to improve the system using recommendation techniques. And the usefulness of these techniques will be evaluated through methodologies for algorithm evaluation.

Keywords: ***data-driven, data, application, mobile, algorithm.***

Abstrakt

Në ditët e sotme, telefonat mobil janë bërë një nga mënyrat kryesore në të cilat njerëzit kërkojnë çdo lloj ndihme për problemet e tyre. Një nga problemet e gjeneratës së re këto ditë është gjetja e një ofruesi të shërbimit për një çështje specifike që ata kanë. Gjithashtu, ofruesit e shërbimeve luftojnë për të gjetur klientë të rinj për të vazhduar dhe rritur biznesin e tyre.

Pra, një platformë mobil e drejtuar nga të dhënat që do të lejojë ofruesit e shërbimeve të lidhen me klientët është shumë e nevojshme në vendin tonë. Të dhënat e krijuara nga kjo aplikacion do të përdoren për të analizuar nevojat e tregut në Maqedoninë e Veriut dhe të vendoset si të vazhdohet me zhvillimin, sipas rekomandimeve.

Kjo tezë e magjistraturës përdorë metoda për grumbullimin e të dhënave sasiore për të analizuar nevojat e tregut në Maqedoninë e Veriut (në formën e anketave dhe intervistave) dhe të ndërtojë një aplikacion mobil që do t'i ndihmojë njerëzit të gjejnë ofruesit e shërbimeve më lehtë dhe në mënyrë cilësore. Në anën tjetër, për të ndihmuar ofruesit e shërbimeve të gjejnë klientë të rinjë, të risin biznesin e tyre, të organizojnë punën e tyre, etj.

Gjithashtu, të dhënat e gjeneruara nga ky aplikacion do të modelohen për zhvillimin e metodave parashikuese. Ne do të analizojmë nevojat e tregut në Maqedoninë e Veriut. Dhe pastaj të dhënat e mbledhura do të përcaktohen për të mbështetur teknikat e klasifikimit që do të përdoren për të përmirësuar sistemin duke përdorur teknika të rekomandimit. Dhe dobia e këtyre teknikave do të vlerësohet përmes metodologjive për vlerësimin e algoritmit.

Fjalët kyçe: ***data-driven, data, application, mobile, algorithm.***

Апстракт

Денешните денови, мобилните телефони станаа еден од главните начини на кои луѓето бараат секаков вид на помош за своите проблеми. Еден од проблемите на новата генерација деновиве е изнаоѓање на давател на услуги или мајстор за специфичен проблем што го имаат. Исто така, давателите на услуги се борат да најдат нови клиенти за да продолжат и растат нивниот бизнис.

Значи, мобилна платформа режирано од податоци што ќе им овозможи на давателите на услуги да се поврзат со клиентите е многу потребна во нашата земја. Податоците генерирани од оваа апликација ќе се користат за анализирање на потребите на пазарот во Северна Македонија и одлучување како да се продолжи со развојот, според препораките.

Оваа магистерска теза користи квантитативни методи за прибирање податоци за да ги анализира потребите на пазарот во Северна Македонија (во форма на анкети и интервјуа) и да изгради мобилна апликација која ќе им помогне на луѓето да најдат провајдери на услуги полесно и квалитативно. Од другата страна, да им помогне на провајдерите на услуги да најдат нови клиенти, да го зголемуваат својот бизнис, да ја организираат својата работа, и тн.

Исто така, податоците генерирани од оваа апликација ќе бидат моделирани за развој на предвидливи методи. Ние ќе ги анализираме потребите на пазарот во Северна Македонија. И потоа собраните податоци ќе бидат утврдени да ги поддржат техниките на класификација што ќе се користат за подобрување на системот користејќи ги техниките за препорака. И корисноста на овие техники ќе се процени преку методологии за проценка на алгоритми.

Клучни зборови: ***data-driven, data, application, mobile, algorithm.***

1. Introduction

Last two decades, the IT industry grew exponentially, especially web and mobile computing, which has played a significant role in developing today's digital age.

Nowadays, smartphones are prevalent and have become a part of our lives. Also we live in the period of big data where the data produced every day is way more than we can think. However, suppose this enormous amount of information can be organized and used effectively. In that case, it can be beneficial for analyzing and predicting the future by using techniques from Machine Learning, Data Mining, Statistics, Artificial Intelligence, etc.

So, a data-driven mobile platform that would allow service providers to connect with customers is much needed in our country. Data generated from this application will be used to analyze the market needs in North Macedonia and decide how to proceed with the development according to the recommendations. Also, the data generated from this app will be modeled for developing predictive methods.

Methods for quantitative data gathering will be used to analyze the needs of the market in North Macedonia. Also, we will build a mobile app with React Native that will help to find service providers much more straightforward and in a quality way. Also, on the other side will allow service providers to grow their business by making possible to have access to more new customers and organize their job.

2. Background Research and State of the Art

2.1. Background Research

The evolution of the internet and technology has played a considerable role in developing the present digital age. The utilization of the internet has moved beyond desktop computers to smartphones. Consistent with ITU (International Telecommunication Union), cellular network coverage has reached 96.8% of the planet's population, and this number even matches 100% of the people within the developed countries, according to [1].

According to [2], the author says that with the development of technology and science, the smartphone industry is overgrowing. As a result, the trend of smartphones has made exponential growth in the mobile apps market.

These days smartphones are used by anyone and have become one of the main ways how everyone searches for any help for their problems. One of the problems of the new generation these days is to find a service provider for a specific issue they have. Also, service providers struggle to find new customers to continue and grow their business.

According to [3], mobile apps and smartphones represent an ample opportunity for serious growth for so many businesses. Tools that enable a far better UX (user experience), notifying in real-time with notifications on their pocked, consistent and fast communication with customers, and organizing their jobs are already making many businesses a lot of money also helps to grow faster.

As it is said in [4], data-driven applications have become a role player for the growth of the worldwide software market. Smart computing in the future will become a multi-billion dollars market, analysts claim; they also say that “we live in the period of data-driven marketing and sales.”

So, a data-driven mobile platform that would allow service providers to connect with customers is much needed in our country. Data generated from this application will be used to analyze the market needs in North Macedonia and decide how to proceed with the development according to the recommendations.

This master thesis uses methods for quantitative data gathering to analyze the needs of the market in our country and build a mobile platform to end time losing while searching over the internet and make appointments blindfolded; instead, with our app, the appointment will be made in qualitative way. On the other hand, service providers do not have to have a physical location in the center of the city to be easily found, instead, they will have a virtual store in the app which will help them grow their business by easily reaching to the new customers.

2.2. State of the Art

The web has now formed the backbone of recent communication. Nowadays, the utilization of the web has moved towards smartphones. As a result, according to [5], cellular network coverage has reached almost all the world's population, especially the population of the developed countries.

This trend also continues in North Macedonia too. Almost everyone has access to the internet and uses smartphones, especially the young generation.

In recent years, the large migration of youngsters and service providers from our country has caused the lack of workers in the country, besides except the younger generation, everyone has difficulties in finding service providers in an easy and quality way.

So, in North Macedonia, a mobile platform that will act as a hub for both service providers and customers is much needed. It will save time, help pick the best quality service for customers, and support service providers to grow their business.

To perfect this application, we will model our system to support classification, recommendation, and clustering.

3. Motivation, Problem Statement, and Proposed Solution

3.1. Motivation

The tremendous growth of the technology has reflected to an extensive widespread of smartphones, smartphone apps, and smartphone users around the world. Smartphone users across the globe reach 3.5 billion, and it's growing every year by a considerable margin. The number of apps in the two biggest app markets goes to 5 million, according to [6].

As technology grew, also, the internet has improved itself. The new evolution of the internet began in November 2018 with the "Starlink" project of "SpaceX" under the leadership of Elon Musk. "SpaceX" with "Starlink" plans to build an interconnected network with thousands of satellites designed to deliver high speed, low latency, and affordable internet to anyone everywhere in the world, says [7].

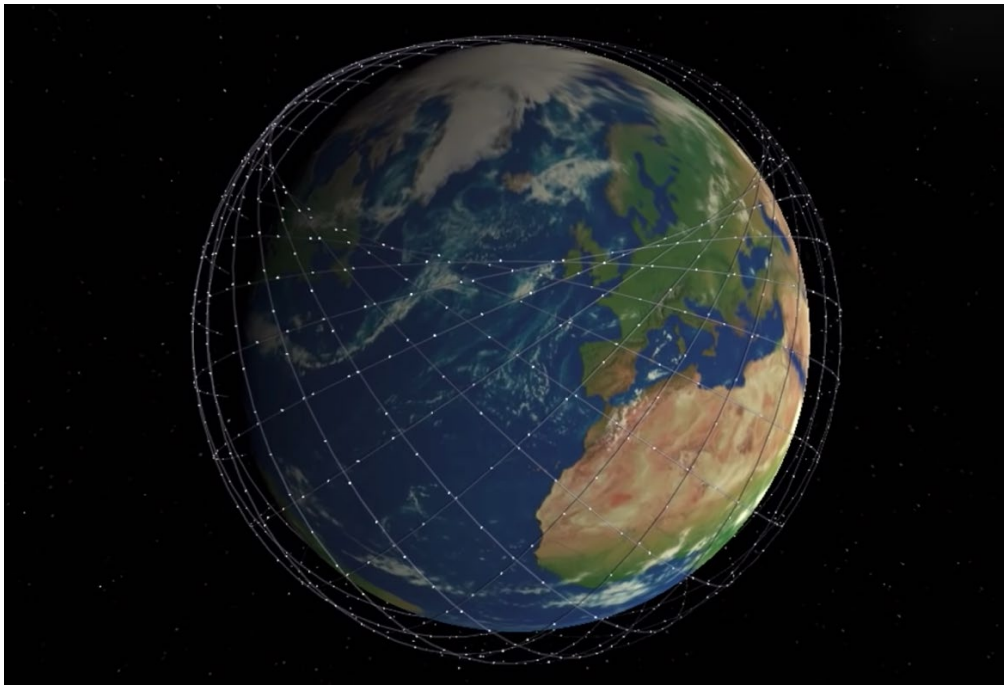


Figure 1. A simulation of SpaceX's Starlink satellite internet service coverage.
Source: Retrieved from [8]

Also, according to the author, 11,943 satellites have been approved to launch in late 2018, with a possible later extension to 42,000. So far, 1,023 satellites have been deployed

in orbit. Satellites are launching with the Falcon 9 rockets 60 satellites each time, and last year in summer, they've said that they are building around one hundred and twenty satellites a month, and to match this massive production, the launch rate will move up as well.

Also, apps like "Reklama5" and "Pazar3" have played a significant role in my motivation for this thesis. These apps are popular in North Macedonia and are used to buy and sell things from cars to houses and more.

That actually is a brilliant idea because everyone can offer something to sell even if they don't have a physical place/selling point. For example, car dealers can be a company with many cars. Still, it is located 100km away from a potential buyer or even a person who wants to sell their car, but the potential buyer could be in another city. And the buyer or Customer would like and want that car from that dealer or seller far away, but he or she won't want to risk and go around the country to try their luck if they can find their dream car somewhere. In this case, these apps are a hub where the sellers and buyers meet each other easily. Sellers find unique customers, and buyers explore more from their smartphones.

3.2. Problem Statement

The new generation these days in North Macedonia encounter difficulties finding a service provider for a specific issue they have. The mass relocation of service providers abroad has made it even harder to find a service provider in our country. So, the idea is to help people, especially the new generations, with their everyday problems by offering smart and quality solutions right in their pocket.

On the other hand, to help service providers find new customers easily by opening a virtual location in the app where they will have immediate access by everyone even if they don't have a physical location in the city center. The app will help them grow their business, organize their job requests, show their work quality, etc.

And according to this analysis, it is seen that a mobile platform that would allow service providers to connect with customers is a necessity in North Macedonia.

3.3. Proposed Solution

From this starting point, the idea behind this project was to create a hub where both service providers and customers will meet. Customers will save time and pick quality service providers, while service providers will be able to grow their business by having access to more customers, even the new ones. It will be a platform where they will compete with their quality.

Scenario 1:

Berat is from Struga. He is working in Tetovo as a professor. One day Berat spotted a water leak from the water pipe in the kitchen, so as there is an issue with the pipes, he decides to call a Plumber to fix it. Berat is new in Tetovo, and he has no Plumber friend or neighbor. In this case, Berat grabs his smartphone and enters our app, where he searches for a good plumber by selecting the category and the rating of the plumbers. He finds one that suits his needs and makes an appointment for 30 minutes later because the plumber was available at that moment. The issue gets solved immediately.

Scenario 2:

Agim is a carpenter, and his work is brilliant, but unfortunately, Agim couldn't afford to open a workshop in the city to be near potential customers. His workshop is outside the city, but his work is so good, and he needs to show and convince potential customers about his work and talent. In this case, Agim opens a virtual location in the app where he shares his work uploading photos, and for his fantastic work, his customers rate him with five stars, so he ranks upon filtering. With this move, Agim is much nearer to his customers even if his workshop is outside the city and his fantastic work and perfect rating make it to find more new customers and grow his business. He also can organize his job requests and contact with customers quickly and easily.

3.4. Hypotheses

For this research, we will present some hypotheses which during and elaboration of the thesis will be proven in their completion or non-completion. Hypotheses of this master thesis are as follows:

H1. A mobile platform that would allow service providers to connect with customers is much needed in North Macedonia.

H2. Data generated from such application can be modeled for developing predictive methods.

H3. Applications can be developed where business process flow will be governed by data.

3.5. Research Methodology

To research and study this problem comprehensively, we will use the following research methods:

- We will use methods for quantitative data gathering to analyze the needs of the market in North Macedonia. Such methods will be in the form of surveys and interviews. Using such research, we can verify the necessity of such applications in the market in North Macedonia and further improve our proposed solution.
- Quantitative methods will also be used for developing predictive methods. Finally, the data gathered from our application will be designated to support classification techniques that can be further used to improve the system using recommendation techniques. The novelty in this approach is that the application will be designed so each process flow is governed by data.
- The usefulness of these techniques will be evaluated through well-established methodologies for algorithm evaluation: cross-validation, ROC, and similar.

4. Smartphones and Smartphone Applications

4.1. The birth and popularity of Smartphones and Smartphone Apps

The author in [6] claims that the smartphone industry had moved forward very quickly since 2007 when Apple introduced their first customer-friendly iPhone with a touch screen interface that revolutionized the smartphone industry.

Roughly 45% of the world's population was projected to own a smartphone by the end of 2020 [6] says, also by the end of 2023, global smartphone shipments are projected to reach around 1.48 billion units. Thus, as shown in Figure 1, in 2020, the number of smartphone users is approximately 3.5 billion, and it is growing by 300 million users every year.

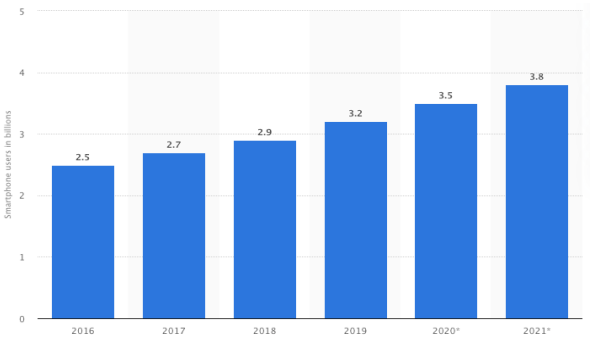


Figure 2. Number of smartphone users between years: 2016 to 2021 (in billions)
Source: *Retrieved* from [55]

On the other side, as shown in Figure 2, in 2019, 204 billion mobile applications were downloaded worldwide. Actually, the number of apps available in the two biggest app markets, in Google Play Store and Apple App Store is around 5 million apps. So, as well as application downloads, the trend of creating new applications is growing every day.

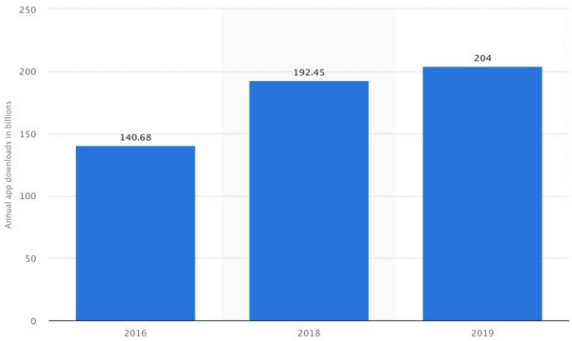


Figure 3. Mobile app downloads worldwide between years: 2016 to 2019 (in billions)
Source: *Retrieved* from [56]

4.2. Smartphone Applications

A smartphone app is a software and a kind of app developed to work on a smart device like mobiles and tablets. Even if apps are usually small with limited functionality, they provide very high-quality service and experience. [9]

So, as it is mentioned above, there are two major platforms, Android and Apple. Applications for these platforms can be developed in two ways. The first way is Native App Development which means different developments for different platforms, and the second way is Hybrid App Development which means single development for different platforms.

