UNIVERSITETI I EJL
ЈИЕ УНИВЕРЗИТЕТ
SEE UNIVERSITY

South East European University **Faculty of Contemporary Sciences and Technologies**

Third Cycle of Academic Studies – Doctoral Studies

Doctoral Dissertation Topic:

"Aspect-based Sentiment Analysis of Albanian Texts"

Candidate

Majlinda Axhiu

Mentor

Assoc.Prof. Azir Aliu

______, 2021

Acknowledgments

Abstract

Currently, we are not only static observers and recipients of information, but successively, we dynamically change the information content and produce new pieces of information. Since many sentiments that express the authors' attitudes, opinions or emotions on a topic are generated, sentiment analysis has a wide range of real-world applications.

The purpose of sentiment analysis is to get the opinion from a given document, paragraph or sentence and to predict the overall sentiment polarity, which may be positive, negative or neutral. In this process, it is assumed that a given piece of text has only one aspect-term and one sentiment polarity. A more detailed and thorough task would be to find all the aspects mentioned in a text and then to predict the sentiments related to each of them. This comprehensive task is called aspect-based sentiment analysis (ABSA).

In the first stages of ABSA the tasks were mostly computed by supervised learning approaches. However, building a labeled training set and testing set for different languages and domains is very costly and time consuming because it can only be achieved by manual labor. Because of this we chose the semi-supervised approach aiming to adapt a system which would be flexible in language and domain, and would be based on novel machine learning techniques. Through this system we will have the opportunity to make use of Albanian unstructured data from the digital world.

The model we are using within the system for implementing the ABSA tasks relies on a topic-modeling approach, namely on Non-Negative Matrix Factorization (NMF) technique and Maximum Entropy Classifier. The inputs needed for the model include an unlabeled corpus, list of stop words and two short lists of seed words for aspect terms and for sentiments' polarities,

while the outputs are: ranked list of aspect-terms, a classification of aspect-terms in aspect categories (in our case in 3 aspects) and sentiment classification polarity document. For both tasks of ABSA, Aspect term extraction and Sentiment polarity classification, we have obtained an acceptable precision for the textual data in Albanian language.

Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

ABSA	Aspect-based Sentiment Analysis
ADT	Aspect detection threshold
ATE	Aspect Term Extraction
BCD	Block Coordinate Descent
BERT	Bidirectional Encoder Representations from Transformers
CBOW	Continuous Bag of Words
CNN	Convolutional Neural Network
FURIA	Fuzzy Unordered Rule Induction Algorithm
HALS	Hierarchical Alternating Least Squares
LDA	Latent Dirichlet Allocation
ME	Maximum Entropy
ML	Machine Learning
NLP	Natural Language Processing
NMF	Non-negative Matrix Factorization
NMF4ABSA	Non-negative Matrix Factorization for Aspect based Sentiment Analysis

NN(s)	Neural Network(s)
PLSA	Probabilistic Latent Semantic Analysis
POS	Part of Speech
SA	Sentiment Analysis
SE	Sentiment Extraction
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
Word2Vec	Word to Vector
W2VLDA	Word to Vector-Latent Dirichlet Allocation

PART I

Introduction

People regularly and publicly express their opinions on the Internet on a daily basis. Their opinions correspond to different social, individual, organizational issues and topics. Since the usage of e-commerce is exponentially increasing, the amount of online reviews also increases. Proceeding as a chain, now almost every "online" buyer first checks for the others' opinions and reviews.

As (Wright, 2009) claims, for many businesses, online opinions have turned out to be a kind of virtual currency that can make or break a product in the marketplace. So, producers and servitors are willing to check the acceptance trend of their services and products, in order to analyze their strong and weak points and try to improve.

When looking into the social media, opinions can be found on many different locations starting from fora and blogs trough to social networks. However, it is very difficult for a human being to find proper resources, extract the needed data, analyze, interpret and visualize them in understandable and useful form. So, analyzing all the reviews and trying to get proper information is next to unmanageable. That is why automated sentiment analysis emerged as a need, with its application growing wider and wider.

Sentiment analysis, also referred as "opinion mining", is a set of methods, usually implemented in computer software, that perceive, measure, report and utilize attitudes, opinions, and emotions, which are generally called sentiments and can be found in online, social, and enterprise information sources (Liu, 2012).

Sentiment analysis is a form of texts' semantic analysis, and its goal is to extract opinions, attitudes or emotions for different objects of interest. For example, producers or businesses might be interested in consumers' opinion about their products and services, political parties might want to know the voters' attitudes, or educational institutions might be interested in course and lecturer evaluation. The access to huge amount of different textual data has become very easy due to news, blogs, tweets, Facebook comments and reviews, etc.

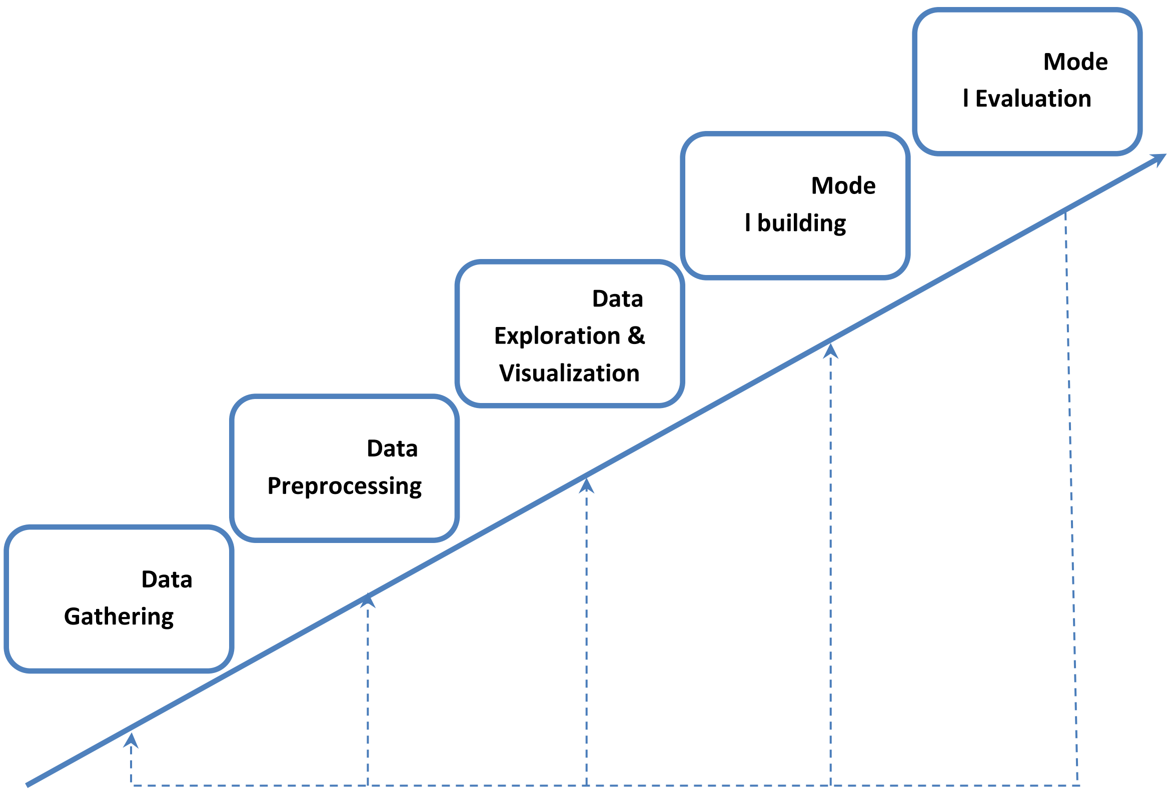

Figure 1. The consecutive phases of Text Mining

The text mining processes which are shown in the above figure are Data collection, Data preprocessing, data exploration and visualization, model building and at last model evaluation.

Even though it looks like a linear process, in practice it may have different kinds of iterative variations. The very first process is gathering the needed data, from which we try to extract information and knowledge. The intention is that the data fits into a particular domain for which people generate data. So, in this phase the web scraping is done from different sources, which may be blogs, social network, news, etc.

The next phase, data preprocessing, refers to cleansing the data and bringing them to the looked-for shape. Even though some non-sense symbols or other non-textual data have been meaningful in the time provided, at the moment we try to mine value out of it, however, they are not needed, or they are even considered as noise data. So, in order to have quality data, we must go through this process.

Afterwards, the step of exploration and visualization refers to understanding and creating a sense of the data with which we are working. This is the moment when we begin to harness information from the given data, to control whether the dataset is balanced or not and to start using them in the decision making process.

The next and maybe one of the most important steps is the model building. In this process the authors try to design and build a specific model, by including several algorithms from different categories. If it is chosen a machine learning approach, we should choose an appropriate model from that approach, and an appropriate algorithm. After the model is built, it should be evaluated. This covers the last phase, where we should conduct a lot of experiments. After we have gained some results, we should analyze and compare them with the results generated from other models or algorithms. When the comparison is done, we can see if our approach has met our requirements and goals. Two different approaches to overall sentiment analysis have been developed: lexicon-based and machine learning.

Researchers inquire sentiments at various linguistic levels, such as: words, phrases, sentences and documents. These tasks analyze sentiments at a fine-grained level and can be used to improve the effectiveness of sentiment classification.

Focusing solely on the overall sentiment was not sufficient for making a good decision, so the research has been broadened and the aspect-based sentiment analysis has been introduced (Thet, et al., 2010). Aspect-Based Sentiment Analysis (ABSA) systems aim to detect the opinions or sentiments which refer to different features or aspects for an entity. In other words, they aim to detect all the aspects of an opinion and to predict the rating of each detected aspect. Aspects are attributes or features of a specific topic and ratings are numerical values for the evaluated interpretation of the user (Axhiu & Aliu, 2020).

An ABSA task is composed of a number of sub-tasks, such as detecting valid entities and aspects, and extracting the related sentiments and their polarities. An ABSA system receives as an input a set of texts referring to a specific entity, and tries to extract the aspect-terms and to evaluate the sentiment for each aspect. Having a subjective content for a specific entity, the aim would be to find all the opinions with the following information: Aspect Category Detection, Opinion Target Expression, and Sentiment Polarity. The main issues, as Madhoushi (2015) states, are that the most used techniques are unable to work well in different domains, sentiment classification based on insufficient labeled data is challenging, and there is lack of sentiment analysis research in languages other than English.

Objectives and Motivation in Research

Regarding the popularity of sentiment analysis for detecting the opinion of the beholder for different topics and in the same time the isolation of “non-popular” languages, we have decided to focus on language-independent sentiment analysis, mainly on aspect-based sentiment analysis approaches. The main objectives of the research are:

- Adapt and implement an ABSA system to automatically determine the sentiments of different language texts with minimal supervision.
- Use of novel machine learning models to extract aspect categories, opinion target expressions and to determine sentiment polarity of each aspect.
- Perform ABSA over an unlabeled Albanian corpus independent of the domain.

The motivation of this research is to increase the sentiment analysis usage for less popular languages such as our local languages, Albanian or Macedonian.

State of The Art

The notion "sentiment analysis" appeared for the first time in the 2000s, which was at the time when the research for this new field started. Sentiment analysis at its first stage focused on determining the overall sentiment of documents (Pang, Lee, & Vaithyanathan, 2002; Turney, 2002; Pang & Lee, 2008). Rule-based (Turney, 2002) and machine learning approaches (Pang et al., 2002) were used for document-level sentiments.

In the first approach the phrases containing sentiments were extracted and then compared to the sentiment of known-polarity words. Dodds et al. provide an example of a lexicon- based approach, which contains a huge size of labeled words. They have worked on nearly five million opinionated texts in 10 different languages. Another well-known and mostly used sentiment lexicon is SentiWordNet, which is built semi-automatically, for English language.

In the second approach, a sentiment classification model from a large set of sentiment labeled texts is constructed first and then tested and applied to the corpus of unlabeled texts (Mozetic et al., 2016). On the other hand, the supervised methods were used with features like unigrams, bigrams, part-of-speech tags, and word position information (Sauper & Barzilay, 2013). In both approaches, considerable involvement of humans is initially needed.

The model maps the extracted aspects into corresponding sentiments, which may have three different polarities, such as positive, neutral and negative. Determining the sentiments and their polarity of the expressed opinions is not an easy task, and depends on subjective judgment of human annotators. Annotators often disagree between themselves as a result of relative difficulties they may face, or the variety of vocabularies in different domains. Sometimes even a single annotator has contradictory annotations.

Wiebe et al., (2004) in their work evaluate the linguistic aspects of expressing opinions and assessments. Wilson et al., (2009) have considered the importance of context in defining the sentiment polarity. Later, the consideration of negation has been a target for Pang and Lee (2008), while Raychev and Nakov in 2009 started to examine the role of the position for the sentiments.

In document-level analysis, the resulting sentiment is an overall sentiment, which corresponds to the general opinion. However getting into sentence-level analysis yields a more fine-grained analysis. In the scope of sentence-level the first task is to detect the opinionated sentences and then extract their polarity (Yu & Hatzivassiloglou, 2003; Dave, Lawrence & Pennock, 2003; Kim & Hovy, 2005, 2006; Pang & Lee, 2008). Summarized by Sauper (2013) the detection of opinionated sentences and polarity analysis can be conducted by supervised classifiers. Another statement is that sentiment analysis can be advanced if the relations of the text-pieces are detected (Pang & Lee, 2004; McDonald, Hannan, Neylon, Wells, & Reynar, 2007; Pang & Lee, 2008). Succeeding the granularity work at deeper levels, namely at sentence level, the need for aspect – based sentiment analysis has arisen.

Aspect-Based Sentiment Analysis (ABSA) refers to systems that determine the opinions or sentiments expressed on different features or aspects of the products or services under evaluation (e.g., food or service for a restaurant). An ABSA model should categorize the reviews or opinions according to the proper aspect, depending on the domain; and to classify the sentiments into their three possible polarities. The used techniques for the whole process are: lexicon-based, features based and machine learning approaches, which may be supervised, semi-supervised and unsupervised methods, and as of lately, deep learning approach. A summarized description of all techniques applied to sentiment analysis is done by Rana and Cheah (2016). However, sometimes hybrid approaches which include two or more techniques are used in practice. Based on the reviewed work, the machine learning techniques are the most- used and sometimes they are combined with the lexical techniques.

Based on the used techniques, there exist three types of systems: systems that use fixed-aspect approaches or data-mining techniques for aspect selection or sentiment analysis, systems that adapt techniques from multi-document summarization, and systems that jointly model aspect and sentiment with probabilistic topic models. (Sauper & Barzilay, 2013). Regarding the tasks of aspect based sentiment analysis, the proposed research works contain between two to four subtasks. Hamdan, Bellot and Bechet (2014) propose the steps: aspect term extraction, aspect term polarity detection, category detection and category polarity detection. On the other hand, in the same year Pavlopoulos recommended to cover the whole task through three steps: The extraction of aspects and its terms for each one, the aggregation of the terms, and at last the prediction of the sentiment polarity for the defined aspect terms. Later, in 2015, Singh and Ullah performed the same process through two aggregated steps: aspect term detection and aspect polarity detection. Except for the category detection and its polarity, all the other tasks cover the same functionalities, even though they are organized differently (decomposed or aggregated steps).

ABSA systems with best performance mainly use manually labeled data and language specific resources for training on a particular domain and for a particular language (Pontiki et al., 2014, 2015, 2016, as cited in Pablos et al., 2017). The recent work of Ruppert et al. (2017) is also in-domain and has achieved competitive results both in document-level and aspect-based sentiment analysis. Although among the submissions of the past years at SemEval, the mostly chosen (Wagner et al., 2014; Kiritchenko et al., 2014; Brychcin et al., 2014; Brun et al.,2014) and winning models use support vector machines, or conditional random field classifiers (Toh & Wang, 2014; De Clercq et al., 2015; Toh & Su, 2015; Hamdan et al., 2015), deep learning has also been explored in this area by Kim (2014). At SemEval-2015 it is used as Convolutional Neural Networks (CNNs) and it has reached a high top performance for the specific tasks in 2015. The phrases that had to be specified manually, now, by CNNs they are automatically detected. For example CNNs learn that ’not good’ is negative, ’not bad’ is positive and ’very good’ is more positive than ’good’ (Deriu, 2016). Zenun et.al (2020) used the Convolutional Neural Network to detect the aspect categories and label the aspects for the next steps.

In 2018 Zhao et al. improved the Convolutional Neural Network with gating mechanisms and resulted to a new approach which was named as Gated Convolutional Network with Aspect Embedding. The authors claim that this approach is simpler than the recurrent network-based techniques. However, the recurrent network based techniques have also been quite successful. In 2018 Hazarika et al. used this technique for real-time classification of aspects in sentences together with the time-based dependency processing of their matching sentence representations.

In the process of identifying the sentiment of aspects of an entity, Nguyen and Shirai (2015) have suggested a model which is based on Recursive Neural Network, extended with the dependency trees and constituent trees of a sentence. Besides Deep Networks, a new technique based on Long Short-Term Memory has been accepted amongst machine learning models. In 2018, this network was extended by Ma et al., with a hierarchical mechanism, including target-level attention and sentence-level attention.

Considering the hierarchical approaches, in 2016 Ruder et.al have used a hierarchical Bidirectional Long Short-term Memory Network for modeling the interdependencies of sentences in a piece of text or review,. This model is also useful for a language-agnostic approach which in 2019 was suggested by Akhtar et.al. Here, the Bidirectional Long Short-term Memory Network was further supported by additional handcrafted features.

Taking advantage of neural network-based techniques in 2019, Sun et.al have used BERT (Bidirectional Encoder Representations from Transformers) technique to create secondary sentences from the aspect and transform Aspect-based Sentiment Analysis to sentence-pair classification task.

All of the techniques and models used in the conduction of experiments undergo a process of evaluation. The evaluation metrics that are used in this process are the F-1 score, recall, precision and accuracy. They are described and used in the experimental section of our work. According to the above compared techniques and according to Chen et al., (2019) and Araque et al., (2017), deep-learning based systems that offer very good performance also require a weighty amount of labeled data for training. On the other hand, less-supervised systems do not require labeled data for training, but they usually need some language specific resources, such as carefully curated lists of seed words or language dependent tools to preprocess the input (Lin et al., 2011; Jo and Oh, 2011; Kim et al., 2013, as cited in Pablos et al., 2017).

Since most of the research on sentiment analysis and specifically on ABSA focuses on texts written in the English language, most of the resources developed, such as corpora and lexicons, are mainly in English. In addition, the provided language-independent models are not sufficiently focused on the issues of languages from other families, which may result in low accuracy of ABSA. This problem has been discussed also by Hercig et al. (2016), for Czech language. Consecutively, very little research has been conducted on less spoken languages, such as our local ones, more specifically on the Albanian language. Among them, Biba and Mane (2014) have worked on topic-based sentiment analysis, and others (Caka (2011), Hasanaj (2012), Kadriu(2013) and Kabashi (2016)) have worked more on part-of-speech tagging models. That is why recent research trends focus more on semi-supervised or unsupervised approaches that need less manual annotation, as well as domain and language independent techniques to get the same or better accuracy compared to supervised methods.