4.3. Native Applications

According to [10], native app development relies on tools and programming languages designed specifically for one platform. For example, languages that dominate developing Native iOS Apps are Objective-C and Swift. On the other hand, for Native Android Apps are Java, Kotlin, or C/C++. So, if it's needed to create an application for different platforms or operating systems with native application development, it would happen by building different apps for different platforms. Eventually, it would double the effort of developing the app in question.

4.4. Hybrid Applications

As it is said in [10], Hybrid App Development or Cross-Platform apps development costs are relatively low since it doesn't need too many programmers to work in two separate teams for different platforms. In addition, development is faster, and the code is sharable between platforms, so it aims that a single app to run identically on several platforms. Languages that dominate developing Cross-Platform apps are React Native, Xamarin, Ionic, etc.

4.5. React Native as a Hybrid App Development Platform

As [11] says, React Native is an open-source framework from Facebook that allows developers to create cross-platform apps using JavaScript. React Native works differently when comparing to other cross-platform frameworks that use JavaScript.

According to [12], the logic is written in JavaScript, and it compiles the JS code to native components, thereby using APIs and modules of that specific platform. And developers as well can create new components by using native components as View, Text, etc.

Thus, react Native plays the role of a bridge between the developed app and the target platform. React Native's bridge system kicks in when a native code runs alongside the JS code and uses the React library, and transfers components to the device's view.

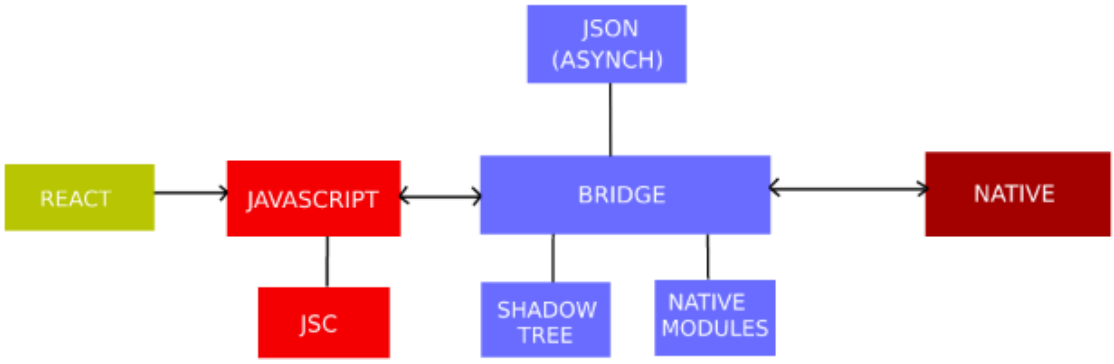


Figure 4. React Native Actual Architecture
Source: Adapted from [12]

According to [12], the future of React Native is bright. The React team is currently working hard to improve each platform's sections individually. The author says that Facebook announced the new architecture that looks like Figure 5.

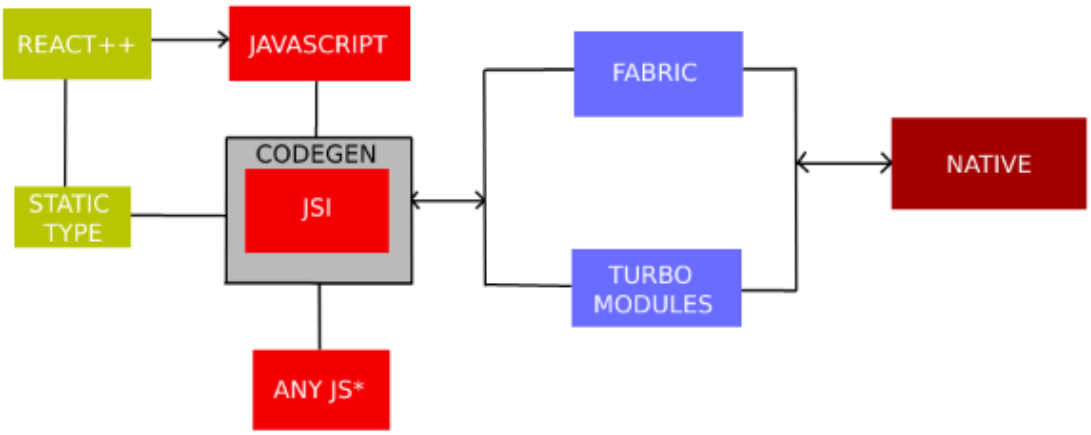


Figure 5. React Native New Architecture
Source: Adapted from [12]

React Native is worthy because it offers faster development, costs less, provides cross-platform development, and the efficiency of developing native apps increases many-fold. It is among the leading frameworks for cross-platform development, and it focuses on UI. In addition, it has access to native API out of the box.

It also provides a single code base for both Android and iOS. The developed app performs like a native app on any platform. It has compatibility with so many third-party plugins and so on.

So, the above-mentioned reasons played a big role in why we decided our mobile app to be built with React Native.

5. Literature Review - Mobile App Development and Best Predictive Analytics Models and Algorithms

5.1. C2C Mobile Application Development

According to [13], the author claims that because of the popularity of smartphones nowadays and context-aware technologies, different contextual data related to various user activities with smartphones are out there around us. This also allows the study of mobile data for the aim of building data-driven mobile apps in a distributed environment for the advantage of end-users.

We live in the period of data science. On the other side, computing is increasing the lead by using smartphones which also support data-driven mobile apps, according to [14].

As it is said in [15] for the companies adopting a data-driven approach to their development, the web industry has become very capable of gathering insights on how end-users behave when visiting their websites, which has led to increased user satisfaction, revenue, also has lowered development costs. However, the author also says that the mobile apps development business is still in its fancy. Many apps are developed, released, and then partly updated with content based on gut feelings, and he claims that there is much to be learned from the web industry.

Mobile devices have changed how we understand the software and how we conceive it, says the author in [16]. However, the author also says that there is a range of development alternatives. Also, in his paper in [16] [17], the author speaks about the development environment explicitly designed for data-oriented mobile apps called Applause [18]. It uses a specific domain language based on the “Xtext” framework and generates source code in Java, C#, Python, or Objective C. Also, the author in [19] claims that we can have great advantages by using the Hybrid way of creating applications because of code reuse for various platforms as well as access to device hardware and distribution through application stores.

According to [20] development of mobile apps in recent times have an exponential growth. Nowadays, everyone in the world has a smartphone in their pocket. The authors also claim that smartphones help us in many ways through the application development procedure.

One of the application developments models that have advantages in many ways is the Cross-Platform Application Development model. Authors say that this model of app development is the development of mobile-based applications so that the development of these types of apps can be made platform-independent.

According to [21], mobile apps are having a huge impact on many sectors of society. Mobile apps can be classified into three categories such as native, mobile web, and hybrid apps, and there are available a lot of tools and SDKs to develop these apps, authors say. They also say that apart from native tools and SDKs, there are many open-source tools that allow the development of cross-platform apps so that the code can work, including small changes or even with no change at all, in more than one operating system such as Android and iOS.

They also in [21] claim that languages such as Objective C for iOS and Java for Android that are used in native mobile app development must be used to develop both the UI and logic parts of the app. On the other hand, cross-platform app development uses HTML5, JS, and other open-source libraries such as jQuery or jQtouch. So, front-end developers also can apply their experiences to create smartphone apps.

Next, the authors in [21] continue to explain about the three development platforms mentioned above. About native applications, they say that these apps are developed using tools, languages, and SDKs natively that belong to a particular OS. They give an example with iOS and Android. IOS apps are developed using iOS SDK, X-Code, and Objective C and Android apps, which are developed using Android SDKs, IDE such as Eclipse and Java. About mobile web apps, they say that the app uses an instance of a mobile web browser to run the app. For example, apps that use this method are: m.yahoo.com, m.facebook.com, etc. These applications, they say, are build using cross-platform SDKs and open-source libraries. The UI is developed using JavaScript and HTML, also JS is used for the

logic. Finally, the app is hosted on a web server and can be accessed by using an instance of a web browser. About the hybrid applications, the authors claim that they are basically the combination of the native mobile apps and mobile-web apps. Also, they say that these apps are built using open-source libraries. For this reason, they also have access to the phone's camera, GPS, file system, and more. The User Interface of the Hybrid Apps is developed using HTML or JavaScript, also JavaScript is used to create the logic. These tools usually are used to convert a mobile web application to a native application, authors claim in their paper. The authors have made a table where they show the overall comparison between three mobile application types in Figure 6.

Type of Application/ Features	Native mobile Apps	Mobile-Web Apps	Hybrid Apps
Programming Language	Java(Android), Objective C(iOS), J2ME(BB)	HTML5 and JavaScript	HTML5 and JavaScript
Executable	Binary (.apk,. ip,. jar)	HTML + JavaScript	Binary
Distribution	Appstore, Market	Hosted on a web server, hyperlink	Either appstore or hyperlink
Execution of Apps	Directly by Operating System	By Web browser	By operating system, web portions executed by browser
API Usage/Access	Full Access to operating system API's	No Access to OS API's	Limited access to OS API's
Device Access	Full Access	Limited Access	Full Access
Speed of Apps	Very fast	Medium	Very fast
Development Cost	Expensive	Reasonable	Reasonable
Approval Process	Mandatory	None	Low overhead
SDK's, Tools	Android, iOS, BB, Symbian, Windows SDK's	jQuery, jQtouch, Sencha	Titanium, PhoneGap, Rho Mobile

Figure 6. Comparison Between Three Mobile Application Types
Source: Retrieved from [21]

According to [22], the authors claim that mobile applications nowadays are targeted at specific mobile platforms. But that comes with heavy restrictions, some of which are diverse technologies, APIs, development environments, and as they claim, this leads to a significant loss of creation time also increases maintenance costs as well. In their paper, authors also discuss the most crucial cross-platform app types such as web, hybrid, interpreted, and generated.

Web apps are apps based on browsers and are based on technologies such as HTML and JavaScript, they say. The authors in [22] also speak about the disadvantages of web applications which are: the pretty limited access to the device's hardware and limited

access to devices data as well. So, it means that there is an extra time required to render the page, and therefore the additional cost to get the webpage downloaded from the internet. But on the other hand, web applications do not require installation and upgrades.

Hybrid apps are the combination of advantages of the web and native apps, authors claim. They are primarily built by using HTML and JavaScript. Similar to web applications, the source code of the hybrid apps is still executed by the browser but can be packed with the application as well says the author in [22], unlike the web apps where the code is downloaded from the web. Hybrid apps have access to the device's hardware and data, and the reason for this is because they are installed directly on the device.

In interpreted apps, the authors claim that the native code automatically is generated to implement the UI with the native components, while the logic is implemented independently using programming languages like Java, Ruby, XML, etc. [22]. The efficiency of the native user interface is the main advantage of this application type, but on the other hand, the complete dependence on the development environment is the downside.

For generated applications, authors in [22] claim that these applications are compiled as a native application, and for each platform, a specific version is created. Applause [18] is the widespread development environment for creating these types of apps, they say. Because of the generated native code, Generated Apps achieve high overall performance as well.

The authors also have created a table where they show the result of the comparative analysis in Figure 7.

	Web	Hybrid	Interpreted	Generated
Marketplace deployment	No	Yes, but not guaranteed*	Yes**	Yes**
Widespread technologies	Yes	Yes	Yes	No
Hardware and data access	Limited	Limited	Limited	Full access
User interface and look & feel	Simulated	Simulated	Native	Native
User-perceived performance	Low	Medium	Medium	High

Figure 7. Comparative analysis of cross-platform development approaches
Source: Adapted from [22]

5.2. Data Modeling Approaches for Predictive Analysis

According to [23], we live in an era where the data has become the new oil, and it's vital to possess the proper technique, tools, and models to process 2.5 quintillion bytes of data produced regularly. Predictive data analysis is a technology that can predict the future trends using historical data and by analyzing them and using that information to draw up predictions about the general direction of the industry.

According to [24], predictive analysis models are created to gauge historical data, observe trends, discover patterns, and use that information to make predictions.

The author also says that predictive analytics models aren't single-shaped. There are different models developed for design-specific functions, some of which are: Clustering model, Forecast models, Outliers models, Classification models, Time series model. The author also speaks about how predictive analytics models work. He says that they also have advantages and disadvantages and are best used for specific users. The most important advantage of using predictive analytics models is that they're reusable also adjustable. Analytical models execute one and sometimes more algorithms in a dataset where the prediction will be made. In fact, it is a process that repeats because it covers the training of the model. Also, the author claims that the predictive analytics models work in a repetitive process. The order of the processes goes like this: the process starts with the pre-processing, then for understanding business objectives, it continues with data mining, and then with data preparation. After that is done, data is modeled, evaluated, and deployed. And in the end, after the process is completed, the whole process is repeated.

In analysis, data algorithms are playing an enormous role since they're utilized in data mining and analysis to assist in determining data models and patterns [24]. There are many algorithm types in analytical models combined to perform specific functions. Some algorithms listed by the author in [24] are: "regression algorithms," "association algorithms," "clustering algorithms," "time-series algorithms," "outlier-detection algorithms," "neural-network algorithms," and "decision trees." So, each algorithm is used for a specific purpose. For example: for detecting anomalies within the dataset, outlier detection algorithms are used.

According to [25], predictive modeling is perhaps the most used subfield of data mining, utilizing machine learning, optimization techniques, pattern recognition, statistics, and DB techniques.

The author also talks about what data mining is. He in [25] says that it is the procedure by which we extract precise and unidentified info from big sizes of data. This material must be in a shape that can be understood and used for refining decision-making processes in the business and industry. Algorithms for data mining or data analysis can be divided into three categories author says in [25], and they are: “frequent pattern extraction,” “predictive modeling,” and “clustering.”

As claimed in [26], “Predictive Analytics” is a term mainly utilized in statistical and analytics techniques, and it is utilized to predict future events. So, to make predictions about the future events, it analyzes the present and historical data using techniques from Data Mining, ML, AI, and Statistics.

The author says that there are several steps that predictive analytics involves through which data analysts can predict the future events based on historical and present data, and below is given a representation of this predictive analytics process in Figure 8.

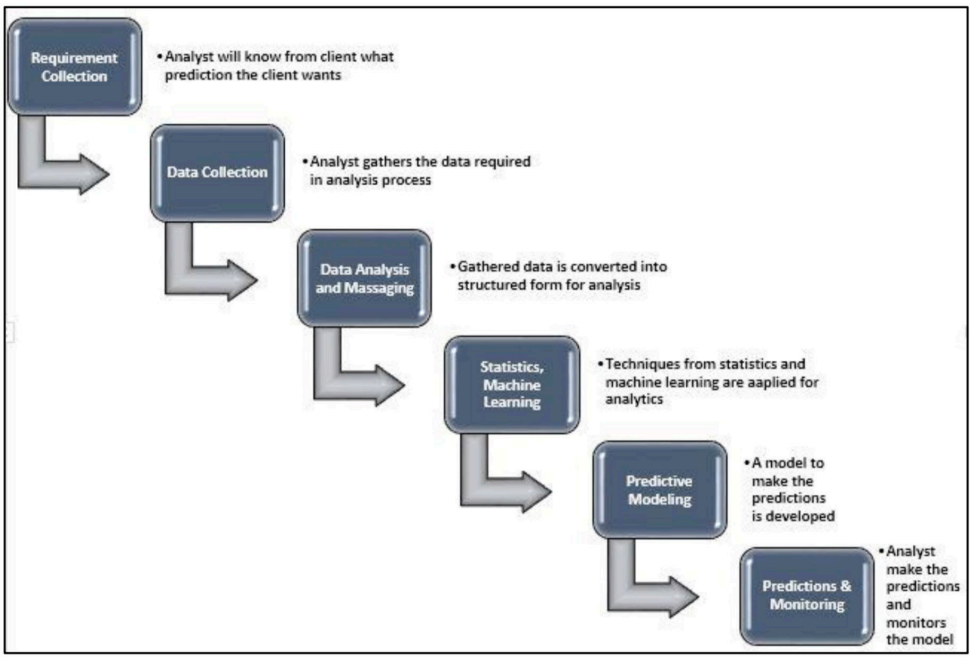


Figure 8. Predictive Analytics Process
Source: Retrieved from [26]

According to [27], everyday petabytes of data are generated, mostly from social media and business organizations. And always there was a question of how to manage effectively and efficiently this gigantic amount of data. The author also claims that each day the amount of data generated from Google is 25 petabytes, 7 petabytes are produced from Twitter, and over 500 terabytes of data are generated at Facebook.