Hypotheses

The research will focus on the following two hypotheses:

- Current machine learning approaches for aspect extraction and on determination of aspect sentiment polarity have better performance than traditional rule-based approaches for language-independent and domain-independent sentiment analysis tasks.
- Semi-supervised or unsupervised learning methods for ABSA can achieve high accuracy (compared to supervised learning methods) in aspect categorization and sentiment polarity.

The result of the hypotheses will be tested during the experimental phase of the research, where the algorithms and the chosen approaches will be tested.

Research Questions

The research will try to answer the following questions, after which the system should be built.

1. What are the challenges of applying methodologies to the Albanian language?
2. What type of topic-modeling algorithms works best for ABSA of Albanian texts?
3. What are the challenges of building a language-independent and domain-independent ABSA system?
4. Are there innovative application areas for ABSA in the Albanian ecosystem (e.g., political opinions, medical prescriptions, company reputations, product reviews, marketing)?

5. How does the language affect the processes of ABSA? (A comparison of English and Albanian/Macedonian language)

Methodology

The first methodology of our research was based on the model of W2VLDA (Pablos, 2017), which is an unsupervised system based on topic modeling (LDA), that combined with some other methods (continuous word embeddings) and a minimal configuration, performs aspect/category classification, aspect-terms/opinion-words separation and sentiment polarity classification (with Maximum Entropy algorithm) for any given domain and language (Pablos et al., 2017).

However while analyzing both LDA and NMF methodologies, which are considered to be state of the art for topic modeling approach, the result obtained was that in the phase of aspect term extraction for Albanian texts, NMF algorithm is slightly better than LDA. That is why we changed partially the methodology of our work, which now is based on the model NMF4ABSA (Purpura, 2018). So, all the algorithms and techniques that will be used are: NMF algorithm, word2vector model and Maximum Entropy classifier.

Besides the unlabeled corpora that we will use, we should give three inputs to the system. First, the list of seed words for aspect-terms, which will be categorized in three aspect topics, second, a list of seed words for sentiments which will be classified as positive or negative, and at last a list of stopwords for Albanian language, which we have built manually. The output of the system will be a list of extracted aspect-terms, their categorization through 3 aspect classes and a sentiment polarity for each detected aspect.

Research Contribution

The design of a framework and a system that will perform ABSA in many languages and with no domain specific requirements will contribute to the wide application of sentiment analysis even in less spoken languages. Taking into consideration the great challenges of creating the annotated corpora for any language, with our work we will contribute to skipping those steps and yet gaining accurate results. As to the Albanian language, besides the published and unpublished tag-sets, there are still no published annotated corpora. So, the limitation of using supervised learning methods can be surpassed with unsupervised or semi-supervised methods. Besides the contribution of creating new Albanian unlabeled corpora for unsupervised training for the ABSA task, we will also achieve new state-of-the-art results for Albanian language.

Structure of the Thesis

The thesis is structured in three main parts. The first part will cover the above mentioned sections, namely the State of the Art for Aspect-based sentiment analysis, the objectives and motivation of the research, the Hypotheses and Research Questions, as well as the Methodology that will be used. In the second part, there are eight chapters included in total, which first contain overviews of ABSA tasks and algorithms followed by several comparisons. The data gathering process and the data cleansing is described thereafter, succeeding with the language challenges we have identified through the process of data collection.

The third part contains the experimental results for all ABSA tasks. Firstly, the preprocessing of the dataset is done and then the lists of “aspect-term seeds”, “sentiment seeds” and STOP-words are built. Successively the steps of aspect extraction and sentiment polarity are conducted for Albanian texts and the interpretation of the results is carried out. The final chapter provides a summary of the research findings, conclusions and suggestions for future work.

PART II

Aspect Based Sentiment Analysis

In aspect-based sentiment analysis (ABSA), an opinion is represented as a group of five elements: **entity**, the **aspect** of the given entity, the **sentiment** of an aspect (for the proper entity), which is given by the **opinion holder** at a specific **time.** All these elements as a structure are denoted with (E, A, S, H, T). So, the opinion target is denoted by the entity (E) and aspect (A). The sentiment (S) can be expressed either by positive, neutral, or negative polarities or by other measurable scales such as stars (from one to five). This depends on the data and on the model used.

It is very important to emphasize that all of the five above-mentioned elements are necessary and must relate to one another, because any inconsistency may give rise to an issue. For example, in the sentence “Spanjolli e adhuron atë, por Italiani e urren.” (The Spanish adores him, but the Italian hates him), the two opinion holders must be distinguished. In this case the time element appears to be less significant, although in practice it is very possible that the stated opinion may differ in a specific period of time. For instance, in this example the feelings of adornment and hate may change after some time. Although this simplified definition/structure may be satisfactory for most of the applications, it does not consider all the possibilities of stating an opinion.

The drawbacks of this simplification may be noticed in the case of the importance of the context of the opinion. For example if we consider the statement “Dera është shumë e ulët për njerëz të lartër” (in English: “The door is too short for taller people”), we can understand from the context of the opinion that it isn’t sad that the door is too short for everyone. Another case when the simplified structure isn't satisfactory is in the case of comparative opinions, when there are expressed relations either of differences or of similarities among two or more entities.

Lastly, the hierarchical structure of the entities should be considered. If there is a need for different aspect-terms of a given aspect (e.g. restaurant food and its taste and price) to be studied, then the aspects (in this case food) of a given entity (restaurant) must be treated as separate entity (which in addition will have the aspects of taste and price). In order to emphasize the accuracy of the sentiment polarity in aspect based sentiment analysis in addition to a single-aspect sentiment analysis, we have given an example in the table below which contains a customer review and the polarity results after implementing a single-aspect sentiment analysis and ABSA.

Customer review about a restaurant	Single-Aspect Sentiment Analysis (Basic SA)	Aspect-based sentiment analysis
Kamarieri ishte shumë i kujdesshëm. Por, bifteku ishte krejtësisht pa shije. Shumë shtrenjt gjithsesi. (In English) The waiter was really diligent. However, the steak was completely tasteless. Too expensive anyway.	66% negative 33% positive	Service: positive Food: negative Price: negative

Table 1. Example of Basic Sentiment Analysis and Aspect-based Sentiment Analysis

The field of the review analysis was in fact the one that pushed the work to succeed from single aspect in document level and sentence level sentiment to aspect based sentiment analysis. All of the models used in this direction, based on their techniques, can actually be divided into three kind of approaches that work through probabilistic techniques, fixed-aspect and multi-document summarization. In the following sections, we examine each proceeding of the work.

Data Mining and Fixed Aspect Methods for Sentiment Analysis

Some methods for aspect-based sentiment analysis use traditional methods of data mining and they even use rule based techniques or other additional resources (e.g. WordNet) for identifying aspect terms and for determining the sentiment polarity on both document level or sentence level (Hu et al., 2004; Liu et al., 2005; Popescu et al., 2005, as cited in Sauper and Barzilay, 2013). An alternative method is to create a set of predefined aspects and then aim to find the opinion or sentiments that are related to the aspects (Snyder & Barzilay, 2007).

Liu et al. (2005) developed a system which worked on three steps: First, by using an association miner the collection of aspects is done, and then they are cleaned up by proper rules. In the second stage, by using the word order and position for each aspect term the associated opinions are identified. After that their polarity is identified by using seed words. At last at the third stage, extra aspects based on the position of the chosen polarity words are identified.

There are also additional methods that use the properties of the entities to identify the aspects of that entity and the sentiment polarity. These are also based on WordNet relationships. Popescu et al. (2005) has first used this feature and generated a group of aspect-terms for a specific aspect class of a product (e.g. laptop). Following this, by using relaxation labeling, they have expanded the sentiment, starting from the single-words continuing to aspect terms and finishing to sentences.

In parallel to this, in order to improve the sentiment analysis prediction, some other studies have been focused on predefined aspects. For instance, Snyder et al. in 2007 have defined a group of specific aspects for the domain of restaurants. More precisely they have built a rating model for each aspect that corresponds to an entity, as well as a model of agreement for evaluating the rating results. Supervised approaches for training these types of models have been used.

Multi-Document Summarization

Multi-document summarization techniques are usually looking for reappearance across the texts and finding the data with higher frequencies, so that they can categorize the important information (Mani, 2001). In the scope of ABSA the focus is on expanding these techniques and by using NLP of text extraction getting unambiguous outlines.

In the system of Seki et al. (2005, 2006) by using subjectivity component via supervised SVM (Support Vector Machine algorithm) with lexical features, they have created opinion focused summaries. Yu et al. and Dave et al. in 2003 have also worked along these lines and to that end. On the other hand, Hu and Liu (2004) have selected the aspects by data mining approaches for expanding the summarization system, instead of using single aspect analysis.

Along these lines and to that end, Carenini et al. (2005, 2006) have expanded the aspect term selection with a user pre-defined hierarchical association over aspects (for instance, performance if part of quality). In order to come up with the final summaries, these extended aspects are included into the existing systems for summarization. This approach however may be not feasible for large corpus of data because of the supervised approach it uses.

At last, some studies count on the old summary techniques of identifying ambiguous or conflicting sentences. Kim and Zhai in 2009, (as cited in Sauper and Barzilay, 2013) were detecting sentences that have opposite opinions for a product characteristic, and generated contrastive summaries. In order to achieve this, they defined metrics for similarities and differences. However, when the selected pairs have a disagreement in scores, there is no information on the exact scores for each particular side. Contrary to the above, our work is focused on defining a specific list of aspect terms and their corresponding sentiment polarities.

Probabilistic Topic Modeling

In the area of probabilistic topic modeling for sentiment analysis, Lu et al. (2008) have presented a system with semi-supervised PLSA which has the job to identify the sentiment-bearing aspects and to extract related or supplementary information for each aspect term. The aspect extraction is done according to the rule-based approaches, and as a general rule, all the aspect terms must be nouns. In contrast to our work, where except the aspect term extraction we want to find the polarity of each opinionated aspect, in this work they only identify and aggregate the aspects. Another thing that we don't find very reasonable is that the defining of the aspects is left to the expert reviewers, who don't guarantee that all of the aspects will be mentioned.

In this context there are studies which focus on identifying the aspects from the entire content, namely from the blogs. To this end, Mei et al. (2007, as cited in Sauper and Barzilay,

2013) have used LDA to model the aspects and sentiments, and they have then additionally used Markov model to find the sentiment trend of the aspect terms. An overall sentiment polarity is calculated by considering several distributions from different labeled datasets (e.g. restaurants, tourism, etc.). In their work the sentiment is calculated on document level.

Compared to this, we work on identifying fine-grained aspects and finding the sentiment polarity for each different aspect. Furthermore, by relating the proper sentiments to aspects, we can extract other sentiments together with their polarities without the need of extra labeling for training the model. The prediction of sentiments can be further estimated by an extension of LDA algorithm.

Blei et al. (2008) have proposed a method with supervised LDA, which contains an extra answer variable that would represent a sentiment. One method of application of this is the rating with stars, which is usually used in reviews' ratings. The advancing of these approaches goes into a deeper level of aspect extraction and sentiment polarity. Titov and McDonald (2008a, 2008b) have introduced a multi-grain unsupervised topic modeling technique by extending the LDA algorithm. This model does a mixture of general and specific topics. Since the topics are compared with all reviews of dataset, they would be more general. As a result it is challenging to find related topic for each aspect. For example, if we consider a dataset from Albanian restaurants and English ones, taking into account the diversities, we should not use the same group of aspect-terms. As Sauper and Barzilay (2013) have stated, the greater number of topics may result in detecting more aspects, but since they are still general aspects, it is difficult to detect specific aspect-terms.

During SemEval-2014 and SemEval-2015 (International workshop on Semantic Evaluation), it stood between SVM based models and deep learning models. Although in the work of Kim (2014) deep learning methods have shown great power in the tasks of sentiment analysis, still none of the teams who took the top places in SemEval-2014 and SemEval-2015 used the deep learning techniques. However, this is not an indicator of value and cannot question the power of deep-learning, especially in aspect-based sentiment analysis.

Aspect-Based Sentiment Analysis Tasks

The objective of Aspect Based Sentiment Analysis is to detect the aspects of a given entity and to find out their sentiments. To achieve this goal of Aspect-Based Sentiment Analysis, the process can be summarized in two main phases (Laskari, 2016). In the first phase the extraction of the aspect terms will be done, as will their aggregation in aspect categories. In the second phase, the sentiment polarity of both aspect terms and aspect categories should be identified. So, these two phases are decomposed into three subtasks, such as: Aspect Term Extraction, Aspect-term and Opinion Separation and Sentiment Polarity Classification. This division in different studies has been made differently. However, all the functionalities of the tasks remain the same, irrespective of how they are grouped or organized.

In our case we have divided the procedure into two main phases: the extraction of aspect-terms and the classification of sentiment polarities. Within the task of Aspect Term Extraction, the detection of categories and aspect terms that correspond to those categories within the texts is done. After the identification of aspects, a polarity needs to be assigned to each aspect. That is why the process continues to the task of polarity classification, where a positive, negative or neutral polarity is assigned for each aspect. Both tasks are explained in more detail in the following sections.

Aspect Term Extraction

There exist two major disadvantages when the sentiment analysis is conducted on sentence and document level. The first one refers to the general sentiment that may result from the overall text. The final polarity will apply to the general topic, and it is very difficult to get the right sentiment polarity since they may be a lot of different polarities (positive, neutral or negative) in the same text. They may appear in similar number however, the most applications present only one general polarity as a result (not showing the percentages, which may have only minor difference). The second problem is that in most of the cases a piece of text may discuss more than one topic, and each of them may have different polarity. This is also affecting the validity of the final result in a great manner. So, the necessity for identification of different aspects and their corresponding polarity gave rise to ABSA, where the first task is to extract aspects or features of an entity from a given text.

Aspect detection can be conducted by three different techniques: supervised, semi-supervised and unsupervised algorithms. As we have discussed in our previous work (2020), supervised aspect term extraction using human annotated datasets reaches high performance for the task itself, on previously unseen data. But, this method has two major drawbacks. First, the process of manual labeling is very costly and time consuming. The other drawback refers to the size of the annotated datasets which are relatively small. The small size of the corpora as a result may reduce the performance of the classifiers.

The above mentioned difficulties may be surpassed by extracting the aspect terms with unsupervised techniques. The first drawback can be exceeded by the usage of an automated data annotated process which will not be as costly as manual annotating process. Our goal is to analyze Albanian language data, but since there are no published labelled corpora in this language, we chose the semi-supervised approach for the task of aspect extraction. In order to decrease incorrectly annotated aspect terms in the automated labelling, similarly to Giannakopoulos, Musat, Hossmann and Baeriswyl (2017), we are considering only nouns and noun phrases.

In contrast, taking into account the size of the used corpora, by making use of reviews, news of other public data, considerable increase is possible. Considering that the online opinions expressed through comments, reviews, posts, etc., contain many different aspects, they become good resource for creating new datasets for aspect based sentiment analysis, more specifically, in this case for aspect-term extraction. Focusing on the task of aspect term extraction, new special target datasets were constructed, which have also been accepted by the well-known international sentiment analysis challenges (SEMEVAL 2014 and 2015) co-organized by Pontiki et al., (2014).

Having the pieces of texts for a specific entity or for several entities of the same type, at this phase the aspects and their terms are extracted and sometimes they are also ranked by importance. The two issues that should be considered at this stage are the enclosure of multi-word aspects (beside the single-word ones) (e.g. “screen resolution”, “hard disk”) and the aggregation of similar aspects (e.g. “color” and “design”; “price” and “cost”).

Aspect Aggregation

While the decomposition of tasks occur in aspect based sentiment analysis systems, in most of them aspect aggregation is considered as a task on its own. However in few systems it is considered as a subtask of aspect term extraction. The aim of aspect aggregation is to aggregate/cluster similar aspect terms so that there won’t be reports of different sentiment polarity scores for similar aspect terms. At the first work done on aspect aggregation by Liu (2012), the aggregation occurred only on synonyms of aspect terms. However by considering only the synonyms of the aspects, the aspect granularity cannot be chosen and also the hierarchical relations of the aspect terms are discounted. Taking into account these disadvantages, Pavlopoulos (2014) has proposed multi-granular aspect aggregation, by which a specific set of aspect terms gained from the process of aspect extraction is partitioned into “k” non-overlapping clusters, for various values of “k”. It is very important for the clusters to be reliable for all different values. Namely, if any two aspect terms are positioned in the same cluster, they must stay together also for every coarser grouping (e.g. if aspect terms are clustered together in the cluster “k”3, they must appear together also in clusters “k”2 and “k”1).

The whole process is conducted in two stages. In the first stage the system tries to fill and adapt the similarity matrix, where the values in each cell will show the semantic similarity among aspect terms. In the next stage the system uses the produced matrix from the previous stage, along with relation criteria, and then executes hierarchical agglomerative clustering. After this stage the hierarchy of aspect terms is created, and by intersecting the hierarchy at different depths, different numbers of clusters are produced. The following figure shows an example of aspect hierarchy for the "laptop" entity.

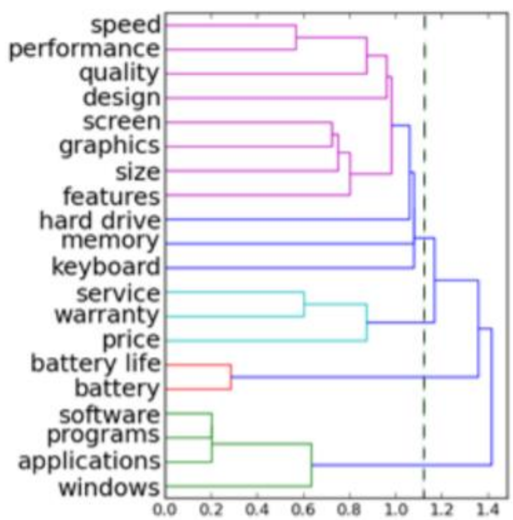

Figure 2. The aspect-terms' orders created by agglomerative hierarchical clustering for the Laptop topic. Adapted from "Aspect Based Sentiment Analysis", by Pavlopoulos, I., 2014, Athens University of Economics and Business

Single-words and Multi-words aspect terms

In the process of Aspect Term Extraction, considering the language and the domain that we are working on, in addition to the single-word aspects we must take into account also the multi-word aspect terms, because they may be in a considerable number. In the analysis of Pavlopoulos (2014) conducted within their datasets, the result was that the number of single-word aspect terms is greater than multi-word aspect terms. However, the number varies between different domains. The following figure shows the percentage of multi-word aspect terms within the m different aspect terms, and as it can be seen they appear more in the domains of electronics and laptops, rather than in restaurants and hotels.