So, this big data, if it's not structured and organized it means nothing, but if it's organized into a recognizable structure, it could be very useful for analyzing and predicting the future events. But unfortunately, as the authors claim in [28] [29], if we talk about the future, the IT industry didn't have the capacity to handle such enormous amounts of data. Still, with the exponential growth of the IT industry nowadays, there are many ways and tools to handle the data available in effective and efficient ways. The author also says that Apache Hadoop is one of the best and successful tools that can be used for handling and managing big unorganized data.

Nowadays, systems based on predictive modeling are becoming very popular because of the big data collections that decision-makers cannot handle manually. Predictive models are usually used to find valuable information or to predict future events. There are many predictive techniques, from simple techniques like linear regression to more complex predictive techniques such as artificial neural networks. If the case is to get better performing predictions, then complex predictive techniques are needed to be used, but as they are more complex, they cannot be used to explain predictions or discovered patterns. Choosing the proper predictive technique is very hard because no technique is better than others over a large set of problems. Every technique has their strength and weaknesses. Even the parameter values are problem dependent which makes it harder to find the best settings. The only way or solution for this problematic choice is to combine several models by creating different techniques and settings for different kinds of problems. So, this confusion will be eliminated by gathering everything in a powerful and robust ensemble, and the diversity there will help with estimating the uncertainty associated with every prediction, says [30].

5.3. Best Predictive Analytics Models and Algorithms

5.3.1. Top Predictive Analytics Models

5.3.1.1. Classification Model

The classification model is the simplest predictive analytics method; it categorizes data based on what it learns from historical data. These models are best for answering yes or no questions, so it provides an exhaustive analysis that is helpful for directing decisive action, says the author in [31]. An example of a Classification model with Decision trees algorithm can be found in Figure 9.

Figure 9. Decision Tree Algorithm with Yes/No Questions
Source: Adapted from [32]

According to [32], the "Classification Analytics Model" is the process of identifying, understanding, and grouping ideas and objects into pre-established categories or "subpopulations." Pre-classified training datasets are used by machine learning programs using numerous algorithms to classify prospective datasets [32].

Machine learning classification algorithms use training inputs to predict the probability into which specific preset category the following data will fall [32]. One of the most common classification is email filtering as "spam" or "not spam," as it is said in [32].

Long story short, "classification" is a form of "pattern recognition" that applies a classification algorithm to training data to find patterns that will be the same in a future dataset, says the author in [32].

Also, the author in [33] claims that the detection of spam emails by email service providers can be classified as a classification problem. In this case, there are two classes, so as the classes are spam and not spam, we can call it a binary classification. Classifiers use training data to understand how a certain input variable relates to a class. So, the emails

recognized as spam and not spam should be used as a training data. Once the classifier has been correctly trained, it can be used to spot unidentified emails [33].

A simple example of the Classification model can be found down in Figure 10.

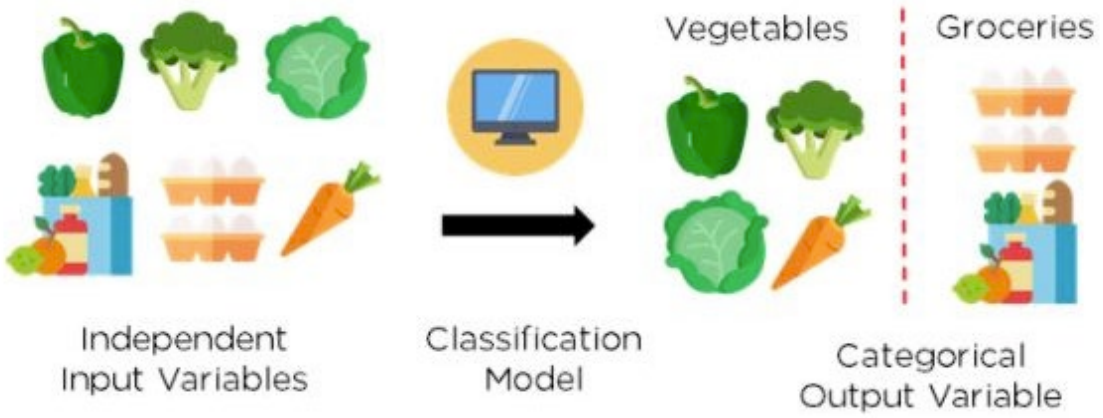


Figure 10. Classification of vegetables and groceries
Source: Retrieved from [34]

According to [33], there are 2 types of classification learners. Lazy Learners and Eager Learners.

- **Lazy Learners:** in this case, they store the training data and then wait for the testing data to appear. And when it happens, the classification is managed based on the most relevant data from the stored training data. As a result, "Lazy Learners" when comparing to "Eager Learners", they have less time, but they have more time on making predictions, says the author in [33].

- **Eager Learners:** builds a classification model based on the training data provided before the classification data has arrived. The author in [33] claims that it should stick to one hypothesis covering the entire instance space. Due to the construction of the model, "Eager Learners" have more training time and have less time on making predictions, the author adds in [33].

5.3.1.2. Clustering Model

Clustering is a predictive analytics model that divides data into separate and smart groups. Such data in the same group have similar attributes and assign them into clusters, as it is said in [35].

According to the authors in [36], the main idea of clustering is to organize datasets into clusters so that the items in a group are more similar to each other when comparing to elements on different clusters.

Some use cases for Clustering, according to the author in [37], are as follows:

- "Recommender systems (grouping together users with similar viewing patterns on Netflix, in order to recommend similar content)."
- "Anomaly detection (fraud detection, detecting defective mechanical parts)."
- "Genetics (clustering DNA patterns to analyze evolutionary biology)."
- "Customer segmentation (understanding different customer segments to devise marketing strategies)."

An example of this model would be, as it is said in [38], let's consider a dataset of different types of fruits in a basket. So, we have a basket with different types of fruits such as apples, pears, bananas, and lemons. Fruits are the data item in this case. The data clustering process clusters or groups fruits of similar type from this dataset, an example shown in Figure 11.

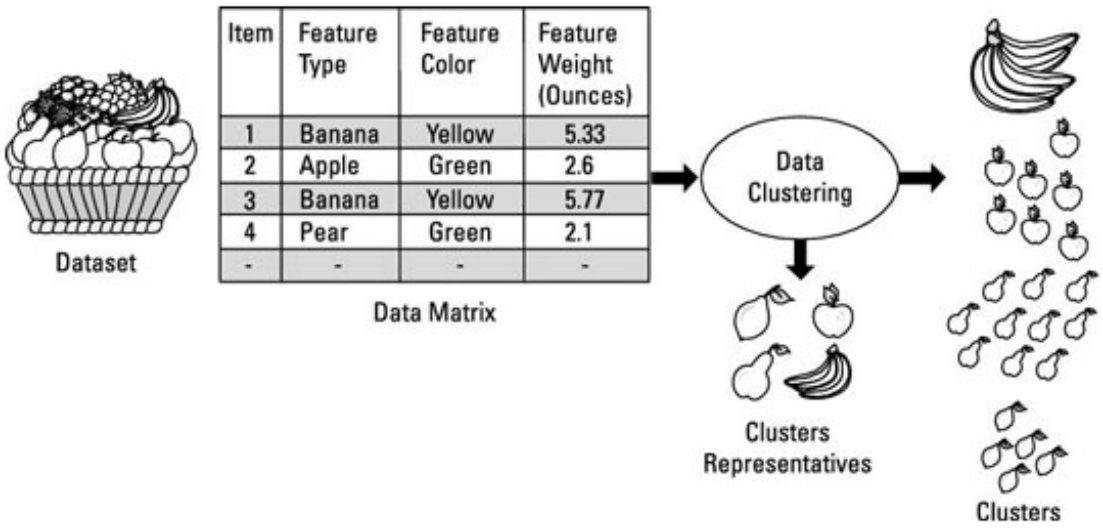


Figure 11. Clustering process example
Source: Retrieved from [38]

The initial step of the data clustering process is to transform the dataset into a matrix, says the author in [38]. An option to model the above-mentioned dataset would be as the rows represent the elements. In this case, fruits and columns represent the characteristics of the elements.

For example, the characteristics of a fruit can be the type of the fruit as well as price, weight, or color. In this sample data set, our item has three characteristics: type, weight, and color.

As it is claimed by the author in [38], in most cases implementing the clustering technique to a dataset as mentioned above allows us to: "Retrieve groups (clusters) of similar items" and "Retrieve cluster representatives of each group." And when we're done with clustering, our dataset will be organized and divided into natural groups.

Another example of Clustering, according to [31], would be if a company is thinking of implementing a targeted marketing using this model of predictive analytics, based on common features, they'll separate customers into clusters and devise strategies for every group at a larger scale.

5.3.1.3. Forecast Model

Forecast Model is one among the foremost used predictive analytics model. This model is mostly used to deal with metric value predictions. It guesses numeric values for new data based on analytics from historical data.

So, this model can be used everywhere where numeric historical data is available, says the author in [31], and gives some examples of in which situations this kind of predictive analytics model can be used.

- *"It can be used in a call center where they can predict how many calls they will receive within an hour or day, and so on."*
- *"It can be used in a shoe store where they can calculate the inventory, they need to keep on hand in order to meet requests throughout a specific sales period."*

According to [39], there are 2 categories of Forecasting: Quantitative Forecasting and Qualitative Forecasting.

Quantitative Forecasting: this includes consuming old data to give accurate predictions and distinguish consonant patterns in sales data. Quantitative methods cover:

- Moving averages
- Regression analysis
- Exponential smoothing
- Adaptive smoothing
- Graphical methods

Qualitative Forecasting: on the other side, “Quantitative Forecasting” is a personal interpretation of data based on the ideas of market leaders and customers, claims the author. Qualitative data is generally used when historical data is insufficient and is best suited for small businesses. Qualitative methods cover:

- Delph method
- Expert opinions
- Market research
- Focus groups
- Historical analogy

5.3.1.4. Outliers Model

The Outlier model catches unusual data entries in a dataset. This model can detect irregular data or records by themselves or in relation with other sections, as it is said in [31].

This model is especially useful in finance and retail predictions. An example would be when identifying fake transactions. Outliers model can estimate the purchase quantity as well as some other information like the history of the purchased item, the location where it was bought, also the nature of the purchase, for example, an $800 purchase for a laptop can not be as misleading as a purchase on a bag of bananas of the same amount of money.

As it is shown in Figure 12., the blue dots are close to each other, while in the other side the two black dots are far from the others at the top right corner. Therefore, these black points are outliers, according to [40].

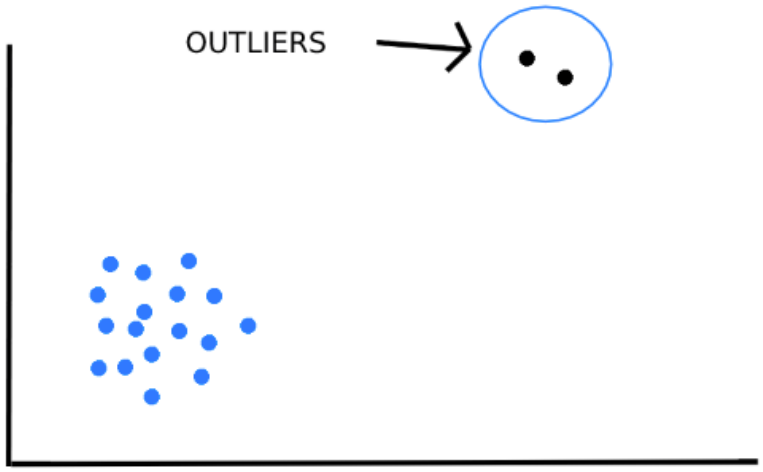


Figure 12. Outliers example
Source: Adapted from [40]

There are three main types of Outliers, claims the author in [40].

1. **Point or global Outliers:** if the data point values are outside the entire dataset, it is considered as a global outlier. For example: student's age in a class will be approximately similar or the same, but if there is a record of a student with an age of 100, it is an outlier.

2. **Contextual or Conditional Outliers:** If the value of a data point is significantly different from the rest of the data points in the same context, it is considered a context outlier. This means if the same value appears in different contexts, it may not be considered an outlier.

3. **Collective Outliers:** if data points altogether deviate significantly from the dataset, it is considered anomalous, but the values of each data point itself are not anomalous in either global or contextual sense.

5.3.1.5. Time Series Model

This predictive analytics model includes an order of data points taken using time as the input parameter. It uses the data from last year to create a metric and predict the next few weeks of data using that metric, claims [31].

Also, the author in [31] talks about the use cases of this model, he says that it can predict the number of the calls made in every day for a few months in a row, how many sales a company had last few months, or the number of patients that went to a specific hospital in the last few weeks. It also takes under consideration seasons or events that would impact the metric.

According to [41], a collection of data points that are assembled uniformly at time intervals and classified in chronological order is called Time Series Data. And Time Series Frequency is the time interval in which the data is collected.

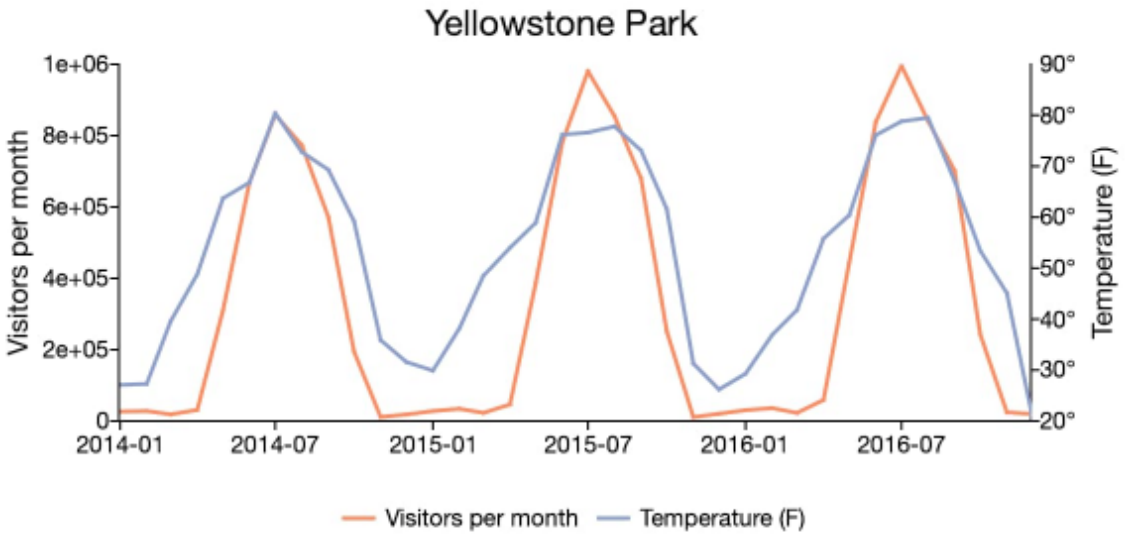


Figure 13. Time Series example graph
Source: Retrieved from [41]

The author gives an example of Time Series in Figure 13. above, it shows the number of visitors per month to "Yellowstone National Park" with the avg. monthly temperatures. The data ranges from the beginning of 2014 to the end of 2016 and is collected every month [41].

This predictive analytics model is an option beyond simple averages, with a higher level of accuracy of understanding how a metric is changing over time, claims the author in [31].

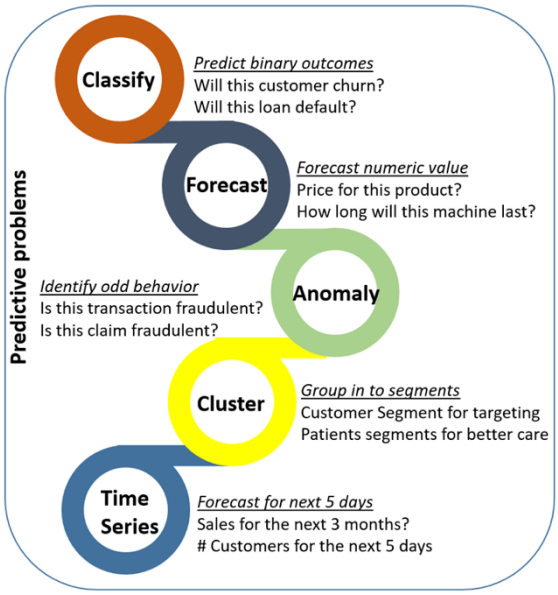


Figure 14. Predictive Problems
Source: Retrieved from [31]

5.3.1.6. Association Rule Discovery

In the last decade, "Association Rule Discovery" has become a central topic in data mining. Due to its wide applicability, it has attracted more attention. "Association Rule Mining" is generally done in the generation of frequent itemsets and rules, and many researchers have proposed several effective algorithms as well [42].

According to [43], "Association Rule Discovery" is typically a data mining method used to explore and interpret large sets of transaction data to identify unique rules and patterns. Those patterns represent interesting connections and cooperation among various elements through transactions [43]. Additionally, association rules discovery is often referred to as "market research," which can be used to analyze the buying habits of customers. Association rules can help identify and predict the behavior of the transaction based on information obtained from examining transactions that use useful properties. So, this method enables us to know some things like, for example, which product the buyer will buy together with another product, representing common groups of goods [43].

Numerous techniques have been established to measure the "interest" of the "Association Rule Discovery." Three common methods are: "***support metric,***" "***confidence metric,***" and "***lift metric.***" [44].

Support: This metric shows us the popularity of a set of items or about frequently purchased items or item combinations so that we can filter out less frequent items [44]. In Figure 15. we can see that the "support of apple" is 50% or 4 out of 8. Also, itemsets can contain more elements for example, the support of "Apple, Beer, and Rice" is 25% [45].

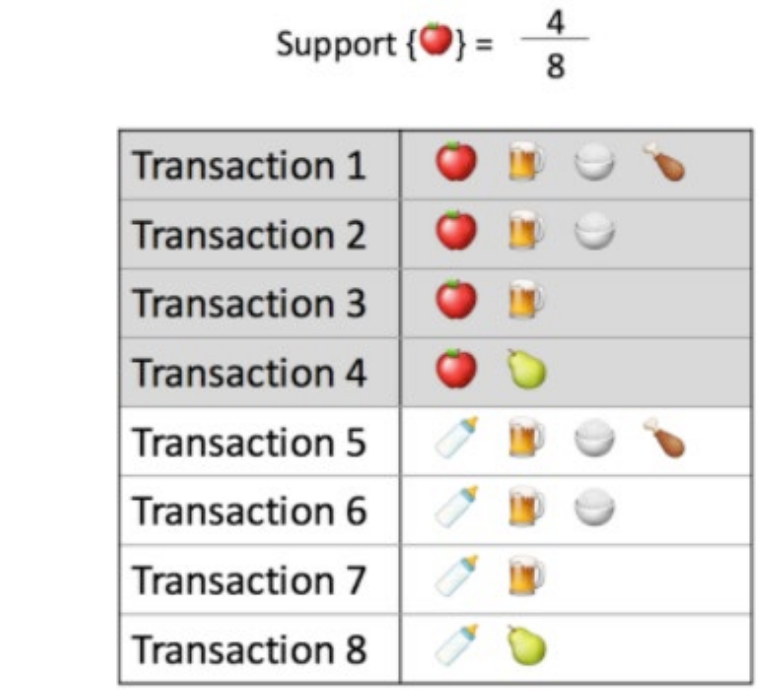


Figure 15. Support Metric
Source: Retrieved from [45]

Confidence: This metric shows how often items "x" and "y" appear together or, with other words, how likely will item "y" be bought when item "x" is bought [44] [45]. In Figure 16. we can see the "confidence of apple -> beer" is 75% or 3 out of 4.

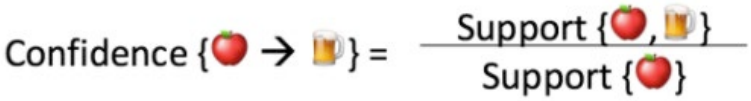


Figure 16. Confidence Metric
Source: Retrieved from [45]

Lift: This metric shows the strength of a rule over the casual event of “x” and “y”, so the strength will be higher if the “Lift” is higher [44]. In Figure 17. we can see that the “lift of apple -> beer” is one which indicates there is no association between these two items.

So, if it’s higher than one, it means that this item “y” is more likely to be purchased, and if It’s less than one, it means that this item is less likely to be purchased alongside the “x” item [45].

$$\text{Lift}\{🍎 \rightarrow 🍺\} = \frac{\text{Support}\{🍎, 🍺\}}{\text{Support}\{🍎\} \times \text{Support}\{🍺\}}$$

Figure 17. Lift Metric
Source: Retrieved from [45]

5.3.2. Top Predictive Algorithms

5.3.2.1. Random Forest

According to [31], the Random Forest Algorithm is the most popular algorithm, also it is capable of both classification and regression. Moreover, it also can deal with and classify accurately large amounts of data. This algorithm is a combination of more decision trees, and in fact, this is the reason why this algorithm is named Random Forest, claims the author in [31].

According to [46], the Random Forest is a learning method that works by building multiple decision trees. And the final decision is chosen by the random forest based on the majority of the trees.

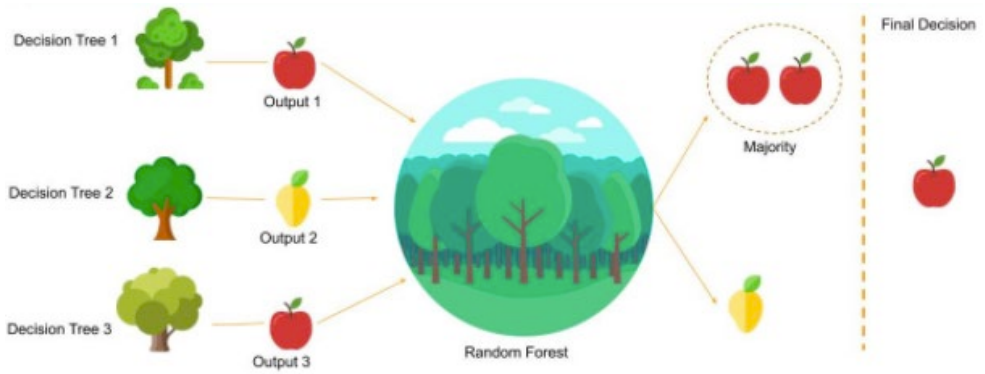


Figure 18. Random Forest example
Source: Retrieved from [46]

The algorithm used by Netflix to recommend movies is an example application of the Random Forest model. Finds people with similar tastes for movies and then recommends movies to them in this way. It uses randomness to remove outliers and prevent incorrect answers from being submitted, claims the author in [47].

Predictive analytics algorithms attempt to reach the lowest error possible by either using "boosting" or "bagging". Boosting is a technique that does the analysis based on the last classification. Bagging, on the other hand, generates subsets of data from examples picked randomly with replacement, so this method is used by Random Forest. An example of a Random Forest is shown in Figure 18.

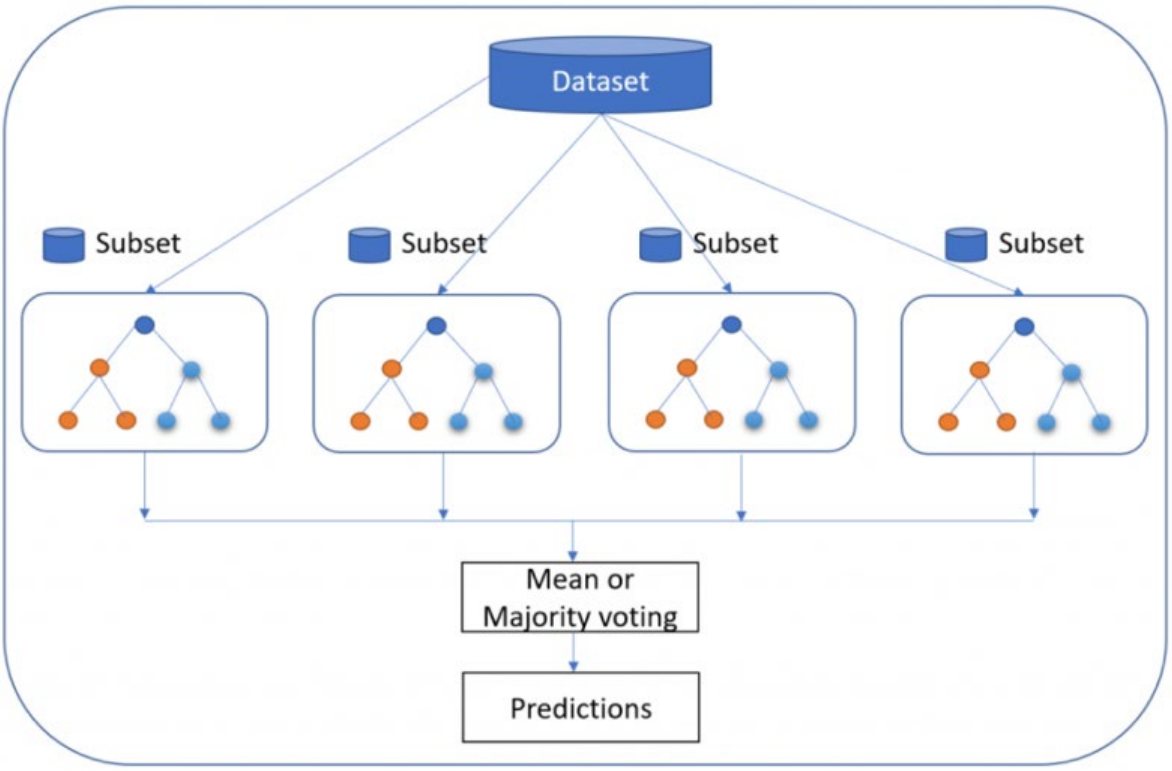


Figure 19. Random Forest Algorithm
Source: Retrieved from [31]

According to the author in [31], these are some advantages that make Random Forest so popular:

- *"It is very efficient as well as accurate when running on large databases."*
- *"Without removing any variable, it can process thousands of input variables."*
- *"Multiple trees lower the inconsistency and bias of a smaller set or single tree."*

- *“For guessing missing data, it offers effective methods.”*
- *“Still keeps its accuracy even when a huge part of data is missing.”*
- *“Guesses which variables are important in classification.”*
- *“Resist to overfitting.”*

Despite these advantages, the Random Forest algorithm also has some disadvantages, claims the author in [48].

- “A lot of computing power and resources are required because it has to build many trees to combine their outputs.”
- “Since many decision trees are combined to determine the class, a lot of training time is required.”
- “Due to the set of decision trees, interpretability is also problematic and does not determine the significance of each variable.”

5.3.2.2. K-Means

According to [31], K-means is a viral and fast algorithm that involves placing unlabeled data points into distinct groups based on similarity.

This algorithm is used in clustering models. The author also gives an example of how this algorithm works: Arben and Ariana are in the first group, and Adrian and Haris are in the second group. This is because the characteristics of Arben and Ariana are so similar, but the characteristics of Ariana and Adrian are different. K-means tries to find out the common characteristics and put them together in the same group.

This is especially useful when we have an extensive data set, and we want to implement a custom plan, which is really hard to do for millions of people.

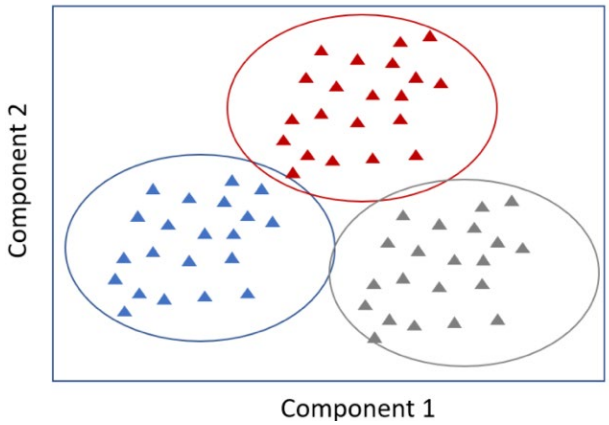


Figure 20. K-means Algorithm
Source: Retrieved from [31]

According to [37], K-Means Clustering is a method of dividing data points into groups or "clusters" categorized by their midpoints or, as it's called, "centroids".

The author gives an example of how K-Means Clustering works:

1. First, we should select the number of clusters (K) that we want to identify. Let's select K=3.

2. Then we randomly generate "K" points on the chart. These will be the midpoints of the initial cluster.

3. Then we measure the distance between centroids and data points and assign data points separately to their closest centroids and matching cluster.

4. Then recalculating centroids of each cluster is needed.

5. Finally, we should repeat steps 3 and 4 to reassign data points to clusters based on the new midpoint locations. We should stop either when:

    a) If the midpoints have been stabilized (if no data points are reassigned after completing the computation of the midpoints of a cluster).

    b) If it has reached the predefined max number of iterators.

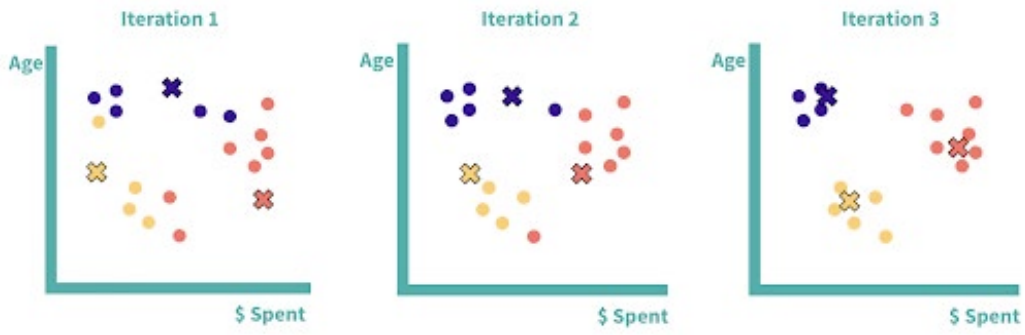


Figure 21. Three iterations of K-means clustering (example)
Source: Retrieved from [37]

5.3.2.3. Generalized Linear Model

The Generalized Linear Model is farther sophisticated variant of the General Linear Model. It requires the last model to compare the effects of multiple variables on continuous variables and then extract the "best fit" model from a set of matrices distributed differently, adds the author in [31].

The author gives an example for this model: Person X wants to know how customers buy winter coats. Regular linear regression can indicate that 300 winter coats were bought for every negative temperature difference. So, if it is lowered from 9 degrees to 3 degrees, it seems like another 2000+ coats can be sold, which seems relevant, but if the temperature is lowered to under twenty degrees, it is less relevant if the amount grows to specifically the same degree.

The generalized linear model will reduce the list of variables, which may indicate that sales will grow when a certain temperature is exceeded, and sales will reduce or stabilize when another temperature is approached.

There are three components to any GLM, according to [49]:

- **Random Component:** represents the probability distribution of the response variable (Y). E.g., the normal distribution of Y in linear regression or the binomial distribution of Y in the binomial logistic regression [49]. It is also known as the error model. How to add random errors to the predictions?

- **Systematic Component:** identifies the descriptive variables such as (X_1, X_2,…, X_k) in the model. More precisely, their linear combination in generating the so-called linear predictor, for example: $\beta_1 + \beta_{1x1} + \beta_{2x2}$ as it is in linear regression [49].

- **Link Function η or $g(\mu)$:** represents the relationship between random and systematic components. It illustrates the relationship between the estimated rate of the response and the linear predictor of the descriptive variable, for example: $\eta = g\left(E(Y_i)\right) = E(Y_i)$ for linear regression, or $\eta = logit(\pi)$ for logistic regression [49].

The main benefit of this algorithm is its quick learning ability. The response can be any form of exponential distribution. The GLM can also handle batch predictions, and they are relatively easy to interpret. In addition, it provides a clear understanding of each of the predictors and is somewhat resistant to overfitting. However, to do so, we need a relatively large data set, adds the author.

5.3.2.4. Gradient Boosted Model (GBM)

The Gradient Boosted Model, before generalization it creates a prediction model consisting of a set of decision trees. As the name suggests, it uses “boosted” machine learning techniques instead of the bagging that was used by Random Forests, according to [31].