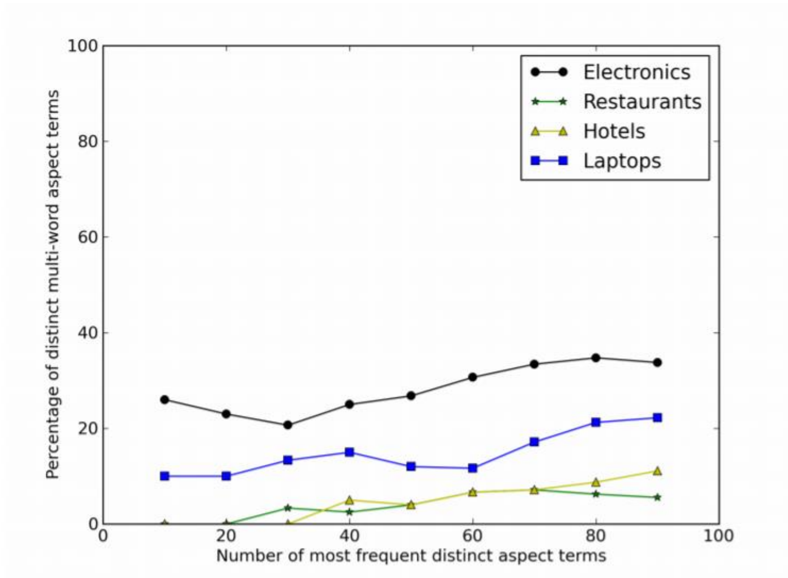

Figure 3. The ratio of multi-word aspect terms from the group of top aspect terms. Adapted from "Aspect Based Sentiment Analysis", by Pavlopoulos, I., 2014, Athens University of Economics and Business

Same as in English language, in Albanian there are also a lot of multi-word aspect terms (e.g. mish viçi, mish qingji, qymyr druri, tryezë buke, bukë misri, etc.) that may affect the accuracy of aspect terms and should not remain unconsidered.

Aspect-Term and Opinion-Word Separation

Most of the published work has been focused on the overall sentiment for a product or service; however the consumers, or the prospective ones, are lately interested in specific opinions for specific characteristics that refer to an entity. That is why the later work has proceeded to all-inclusive approaches for analyzing the opinion expressions that contain both different aspect terms and sentiments, and consequently classifying them throughout logical aspect categories. Consumers have always been interested in discovering the good and not too good, not to say bad characteristics of a product or service. They also like to compare the key points of several similar products. That is why the sentiment for each aspect is more important than just to have a specific sentiment for a general topic, because that sentiment will not correspond to the proper context (Laddha and Mukherjee, 2016). For these aims the main source are the reviews.

Taking into consideration this issue, the later work by Titov and McDonald (2008) has been concentrated on detecting and classifying the aspect terms and opinions. This has been done by generative models but, the problem was that they still didn’t take into consideration the direct relation among the opinion for a specific aspect. Later, Wang et al. (2016) and Fei et al. (2016), (as cited in Laddha and Mukherjee, 2016) worked on aspect specific opinion unigrams but not in the scope of whole phrases.

Alternatively, the W2VLDA model (on which our research is partially based on) uses biased hyper parameters which will lead the Dirichlet distributions to get the required results for aspects and opinions (Pablos et al., 2017). In a standard LDA setting those hyper-parameters are symmetric because topics or aspects and sentiment distribution are not presumed. As the authors have stated, their model uses biased parameters by applying a similarity calculation among the words. The similarity is measured by the cosine distance among the vector representations of the gold/seed aspect terms and every word of the vocabulary/dictionary. For the vocabularies in this, the word to vector (word2vec) approach it is used (Mikolov et al., 2013).

Sentiment Polarity Classification

In this task of ABSA system the aim is to estimate the sentiment of all detected aspects for a specific entity. All of the extracted aspect terms (previously described) that are found in the textual data, are supposed to be correctly extracted; and for each of them the goal is to detect the sentiment polarity. The main polarities that aspect terms may have are positive, negative or neutral. The opinions that have neither positive nor negative sentiment fall into the group of neutral polarity. However, there is another possible category which is known as conflict polarity, which holds both positive and negative opinions. For instance, within our data (2020) we have found the following statements about a politician:

- *"Më pëlqejnë veprimet e tij"* – "veprimet" has a positive polarity (in English: "I like his actions")
- *"I dua mendimet e tij, por s'më pelqen qëndrimi i tij!"* – "mendimet" has a positive polarity, while "qëndrimi" has a negative polarity. (In English: "I love his thoughts, but I don't like his posture!")
- *"Mendimet që i shprehu ishin të mira, por në kohë të gabuar."* – "mendimet" has a conflict polarity, since it has both positive and negative polarity. (In English: "His expressed thoughts were good, but with a bad timing.")
- *"Ai është një nga politikanët tanë! "*– "Ai-politikani" has a neutral polarity. (In English: He is one of our politicians!")

Regarding the measures of the evaluation at previous studies, Hu and Liu (2004) have measured the precision of classifying the sentiment of aspect terms in two polarity categories:

positive and negative. Moghaddam and Ester (2010) (as cited in Pavlopoulos, 2014) rank the different aspect terms by their polarity scores. In order to measure the distance between the proper sentiment and the predicted one, a ranking loss coefficient is used. The measures based on rankings are appropriate for approaches that try to rank the diverse aspect-terms of a given entity by its opinion, for instance, starting from the most positive aspect term and ending to the most negative, without assessing precisely how positive or negative the measures of the opinions are, and without pre-defining polarity labels to aspect terms. In our work we are going to use only the two main polarity classes and separate the sentiment polarities as positive or negative.

In the first studies the sentiment polarity task focused on assigning sentiment polarities or scores to the entire documents, sentences or messages. Further on, with the usage of Aspect-based sentiment analysis the sentiment polarities were extracted for each aspect term separately. However this shouldn't be an issue and if required, the average of all aspect polarities can be obtained for the whole text. This is valid also for the cases when we have aggregated several aspect terms in specific clusters.

Topic Modeling Algorithms

Currently available approaches for the tasks of ABSA can be divided into supervised learning approach, semi-supervised or weakly-supervised (Pablos et al. and Purpura et al., 2018) learning approaches, and unsupervised learning approach. For the first technique, the most commonly used approaches are: Support Vector Machine, Naïve Bayes classifier, Maximum Entropy classifier and lately Neural Networks. In order to get good results, these techniques need a lot of annotated data. However, this is the main issue for the languages with low resources. This is the reason why in these cases the semi-supervised or unsupervised techniques are suggested.

Topic modeling algorithms are used for aspect-term extraction and sentiment extraction with the unsupervised technique. The biggest difficulty in this process is to extract both implicit and explicit terms, where the first ones appear more and more due to the use of informal language. Since they cannot be detected and extracted by simple syntactic analysis (Chen et al., 2019) an analysis which compares the topic modeling algorithms is done abundantly.

In the field of Natural Language Processing, topic modelling represents a method of detecting hidden structures in a group of texts. Topic modelling can appear as a form of dimensionality reduction, unsupervised learning approach and as a form of tagging. When considering the dimensionality reduction a text T can be represented in its topic space rather than in its feature space. Namely, instead of denoting as {Word_i: count(Word_i, T) for Word_i in V} where V is for vocabulary (set of unique words in a specified collection) , we can denote it as {Topic_i: weight(Topic_i, T) for Topic_i in Topics}.

Regarding the phases of unsupervised learning approach, topic modelling can be used for clustering aims. Similarly to clustering, the number of topics equivalent to the number of clusters is a hyper-parameter. By using topic modelling, clusters of words, rather than clusters of texts, are created. Consequently a text is a mixture of topics, where each of them has a certain weight. After clustering is done, topic modelling appears in a form of tagging and it assigns multiple tags to a specific text. The resulting topics then are labeled by experts and various heuristics are applied so that the weighted topics are converted to a set of tags.

The most-used techniques that are considered to be SOTA for topic modelling algorithms are Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). They are both capable of extracting the topics automatically, but they present the content in different ways; LDA uses the probabilistic generative process, while NMF uses geometrically linear combinations. Therefore it is challenging and in the interest of researchers to do a comparative analysis of both models (Chen, et al., 2019). Even though these approaches are efficient and detect the correlations in a successful manner, they still have some limitations. The first drawback is their assumption that the words are independent from each other, so they extract only unigrams. The second limitation for current topic modelling techniques is the statement that the word order is not important (Bagheri, et al., 2014).

Both of the most-used topic modelling algorithms (LDA and NMF), have several common characteristics. Both of them expect the number of topics as an input parameter, because none of them can predict the number of topics that may be found in a collection of documents. Another common input is the Document_Word Matrix or the Document_Term Matrix. The Document-Word Matrix represents the number of occurrences of specific words in specific document. The algorithms also are similar in the outputs. Both of them as a result have the Word-Topic Matrix and the Topic-Document Matrix. The result of the multiplication of these two matrices should be as close as possible to the Document-Word Matrix (which was previously mentioned in the input matrices).

Sentiment analysis is still a rarely implemented process when considering Albanian language. Until now only Biba and Mane (2014) has worked on supervised learning approaches, however, to the best of our knowledge there is not a single study of unsupervised learning algorithms,. That is why we use unsupervised approaches for the task of aspect extraction.

Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a probabilistic technique which is one of the top used algorithms for topic modeling. It uses two probability measures, such as P(word/topics) and P (topics/documents). At first each word is randomly categorized in one of the topics. Then with an iterative procedure the probabilities are calculated multiple times up to the convergence of the algorithm (Chawla, 2017). To ensure a better classification of the topics a considerably high number of iterations should be conducted, because the documents' information will be updated in each iteration. In this case the documents are considered to be a mixture of latent topics which include words that in ABSA are stated as aspects.

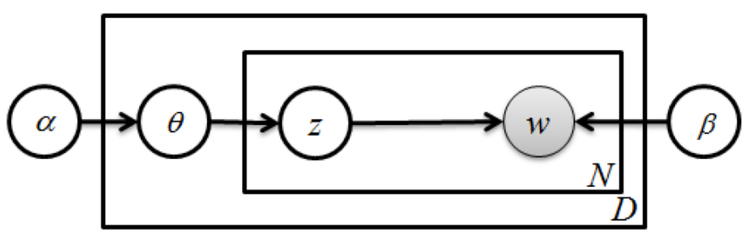

Figure 4. Graphical annotation of LDA. From "ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences", by Bagheri, A., Saraee, M., de Jong, F., 2014, Journal of Information Science 40 p. 621-636.

The above figure shows the graphical model of LDA. The nodes represent the variables, while the edges denote the dependencies. There are two types of nodes; the shaded ones symbolize the observed variables, while the others symbolize the latent variables (Bagheri, 2014). Also, there are two frames which denote the repetitions. The inner frame is for the words, while the outer frame is for the reviews. D shows the number of reviews and N shows the number of words in each review. When we have a corpus of documents with size D, each word w is linked with a latent topic z. From a dirichlet distribution Dir(α), a random sample representing the topic distribution (noted as θ)of a specific document is drawn. From a multinomial θ fo reach word w, a specific topic z is selected. After that from another drichlet distribution of BETA, a random sample which represents the word distribution of topic z is selected. And then from that distribution the word w is chosen. (Xu, 2020)

By using LDA we can detect proper topics which are accepted and interpreted by humans as well, and they are described and represented by the most closely related words. So, the corpus plays the role of training data for the dirichlet distribution of document-topic distributions. Even if we haven't seen a document, we can easily sample from the dirichlet distribution and move to the next steps from there (Xiang & Zhou, 2014).

Since for the anonymous topics generated from LDA an additional step is needed to map them to a meaningful topic's category, some use manual reviews by a professional and others do a mapping calculation to existing resources (Pablos, 2018). Pablos et al. use the combination of seed words and word embeddings for the similarities in the semantic context. So, in the next steps of ABSA, LDA in combination with other techniques and models successfully detects the aspect-terms and sentiments, and then proceed with the sentiment polarity classification.

Non-Negative Matrix Factorization Algorithm

Non-Negative Matrix factorization (noted as NMF) is a linear algebraic optimization algorithm, which due to the wide application in the field of natural language processing, is stated to be one of the top used algorithms for the chore of topic modeling. It detects significant and implicit aspect topics from a text, without any prior knowledge.

The goal of NMF is to factorize the matrix A, which is considered to be the input matrix, into two non-negative matrices W and H, in such manner that their product would be as close as the matrix itself. For topic modeling, the input matrix of choice is the document-word matrix (Suri, 2017). Matrix W represents the topics, namely the clusters that are extracted or discovered from the documents, and matrix H includes the coefficient weights for the topics in each document (Chawla, 2017). W and H are calculated by optimizing over an objective function and both of the matrices are updated by iteration until convergence.

$$\frac{1}{2}\,||\mathbf{A}-\mathbf{WH}||_{\mathsf{F}}^{2}=\sum_{i=1}^{n}\sum_{j=1}^{m}(A_{ij}-(WH)_{ij})^{2}$$

In the above written objective function, the error of reconstruction between A and the product of its factors W and H is measured, based on Euclidean distance. By using the objective function, the update rules for matrices W and H can be derived as follows:

$$W_{ic}\leftarrow W_{ic}\frac{(\mathbf{A}H)_{ic}}{(\mathbf{WH}H)_{ic}}$$

$$H_{cj} \leftarrow H_{cj} \frac{(W\mathbf{A})_{cj}}{(W\mathbf{WH})_{cj}}$$

The next calculations of the error are conducted by the new values of the matrices W and H. This calculation and update of results will be repetitive up to its convergence. The loss of some mathematical precision due to the non-negativity restriction is compensated by a meaningful and coherent representation (Ho, 2008). Considering both the explicit and implicit aspects, NMF tends to be mostly preferred for the implicit ones, because it has better results than other algorithms for topic modelling, especially from LDA. (Xu, et al., 2019)

Comparison of W2VLDA and NMF4ABSA

W2VLDA

As mentioned previously W2VLDA is a weakly supervised system for multilingual and multi-domain Aspect-based Sentiment Analysis, which works on unlabeled data with a minimal configuration. The main goal of the model is to execute the three tasks of Aspect-based sentiment analysis simultaneously. Namely, throughout this method the texts will be classified in previous known aspect categories/domains and the polarity of the sentiments will be positive or negative. Furthermore, the model in an automatic way detects the expressed opinions and separates them from their corresponding aspect-terms. With this the extraction of the sentiments is achieved. This system essentially comprises the LDA algorithm with some extensions and modeling hyper-parameters based on word embedings. Then it is united with an unsupervised classification model for aspect and sentiment extraction.

W2VLDA needs only a minimal topic configuration for each language and domain. The configuration contains a list of aspect seeds as well as sentiment seeds. This simple configuration is the only information that depends on language and domain. At last a list of stop-words for the used language is necessary. Stop-words are selected to be the most commonly used words of a language which do not contain sentiment. The list of Albanian stop-words that will be used in this model is built manually and contains about 160 words. (See Appendix 1)

Imagining a scenario where an owner of a restaurant wants to know the opinion of their customers with respect to several aspects, such as: food, service, location, price, etc., the system will need a dataset of customers' reviews and an example for each aspect. The following below describes the flow of inputs and outputs of the system. After providing the input (the reviews and the examples for each aspect; such as "burger" for the aspect "food", "rooftop" for the aspect location, etc.), the system as an output will provide a ranked list of words per aspect (e.g. burger, salad, beef, and chicken for "food"), a list of positive words (generally adjectives) and a list of negative ones for each chosen aspect. So, W2VLDA at fine-grained level executes three subtasks: aspect classification, aspect-term and opinion-word separation, and sentiment polarity classification. The system also gives a weighted list of sentences for every chosen aspect category and polarity.

.

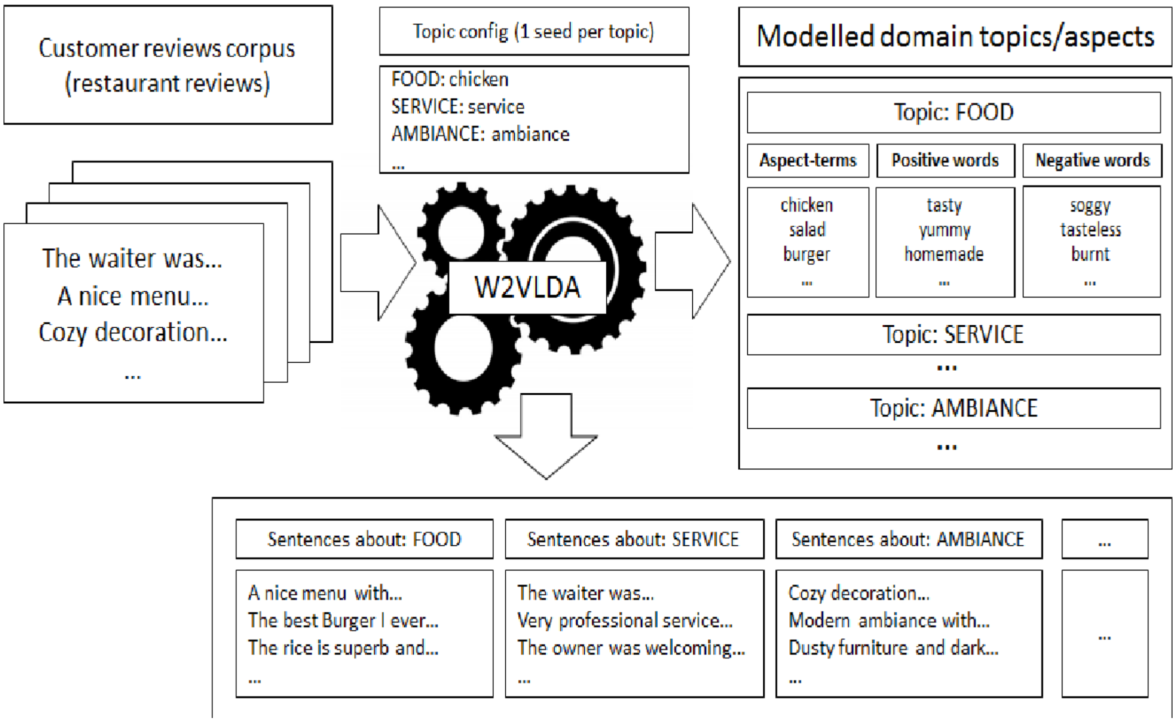

Figure 5. A schema of W2VLDA system. From "W2VLDA: Almost Unsupervised System for Aspect Based Sentiment Analysis" by Pablos, A., Cuadros, M., & Rigau, G., 2017, Arxiv.org.