The distinctive feature of GBM is that it only builds one tree at a time. Thus, each new tree helps to correct the mistakes made by the previously trained tree, which is different from the “random forest” model, in which trees have no relation. Similar to the Yahoo and Yandex search engines that often use this technique for machine-learning rankings, as it is claimed in [31].

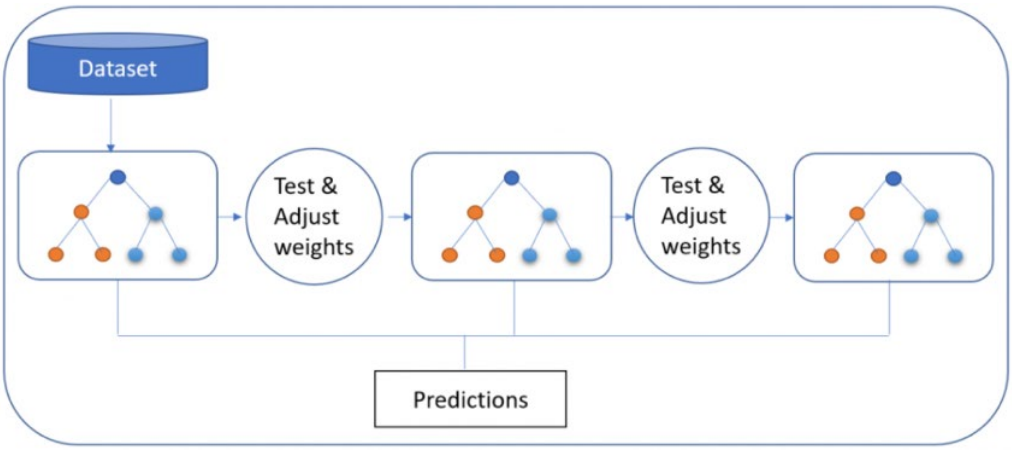


Figure 22. Gradient Boosted Model Implementation
Source: Retrieved from [31]

Gradient boosted model involves three elements, according to [50]:

- **Loss Function (to be optimized):** this function varies on the kind of the issue being solved. As the author says in [50] it should be distinguishable, but numerous standard "loss functions" are supported so we can define our own. E.g., classification can use logarithmic loss, and on the other hand, a regression can use squared error [50].

- **Weak Learner (to make predictions):** in the Gradient Boosted Model, decision trees are used as weak learners. As the author says in [50] precisely, the regression trees are used to generate the actual value of the division, its results are grouped, letting the following model results to be included and fix the residues in the prediction. They are greedily designed, choosing the best split point based on purity score (E.g., Gini) or minimize loss. Extremely short decision trees called "Decision Stump" were used that only had one split. Larger trees can be mostly used with 4-8 levels. "Weak Learners" are often limited in some ways, e.g., the max number of layers, as well as nodes, leaf nodes, or splits, to ensure that learners remain weak but can be built greedily [50].

- **Additive Model (minimize the loss function):** in this model, trees are being included one by one while the current ones are not modified. When adding trees to minimize the loss, a gradient descent procedure is used. "Gradient

Descent" traditionally is used to reduce a set of params. Weights are updated to minimize errors after the calculation. Instead of parameters, there are decision trees. After calculating the loss, it's needed to add trees to the model to perform the gradient descent model and reduce the loss. To do so, it needs to create the parameters of the tree, change the parameters on the tree and move them in the correct direction [50].

5.3.2.5. Prophet

Prophet Algorithm is an algorithm developed by Facebook for internal use by their company. It also is used in the time series and forecast models as well.

This algorithm is very useful for capacity planning, such as allocating resources and establishing sales objectives. However, this process is difficult to automate due to the inconsistent and inflexible level of performance of fully automated forecasting algorithms.

On the other hand, manual forecasting requires the work of an experienced analyst, the author claims in [31].

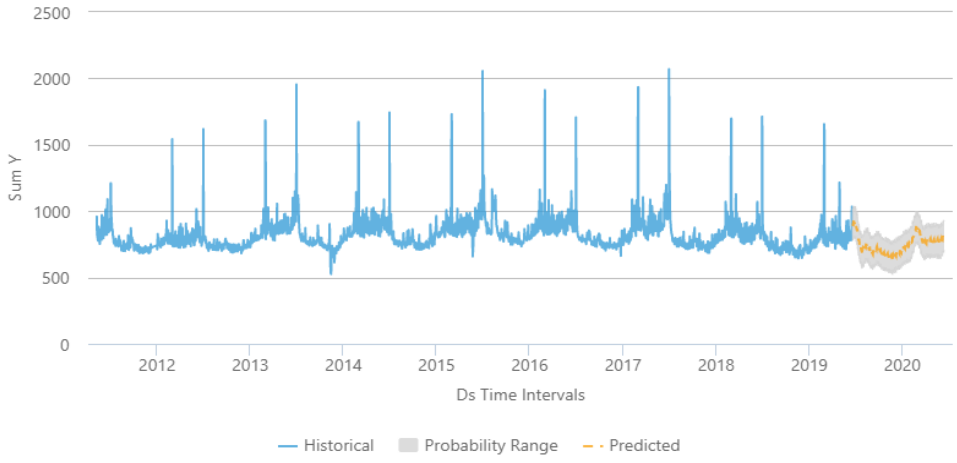


Figure 23. Prophet Algorithm
Source: Retrieved from [31]

The prophet algorithm was flexible enough to consider useful heuristics and assumptions. The speed, stability, and reliability of this algorithm for working with disordered data have made it a successful alternative to time series and forecasting models. Other analysts find it very useful as well.

5.3.2.6. *Apriori Algorithm*

Back in 1994, to find frequent itemsets for the “Association Rule,” “R. Agrawal and R. Srikant” proposed the “Apriori Algorithm.” Why it's called the "Apriori Algorithm”? So, because it uses prior knowledge of the props of frequent groups of items. Then, an iterative method or step-by-step search is applied, where k sets of frequent elements are used to find k + 1 itemsets, says the author in [51].

According to [52] [44], they say one of the most significant Data Mining difficulties nowadays is searching and finding frequent patterns in datasets. Thus, the Apriori Algorithm comes into action using “frequent itemsets” to generate association rules and complete this task with success.

As it is said in [53], there are several ways to perform Association Rule Discovery, and the Apriori Algorithm is the simplest and most direct approach.

The author also says that the main idea of Apriori is: “All non-empty subsets of a frequent itemset must also be frequent.”

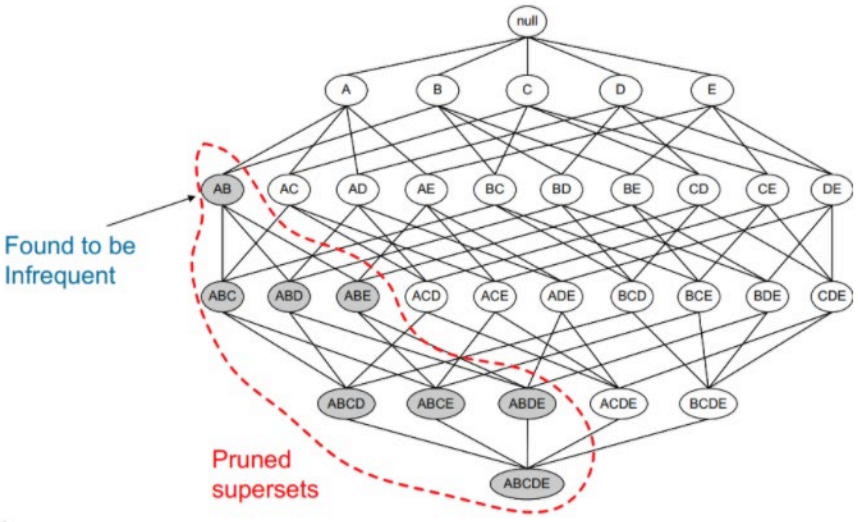


Figure 24. Dataset illustration for applying Apriori Algorithm
Source: Retrieved from [53]

The author explains how the technique used in the dataset above (Figure 24.) for illustration is actually working. They say that it is a bottom-up approach. It starts with each item in the group. The candidates are then generated by the self-join function. The length

of the set of elements extends one by one or, in other words, one step at a time. Subset testing is done at each level and removed or cleaned up item sets that contain infrequent subsets. The process is repeated until we no longer get successful itemsets from the data [53].

The official pseudocode of the Apriori Algorithm:

- L_k: frequent k-itemsets, satisfy the minimum support
- C_k: candidate k-itemsets, possible frequent k-itemsets

```
L1 = { frequent 1 – itemsets};
for ( k = 2; Lk-1 != 0; k++) do begin
    Ck = apriori – gen(Lk-1);
    // scan database
    foreach transaction t ∈D do begin
    Ct = subset(Ck, t)
    foreach candidate c ∈Ct do
    c.count++;
    end
    Lk = {c ∈Ck | c.count >= minsup}
end
Answer = ∪kLk;
```

Official pseudocode of Apriori
Source: Adapted from [53]

And here, we have the author's version of the above official pseudocode in an easier way to understand.

```
L[1] = { frequent 1 – itemsets };
  for (k = 2; L[k – 1] != 0; k++) do begin
      // perform self-joining
      C[k] = getUnion(L[k – 1])
      // remove pruned supersets
      C[k] = pruning(C[k])
      // get itemsets that satisfy minSup
  L[k] = getAboveMinSup(C[k], minSup)
  end
  Answer = Lk(union)
```

Easier way of official Apriori pseudocode
Source: Adapted from [53]

So, finally, if we simulate the algorithm in our head and implement it with the example below given by the author in [53], he claims that it should be very clear.

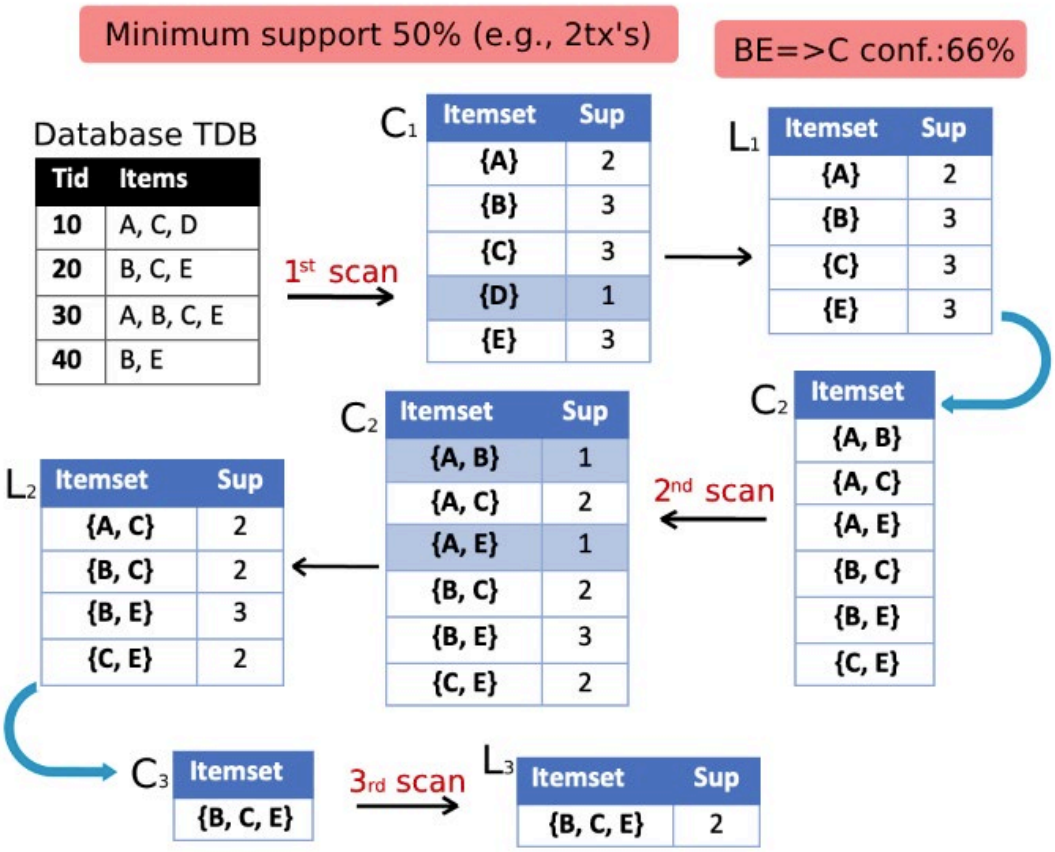


Figure 25. Implementing Apriori Algorithm in a Database table
Source: Adapted from [53]

5.4. Conclusion

To conclude, we will model and design our system to have support for classification, recommendation, and clustering, and the reason for this is because our application will be a data-driven mobile platform that would allow service providers to connect with customers. The data generated from this application will be used to analyze the needs of the market, and data will be modeled for developing predictive methods. Also, the data gathered will be designated to support classification techniques that will be used to improve the system using recommendation techniques.

6. Database Design

According to [54], entity-relationship diagrams (ER Diagrams) are graphical representations of information systems that show the relationships between people, objects, places, concepts, or events in that system also provide a visual starting point for database design.

6.1. The General Database Structure

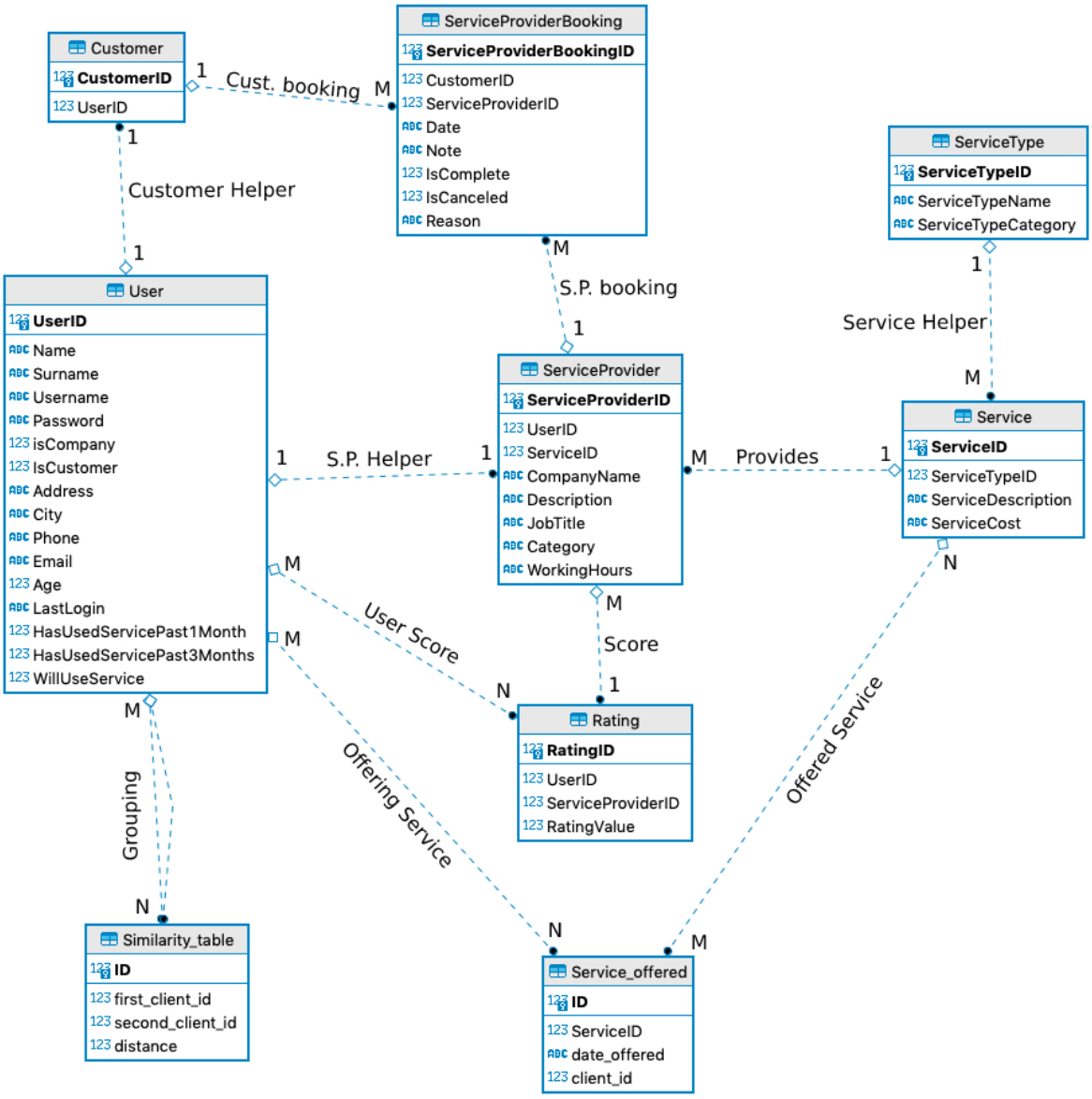


Figure 26. Entity-Relationship Diagram

The E-R Diagram above in Figure 26. with nine tables and relationships between them, shows the general structure of our database.