Another output of the model is the aspect-term and opinion-word separation into different categories. For achieving this, first the Brown clusters are used for the first stage and then for the extraction of polarities and their classification, it is used a Maximum Entropy classifier. Brown clustering is a hierarchical agglomerative clustering issue, which is based on distributional information. It is normally used for texts, and its aim is to group the words into semantically related categories. Brown clusters are used as unsupervised features which give high results in Part-of-Speech tagging (Turian et al., 2010, as cited in Pablos et al., 2017) and Named Entity Recognition (Agerri and Rigau, 2016, as cited in Pablos et al., 2017). They are extracted from the unlabeled dataset without any extra supervision. The training instances are gained with the initial configuration of seed words, supposing that topic words are aspect-terms seeds and polarity seed words are opinion-words (Pablos et al., 2017). The figure below describes the process of obtaining the classification model, where the topic seed words are defined and used first as gold aspect-terms and polarity seed words are used as gold opinion words. The authors have used the Liu's polarity lexicon to specify gold opinion-words, and additionally for gold aspect-terms they have used the labeled dataset of SemEval 2016.

To the best of our knowledge there is no available lexicon or labeled dataset in Albanian language. Following that, the appearances of the words are detected from the corpus and adapted to their context. The next phase is the replacement of those words by their corresponding Brown cluster, so that each training instance is built. And at the very end of the process by using the generated training instances a Maximum-Entropy model is trained.

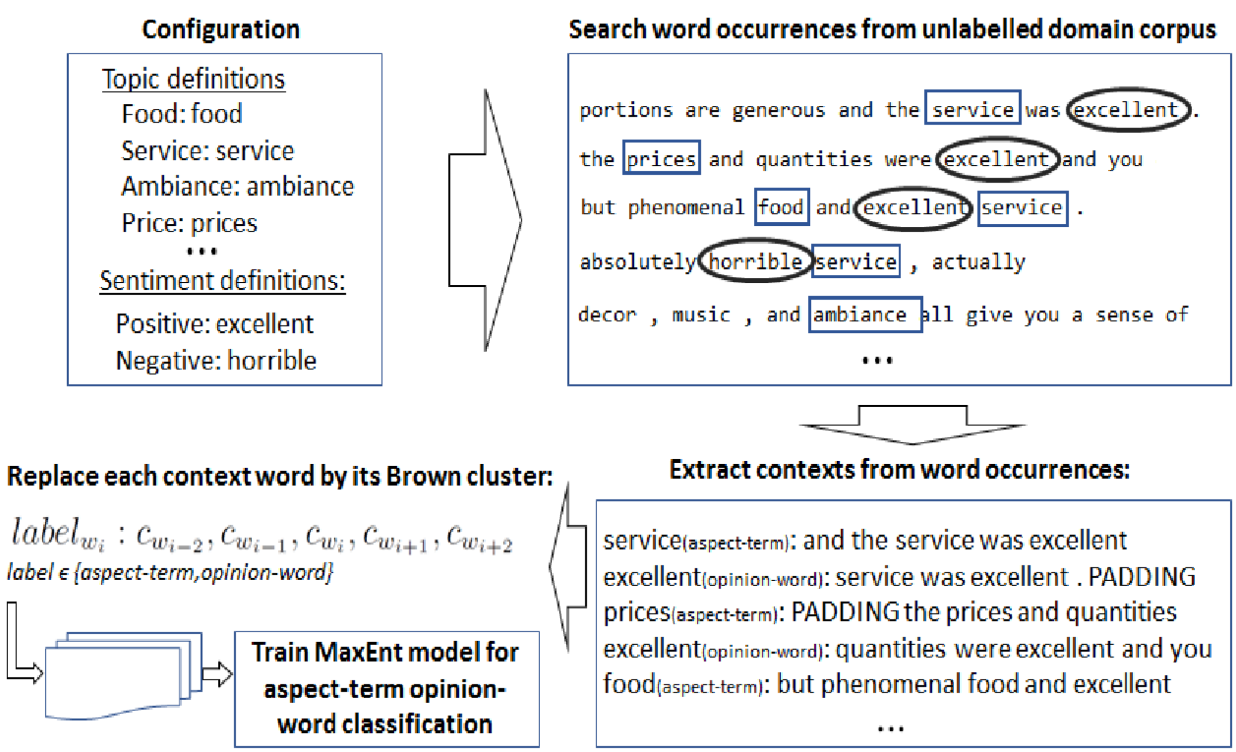

Figure 6. The process of aspect-term and opinion-word separation. From "W2VLDA: Almost Unsupervised System for Aspect Based Sentiment Analysis" by Pablos, A., Cuadros, M., & Rigau, G., 2017, Arxiv.org.

The main disadvantage of this method is that every word in the vocabulary of the corpus will be categorized as an aspect-term or as a sentiment, even though not all words belong to these classifications. It would be useful to include a third class, but additional labeled training instances would be required for that class. Namely a manual supervision, of which this system wants to rid itself, would be needed.

NMF4ABSA

Purpura (2018) has proposed Weakly-Supervised Approach for Aspect-based sentiment analysis (WS4ABSA), which is capable of execution and completion of the main tasks of ABSA, and most importantly it can be easily modified to deal with different languages and different domains. WS4ABSA system divides the tasks of ABSA in two parts: the aspect extraction and sentiment polarity classification. For the aspect classification they are using NMF technique. This method lets the user give some domain knowledge in an almost unsupervised way, in this case the linkage of discovered topics and aspect terms. In this manner this model takes as an input the list of seed words that will lead the system to a more substantial topic definition. Regarding the sentiment polarity classification, WS4ABSA uses again a list of positive and negative seeds for the defined sentiments. These lists are further on extended by using Word2Vec (Mikolov, 2013) and used to give a polarity to each aspect-term that is identified in the first task. According to Purpura (2018) this system differentiates itself from the others because it does not depend on any additional resources or annotated corpora, but only on the before-mentioned lists and optionally some simple grammar rules, which depend on the language. That is why, as stated by the author, WS4ABSA can be easily applied for the analysis of texts from any domain and in any language. In addition, this system allows the user to include some prior knowledge on the case and to continuously progress the results thanks to the adaption of NMF algorithm. So, the overall idea of the model is to extract the aspect-terms and categorize them in several aspects through aspect seeds, and then the list of sentiment seeds is used to classify the sentiments in their polarities,.

Considering a non-negative $m \times n$ matrix A (i.e. a matrix where each element $A_{ij} \geq 0, \forall i, j$), non-negative matrix factors $W \in \mathbb{R}_+^{m \times k}$ (*term-topic matrix*) and $H \in \mathbb{R}_+^{k \times n}$ (*topic-document matrix*) should be found, for a given number of aspects $k \in \mathbb{N}_+$, such that $A \approx WH$. In their statement of the problem, the matrix A represents the collection of n documents that they want to analyze, with respect to the m distinct terms contained in the collection. W signifies the relations between the terms included in the collection and k the considered aspects, and H represents the associations between the aspects and each document in the indexed collection. The factorization problem based on the Frobenius Norm is denoted here:

$$\min_{W \geq 0, H \geq 0} f(W, H) = ||A - WH||_F^2,$$

where, with $W, H \geq 0$, it is imposed that the constraint on each element of the matrices be non-negative, and with $|| \cdot ||F$ i the Frobenius norm is indicated. Although NMF is an NP-hard problem (Vavasis, 2009), there is still an attempt to approximate a local minimum. In their study the focus was on the Block Coordinate Descent (BCD) method that is an algorithmic framework to optimize the above objective function. BCD distributes variables into several subclasses and iteratively minimizes the objective function with respect to the variables of each subclass at a time. In order to solve the NMF issue with the BCD technique, it refers to a method called Hierarchical Alternating Least Squares (HALS) (Cichocki, 2009). If the matrices *W* and *H* are separated into 2*k* blocks (*k* blocks are respectively the columns of *W* and the rows of *H*), in this case the problem can be seen in the objective function in first equation as

$$||A - WH||_F^2 = ||A - \sum_{i=1}^{k} w_{\cdot i} h_{i \cdot}||_F^2.$$

To minimize each block of the matrices it is solved

$$\min_{w_{\cdot i} \geq 0} \|h_{i\cdot}^T w_{\cdot i}^T - R_i^T\|_F^2, \quad \min_{h_{i\cdot} \geq 0} \|w_{\cdot i} h_{i\cdot} - R_i\|_F^2,$$

where

$$R_i = A - \sum_{\tilde{i}=1,\ \tilde{i} \neq i}^{k} w_{\cdot \tilde{i}} h_{\tilde{i} \cdot}.$$

The promising aspect of this 2*k*-block partitioning is that each sub-problem in the third formula has a closed-form solution using the second theorem from Kim. The convergence of the algorithm is assured if the blocks of *W* and *H* remain non-zero throughout all the iterations and the minima of third formula are attained at each step (Kim, 2014). Lastly, a regularization factor for *H* is included in the first equation to induce sparse solutions, so that each document is modeled as a mixture of just a few topics. At the same time, a regularization term on *W* is added, to stop its inputs to grow too much, and make use of available knowledge of the user. Namely, for each topic, the user can ascertain some terms as appropriate, or exclude some others. Taking into account all of this, the following expression is obtained:

$$\min_{W,H\geq 0} ||A - WH||_F^2 + \phi(\alpha_p, W, P) + \psi(H)$$

Where $\psi(H)$ =β $\sum_{i=1}^{n} \quad ||h||^2{}_1$ and $\varphi(\alpha p,\ W,\ P)$ = $\sum_{i=1}^{m} \quad \sum_{j=1}^{k} \quad a_{Pij}\big(w_{ij-} P_{ij)}\big)2+ \propto$ $||W||_F^2$

The notation *hi* ・ is used to represent the *i*-th row of matrix *H*, the *l*1 term denotes sparsity on the rows of matrix *H*, while the Frobenius norm stops the growing of the values in matrix *W*. The term *P* is an *m*×*k* matrix in which inputs are either 1 for aspect-terms that according to the prior knowledge about the domain should be categorized to a certain aspect and 0 for the aspect-terms that do not belong to any of the given aspetcs. For instance, if the first terms list of the aspect Food includes the terms 'beef' and 'pizza', the values corresponding to the aspect are noted with 1 (one). The lists with seed-words are generally short, since the aim is to have minimal human supervision.

The values in *αp* help to normalize the difference between the matrices W and P, as well as to enable or disable the previous state of a given aspect-term for each k aspect. Namely, the matrix P will be affected in such way that if the element *pij* = 1, the i-th term is associated with j-th aspect. Conversely, if the value of *pij* = 0, the *i*-th term will not be associated with the given aspect, which is in position j. This gives the chance to manipulate with the values in *αp* and *P*, to choose the limit of influencing the correlation between certain aspect topics and the seed terms. So, this change and adaption of matrix P affects the formula for the matrices W and H, which is stated as:

$$
\begin{aligned}
& w_{\cdot k} \leftarrow \frac{[\nu]_+}{\alpha_{p\cdot k} + (HH^T)_{kk}}, \quad h_{i\cdot}^T \leftarrow \left[h_{i\cdot}^T + \xi\right]_+, \\
& \nu = (AH^T)_{\cdot k} - (WHH^T)_{\cdot k} + W_{\cdot k}(HH^T)_{kk}\alpha_{p\cdot k} \odot P_{\cdot k} - 0.5\alpha 1_m, \\
& \xi = \frac{(A^T W)_{\cdot i} - H^T((W^T W)_{\cdot i} + \beta 1_k)}{(W^T W)_{ii} + \beta}.
\end{aligned}
$$

Here, ʘ shows a product, $[x]_+ = \max(0, x)$, while $1t$ shows a vector of ones (1s) in a length of t. Once the factorization of matrix *A* is done into the factors *W* and *H,* the normalization of the columns of matrix *H* is done next. In order to proceed with identification of relevant topics in a given text, a threshold for the detection of aspects (ADT) should be set to a proper value, considering the pre-defined number of the topics. In this case if the value in matrix H is equal or greater than the specified threshold, the text is associated to a specific topic.

As it is mentioned previously the initialization of the document-term matrix is a crucial aspect on NMF. It should be assured that the prior knowledge on the topics should appear in that matrix. The proposed model in NMF4ABSA as it is mentioned previously uses short seed lists of terms which will affect the process of aspect classification. Namely, they select a set of terms D for indexing the collection through several steps. Firstly, all the seed words that appear in a specific collection are added to D. Then, the seed list is extended automatically by using Word2Vec. And at last the TF-IDF weight is calculated for all of the terms of the collection and the selected top few hundreds are also added to D.

After the detection and extraction of the aspects for each sentence, their polarity is calculated and the sentiments are classified as positive or negative. Similarly, a weakly supervised approach is proposed in this stage as well. First of all two lists of seed words are created, both for positive and negative sentiments for each detected aspect. Then again the previously created lists are extended by using Word2Vec model which is described in the next section. The process of pre-processing of the data set continues with stemming and the removal of the stop-words. And lastly, the identification of sentiments for each document is done.

In order to compute the polarity for each aspect, we calculate the average of all aspect-terms' polarities which correspond to the same aspect, and the result will be between -1 and 1. For achieving the highest performance the authors give the possibility to extend the seed lists, as well as to include some grammatical rules, especially concerning the negation. They search for negation terms in a sentence and if they are detected within a range of 3 tokens before a sentiment term, the polarity would be changed to the opposite one.

Maximum-Entropy Algorithm

Maximum Entropy Algorithm is a Machine Learning Algorithm that is used for the tasks of Natural Language Processing. The Max Entropy classifier takes part in the class of exponential models. It is a probabilistic classifier and its main rule is that the probability distribution should be uniform in case we do not have pre-knowledge about the previous distributions and successively when it is risky to make any assumption. Furthermore, this classifier is needed when we can't assume the conditional independence of the features, as it may appear in the Text Classification problems where the features are typically dependent.

The probability distribution that is as close to the uniform distribution which is used by the Maximum Entropy algorithm affects the results which are better than Naive Bayes. Compared to Naïve Bayes, ME algorithm needs more time to train. The reason is mainly because of the optimization problem that needs to be solved with the purpose of estimating the parameters of the model. However, after having these parameters, the technique provides strong results and it is competitive in terms of CPU and memory consumption (as it is stated by Vryniotis, 2013). The aim is to use the appropriate information of the document, namely the specific features like unigrams, bi-grams or other characteristics, so that we can categorize it to a particular class, which depending on the polarity may be positive, negative or neutral class, and depending on the opinion holder the information can be categorized into a subjective or an objective class.

The polarity is based more on the position of words rather than their frequencies. Maximum Entropy supports the rule ``the more the merrier” so, the higher the entropy - the higher the uniformity (Patel, et.al. 2016). This rule is based on choosing the most uniform distribution, which may be detected and selected by the one which has the maximum entropy. Considering the typical bag-of words framework which is normally used in natural language processing and retrieval of information; by the proposed approach of Pang and Lee (2002) the set {word1,…wordm} represents the m words that can be found in a document. Successively for representing a document an array is used with 1s and 0s which indicate the existence of wordi in the context of the document.

The model of Vryniotis (2013) accepts the x contextual information as an input and creates the output y. Similarly to the case of Naive Bayes, the first step of building this model is to gather a large number of training data which contains samples presented on the format of (xi,yi), where xi contains the contextual information of the document, which previously has been represented by the sparse array, and yi includes its class. Following this, the training sample is summarized by the following formula:

$$\tilde{p}(x, y) \equiv \frac{1}{N} X \; number\ of\ times\ that\ (x, y)\ occurs\ in\ the\ sample$$

where N is the size of the training dataset. The empirical probability distribution stated above is used for constructing the statistical model of the random process which allocates texts to a specific class by taking into account the contextual information.

Word Embeddings

A Word Embedding is one of the most popular representations of document vocabulary which attempts to plot a word in a vector by using a dictionary. It can capture the context of the words, and detect the semantic similarities with other texts. The need of word embeddings appeared because many Machine Learning algorithms and almost all Deep Learning Architectures in their original form are unable to process plain texts. In order to perform classification, regression or other functionalities, they need numeric values. Below is an elaborated example of word embeddings, namely, the texts' conversion into numbers is shown. It is important to mention that one piece of text may have different numerical representations (Vidhya, 2017).

If we consider the sentence "Word embeddings are vector representations of a word", as words can be extracted "embeddings", "vector", etc. A dictionary D of the stated sentence will be the list of all unique words. Namely it will be like D=[“word”, “embeddings”, “are”, ‘vector”, “representations”, “of”, “a”]. A vector representation of a word at its simplest form will be a one-hot encoded vector where 1 stands for the position where the word exists and 0 everywhere else. The length of our one-hot encoded vector would be seven. Here are the some examples of vector representations of some word from our dictionary.

‘embeddings’=[0, 1, 0, 0, 0, 0, 0]

‘vector’=[0, 0, 0, 1, 0, 0, 0]

After visualization of the encodings, a seven-dimensional space is created where each word is placed in different dimension and does not have any relation with the other words. Our goal is that words which are found in a similar context take place in close spatial points. Namely, in the mathematical aspect, the cosine value of the angle between those vectors should intend to go to 1.

There are several kinds of word embeddings, which are separated in two main categories: Frequency based embeddings and Prediction based embeddings. In the category of Frequency-based embeddings there are three types of vectors: Count Vector, TF-IDF Vector and Co-Occurrence Vector. On the other hand within the scope of Prediction-based embeddings there are two main vectorization methods: Continuous Bag of words (CBOW) and Skip-Gram.

Word2vec

Word2Vec (developed by Tomas Mikolov in 2013 at Google) is one of the most popular prediction-based techniques for word embeddings. It is not one algorithm in its own, namely it is consisted by two different models: CBOW and Skip-gram. They are both shallow neural networks which do the mapping of other words to the target word. These methods are prediction based which proved to be the SOTA for word similarities and for word analogies.

Continuous Bag of words

Continuous Bag of words (CBOW) is a vectorization method, which predicts the probability of a word given e context. The context can be either a single word or a group of words. In our case we have taken a single context word and tried to guess another word corresponding to the context. Assume that we have the corpus C= “Ky është shembull korpusi që përdor vetëm një fjalë konteksti.” and we have defined a context window of 1. The following table shows the gaining of training set for the CBOW model.

Input	**Output**	Datapoint	Ky	është	shembull	korpusi	që	përdor	vetëm	një	fjalë	konteksti
Ky	është	1	1	0	0	0	0	0	0	0	0	0
është	Ky	2	0	1	0	0	0	0	0	0	0	0
është	shembull	3	0	1	0	0	0	0	0	0	0	0
shembull	është	4	0	0	1	0	0	0	0	0	0	0
shembull	korpusi	5	0	0	1	0	0	0	0	0	0	0
Korpusi	shembull	6	0	0	0	1	0	0	0	0	0	0
korpusi	Që	7	0	0	0	1	0	0	0	0	0	0
Që	korpusi	8	0	0	0	0	1	0	0	0	0	0
Që	përdor	9	0	0	0	0	1	0	0	0	0	0
përdor	Që	10	0	0	0	0	0	1	0	0	0	0
përdor	vetëm	11	0	0	0	0	0	1	0	0	0	0
vetëm	përdor	12	0	0	0	0	0	0	1	0	0	0
vetëm	Një	13	0	0	0	0	0	0	1	0	0	0
Një	vetëm	14	0	0	0	0	0	0	0	1	0	0
Një	Fjalë	15	0	0	0	0	0	0	0	1	0	0
Fjalë	Një	16	0	0	0	0	0	0	0	0	1	0
Fjalë	konteksti	17	0	0	0	0	0	0	0	0	1	0
konteksti	Fjalë	18	0	0	0	0	0	0	0	0	0	1

Table 2. CBOW example

The matrix, on the right holds the one-hot encoded of the left word given as an input. This matrix is sent to neural network through three layers: the input layer, the hidden layer and the output layer. The softmax layer sums up the probabilities gotten in the output layer. Considering the fifth datapoint as an example where the input word is “shembull” the aim is to predict the target word “korpusi” by using single-context word. Namely, by using the one-hot encoding of the input, the output error is measured. While predicting the target word, we obtain the vector representation of the target word. For example the vector representation of the target word “korpusi” (datapoint 5) is as follows:

shembull	korpusi	Datapoint 5										

The following diagram shows the architecture of the CBOW model where the input is the vector of size V; The hidden layer has N neurons; and the output layer similarly to input layer is a V- length vector, the elements of which are the softmax values.