Provides Relationship: is a M:1 (many to one) relationship where a Service Provider can provide one Service and the same Service can be provided by many Providers.

Service Helper Relationship: is also a M:1 (many to one) relationship that acts as a helper between Service and ServiceType tables. Where a Service can have one Service Type and a Service Type can be related to many Services of the same category.

S.P. booking Relationship: is a 1:M (one to many) relationship where a Service Provider can have many bookings, and a specific booking can be related to only one Service Provider at a time.

Cust. booking Relationship: is also a 1:M (one to many) relationship where a Customer can book many service providers, and a specific booking can only be related to one Customer.

S.P. Helper Relationship: is a 1:1 (one-to-one) relationship that acts as a helper between the User and Service Provider tables. In the User table, we store general user (service provider) data.

Customer Helper Relationship: is a 1:1 (one-to-one) relationship that acts as a helper between User and Customer tables. In the User table, we store general user (Customer) data.

Score Relationship: is a M:1 (many to one) relationship where a Service Provider can have only one score value and the same rating score value can behold by many different Service Providers.

User Score Relationship: is a M:N (many to many) relationship where a user or Customer can give many ratings and a rating value can be given by many different users.

Offering Service Relationship: is a M:N (many to many) relationship where a user can have many offered services by the app, and many services can be offered to many users at the same time.

Offered Service Relationship: is a M:N (many to many) relationship where Service_offered table can have many services and many services can be offered to many clients simultaneously.

Grouping Relationship: is a M:N (many to many) relationship where a user can show up in many similarity tables, and Similarity_table can have many users in one group.

6.2. Database Structure for Recommendations

For recommendation purposes, we will use the Apriori algorithm. As discussed earlier, Apriori needs itemsets which are then filtered to frequent itemsets, and upon them, we can generate rules of the type A=>B.

In the traditional approach, Apriori is used for market basket analysis. In our case, the products offered are services to our clients. We assume that some of these services can be used jointly or even affect one another (causation). E.g., a client might need drywall services, and after that, he can request a painting service.

To support frequent itemsets in our database, we propose the following design:

Services_offered(Id:int, ServiceID:int, date_offered:datetime, client_id:int)

Here, each service that was used by a client will be recorded. As such, we can generate itemsets by just querying this table:

SELECT service_name FROM services_offered GROUP BY client_id

Different database management systems offer support for such queries, notably SQL Server provides the COALESCE keyword, MySQL also offers COALESCE, etc.

Plumbing, drywall, painting, 3

6.3. Database Structure for Clustering Support

When clustering information, it is important to have a method to calculate distances. Therefore, some kind of numerical measure is to be used. An example of clustering would be to group similar customers together. This would allow us to improve recommendations and provide the so-called collective intelligence recommendation.

Similarly, to the Apriori algorithms, with the difference that this approach tries to detect similar users and groups them together. The database design, in this case, can be solved by just querying the previously designed table. The results we want are:

Client	Service_1	Service_2	Service_3	…
1	Yes	No	No	
2	No	Yes	No	
3	Yes	Yes	No	
…				

Having this table, we can generate similarity measures using **Simple Matching Coefficient** or **Jaccard Distance**. The result would be a new distance table that should be updated often and kept in the database for future querying:

	Client_1	Client_2	Client_3	…
Client_1	0	0,3	0,2	
Client_2		0	0,4	
Client_3			0	
…				0

To design this table, we use the following structure:

Similarity_table(id, first_client_id, second_client_id, distance)

As such, querying this table for similar clients would be as simple as:

SELECT TOP(10) first_client_id, second_client_id FROM similarity_table WHERE first_client_id = [SOME NUMBER] ORDER By distance DESC

7. Application Design

Our application is developed using React Native, which is a framework for building cross-platform apps using JavaScript. Although it is a relatively new framework created in 2015 by Facebook, it is a trendy and robust framework.

Some of the pros of React Native are that it provides cross-platform app development, offers faster development, costs less because it's not needed to hire two groups of developers to create separate apps for Android and iOS, etc.

As can be seen down in Figure 27. we have three basic screens for "Login," "Create account," and "Forgot password." When a user wants to create a new account in the app, the user type must be selected. It is for determining whether the user is a "Customer" or a "Company" and based on the user type, other fields appear. If users forget their password, they can reset it by entering their login email and receive a password reset link.

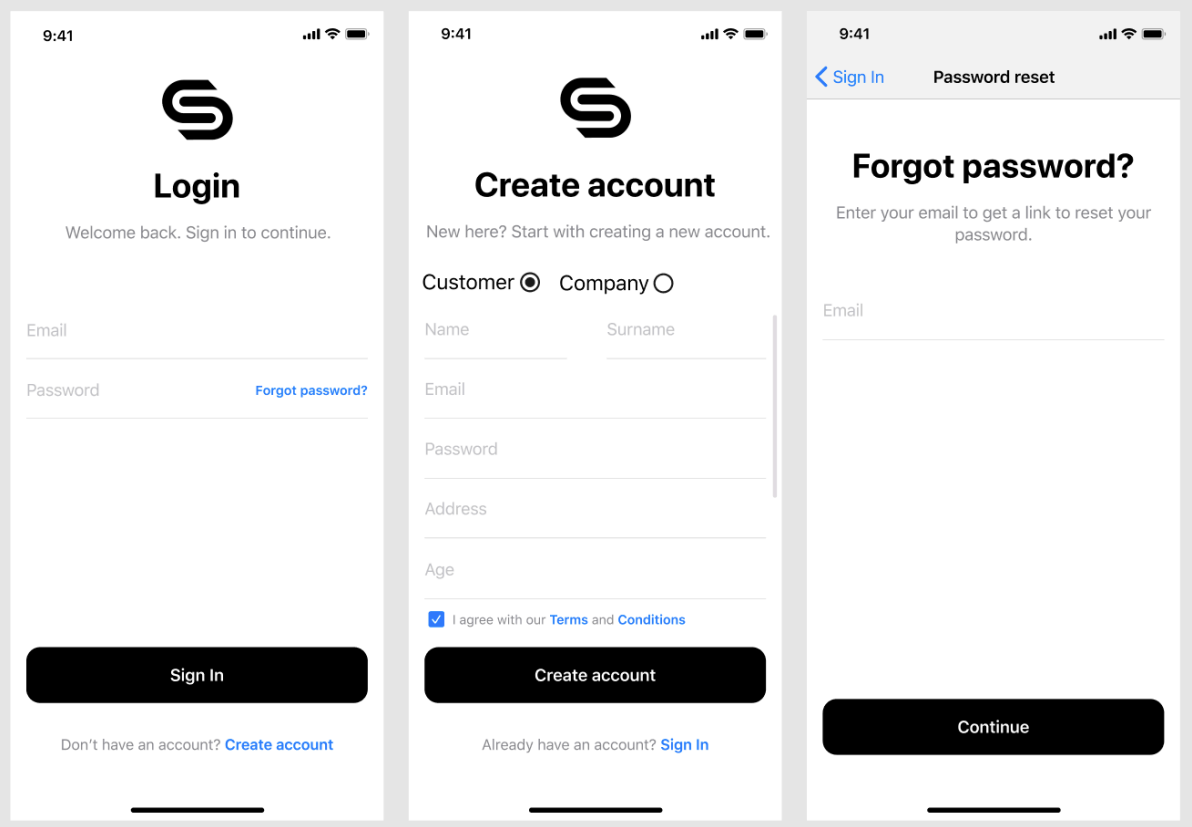


Figure 27. Login, Create account, and Forgot password screens

Following six screens are how the app looks like when a “Company” user enters the app. The first screen that will appear when a user signs in is the “Home screen,” as shown in Figure 28. Here the user will see new job requests that they have as a company. The next screen is the job requests detail screen. Here the company user can see the customer’s message, requested time/date, and the customer’s location. The company user can edit the requested date, accept the request, and decline it.

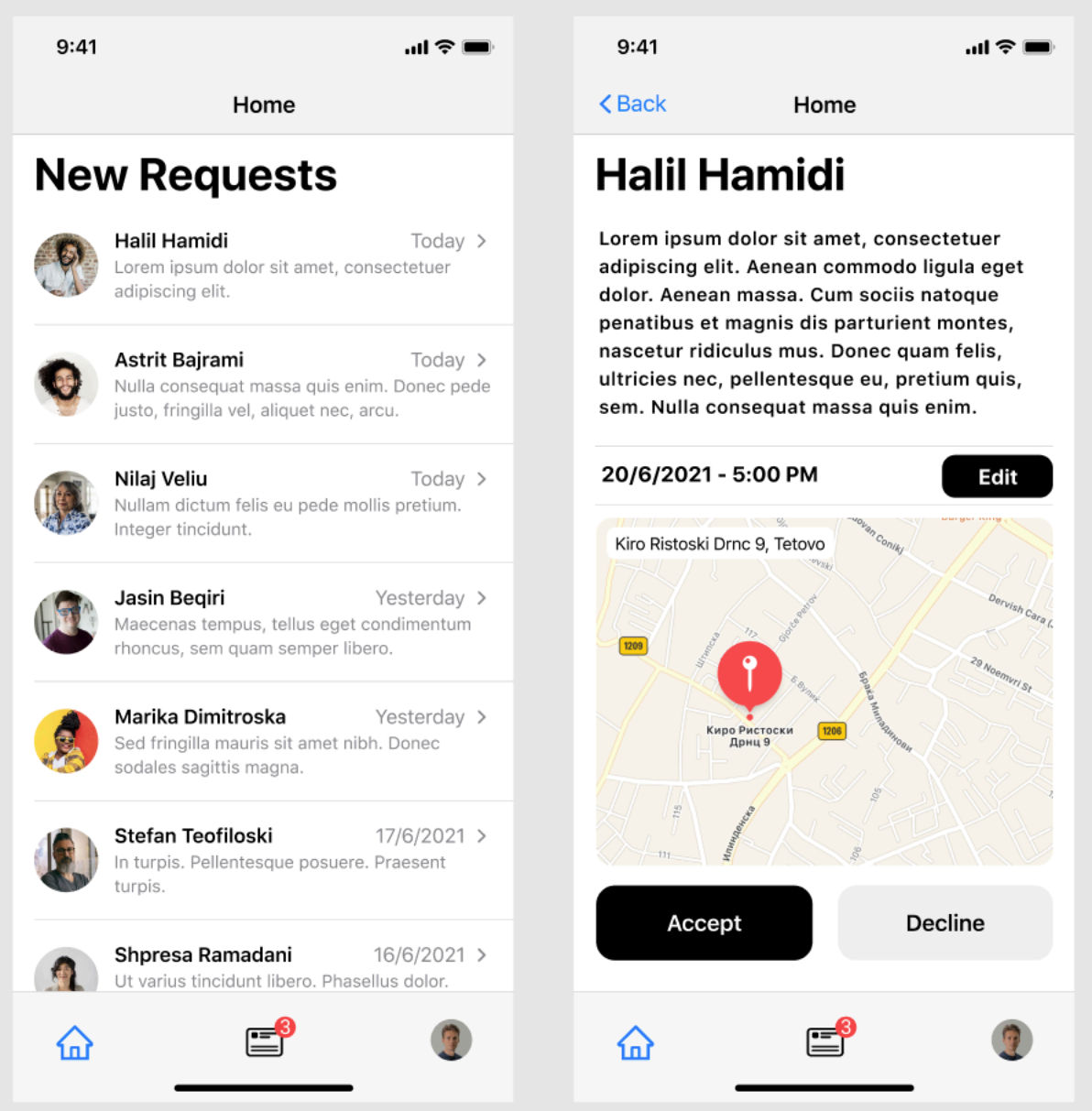


Figure 28. Company home screen (new requests and request details)

In Figure 29. down below, we can see the “Active Jobs” tab. Here are the jobs the company user has accepted. If the job is done, the company enters that row that will lead to a details screen, as shown in the second screen of Figure 29. where they can see customers name, message, time/date, and their location. There also is a “Done” button at the end where the company can end this job after finishing it.

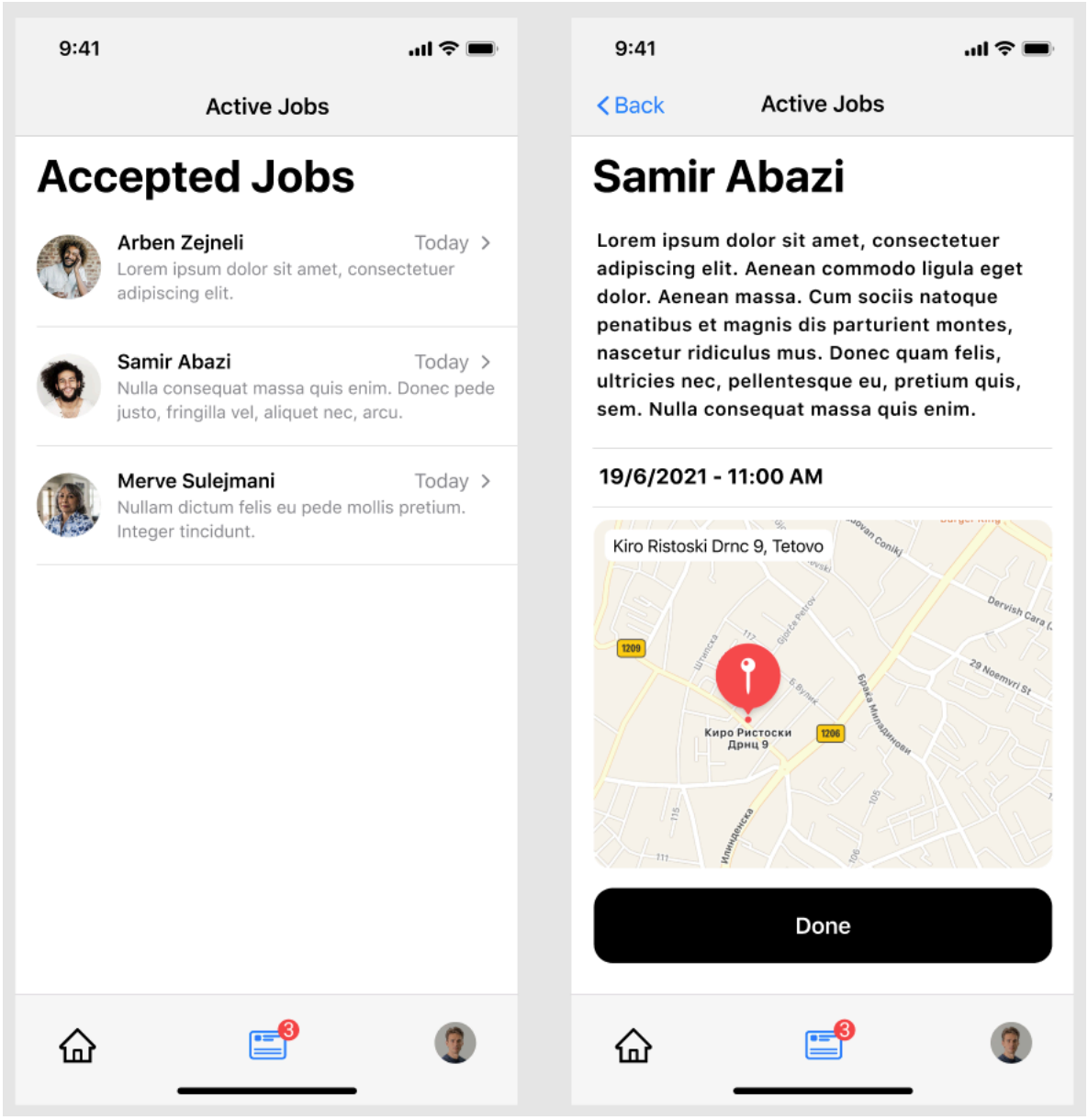


Figure 29. Company active jobs (accepted jobs and job details)

In Figure 30. are two screens of the last tab, “Profile.” Here in the profile tab, users can upload or update their profile image and see some basic information about them like the occupation, city, and country. In addition, the rating number of the company is shown here alongside some statistics like finished jobs, active jobs, profile views, and how many times they have been rated. Here the company user also has some settings like allowing the app to use their location and to send them notifications. Also, here they can read about the terms of service and contact the support team for an issue. And at the bottom of this screen, there are two buttons, “Logout” and “Edit Profile” buttons. Edit profile is a screen where company user can edit their information and/or change their existing password, as we can see in the second screen in Figure 30.