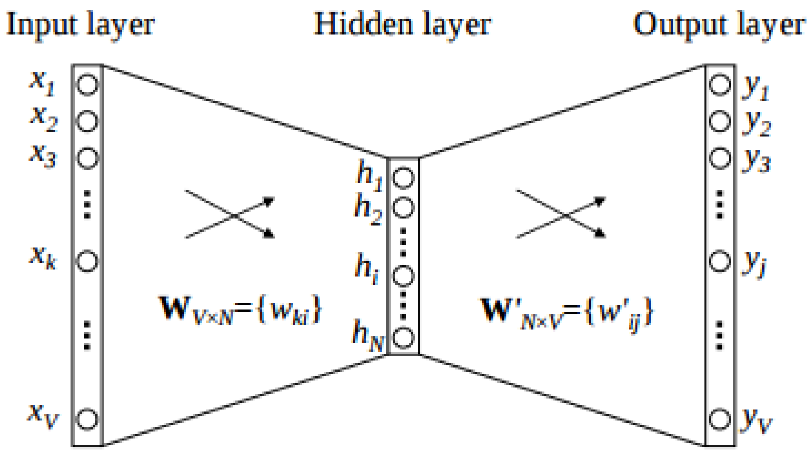

Figure 7. CBOW for single data point. From "Continuous Bag of Words (CBOW) - Single Word Model - How It Works" by Naskar, n.d., Think Infi

The matrix representation considering the above architecture for a single data point is shown below.

Context

Datapoint5	0	0	1	0	0	0	0	0	0	0

Input Hidden Weight

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16
17	18	19	20
21	22	23	24
25	26	27	28
29	30	31	32
33	34	35	36
37	38	39	40

Hidden Activation

9	10	11	12

Figure 8. Example for one context word

Considering the above mentioned example, here are the steps of the flow that were specified by Sarwan (2017):

- Both input layer and the target are one-hot encoded vectors of size [1xV], where the value of V in our example is 10.
- There are two sets of weights. The first set of weight is between the input layer and the hidden layer, while the second set is among hidden layer and the output layer.
- The size of matrix for the input-hidden layer is [V x N] and opposite to that the size of the hidden-output layer matrix is [N X V]. Here N represents the number of dimensions that we have chosen to represent the selected term. It is a hyper-parameter for a neural network. N is also the number of neurons that appear in the hidden layer. In our case it is N=4.
- The input is multiplied by the input-hidden weights and called as “hidden activation”. It is simply a copy of the corresponding row in the input-hidden matrix.
- The output is calculated by the multiplication of hidden-input with hidden-output weights.
- The error that appears among output and target is calculated and then turned back to change the weights.
- The word vector representation of the selected word is provided from the weight between the hidden layer and the output layer.

The whole process can be represented by the following figure.

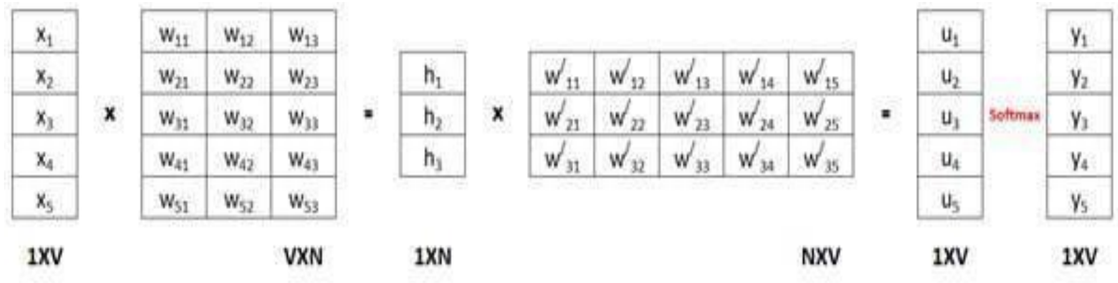

Figure 9. CBOW in vectorized form. From "Continuous Bag of Words (CBOW) - Single Word Model - How It Works" by Naskar, n.d., Think Infi

Besides the above-described model for predicting the target word, which uses single context word, we can also use multiple context words. Figure 10 represents the structure for multiple context words. W_{VxN} calculates the hidden layer inputs; in this case an average of all context word inputs is taken. So, in this structure the input layer will have three input vectors with size [1 x V] and one vector in the output layer (with size [1 x V]). The other part of the architecture is identical to a one-context CBOW.

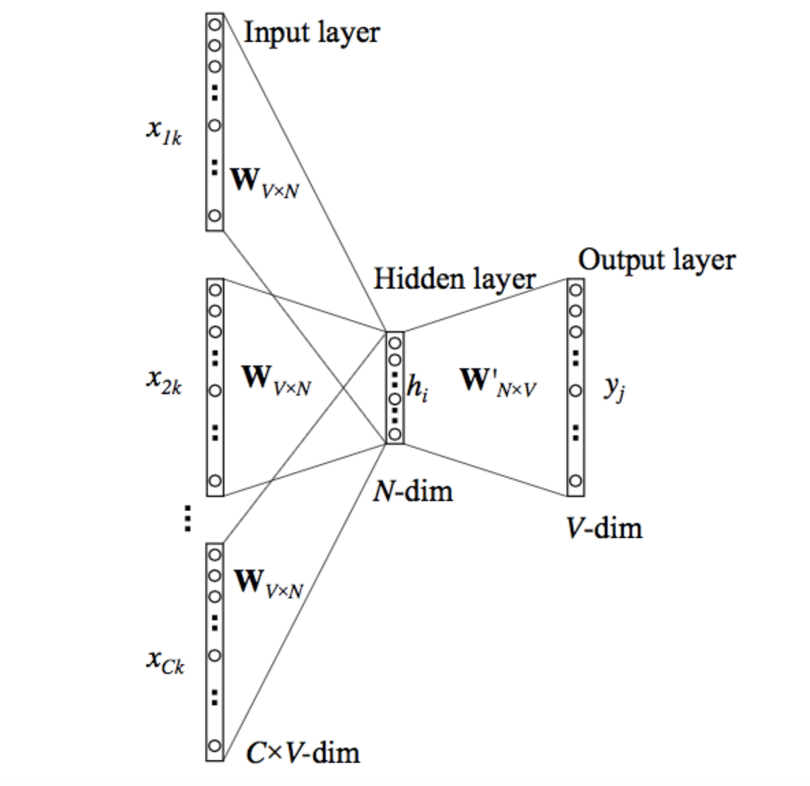

Figure 10. The CBOW model for multiple context words. From "Word2vec.", 2020, Version 4, Devopedia.

There are several advantages and disadvantages of CBOW. One of the main advantages is the probabilistic nature that it has, due to which its general performance is superior to the deterministic methods. Another advantage that is worth mentioning is the low memory requirements needed.

The main drawback of CBOW is found in the homonyms, because in calculating the hidden activation, it takes the average of the words' context. For example, "verë" (in Albanian) can be both wine and summer, however this model takes an average of both contexts and these words are placed in between a cluster for drinks and seasons. Beside the CBOW model for generating word representations by using context words, there is another method that can do the same. Skip-Gram model is a prediction based method as well, which follows the same topology as CBOW. The following section describes its architecture.

Skip-Gram model

Ski-Gram is an unsupervised learning technique, which is used to discover the most associated words for a specific word. The goal of Skip-Gram model is to predict the context by using a target word that we want in order to generate its representation (Doshi, 2019). This model is similar to multiple-context CBOW method, whose architecture is flipped. To some extent this is correct because by giving a target word as input, the Ski-Gram model outputs C probability distributions. This process becomes even more difficult because more than one context word should be predicted. The architecture of the model is demonstrated in the next figure.

Considering the architecture of the model, here are the phases to be followed:

- First, by using one-hot encoding the words are converted into vector. The size of the vector would be [1xV], where the V is the number of neurons.
- Each word w(t) is distributed to the hidden layer from V.
- Afterwards, in the hidden layer is calculated the product between weight vector $W_{[VxN]}$ and the input vector.
- Since, (as we have mentioned before) the hidden layer does not contain activation function, the $H_{[1xk]}$ will be sent straight to the output layer.
- The output layer will calculate the product of H[1xN] and W'[NxV] and will provide the vector U.

- In order to calculate the probability of each vector the softmax function is used. In each iteration there is generated an output vector U.
- The result will be the word with the highest probability.

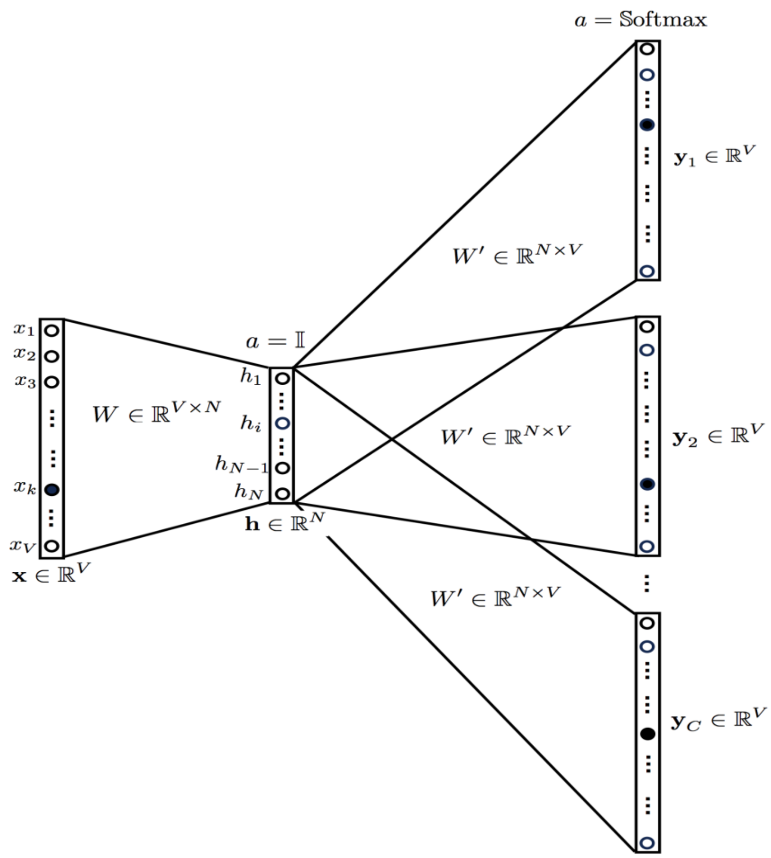

Figure 11. Skip-Gram architecture. From "Skip-Gram: NLP context words prediction algorithm." by Doshi, S., 2019.

These steps will be executed for each word of the vocabulary, which will be passed k times. For getting a better picture we have considered the same corpus that we have used in CBOW model. C=" Ky është shembull korpusi që përdor vetëm një fjalë konteksti." The construction of the training data is demonstrated in following table.

Input	Output (Context 1)	Output (Context 2)
ky	është	/
është	ky	shembull
shembull	është	korpusi
korpusi	shembull	që
që	korpusi	përdor
përdor	që	vetëm
vetëm	përdor	një
një	vetëm	fjalë
fjalë	një	konteksti
konteksti	fjalë	/

Table 3. Example for the construction of the training data

Except for the target variable (which will be two-hot encoded target and two corresponding outputs), both of the models will be the same, until the hidden layer activation. The advantage of Skip-gram model is that it can detect two different semantics for a single word. Namely, it will have two different vector representations for the previous mentioned example (verë-which in Albanian means wine or summer). So, one of the clusters will be for the drinks and another one for seasons. Another advantage is the ability to work on any raw text since it is unsupervised learning method. It also needs less memory compared to the other methods for word to vector representations. On the other hand this model has difficulties on finding the best value for **N** (number of neurons) and **c** (the window size). Also the Softmax function has a high processing and time cost.

In both analyzed models in case that the projected target word for a specific context position is not correct, there is used back-propagation in order to modify the weight vectors W and W'. While comparing both of the models, according to Mikolov (2013) the architecture of CBOW works better on the syntactic tasks in contrast to Skip-Gram architecture. But, Skip-Gram works significantly better for semantics.

Term Frequency-Inverse Document Frequency (TF-IDF)

Term Frequency-Inverse Document Frequency (TF-IDF) is one of the frequency based methods of word embeddings, which has been invented for document search and information retrieval. Compared to count vectorization method, which takes into account the occurrence of a word in a document, TF-IDF considers the words' occurrence in the whole corpus. This statistical evaluation is calculated by the product of the number of appearance of a word in a document with the inverse frequency of the word across the corpus. So, the words such as "if" ("nëse), "this" ("ky/kjo"), "what" ("çfarë"), "a" ("një"), etc. appear relatively in frequent manner, they still have a low rank because of their poor meaning to a particular text.

Regarding the calculation of TF-IDF, as mentioned previously, we should consider two metrics; term frequency and inverse document frequency. There are different ways that apply for calculating the term frequency of a word in a specific document. The first and the simplest form is a raw count of occurrences of a word in a document. Other ways refer to adjustments of frequency by the length of a document or by the raw frequency of the most repeated word in a document. On the other hand the inverse document frequency, which is the second metric, evaluates how common or rare a word is in the corpus (the whole set of documents). This term can be calculated by dividing the total number of documents by the number of documents that contain the requested word and calculating the logarithm. So, if the word appears too many times and is found in many documents, this value will approach zero, otherwise it will approach one. The calculation of TF-IDF score for the term t in the document d from the document set (corpus) D is done by the following formula:

$$tfidf(t,d,D) = tf(t,d) * idf(t,D)$$

where:

$$tf(t,d) = lo\big(1 + freq(t,d)\big)$$

$$idf(t,D) = log(\frac{N}{count(d \in D: t \in d})$$

The calculation of TF-IDF is explained by the following example where four terms from two documents are considered. There are two sample tables below where the count of each term is shown.

Document 1			**Document 2**	
Term	**Count**		**Term**	**Count**
this	2		This	2
is	1		Is	2
about	2		About	1
Covid19	2		tf-idf	1

Table 4. Term Frequency

Now, in order to calculate the first metric TF of the word “this” in document 1, we need to find its frequency first and then to calculate the logarithmic expression as in the above formula.

tf (“this”, Document1)= log(1+(freq(“this”, Document1)) =log(1+2/7)=0.109

idf(“this”, D)=log(2/2)=0

tfidf(“this”, Document1, D)=0.109*0=0.

We do the same calculation for the term Covid in order to see the relevance of the words that are very frequent and the ones that really matter in our corpus.

tf (“Covid19”, Document1)= log(1+(freq(“Covid19”, Document1)) =log(1+2/7)=0.109

idf(“Covid19”, D)=log(2/1)=0.301

tfidf(“this”, Document1, D)=0.109*0.301=0.033.

If we compare the TF-IDF results of both “this” and “covid19”, we can notice that this method gives greater weight to the word “Covid19” even though the word “this” appears in both documents and has double frequency. With this we can conclude that TF-IDF plays a great role in the field of machine learning, especially in text analysis, because in addition to deciphering words by allocating them a numerical value, it also helps to sort data into categories and to extract keywords from a corpus.

Comparison of Rule-Based and other Machine Learning Approaches

The process of detecting and analyzing the expressed sentiment and emotion is very difficult. Generally there are two approaches which are used for this task; the lexicon based approach and the machine learning approach. The first one depends from either a dictionary or a corpus. In our work we are focused on machine learning approaches, that is why in this section it is done a comparison among different algorithms and models from this field. When compared to lexicon-based approach, the machine learning approach has higher efficiency for detecting the aspects, their sentiments and successively their polarity (Syamala, 2019).

In this section our aim is to compare the different approaches used for the task of aspect-term extraction and sentiment extraction and classification. For the task of aspect-term extraction there are three methods, such as: rule based approach, seed-based and topic modeling approach. By the rule-based algorithms the aspects are extracted through certain pre-defined rules that take into account the score and frequency of the occurrence of the terms. In the direction of frequency, Muangon et al. (2014) extracted the terms by lexicon and then ranked them according to the frequency, while Zafra et al. in 2016 applied a method which uses bag of words for the same task.

Some other researches (Wang et al. (2017), Taylor et al. (2013, 2014)) have proposed the usage of part of speech taggers (POS). First the data goes through the process of stemming where the flexional endings are removed and then in the step of tagging, the aspect extraction increases. As an improvement of the existing rule based methods, a fuzzy-based learning approach was proposed. Afzaal et al (2016) have used the fuzzy unordered rule induction algorithm, with which the frequent nouns are extracted from the reviews. In the case of document level sentiment, the ease of rule-based approach is considered a good option. However, this approach has several disadvantages. First, the system based on this approach must have a rule for each word combination in the dictionary. A high amount of human labor is needed for the purposes of defining and preserving these rules, since they are very strict. The rule set cannot include all the abbreviations, ambiguous meaning, homonyms, acronyms, etc. So, since there is a huge amount of barriers in the rule based systems, there is a need for the implementation of other machine learning approaches.

Another approach is seed-based, in which a set of aspect seeds is used that will help in the process of extracting the aspects either by using grammatical or semantic relations. Colhon et al. (2014) have selected several aspects with the highest frequency and in order to create the set of seed aspects, he considered the grammatical connections and then the extended list is taken into account for the further process of aspect extraction. On the other hand Mukherjee et al. (2012) have used the seed based approach by finding the semantic relation among words of the reviews. Kayaalp et al. (2017) have also used the approach based on seeds but they modified the approach by involving index-based extraction. In this model, the most frequent aspects are selected first, and then the evaluation words are indexed and stored. And finally the indexed words are classified under the chosen aspect terms, based on their similarity. The usage of seed-based approach has also been applied by Pablos et al. (2017) and by Purpura (2018) however, they have combined this approach with topic modeling. The limitation of this approach is the domain dependency that can create, since for different domains, the same seeds cannot be valid.