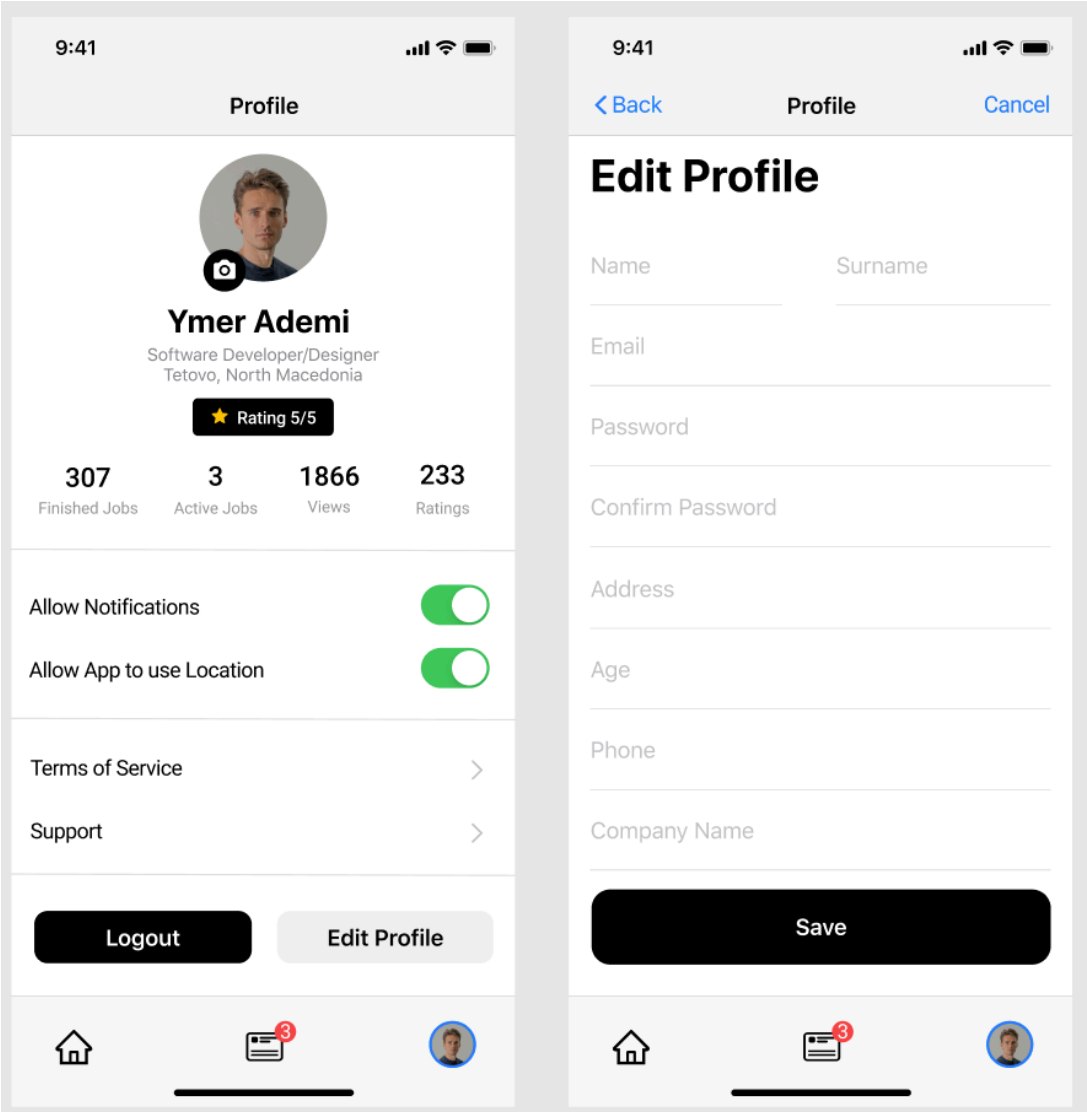


Figure 30. Company profile (profile and edit profile)

Following screens are how the app looks like when a “Customer” user enters the app. The first screen that will appear when a customer user logs in is the “Home screen,” as shown in Figure 31. Here the user can see “Active” appointments or bookings that they have made. Below that, they have the “Suggestions” section. These are suggestions made automatically by the app using recommendation techniques and methods developed for prediction.

In the next screen, customers can see their active or ongoing job request in detail and the “Finish” button to finish the actual ongoing activity. The third screen is to rate the company or worker. This screen appears when customers tap on the “Finish” button, which means the job is completed.

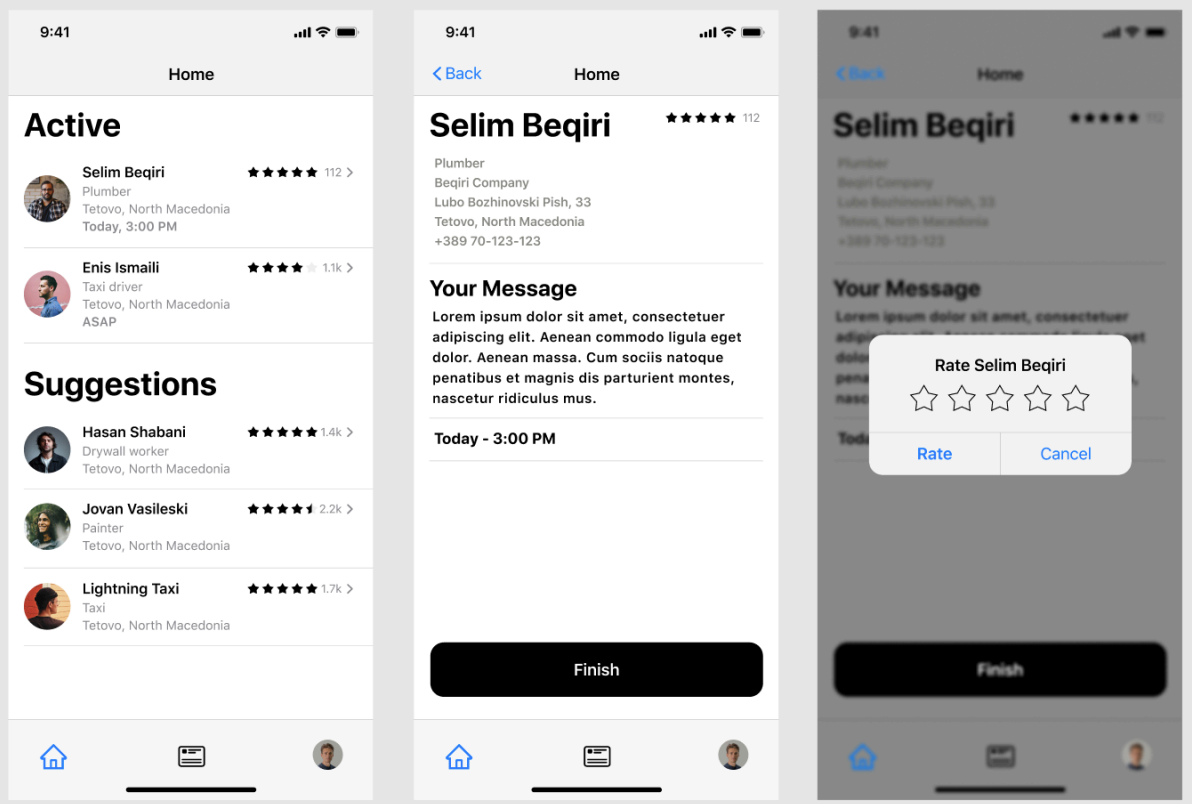


Figure 31. Customer home screen (active bookings and suggestions, job/worker details, and rating)

In the second tab, “Customer users” can find the “Services” screen where all types of services are listed, as shown in Figure 32. The second screen shows when a user enters a service category. In this screen, the customers can see a list of service providers with some information like their occupation, city, and rating.

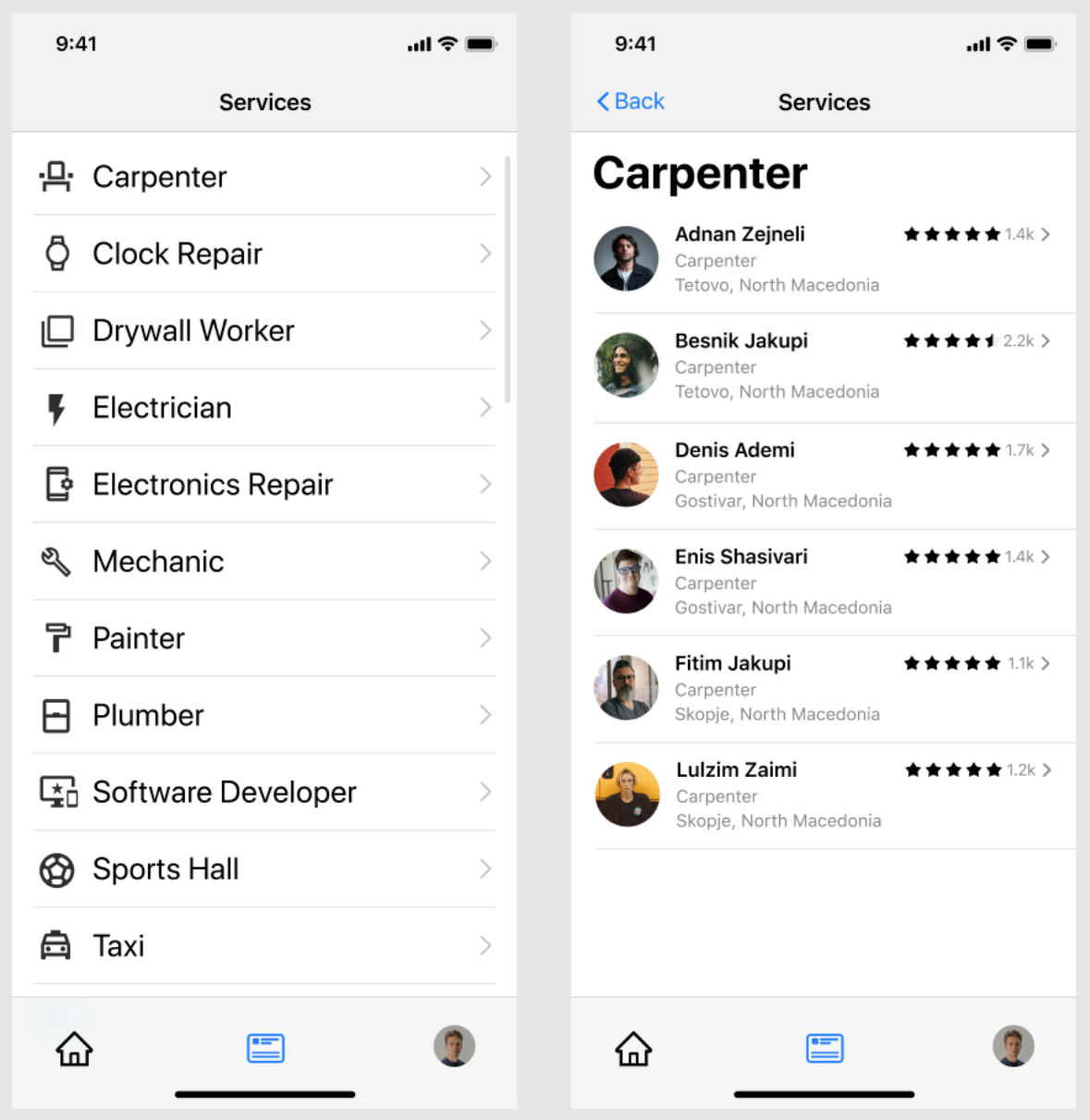


Figure 32. Customer services screen (services list and service providers list)

The following two screens shown in Figure 33. are about making an appointment or booking a service provider. The first screen shows up when a customer taps a service provider row in the service providers list. On this screen, customers can show details about the service provider. For example, their rating, company name, phone number, city, their work time, and the address. At the end of the screen, there is a button called “Proceed,” after tapping that customer goes to the second screen where they can write about their problem or need, select the time that suits them better, and choose their address. Finally, after tapping the “Finish” button, the request goes to the service provider.

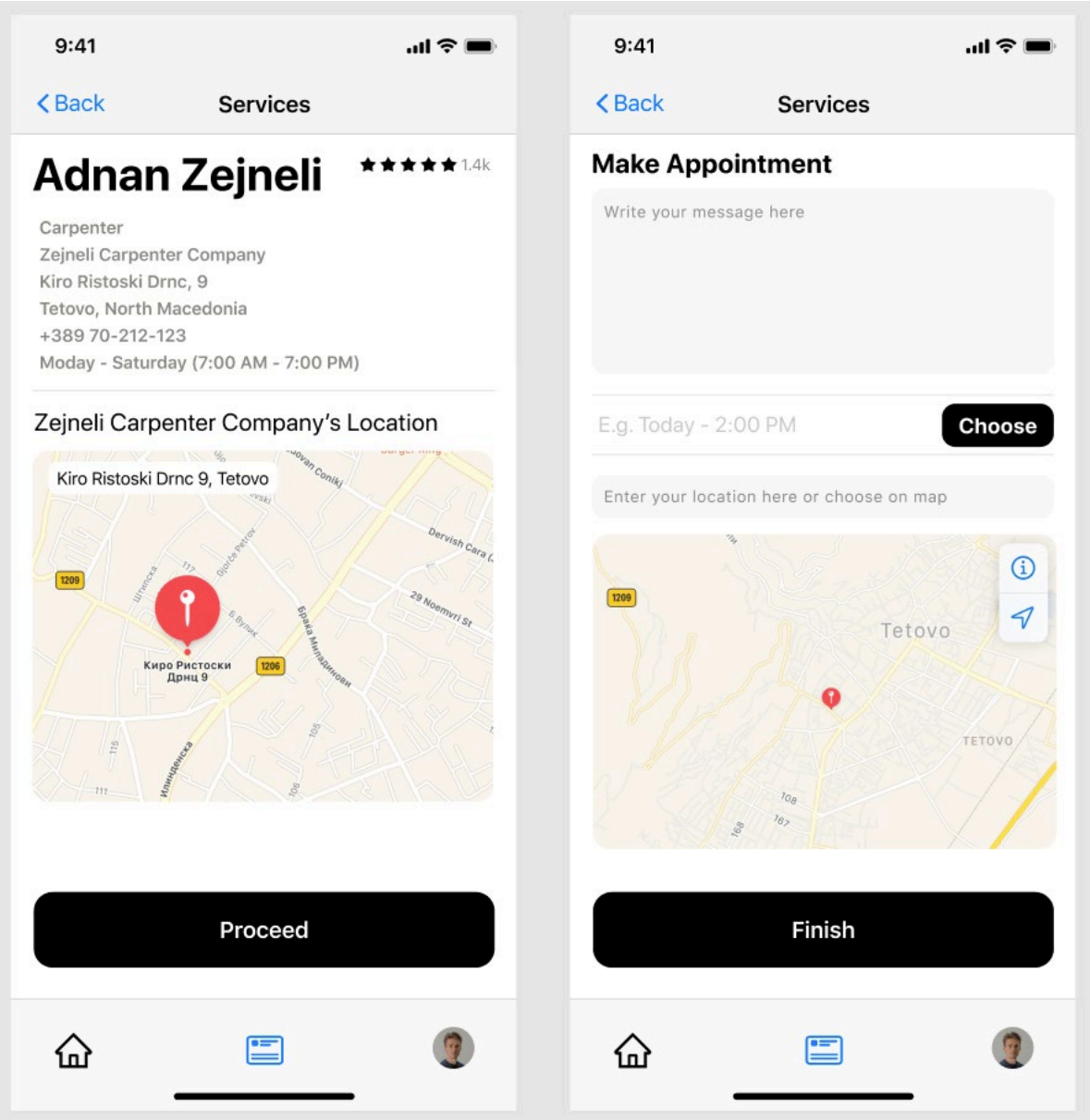


Figure 33. Customer services screen (service provider details and making an appointment)

Last but not least, in Figure 34. are shown two screens of the customer's profile. In the "Profile" tab, customers can upload or update their profile picture. Also, there are some statistics exhibited like all-time bookings of the customer, active bookings, and how many times they have rated others. Also, here the customer has some settings like allowing the app to send them notifications and use their location. They also can find the terms of service and support pages in this tab. And there are two buttons, "Logout" and "Edit Profile," at the bottom of the screen, same as for "Company" user. Edit profile is a screen where customers can edit their information and change their existing password.