The third approach for aspect extraction is based on topic modeling, which considers that each sentiment corresponds to a specific aspect, and according to this the classification is done. Two algorithms (LDA and NMF) of topic modeling approach were described in the above sections. Taking into account all of the mentioned approaches in the following table there is a comparison among them, considering their accuracy of aspect extraction. The size of the datasets is similar, but the domains are different, namely there are reviews from restaurant, hotels and tourist reviews.

Used Technique	Accuracy	Reference
Rule-based	71.2%	Zafra et al.
Seed-based	66%	Kayaalp et al.
Topic-based	75%	Shams et al.
FURIA based	79%	Afzsal et al.
Seed-based + Topic modeling (LDA)	73%	Pablos et al.
Seed-based + Topic modeling (NMF)	74%	Purpura et.al

Table 5. Comparison of Machine Learning techniques

When we take a look at the results of the considered models, we can easily see that most of the models from the last two approaches have higher accuracy then the rule-based model. Taking into account these results and all of the drawbacks that the rule-based approach has, we strongly suggest the usage of other machine learning approaches. In our model we have used a combined approach among seed-based and topic-modeling, which we find it to be more appropriate for language independent and domain independent, unsupervised systems, especially for the less-spoken languages, where there is a lack of lexicons and rule-sets.

Data Collection and Cleansing Of Data

As it was mentioned previously, the phase of Data preprocessing gets the textual data in a desired structure. This task can be conducted through three different actions:

- Tokenization,
- Normalization, and
- Noise removal

As it can be noticed from figure 12, this process is not conducted linearly and the activities do not have a specific order.

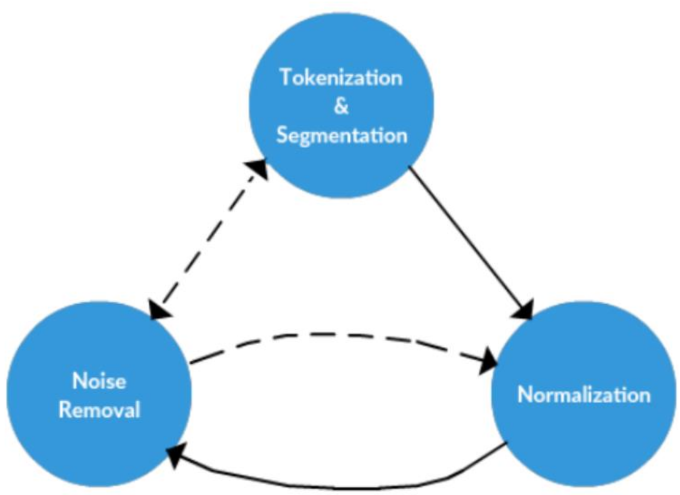

Figure 12. Data Preprocessing

Tokenization

The first activity of the NLP process is tokenization, which refers to the procedure of splitting textual structures starting from the longer strings to shorter ones according to the syntactic rules of a language. Usually, the document as a whole is first split into paragraph, after that into sentences finally into words. This procedure is not as straightforward as it seems because of two main issues; firstly, the end of the sentence is not always specified by the punctuation marks, and secondly the complexity may be even higher when foreign languages are included. In these cases it will be inappropriate to continue with the automatic tokenization, instead of consulting linguists or other native speakers that know the language.

The process of tokenization is conducted with the purpose of easing the process of handling data. Namely, it is easier to adjust single words considering the prior knowledge from the dictionaries, rather than to find and adapt the key points within the documents, paragraphs or sentences. In our case we have conducted a unigram tokenization in order to find the aspects properly.

Normalization

Another important activity in the phase of preprocessing of data is the normalization. In order to adapt and put the data in the same level, there is a group of tasks that should be conducted. Some technical tasks should be done first, such as: converting all textual information to the same letter case, then removing the punctuation symbols and other symbols which may appear in the text, but which are considered to be unnecessary, converting numbers to textual data, deciding for the need of date data, etc. After finishing the above mentioned tasks, other steps which are related to linguistic characteristics are considered. The most important ones are stemming, lemmatization and stop-words removal.

In order to reduce the morphological variation of a word and to get its standard raw form, the processes of lemmatization and stemming are used. By eliminating suffixes, infixes and prefixes, stemming will return the stem of a word, which is not always identical to the morphological root of the word. This may lead to incorrect spelling and meaning. On the other hand, by the process of lemmatization, the lemma that we get is a valid and meaningful standard form.

While the stemmer operates without any knowledge of the context, lemmatization takes into consideration the context, and for each different context of the word, the Part of Speech (POS) tag should be identified. Sometimes one word has several different lemmas, depending on the POS tag that we have mentioned. Some examples of different forms that a word can get after lemmatization and stemming are demonstrated in the following table. The words which either have wrong meaning or root form are specified in red. The groups of verbs that have the same (and correct) lemma and stem after corresponding processes are highlighted in blue color.

WORD	POS TAG	LEMMA	STEM
caring	verb	care	car
stripes	verb	strip	strip
	noun	stripe	strip
felt	verb	feel	felt
walking	verb	walk	walk

running eating …		run eat	run eat

Table 6. Word forms after Lemmatization and Stemming

The process of stemming that we have conducted for Albanian language is based on the implementation of Dine's project (2018). We have taken into consideration the following suffixes: -tar, -tare, -shme, -ishte, -shëm, -iste, -mjet, -sore, -tare, -uar, -ore, -eve, -ave, -ake, -jes, -ste, -nte, -ist, -yer, -et, -ur, -te, -re, -je, -ar, -en, -it, -ve, -es/ës-, -in, -im, -on, -ne, -a, -u, -e, -i, -esë, -si, -ri, -or, -tor, -ak, -as, -an, -it, -iot. Here are some examples that contain the mentioned suffixes: treg**tar**, mrekullue**shëm**, lul**ishte**, mal**ore**, presp**ar**, këshill**tare**, pun**ishte**, pronë**si**, punë**tor**, fush**ave**, ar**tar**, etc. However, by applying the process of stemming, there are a lot of cases that are changing the part of speech. For instance: "bukuri" which is a noun, after eliminating the suffix –i, it becomes "bukur" and it may appear either like adjective ("këngë e bukur") or like adverb ("shkruan bukur"). Another issue in this and other similar cases is the considered suffixes (more than one) that may be applied in the same words. In this case, we have the suffixes "–i" and "-ri", which may be applied in the word "bukuri". We have already discussed the first case, but if we remove the second suffix (-ri) then the resulted stem will be "buku" which does not have any meaning in Albanian. So, the cases like these are the ones that affect the loosing of aspect-terms or sentiments, and successively impact the precision of sentiment polarity.

Lemmatization as a process is computationally more expensive then stemming, because it includes look-up tables and other resources. That is why in large datasets there can be an issue with performance. Stemming is preferred in these cases. However in some circumstances where the accuracy is the most important criteria and the dataset is huge, the process of lemmatization is a must.

The next process of removal of stop-words is another important task for sentiment analysis. Since a lot of techniques that are used in the sentiment analysis tasks depend on word count, the removal of words that do not contain any sentiment value is very important, because they are usually in a big number of appearances. Stop-words removal is conducted in the phase of preprocessing. There are several key points that should be considered in the process of building the list of stop-words (the list that we have built for Albanian language which contains hundreds of words is described in the experimental section).

First, an issue to be considered is their appearance in the multiple word aspect terms, which as a result splits the terms in two different aspects. This is described in more details in the section of language challenges. Second, we should be very careful not to include the negation terms, even though at first sight it may seem that they must be included in the list of stop-words. The reason for this is that they contain sentiment, which in our case is the most important. So, if we exclude the negation terms from a piece of text, the negative sentiment of the opinion would be switched to positive, which of course will be a non-valid result. And the last issue is that sometimes the high removal of stop-words can result in damaging or losing an opinion. For instance, the famous quote “Të jesh, a të mos jesh, kjo është pyetja!” (in English “To be, or not to be, that is the question!”), will totally be distracted after eliminating the words: të, jesh, a, kjo, është (which are included in the list of stop-words), and only the word “pyetja” remains from the whole opinion, which in this case does not contain any sentiment.

Even though the sentiment in this case would be neutral, it is nonetheless not the same with not having any sentiment, because the opinion and its core meaning are destroyed. All of the stop-words are part of curated lists, which are defined or used by the researchers. Since it is a subjective selection of the words, it is of great importance these lists to be publicly available, so that any researcher will be aware of what is neglected and how much it affects the final results.

Another piece of text that we have decided to remove in this phase is the discussable sentiment, expressed by emoticons and emojis. People use them in a great manner in social media, namely in informal language usage. Both emojis and emoticons are used with the aim to somehow manifest the feeling expressions and body language in a textual form. Most of the time, they are used together with texts, and with that the opinion holders aim to emphasize and clarify their opinion, namely the sentiment for a specific issue. Due to the importance of emojis, emoticons and reactions in sentiment analysis, the research has come to a point that they are considered to create a new field of study, named as emoji-based sentiment analysis.

There is a study conducted by Yoo and Rayz (2021) in which it is analyzed and compared the way of usage of emojis and their affect in the accuracy of sentiment polarity. As they state the usage of emoticons has started to be replaced by emojis, and another point is that there are a lot of emojis that cannot be represented by emoticons, like animals' or objects' emojis. Focusing on them, there are two main different ways that they are used in an opinion. First, there can be used only one emoji per sentence, usually at the end; second, there may appear several different emojis in a single sentence and third, there may be used more than one identical emoji, one after another. The most neglected case is the second one, because if they are expressed different sentiments, it cannot be concluded the final and accurate one. On the other hand, comparing the two other cases, it is detected a huge difference in the accuracy of the sentiment, because the more used emojis, the better expressed feeling/sentiment. This can be shown in the next table, which is generated by the above mentioned authors' research.

Token	Sentiment	Ratio
😭😭😭	neg : pos	154.4 : 1.0
😭😭	neg : pos	137.8 : 1.0
😱	neg : pos	135.3 : 1.0
🔥🔥	neg : pos	88.4 : 1.0
💔	neg : pos	86.7 : 1.0
😂😂😂	pos : neg	73.4 : 1.0
😘	pos : neg	61.4 : 1.0
😂😂	pos : neg	52.8 : 1.0
😭	neg : pos	51.7 : 1.0
😢	neg : pos	47.8 : 1.0

Table 7. Sentiment ratios from the usage of emojis. From "Understanding Emojis for Sentiment Analysis", by Yoo and Rayz, 2021.

As it may be easily noticed the ratio of sentiment is almost 3 times higher in the case when there are used 3 same emojis, in contrast to one. However, the difference is not so big in the cases when we have 2 or 3 emojis. Most of the emojis have a single meaning like "lol" for the laughing emoji 😂, 😠 for angry, etc. Their meaning may change very slightly and slowly through the years. However, in some situations they may get several different meanings for a short period of time. For instance in the pandemic time, for only several months the emojis such as: feeling hot, measuring the temperature, wearing a mask, sneezing, etc., got a meaning of “pandemic”, “mask”, “covid-19”, “restrictions”, “infected”, etc. In some cases all of these emojis are “translated” with words and considered them as textual data in the lexicons.

In general, emoji-based sentiment analysis is implemented with lexicon-based approaches (Hogenboom, et al., 2013), (Vashisht and Jailia, 2020) and it is applied for general sentiment detection, which considers that there is only one aspect in the whole piece of text. Since we are working on unsupervised data and analyzing the sentiments for each aspect of the text, by considering the mentioned limitations from the current literature and the fact that in our data they were not used too much emojis and emoticons, we decided to exclude them. Both of them were excluded by removing the quotation marks and symbols.

Noise Removal

The third step of text data processing is noise removal. In this process the conducted activities depend on the request that should be fulfilled and on the dataset. Noise removal includes the removing of special characters, digits and pieces of text that can affect the process of text analysis. As a step of preprocessing phase, it can also be dependent on the domain and resources. For example in tweets among all special characters, hashtags must remain, because they hold concepts that can specify a tweet. HTML tags, people’s tags, greetings etc. are also considered noise. All of these can confuse the algorithm used for text-processing. Although noise removal is an important step and must be conducted, sometimes it may also result in inconsistent outputs, since we can have lost data.

Language Challenges

In one of our previous work (Axhiu, 2019) we have done a comparative evaluation among Albanian and English languages that are part of the Indo-European family. Regardless of the fact that they are part of the same language family; they still have a lot of phonetic, semantic and grammatical structure differences. In this scope we have identified several language challenges that must be considered in the tasks of aspect-based sentiment analysis. The main ones are: inflections, negation, homonyms, dialects, irony, sarcasm and stop-words' presence within aspects.

Inflections

One of the first issues that we have come to understand (2019) will influence the implementation of the system is the fact that Albanian language compared to English is a highly inflectional language. Due to the high varieties of word forms, previous researches have suggested that the processes of stemming and lemmatization are a must in the process of data preprocessing. Another influencing factor for the Albanian language is the flexible word order. Opposite to this, in the English language the word order is fixed and any changes may affect the meaning of a sentence, as well as its grammatical structure.

Negation

Another challenge which considerably affects the accuracy of the sentiment classification is *negation*. There are a lot of differences regarding double and multiple negation between these two languages. In English double negation refers to two logical negations expressed by two negative markers, while multiple negation or negative concord refers to two or more negative markers contributing to one logical negation. On the other hand, since Albanian is a negative concord language, double negation is used to refer to negation realized by two negative elements in a sentence, while multiple negation refers to using more than two negative elements in a sentence. (Bekteshi, 2020). For example,

- *“**S’vjen asnjeri**” - "**No** one comes**",***
- *“**Askush askujt s’i** ka borxh” – “**Nobody** owes anyone”*
- *“**Asnjëherë asgjë nuk** I kam thënë.” - "I **never** said anything to him."*

Another issue is that negation is neither structured, nor expressed in the same way in English and Albanian and the negative markers take a different position. As a case we consider the following: in primary verbal negation the negative marker “**not”** corresponds with the Albanian negative markers **nuk/s’**, while in secondary verbal negation it mostly corresponds with **nuk, mos, pa** and **jo.** This denotes that there are huge differences in terms of Negative Concord; English sentences are structured with single negative markers, while Albanian sentences allow two or three negative markers or indefinites in their structure. All of these structural and expressional differences harden the work of the machines, because the real meaning of the sentences, namely the real sentiment, becomes more difficult to be detected.

Homonyms

The other difficulty that we encountered in the task of aspect-term extraction is the appearance of homonyms. We have detected three different cases when homonyms may lead to wrong data extraction. The first case is the more usual and general one, because even though they appear as nouns they affect the differentiation of aspect terms. Namely, while two or more different terms exist, the automatic extraction will count them as one. For instance, the word: “bari” has three meanings: shepherd, grass and medicament. In the extraction process they will be considered as one aspect-term, which definitely will lead to incorrect results.

The second problematic situation is when homonyms appear as different parts of speech. In our case we have homonyms as nouns and verbs. For example the word “vesh” – means “ear” or “put on” (some clothes). When it is used as noun, the word will be an aspect-term, but if it is used as verb, it shouldn’t be categorized as a term. But, the system will most probably put them in one group.

The third moment of miscategorization is when a homonym may appear as an aspect or as a stop-word. We have an example when one of the homonyms is a noun and the other a conjunction (e.g.: “dhe”- it means soil-noun or and-conjunction). In the first case it should be considered as an aspect. However, if we have added the word in the list of stop-words, then it will be ignored, and in this situation we lose an aspect. On the other hand, if we exclude it from the list of stop-words than we will have a situation where it will be ranked as one of the top aspect-terms, since it is one of the mostly used conjunctions in Albanian language.

Dialects and Regional Speech Varieties

When taking into account the informal language, the implementation of ABSA for Albanian language is even more challenging, since it has two main dialects and tens of regional speech varieties. The chances are really high that some of the aspects (which may be with high frequency) may not be extracted or, may be considered as different aspects due to the different spelling or even completely different words' usage (e.g. qumësht-tomël, rrugë-udhë, atëror-atnor, vatër-votër, grua-grue, etc.).

There are some cases where even a small difference that occurs in regional speech varieties may lose the whole meaning of the sentence. For example, in our data set we have some comments where as a result of a misuse of a single letter, the context of word changes completely. A specific example we have from the Polog region where letter "I" is used instead of letter "Y". So, we have a case where the word "dy", which means "two", in that region is written as "di", which in standard language refers to the verb "know". Because of this misusage the whole sentence loses the meaning and it might be neglected. There exist some researches which focus on the issues of multi-dialects, regional speech varieties and other cases of informal language in ABSA, and they conclude that in order to successfully perform the tasks of ABSA, the system should be language-specific.

Irony and Sarcasm

In the online world most of the time people express their thoughts more freely than in eye-to eye communication. This freedom affects the usage of irony and sarcasm as well. So, in case the author wants to emphasize a wrong situation, he/she uses irony. While if he/she wants to criticize something/someone than he would use sarcastic expressions. The recognition of both of them is difficult even for people. Since most of the time it looks like there are expressed opposite opinions, their handling by natural language processing techniques becomes even harder. In this case the probability of having an opposite sentiment polarity is very high.

For example,

- *"Të gjithë duken normal, deri sa i njofton!"* (In English: "Everyone looks normal, until you meet (them)!" – Here according to the structure and the initial meaning the statement is positive, however the real meaning is negative (none is normal).

Sentences with Ambiguous Meaning

There exist a lot of ambiguous sentences that may have different sentiment polarities in different cases (mainly depending from the opinion holder). For example the opinion "I love kids" may have a positive or negative sentiment, depending if the author is a child abuser or not. Similar to that is the following opinion: "I love sweets, and I can eat them at every opportunity!" which again may have positive or negative sentiment, depending on whether the opinion holder is suffering from diabetes or not.

There exist another group of opinions which according to the structure seem to be affirmative but the meaning is negative. For example,

- *"Ku dihet?"= "Nuk dihet",*
- *"Ku di unë!" = S'di unë.",*
- *"E kush pyet për mua?" ="Asnjë s'pyet për mua."*

The same problem appears when we have the opposite way of expressing the opinions; structurally the opinion looks like it is negative, however it holds a positive sentiment. For example: *"E ç'nuk kishte aty!"= " Kishte nga të gjitha aty!"*

The last group of ambiguous sentences that we have detected uses some particles such as: “sikur”, “mbase” and “ndoshta”, which the positive or negative meaning converts to neutral. For example:

- *“**Sikur nuk** ndjehem…”*
- *“**Mbase nuk** vij as unë.”*

Stop-Words within Aspect Terms

As we have mentioned in another section the list of stop-words consists of words that are used very often within a text, but do not hold any sentiment. That includes propositions, conjunctions, etc. However, sometimes they may have a negative instead of a positive effect. This may happen in two situations: first in the case that one stop-word represents an aspect (the case that we have mentioned in the above sections) and second when they appear in multi-word aspects. Here, the problem becomes greater, because by excluding them, we separate the aspect term, and consequently they would be separated as two different terms. For instance in the cases of “këpucë për fëmijë”, “sapun për tesha”, “baltë e kuqe”, etc., where the words “për” and “e “are already in the list of stop-words, their consideration can be destructive because the structure of aspect terms will be divided and most probably they will be considered as two different aspect terms (e.g. këpucë and fëmijë, sapun and tesha).

Part III

Experimental Results

Text-Processing Steps in Practice

The corpus that we have created for experimental needs, holds Facebook comments, which are extracted (by Facepager) from one of the most followed Albanian portals –Tetova1. The comments are from a one year period and they correspond to several domains. After analyzing the Albanian ecosystem for their opinion expression in the online world we concluded that the main topics are politics, restaurants, sports and showbiz. This can be noticed in world cloud that is shown in the following figure. We suppose that a factor that could affect this result is the pandemic; since due to restrictions, people have discussed mostly the things they have missed, among which were restaurants. So, they have re-reviewed the places they have been.

Figure 13. Word Cloud for the Albanian ecosystem in online world

The first step was to fetch the comments from Facebook posts, and we did this by using the tool FacePager. The parameters that the tool contains for each comment are as follows: id, parent id, level, object id, object type, query status, query time, query type, name, message, type, metadata type, talking about count, likes, likes, likes count, shares count, comments count, created time and updated time. For our needs we are using only ID, parent ID, level, object ID, message and created time. The parent ID and object ID help us to know if the comments are related to the same or different post, while the "created time" will help us analyze the trend of the sentiment for a specific entity in a specified period of time. The initial number of rows in the fetched dataset was above 60000. However in the process of cleansing the data this number was more or less halved. In the cleansing of our data we have considered the issues such as: blank comments, tags, URL and non-textual data. In the following figure there is an example of how the data are shown at this stage of the work.

One of the biggest issues which considerably cut the number of data was the missing data, namely the blank comments. According to the tool's specifics the blank comments may appear in the case of deletion of comments. However the number of these cases has been up to 25000, that is why we are assuming that besides the deleted comments there is another factor which contributes to this situation. We give a high probability to the cases when there are no comments in the posts, and they may be counted as "empty comments". Regarding the issue of the tags, there have been also a lot of comments which contained only tagged people, which in our case aren't helpful at all, because we cannot extract a sentiment polarity from that kind of data. Due to the same reason we have also cleansed our data set from comments which contained only URLs (mainly URLs of some other fan pages, websites, or posts). At last we have cleaned the non-textual comments which contained only numbers or only punctuation marks. After cleansing the dataset the final number of comments has turned out to be around 31.000 comments. We should further preprocess and adapt the dataset according to the input requirements and specifics of the system.

	A	B	C	D	E	F	G	H	I	J
295	135883	35656	2	1213905332048066_	data	fetched (200)	42963.4404205324	Facebook:<post>/comments		Ai e sheh ate qe nuk e shohin te verbuarit nga nacionaliz
296	135884	35656	2	1213905332048066_	data	fetched (200)	42963.4404205324	Facebook:<post>/comments		pe pranon a bravo qetnike
297	135885	35656	2	103882726383671_1	offcut	fetched (200)	42963.4404205324	Facebook:<post>/comments		
298	35657	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		
299	135886	35657	2		empty	fetched (200)	42963.4404310185	Facebook:<post>/comments		
300	35658	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		
301	135887	35658	2		empty	fetched (200)	42963.4404324653	Facebook:<post>/comments		
302	35659	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		
303	135888	35659	2		empty	fetched (200)	42963.4404366319	Facebook:<post>/comments		
304	35660	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		Zjarr në shkollën fillore Kirili dhe Metodi në Tetovë. Fatmirësis
305	135889	35660	2	1213520375419895_	data	fetched (200)	42963.4404416898	Facebook:<post>/comments		Uka Aliti apelo se shkollat na duhen...ne të djegin diçka qe esh
306	135890	35660	2	1213520375419895_	data	fetched (200)	42963.4404416898	Facebook:<post>/comments		Shqiptaret kurri nuk de ken mundesi me uba njerzi me ket injor
307	135891	35660	2	1213520375419895_	data	fetched (200)	42963.4404416898	Facebook:<post>/comments		Ishlla sa gjek □
308	135892	35660	2	1213520375419895_	data	fetched (200)	42963.4404416898	Facebook:<post>/comments		Lili bulica luti bulica
309	135893	35660	2	103882726383671_1	offcut	fetched (200)	42963.4404416898	Facebook:<post>/comments		
310	35661	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		
311	135894	35661	2	121329206544272б_	data	fetched (200)	42963.44044875	Facebook:<post>/comments		Alahu esheroft
312	135895	35661	2	103882726383671_1	offcut	fetched (200)	42963.44044875	Facebook:<post>/comments		
313	35662	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		
314	135896	35662	2	1213287718776494_	data	fetched (200)	42963.4404510995	Facebook:<post>/comments		Genc Ahmeti
315	135897	35662	2	1213287718776494_	data	fetched (200)	42963.4404510995	Facebook:<post>/comments		A Haqim Farizi
316	135898	35662	2	103882726383671_1	offcut	fetched (200)	42963.4404510995	Facebook:<post>/comments		
317	35663	34737	1	103882726383671_1	data	fetched (200)	42963.0100773611	Facebook:<page>/posts		
318	135899	35663	2	1213379798767286_	data	fetched (200)	42963.4404571296	Facebook:<post>/comments		ArsimMurati GzimAmiti □□□□□□□

Figure 14. Example of the fetched data

Beside these data in the system that our work is based on, there is also a need for inputs such as STOP-words and seed words. We have already built the list of STOP-words for Albanian language (currently containing around 200 words, but may be extended additionally) and the topic seed words and polarity seed words, on which our future work regarding the data will be focused. The first APPENDIX shows the list of the stop-words that we have created.

Dataset and Preprocessing For Evaluation

The corpus that we are using for experimental phase contains restaurant reviews which are collected mainly from Facebook fun pages. This testing dataset is created manually since to our best of knowledge there is no publicly available dataset with reviews. The data collection is done by the free tool – Facepager. The reviews are 5 to 180 characters in length. Before continuing to the processing phase, first a data cleansing must be done. This included the tags, unused symbols and emoticons removal. There were some situations where we had to interpret and write the meaning of the symbols, like the currency symbols. Once we had the cleansed data, the next phase was tokenization, where we decided to do a unigram tokenization for aspect extraction. The complete flow of data preprocessing is shown in the next picture.

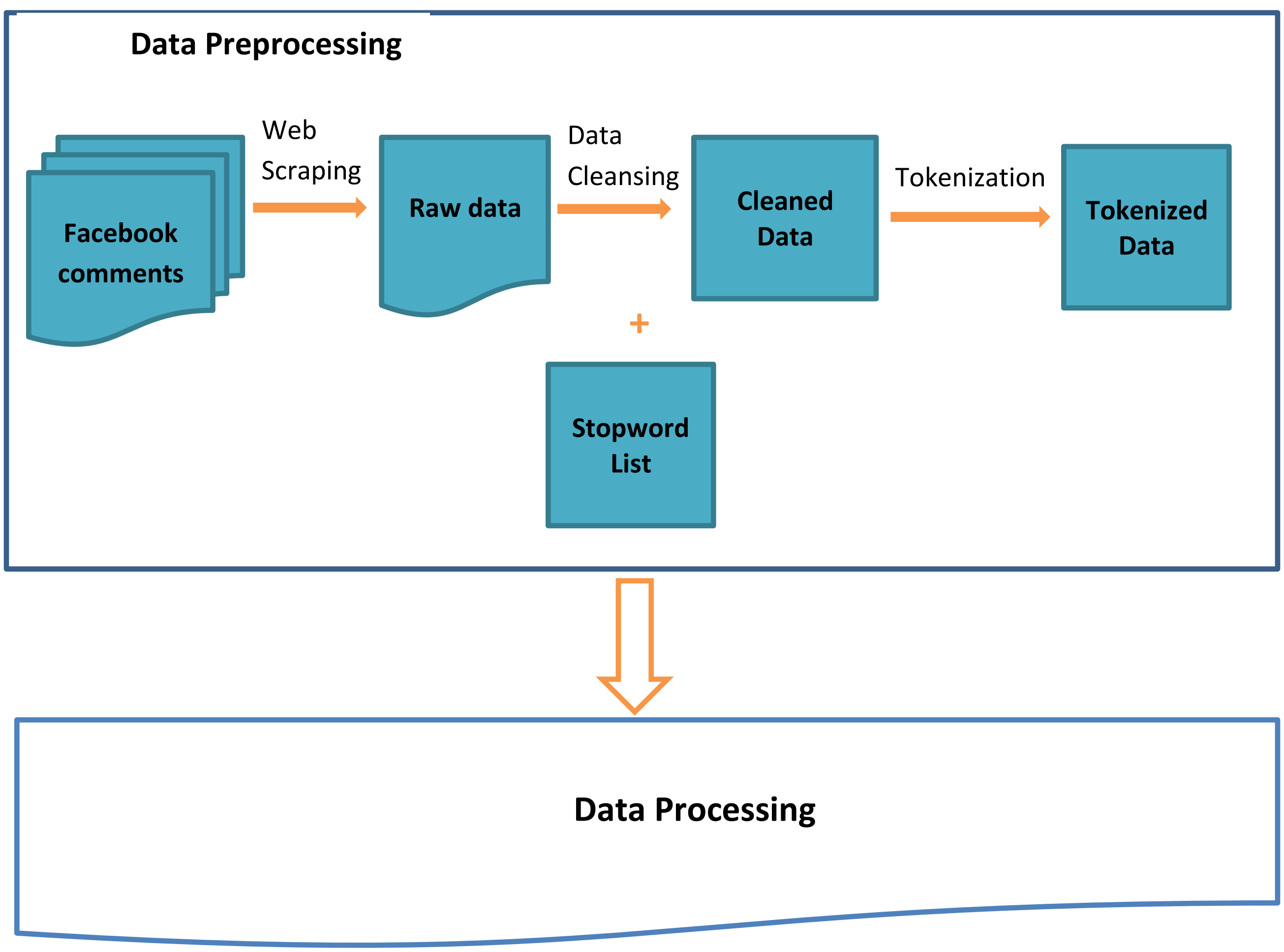

Figure 15. Data Preprocessing

In this phase we have also manually built a list of stop-words, which consists of the most used words which do not contain any meaning in the context of sentiment. Parts of this list are mainly conjunctions, prepositions, etc. It contains about 200 words, which in the process of aspect and sentiment extraction are excluded. The cleansed dataset which has gone through all the above mentioned processes came out to be as in the following picture, where we have taken into consideration only the reviews, without any other metadata of the reviews.

REVIEWS
Jeni numër një
Këtu qenka mire me shku.
Vend ideal për të gjithë!
A do ma jepni be veç një kajmak prej ne saraj do vij
A do shkojm ta hamë një darkë se shumë mirë e paskan bë
Ai moment kur të shkon mendja për kajmak
Aj qe e viziton nje her nuk e harron per jet
Ambient i ngrohte në dite të ftohte
Ambienti fantazi mikpritja ne nivel te lart pastertia gjithashtu po mos flasim per kajmakun
Aty asnjehere ska qene problem cfare do hame, po kush do paguaje.
Aty ku freskia është specialitet ,provoeni skarë të përzier
Aty ku natyra e bukur që të rrethon,shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura,të bëjnë të ndihesh më mirë se kudo tjetër.
Aty ku ne çdo koh ke shije dhe kualitet ....
Beni gati nja 10kg kajmak
Bravo i qofte pronarit
Bravo mrekulli

Bravo pronarit
Bravo! Vend i shkëlqyer dhe stafi i mire
Bukur
Bukur shumë ...e pershendes gjellberesin
Bukur
Bukuri e jashtezakonshme
Eshte pamje mahnitse sot me kte boren qe zbukuron cdo gje,ka lezet ne kohen e vet
File Troft për një vikend fantastik
Krejt eshte mir OK, vec mikepritja eshte 0 aty keni gabu
Pa koment
Pamje fantastike
C'do stinë në ktë ambijent është perfekt.
Kam ngrene njiher, sbonte, shume i bute ishte, se mora vesh a hongra a jo.
Cdo ushqim qe ju pergadisni e beni ne menyre perfekte.
Fileto viçi në gurë. Për një drekë me shumë shije dhe të shëndetshme, kjo është receta e duhur.

Table 8. Cleansed Data

In order to calculate the precision and accuracy of the used model, we have additionally annotated a smaller dataset which we have used for testing. First, we have extracted the aspects and then labeled their sentiments with one of the possible values: negative, neutral or positive polarity. The manual extraction and labeling are described in the following section.

Manual Aspect-Term Extraction and Sentiment Polarity

Before testing the model, in order to evaluate its accuracy we have to manually extract the terms and classify them into three aspect classes such as: food, service and ambiance, as well as to label the polarity of each aspect that may take place in the reviews. For this aim we have considered a smaller dataset that we have worked on. In the process of extraction and labeling we analyzed the accuracy of human annotators. The key point behind this analysis is to estimate the quality manual labeling and the quality of used classification models, by using the same measures for evaluation. So, the performance of the classification model is directly affected by the quality of the labeled data. This can be achieved by the agreement among the human annotators. Determining the expressed sentiment in a review depends on the subjective decision of the annotators. It is not rare that they disagree among themselves, and also they are not always consistent in their own previous statements.

There are several causes for disagreements between the annotators, such as: unfamiliar tasks and domains (e.g. to find the sentiment about the future Bitcoin movement), the differentiation of vocabularies in different domains (e.g. educational vs. financial issues), change of topics during the discussion or even the bad quality of the annotators' work. During the analysis, all the above mentioned issues are observed, and the gap in the process is identified. After that a decision is made about the new annotation process that will be used in the future.

In our case in the process of manual labeling of Facebook comments, a part of the comments was doubled, so that they may be labeled twice, either by the same person or by two different persons. In this case the same annotator can label the sentiments with two different values (depending on the confusion that the annotator would have). Regarding the agreements that we have mentioned, from the repeated annotations of the same person, a self-agreement is generated, while from the several annotations by two or more annotators, the inter-annotator agreement is generated. The result from a self-agreement is the identification of the low quality human annotators, while from the inter-annotator agreement we can estimate the objective difficulties of the tasks (the ones that are mentioned above). For the aspect term extraction, their classification in three classes (food, ambiance and service) and labeling the sentiment polarity for each comment/aspect we involved two annotators, who are native speakers of Albanian language.

Manual aspect-term extraction and their classification in aspect classes

In this process both of the annotators have extracted the aspect-terms for each sentence and then classified each of them in three aspect classes: food, service and ambiance. The manual extraction of the aspects-terms for the given Facebook comments is demonstrated in the table below. Both annotators have almost totally agreed about these extractions. The debatable cases have been mainly the reviews which do not have a clear aspect. As a case we have the comment “Këtu qenka mire me shku.” where the annotators have discussed if as an aspect may be considered the place- “vend”, because the word “këtu” (here) indicates a location. Due to the opposing views of the annotators the specific example and the ones similar to this, is considered to be a review with no aspects. Even if the “place” is considered as a hidden aspect, it is still too general which refers to a restaurant, but it does not correspond to any of the pre-defined categories.

REVIEWS	ASPECT-TERMS
Jeni numër një	/
Këtu qenka mire me shku.	/
Vend ideal për të gjithë!	vend
A do ma jepni be veç një kajmak prej ne saraj do vij	kajmak
A do shkojm ta hamë një darkë se shumë mirë e paskan bë	darkë
Ai moment kur të shkon mendja për kajmak	moment mendja kajmak
Ai qe e viziton nje her nuk e harron per jet	
Ambient i ngrohte në dite të ftohte	ambient dite
Ambienti fantazi mikpritja ne nivel te lart pastertia gjithashtu po mos flasim per kajmakun	ambient mikpritja pastertia kajmak
Aty asnjehere ska qene problem cfare do hame, po kush do paguaje.	problem
Aty ku freskia është specialitet ,provoeni skarë të përzier	freskia specialitet skare
Aty ku natyra e bukur që të rrethon,shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura,të bëjnë të ndihesh më mirë se kudo tjetër.	natyra sherbimi kualiteti shija ushqim
Aty ku ne çdo koh ke shije dhe kualitet	shije kualitet
Beni gati nja 10kg kajmak	kajmak

Bravo i qofte pronarit	pronar
Bravo mrekulli	
Bravo pronarit	pronar
Bravo! Vend i shkëlqyer dhe stafi i mire	vend staf
Bukur	
Bukur shum ...klm e pershendes gjellbersin	gjellberes
Bukur	
Bukuri e jashtzakonshme	bukuri
Bukuri e ralle respekt per pronarin	Bukuri pronar
Eshte pamje mahnitse sot me kte boren qe zbukuron cdo gje,ka lezet ne kohen e vet	pamje bore kohe
File troft për një vikend fantastik	file trofte vikend
Krejt eshte mir ok, vec mikepritja eshte 0 aty keni gabu	mikpritja
Pa koment	koment
Pamje fantastike	pamje
C’do stinë në ktë ambient është perfekt.	stinë ambient
Kam ngrene njiher, sbonte, shume i bute ishte, se mora vesh a hongra a jo.	(ushqim)
Cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	ushqim
Fileto viçi në gurr. Për një drekë me shumë shije dhe të shëndetshme, kjo është receta e duhur.	fileto viçi gurr drekë shije

receta

Table 9. Manual extraction of aspect-terms

As it may be noted from the table, for some comments even though there is no concrete aspect-term, the opinions still hold a positive, negative or neutral sentiment polarity. Here are some of the examples:

1. **“Jeni numër një”** (in English: “You are number one”) – even though there is a positive opinion, the existence of aspect is discussible.
2. **“Ai që e viziton një herë, nuk e harron përjetë”** (in English: “The one who visits it once, will never forget it”) – in this case the polarity of the opinion may be debatable, since it may be interpreted either in a positive or negative context. However, again there is no aspect-term.

In most of the comments, the presence of multiple aspect-terms, which holds different polarities, demonstrates once again the need of aspect-based sentiment analysis. The next stage was to classify the extracted aspect-terms in three aspect classes; in our case food, service and ambiance. In this task the annotators encountered different thoughts and evaluations, and at this point the inter-annotator agreement was needed. The following figure shows the manual classification of the restaurant reviews done by the first annotator. Depending on the detected aspect terms they are classified into three aforementioned aspect classes.