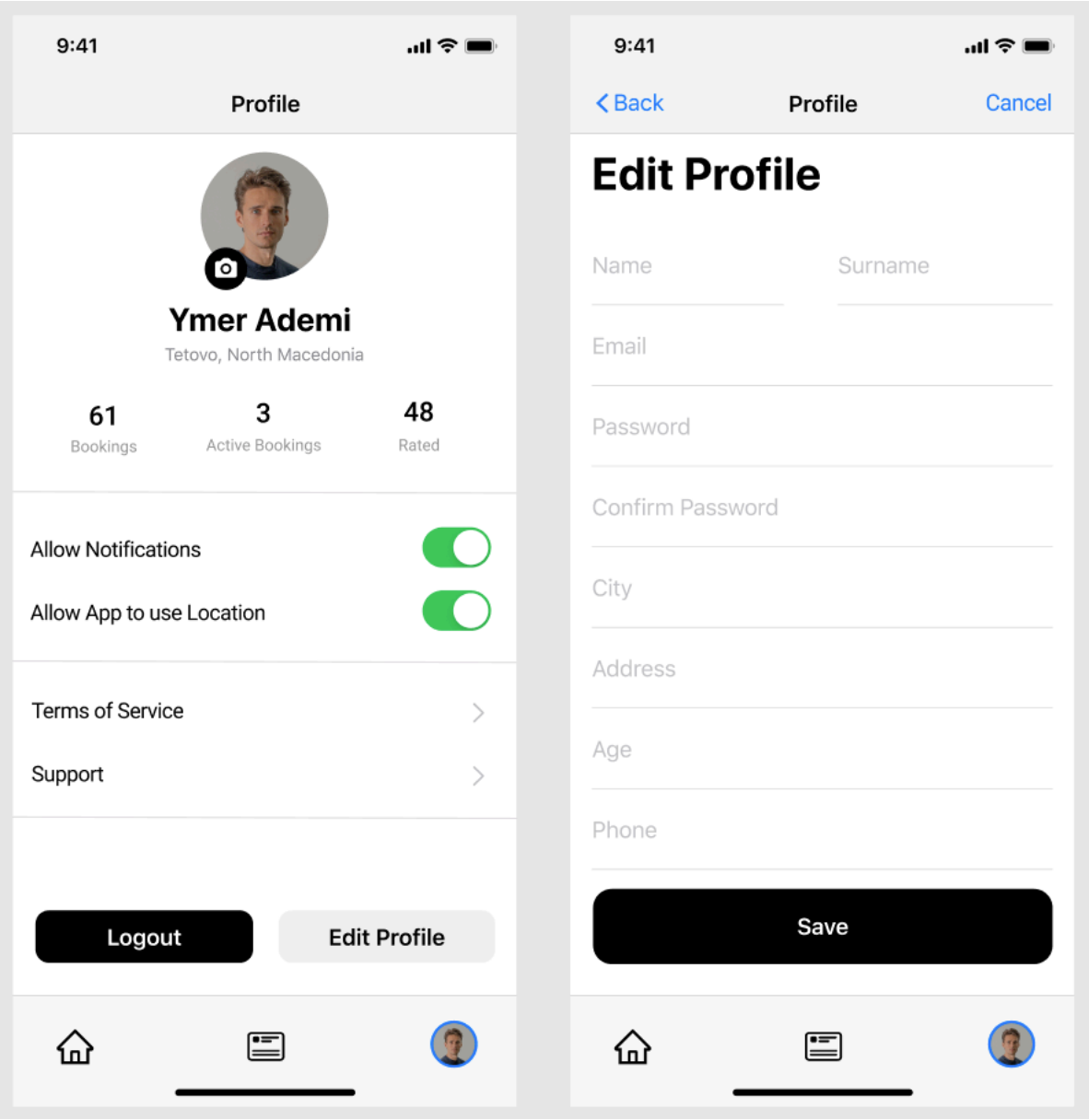


Figure 34. Customer profile screen (profile and edit profile)

8. Conclusion and Future Work

The main idea was to build an app that will act as a hub for both service providers and customers in North Macedonia. This app will allow service providers to provide numerous services to clients and help them grow their business and have easy access to new customers. On the other side, customers, especially new generations that have trouble finding service providers or even if they are living in another city, can find the right service provider for their problem with not much effort just through one app.

So, this app will be a huge advantage for the students as well. If we think of ones that live in another city during the studies, they probably will face difficulties finding service providers and will lose too much time searching through web. Our app will be a perfect solution for them and even the locals that need this kind of help.

Since our app is a data-driven app, we modeled and designed our system to have support for classification and recommendation. The data gathered from the app will be used to analyze the market's needs and will be modeled for developing predictive methods.

The app is developed using React Native which is an open-source JavaScript framework and the reason we went this way is because it provides cross-platform app development, offers faster development, costs less, etc.

For further advancing our app, we will designate the gathered data to support classification techniques that will be used to improve the system using recommendation techniques.

Bibliography

[1] R. A.-S. V. A.-F. S. M. Rocío Aguilar-Illescas, "Mobile Fashion C2C Apps: Examining the Antecedents of Customer Satisfaction.," IGI Global, Malaga, 2020.

[2] B. Sanou, "Measuring the Information Society Report," ITU Publication Production Service, Geneva, 2015.

[3] L. Miller, "Entrepreneur | 5 Data-Driven Reasons You Should Build a Mobile App for Your Business," 25 April 2019. [Online]. Available: https://www.entrepreneur.com/article/332517. [Accessed 20 September 2020].

[4] G. Z. Z. S. J. H. H. W. G. T. Min Peng, "Personalized app recommendation based on app permissions," *Special Issue on Security and Privacy of IoT,* vol. I, no. 21, p. 89–104, January 2018.

[5] R. &. A.-S. R. &. Á.-F. V. &. M. S. Aguilar-Illescas, "Mobile Fashion C2C Apps: Examining the Antecedents of Customer Satisfaction," IGI Global, Malaga, 2020.

[6] S. O'Dea, "Smartphones - Statistics & Facts | Statista," Statista, 26 October 2020. [Online]. Available: https://www.statista.com/topics/840/smartphones/. [Accessed November 2020].

[7] M. Sheetz, "CNBC," 29 January 2021. [Online]. Available: https://www.cnbc.com/2021/01/28/spacex-plans-next-generation-starlink-satellites-with-1000-launched.html. [Accessed March 2021].

[8] M. Handley, *Starlink revisions, Nov 2018 (https://www.youtube.com/watch?v=QEIUdMiColU),* London: YouTube, 2018.

[9] C. Janssen, "Mobile Application (Mobile App)," Techopedia, 7 August 2020. [Online]. Available: https://www.techopedia.com/definition/2953/mobile-application-mobile-app. [Accessed February 2021].

[10] N. Sakovich, "Cross-Platform Mobile Development: Five Best Frameworks | Sam Solutions," Sam Solutions, 25 October 2019. [Online]. Available: https://www.sam-solutions.com/blog/cross-platform-mobile-development/. [Accessed November 2020].

[11] J. Bender, "Cross-Platform Mobile Applications with React Native | Code Magazine," Code Magazine, 1 February 2017. [Online]. Available: https://www.codemag.com/article/1701071/Cross-Platform-Mobile-Applications-with-React-Native. [Accessed November 2020].

[12] S. Khoroshulia, "How React Native App Development Works Under The Hood," MobiDev, June 2020. [Online]. Available: https://mobidev.biz/blog/how-react-native-app-development-works#:~:text=React%20Native%20compiles%20the%20JavaScript,developers%20can%20create%20new%20ones.. [Accessed November 2020].

[13] I. Sarker, "Mobile Data Science: Towards Understanding Data-Driven Intelligent Mobile Applications.," *EAI Endorsed Transactions on Scalable Information Systems, European Alliance for Innovation,* vol. 5, no. 19, pp. 1-11, 12 2018.

[14] L. Cao, "Data science: a comprehensive overview," *ACM Computing Surveys,* vol. 50, no. 3, p. 43, June 2017.

[15] O. Wirén, "Data Driven Development for Mobile Applications," Uppsala Universitet, Uppsala, 2013.

[16] N. G. P. J. T. L. C. P. P. Lisandro Delia, "Multi-platform mobile application development analysis," in *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*, La Plata, Argentina, 2019.

[17] exlipse.org, "Xtext," eclipse, 2 June 2020. [Online]. Available: http://www.eclipse.org/Xtext/. [Accessed February 2021].

[18] H. B. M. F. G. L. Peter Friese, "Applause DSL," Applause, 22 June 2011. [Online]. Available: https://github.com/applause/applause. [Accessed February 2021].

[19] N. G. P. T. P. P. Lisandro Delía, "Un Análisis Experimental de Tipo de Aplicaciones para Dispositivos Móviles," in *XVIII Congreso Argentino de Ciencias de la Computación, CACIC*, La Plata. Argentina, 2013.

[20] S. V. Chaitanya Kaul, " Review Paper on Cross-Platform Mobile Application Development IDE," *IOSR Journal of Computer Engineering (IOSR-JCE),* vol. XVII, no. 1, pp. 30-33, January-February 2015.

[21] C. E. C. Anirudh Nagesh, "Cross-Platform Mobile Application Development," in *ResearchGate*, Syracuse, New York, 2014.

[22] S. X. Spyros Xanthopoulos, "A Comparative Analysis of Cross-platform Development Approaches for Mobile Applications," in *Research Gate*, Thessaloniki, 2013.

[23] M. Dixon, "Selerity," 17 November 2020. [Online]. Available: https://seleritysas.com/blog/2020/11/17/why-use-predictive-analysis-models-for-better-decision-making/. [Accessed March 2021].

[24] M. Dixon, "Selerity," 12 December 2019. [Online]. Available: https://seleritysas.com/blog/2019/12/12/types-of-predictive-analytics-models-and-how-they-work/. [Accessed March 2021].

[25] S. J. H. R. N. E. P. D. P. F. A. T. S. M. W. C. V. Apte, "Data-intensive analytics for predictive modeling," *IBM Journal of Research and Development,* vol. 47, no. 1, pp. 17-23, 27 May 2014.

[26] M. L. G. Vaibhav Kumar, "Predictive Analytics: A Review of Trends and Techniques," *International Journal of Computer Applications,* vol. 182, no. 1, pp. 31-32, 23 January 2019.

[27] P. T. Lakshay Swani, "Predictive Modelling Analytics through Data Mining," *International Research Journal of Engineering and Technology (IRJET),* vol. IV, no. 9, pp. 5-10, September 2017.

[28] A. D. N. S. G. S. Irad Ben-Gal, "Efficient Construction of Decision Trees by the Dual Information Distance Method," *Quality Technology & Quantitative Management (QTQM),* vol. XI, no. 1, pp. 133-147, 9 February 2016.

[29] Y. S. E. W. U. W. Irad Ben-Gal, "Peer-to-peer information retrieval using shared-content clustering. Knowl Inf Syst 39, 383–408 (2014). https://doi.org/10.1007/s10115-013-0619-9," Springer, Tel-Aviv, 2014.

[30] R. König, "Predictive Techniques and Methods for Decision Support in Situations with Poor Data Quality," Digitala Vetenskapliga Arkivet, Borås, Available: 2009 / Updated: 2018.

[31] S. Parthasarathy, "Logi Analytics," 12 February 2021. [Online]. Available: https://www.logianalytics.com/predictive-analytics/predictive-algorithms-and-models/. [Accessed March 2021].

[32] R. Wolff, "Classification Algorithms in Machine Learning: How They Work," 26 August 2020. [Online]. Available: https://monkeylearn.com/blog/classification-algorithms/. [Accessed April 2021].

[33] S. Asiri, "Machine Learning Classifiers," 11 June 2018. [Online]. Available: https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623. [Accessed April 2021].

[34] Simplilearn, "Everything You Need to Know About Classification in Machine Learning," 5 March 2021. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/classification-in-machine-learning. [Accessed April 2021].

[35] C. Model, "Analytics Vidhya," 3 November 2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/. [Accessed March 2021].

[36] M. D. F. M. M. I. T. Chiara Di Francescomarino, "Clustering-Based Predictive Process Monitoring," *IEEE Transactions on Services Computing,* vol. XII, no. 6, pp. 896-909, 1 December 2019.

[37] K. Gross, "Clustering: How It Works (In Plain English!)," 18 June 2020. [Online]. Available: https://blog.dataiku.com/clustering-how-it-works-in-plain-english. [Accessed April 2021].

[38] M. C. T. J. Anasse Bari, "Basics of Data Clusters in Predictive Analysis," [Online]. Available: https://www.dummies.com/programming/big-data/data-science/basics-of-data-clusters-in-predictive-analysis/. [Accessed April 2021].

[39] J. Hyun, "Forecasting Methods for Any Industry- Everything You Need to Know," 27 July 2020. [Online]. Available: https://zipforecasting.com/forecasting-methods.html. [Accessed April 2021].

[40] M. Tripathi, "Knowing all about Outliers in Machine Learning," 16 June 2020. [Online]. Available: https://datascience.foundation/sciencewhitepaper/knowing-all-about-outliers-in-machine-learning. [Accessed April 2021].

[41] Erica, "Introduction to the Fundamentals of Time Series Data and Analysis," 13 September 2019. [Online]. Available: https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/. [Accessed April 2021].

[42] N. R. T. Karthikeyan, "A Survey on Association Rule Mining," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. III, no. 1, pp. 5223-5227, 2014.

[43] A. Subasi, "Machine learning techniques," in *Practical Machine Learning for Data Analysis Using Python*, Jeddah, ScienceDirect, 2020.

[44] S. Maitra, "Association Rule Mining using Market Basket Analysis - Towards Data Science," 15 October 2019. [Online]. Available: https://towardsdatascience.com/market-basket-analysis-knowledge-discovery-in-database-simplistic-approach-dc41659e1558. [Accessed April 2021].

[45] A. Ng, "KDnuggets," 14 April 2016. [Online]. Available: https://www.kdnuggets.com/2016/04/association-rules-apriori-algorithm-tutorial.html. [Accessed April 2021].

[46] Simplilearn, "Random Forest Algorithm," 24 March 2021. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm?source=sl_frs_nav_playlist_video_clicked. [Accessed April 2021].

[47] K. Fuchs, "Machine Learning: Classification Models," 28 March 2017. [Online]. Available: https://medium.com/fuzz/machine-learning-classification-models-3040f71e2529. [Accessed April 2021].

[48] Great Learning Team, "Random Forest Algorithm- An Overview," 19 February 2020. [Online]. Available: https://www.mygreatlearning.com/blog/random-forest-algorithm/. [Accessed April 2021].

[49] Pennsylvania State University, *PennState Eberly College of Science,* Pennsylvania: Pennsylvania State University.

[50] J. Brownlee, "A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning," 9 September 2016. [Online]. Available: https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/. [Accessed April 2021].

[51] Mohityadav, "Apriori Algorithm," 4 April 2020. [Online]. Available: https://www.geeksforgeeks.org/apriori-algorithm/. [Accessed April 2021].

[52] L. Z. Q. H. Z. S. Ning Li, "Parallel Implementation of Apriori Algorithm Based on MapReduce," in *13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, Kyoto, 2012.

[53] T. Chou, "Apriori: Association Rule Mining In-depth Explanation and Python Implementation," 26 October 2020. [Online]. Available: https://towardsdatascience.com/apriori-association-rule-mining-explanation-and-python-implementation-290b42afdfc6. [Accessed April 2021].

[54] J. Biscobing, "Entity Relationship Diagram (ERD)," September 2019. [Online]. Available: https://searchdatamanagement.techtarget.com/definition/entity-relationship-diagram-ERD. [Accessed 29 April 2021].

[55] S. O'Dea, "Number of smartphone users from 2016 to 2021 | Statista," Statista, 20 August 2020. [Online]. Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. [Accessed November 2020].

[56] J. Clement, "Number of mobile app downloads worldwide from 2016 to 2019 | Statista," Statista, 17 January 2020. [Online]. Available: https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/. [Accessed November 2020].

[57] E. Dezhic, "Becominghuman," 26 June 2017. [Online]. Available: https://becominghuman.ai/understanding-decision-trees-43032111380f. [Accessed March 2021].