REVIEWS	FOOD	SERVICE	AMBIANCE
jeni numër një	1	1	1
këtu qenka mire me shku.	1	1	1
vend ideal për të gjithë!	1	1	1
a do ma jepni be veç një kajmak prej ne saraj do vij	1	0	0
a do shkojm ta hamë një darkë se shumë mirë e paskan bë	1	0	0
ai moment kur të shkon mendja për kajmak	1	0	0
aj qe e viziton nje her nuk e harron per jet	1	1	1
ambient i ngrohte në dite të ftohte	0	0	1
ambienti fantazi mikpritja ne nivel te lart pastertia gjithashtu po mos flasim per kajmakun	1	1	1
aty asnjehere ska qene problem cfare do hame, po kush do paguaje.	1	0	0
aty ku freskia është specialitet ,provoeni skarë të përzier	1	0	0
aty ku natyra e bukur që të rrethon,shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura,të bëjnë të ndihesh më mirë se kudo tjetër.	1	1	1
aty ku ne çdo koh ke shije dhe kualitet	1	0	0
beni gati nja 10kg kajmak	1	0	0
bravo i qofte pronarit	0	1	0
bravo mrekulli	1	1	1
bravo pronarit	0	1	0
bravo! Vend i shkëlqyer dhe stafi i mire	0	1	1
bukur	0	0	1
bukur shumë ...e pershendes gjellberesin	0	1	1
bukur	0	0	0
bukuri e jashtezakonshme	0	0	1
Eshte pamje mahnitse sot me kte boren qe zbukuron cdo	0	0	1

gje,ka lezet ne kohen e vet			
File Troft për një vikend fantastik	1	0	0
Krejt eshte mir OK, vec mikepritja eshte 0 aty keni gabu	0	1	0
Pa koment	0	0	0
pamje fantastike	0	0	1
c'do stinë në ktë ambijent është perfekt.	0	0	1
Kam ngrene njiher, sbonte, shume i bute ishte, se mora vesh a hongra a jo.	1	0	0
cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	1	0	0
Fileto viçi në gurr. Për një drekë me shumë shije dhe të shëndetshme, kjo është receta e duhur.	1	0	0

Table 10. Classification of comments in three aspect classes- Classified by the first annotator

Considering the aspect-terms, with 1s we denote their presence in a specific class and with 0s their absence in that class.

1. First to be discussed are the reviews with one or more aspect terms. Here are some examples of classification which differ in the number of aspect-terms:

- **"Ambient I ngrohtë, në ditë të ftohtë"** – this comment contains only one aspect-term – **"ambient"** and it corresponds to the aspect class of ambiance.
- **"Bravo! Vend i shkëlqyer dhe stafi i mirë" –** in this review there are two aspects "**vend"** and "**stafi",** so the annotation is [0,1,1] for food, service and ambiance correspondingly.
- **"Aty ku natyra është e bukur që të rrethon, shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura, të bëjnë të ndihesh më mirë se kudo tjetër."** – In this case there are three aspect-terms which correspond to three

classes. “Natyra” in the class of ambiance, “sherbimi” in the class of service and “ushqimet” in the class of food.

2. The most specific cases are the one that do not contain aspect-terms. The annotation of these cases depends from the annotators’ observations. For instance:

- **“Jeni numër një” –** this review has been categorized in all three classes, because the annotator has considered that it is a positive review for all three aspect classes.
- **“Ai që e viziton një herë, nuk e harron për jetë” –** the same evaluation is done here as well.
- On the other hand, a similar sentence, which does not contain any aspect, **“bravo, mrekulli!”,** by the same annotator is not categorized in the predefined aspect classes (the annotation is [0 0 0]).

3. Another specific and confusing case is with the indirect aspect-terms. For instance:

- **“Vend ideal për të gjithë” –** some of the annotators treat this comment similarly to the previous example, and consider that it is valid for all three categories. However, some other annotators correlate the word “vend” with the ambiance aspect. This happens because the word “vend” (in English “place”) can either be considered just as a location, or refer to the ambiance.

4. Regarding the intentional duplication of the reviews, which may have different classification by the same person, we have taken a single-word sentence as an example:

“Bukur!”.

In the first case it is categorized in the ambiance aspect class, because according to the evaluator, the sentence is referred to the place/location, namely to the ambiance. While in the second case it is not categorized in any of the aspect classes, because it is not considered to contain a hidden aspect-term. These are the situations when a self-agreement of the annotator is needed.

5. Another specific case is when there is a doubt in the number of aspects. In the review **"Bukur shumë…e pershendes gjellbërësin.",** the evaluator detects two aspects, one in the first part of the sentence **"bukur shumë",** where it is thought that the extracted aspect is "vend" and has classified it in the class of ambiance, while the second part of the sentence contains the aspect **"gjellbërës"** and it is put in the class of service.

In the next table the same Facebook comments in the domain of restaurants are classified again in three groups according to the same aspect-terms that we have previously discussed; this time from the viewpoint of the second person. In order to build the final list of all aspect-terms, we must consider the categorization that is done by each of the involved annotators.

REVIEWS	FOOD	SERVICE	AMBIANCE
jeni numër një	0	0	0
këtu qenka mire me shku.	0	0	0
vend ideal për të gjithë!	0	0	1
a do ma jepni be veç një kajmak prej ne saraj do vij	1	0	0
a do shkojm ta hamë një darkë se shumë mirë e paskan bë	1	0	0
ai moment kur të shkon mendja për kajmak	1	0	0
ai qe e viziton nje her nuk e harron per jet	0	0	0
ambient i ngrohte në dite të ftohte	0	0	1
ambienti fantazi mikpritja ne nivel te lart pastertia gjithashtu po mos flasim per kajmakun	1	1	1
aty asnjehere ska qene problem cfare do hame, po kush do paguaje.	1	0	0
aty ku freskia është specialitet ,provoeni skarë të përzier	1	0	0
aty ku natyra e bukur që të rrethon,shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura,të bëjnë të ndihesh më mirë se kudo tjetër.	1	1	1
aty ku ne çdo koh ke shije dhe kualitet	1	0	0
beni gati nja 10kg kajmak	1	0	0
bravo i qofte pronarit	0	1	0
bravo mrekulli	0	0	0
bravo pronarit	0	1	0
bravo! Vend i shkëlqyer dhe stafi i mire	0	1	1
bukur	0	0	1
bukur shum ...klm e pershendes gjellbersin	0	1	0
bukur	0	0	0
bukuri e jashtezakonshme	0	0	1

Eshte pamje mahnitse sot me kte boren qe zbukuron cdo gje,ka lezet ne kohen e vet	0	0	1
File Troft për një vikend fantastik	1	0	0
Krejt eshte mir OK, vec mikepritja eshte 0 aty keni gabu	0	1	0
Pa koment	0	0	0
pamje fantastike	0	0	1
c'do stinë në ktë ambijent është perfekt.	0	0	1
Kam ngrene njiher, sbonte, shume i bute ishte, se mora vesh a hongra a jo.	1	0	0
cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	1	0	0
Fileto viçi në gurr. Për një drekë me shumë shije dhe të shëndetshme, kjo është receta e duhur.	1	0	0

Table 11. Classification of comments in three aspect classes- Classified by the second annotator

Comparing the two tables of classification there are several points that are important to be mentioned.

1. The reviews with one or more visible aspect are classified totally the same by both annotators.
2. The biggest difference is in the reviews which do not contain aspects. As it may be noted in the following sentences:

- **"Jeni numër një.",**
- **"Këtu qenka mire me shku." And**
- **"Bravo! Mrekulli. "**

Unlike the first annotator, the second did not classify them in any of the aspect classes (food, service, ambient), as he/she considers that there are no aspect-terms in the reviews. Although there exists a sentiment polarity, the evaluator found that there is no reference point for polarity. In this case the annotation is [0 0 0].

3. Another case that is evaluated and classified differently is the example that contains an indirect aspect-term. For instance:

    - **“Vend ideal për të gjithë!”**

    As it is mentioned in the evaluation of the first annotator, there are two different dilemmas that appear. Some of the annotators consider that since the sentence is general and contains sentiment polarity, it applies in all three classes. While the second annotator considers that there is an indirect aspect and it should be classified (in this case “vend” corresponds to the aspect “ambiance”.

4. Comparing the work of both evaluators in the case of duplication of the reviews (for instance: **“Bukur!”),** surprisingly the categorization is the same, although there are a lot of different possibilities.

5. Taking into consideration the fifth point mentioned in the annotation by the first person, there is a difference in the evaluation by the second person.

    In the review **“Bukur shumë…e pershendes gjellbërësin.”,** the annotator thinks that there is only one aspect, and that is **“gjellëbërësi”.** According to this, the classification corresponds to the service aspect.

Taking into account all the different points that we have focused on in the work of the two annotators, we conclude that the number of annotators affects the aspect-term manual extraction and classification. First, the number of the aspect-terms that are extracted by the evaluators may differ. As we saw in the last mentioned example, the first annotator detected two aspect-terms, while the second annotator only one. In this case we have added both of the terms in the list. Second, the classification of the aspects, as we noticed in the discussed examples, can vary a lot. For both cases an inter-annotator agreement is a must, because it affects the accuracy of the results. The positive point is that the domain was not a problem for the evaluators, and that was not the reason for different ratings. The diverse extraction and classification of the aspect-terms was simply affected by their subjective evaluations.

Manual Sentiment labeling

After extracting the aspect-terms from the reviews and classifying them into predefined aspects, the next step is to label the sentiment polarity for each aspect-term. Each adjective that is related to the nouns we have taken into account, namely to the aspect-terms is considered as a sentiment-word. In the following table we have extracted the sentiments for each existing aspect.

REVIEWS	FOOD	SERVICE	AMBIANCE
jeni numër një			
këtu qenka mire me shku.			Mirë
vend ideal për të gjithë!			Ideal
a do ma jepni be veç një kajmak prej ne saraj do vij			
a do shkojm ta hamë një darkë se shumë mirë e paskan bë	shumë mirë		
ai moment kur të shkon mendja për kajmak			
ai qe e viziton nje her nuk e harron per jet			
ambient i ngrohte në dite të ftohte			ngrohtë
ambienti fantazi mikpritja ne nivel te larte pastertia gjithashtu po mos flasim per kajmakun		të lartë	fantazi
aty asnjehere ska qene problem cfare do hame, po kush do paguaje.			
Aty ku freskia është specialitet ,provoeni skarë të përzier	(specialitet)		
aty ku natyra e bukur që të rrethon,shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura,të bëjnë të ndihesh më mirë se kudo tjetër.	Shije kualitet mire	profesional	e bukur
aty ku ne çdo koh ke shije dhe kualitet	shije kualitet		
beni gati nja 10kg kajmak			
bravo i qofte pronarit		Bravo	
bravo mrekulli			mrekulli

bravo pronarit		Bravo	
bravo! Vend i shkëlqyer dhe stafi i mire		I mire	shkëlqyer
bukur	bukur	Bukur	bukur
bukur shumë ...klm e pershendes gjellbersin			bukur shumë
bukur			bukur
bukuri e jashtezakonshme			jashtëzakonshme
Eshte pamje mahnitse sot me kte boren qe zbukuron cdo gje,ka lezet ne kohen e vet			mahnitse
File Troft për një vikend fantastik	fantastik		
Krejt eshte mire OK, vec mikepritja eshte 0 aty keni gabu	mire ok	0 (zero)	
Pa koment			
pamje fantastike			fantastike
c'do stinë në ktë ambient është perfekt.			Perfekt
Kam ngrene njiher, sbënte, shume i bute ishte, se mora vesh a hongra a jo.	I butë		
cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	Perfekte		
Fileto viçi në gurr. Për një drekë me shumë shije dhe të shëndetshme, kjo është receta e duhur.	Shume shije shendetshme e duhur		

Table 12. Manual extraction of sentiments for each aspect term

In this process we have encountered in five different cases in the scope of the number of sentiment-words.

1. First, we have some reviews that do not contain any sentiment words for the aspects. This happens in the sentences where the existence of aspect-terms was debatable.

For instance:

- **Jeni numër një.**
- **Aty asnjehere ska qene problem cfare do hame, po kush do paguaje.**

Both reviews don't have any direct sentiment word, however as a whole they hold a sentiment polarity. It has to do with specific expressions that human beings understand. Idioms are also included in this group.

The first example ("You are number one!") is an expression that corresponds to each aspect.

The second example ("- There has never been a problem with what we will eat, but who will pay.") according to the second part of the sentence, the moment of paying, means that they have eaten the food, and then from the first part of the sentence we understand that it refers to the aspect of food.

2. Another issue appears again in the reviews that do not contain aspect-terms, where in this case we do have a sentiment word. For instance in the review **"Bukur"** , we are not sure where to classify it, since there is no aspect-term. However, the sentiment is "bukur". In our case as it is described in the previous section, in one instance we have considered that it refers to all three aspects while in the second annotation, that it refers to none of them. Having agreed with this, we have positioned the sentiment in the proper class.

3. The third case is when we have the aspect-terms but they are not correlated with sentiment words. Here are two reviews that represent this case.

    - **Ai momenti kur të shkon mendja për kajmak!**

- **Bëni gati nja 10kg kajmak.**

In the first sentence we have three aspect-terms (moment, mendja and kajmak), but there is no direct sentiment for any of them.

In the second sentence we have the aspect term “kajmak”, but again we don’t have a sentiment-word. These cases also have hidden meaning understandable for humans but very difficult for the machines.

4. The fourth case is when we have a sentiment for each detected aspect-term. Here are some examples:
   - **“Bravo! Vend I shkëlqyer dhe stafi I mirë**.

     Here for the term “**vend**” we have extracted the sentiment “**shkëlqyer**”, and for the aspect “**stafi**” we have the sentiment “i **mirë**”.
   - **“Çdo ushqim që ju përgadisni e bëni në mënyrë perfekte.”**

     In this example there is only one aspect-term, which is “**ushqim**” and its sentiment is “**perfekte**”.
5. The fifth case is the existence of more than one sentiment for a term. This can affect in the afterword phase, namely in sentiment polarity.
   - **“Aty ku natyra e bukur që të rrethon, shërbimi rakesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura, të bëjnë të ndihesh më rak se kudo tjetër.“**

Here for the term “natyra” which is classified in the aspect of ambiance, we have the sentiment “e bukur”. The term “sherbimi”, which corresponds to the “service” aspect, carries the sentiment “rakesional”. And the term “ushqimet” which takes place in the food aspect, carries two sentiments- “kualitet”, and “shija-shijshme”.

- **Fileto viçi në gurr. Për një rake me shumë shije dhe të shëndetshme, kjo është receta e duhur.**

In this review we have two aspects in the same class – “**rake**” and “**receta**”, which belong to the food aspect. However, similarly to the previous example we have more than one sentiment. The term “**drekë**” holds the sentiments “**shumë shije**” and “**shëndetshme**”, while the term “**receta**” has the sentiment “**e duhur**”.

These are specific cases which should be taken into account very carefully, because they may contain different polarities.

In the following table we have labeling of sentiments, which is a process after the extraction of sentiments. The polarity labeling is done for each sentiment of each aspect-term that a review contains. We have decided to work on the three possible polarities, namely the positive, negative and neutral polarity. Positive polarity is noted with 1, negative with -1 and neutral with 0. Again there are some specific cases that should be elaborated in details, since they will affect the accuracy of polarity. The polarity labeling is checked by the two annotators and they have agreed with the following polarities, which are shown in the table.

REVIEWS	FOOD	SERVICE	AMBIANCE
jeni numër një	1	1	1
këtu qenka mire me shku.			1
vend ideal për të gjithë!			1
a do ma jepni be veç një kajmak prej ne saraj do vij	1		
a do shkojm ta hamë një darkë se shumë mirë e paskan bë	1		
ai moment kur të shkon mendja për kajmak	1		
ai qe e viziton një herë nuk e harron për jetë	1	1	1
ambient i ngrohte në dite të ftohte			1
ambienti fantazi mikpritja ne nivel te lart pastertia gjithashtu po mos flasim per kajmakun	1	1	1
aty asnjehere ska qene problem cfare do hame, po kush do paguaje.	1		
aty ku freskia është specialitet ,provoeni skarë të përzier	1		
aty ku natyra e bukur që të rrethon,shërbimi profesional e mbi të gjitha kualiteti dhe shija që kanë ushqimet e përgatitura,të bëjnë të ndihesh më mirë se kudo tjetër.	1	1	1
aty ku ne çdo koh ke shije dhe kualitet	1		
beni gati nja 10kg kajmak	1		
bravo i qofte pronarit		1	
bravo mrekulli			1
bravo pronarit		1	
bravo! vend i shkëlqyer dhe stafi i mire		1	1
bukur			
bukur shum ...klm e pershendes gjellbersin		1	1
bukur			1
bukuri e jashtezakonshme			1

Është pamje mahnitse sot me kte boren qe zbukuron çdo gje,ka lezet ne kohen e vet			1
File Troft për një vikend fantastik	1		
Ushqimi eshte i mire, vec mikepritja eshte 0 aty keni gabu	1	-1	
Pa koment	0	0	0
pamje fantastike			1
c'do stinë në ktë ambijent është perfekt.			1
Kam ngrene njiher, sbonte, shume i bute ishte, se mora vesh a hongra a jo.	-1		
cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	1		
Fileto viçi në gurr. Për një drekë me shumë shije dhe të shëndetshme, kjo është receta e duhur.	1		
Ushqimi i shijshëm por pak i thatë	1/-1		

Table 13. Sentiment polarity labeling

In the scope of polarity of each sentiment we encountered several different cases regarding the polarity category and the hidden polarity in special expressions.

- First we considered the implicit reviews, where the automatic systems can barely detect. For instance:
    - **"Jeni numër një"** – This as a review has positive polarity, since as in every language, in Albanian as well the expression **"number one"** is used in a positive context.
    - **"Ai qe e viziton një herë nuk e harron për jetë"** – This is also an implicit sentence; because it is has not a clear sentiment and an aspect-term. When speaking of polarity, this and similar cases are debatable, because most annotators value it as a positive review (since it says **"you will never forget").** However, some other

annotators may consider it as negative sentiment, as negative things cannot be forgotten as well.

- Similar issue appears in the example **"bëni gati nja 10kg kajmak",** for which the annotators label it with a positive polarity, since the opinion holder expresses a need (in the part "beni gati"). However, since there is no direct sentiment, the model will categorize it as a neutral review.

- Other group holds the cases where there are explicit sentiments and their polarities are easier to be detected. For instance:
  - **"Është pamje mahnitse sot me kte boren qe zbukuron çdo gje,ka lezet ne kohen e vet."-** Here the polarity of the sentiment "mahnitëse" is easily understood as being positive.
  - **"Ushqimi është i mirë, veç mikepritja është 0 aty keni gabu!"** – in this case the sentiment polarity of "është i mirë" which refers to the food (ushqimi) is positive, while the sentiment polarity of the term "mikepritja" is negative.

3. Another cluster of reviews are the ones that hold definite sentiments but in the same aspect. Since they hold different polarities, the final polarity of the aspect is doubtful. This characteristic may be seen in the following example:
   - **"Ushqimi i shijshëm por pak i thatë"-** here the only term is "ushqimi" which corresponds to the class of Food. There are two different sentiments "I shijshëm" and "i thatë", and they hold opposite polarities. The first one is a positive polarity and the second one is negative. In these cases the machine categorizes the sentiment as neutral, since it considers both sentiments at the same time.

All of these different groups are a real challenge for the automatic aspect based sentiment analysis, since as we saw in the aforementioned examples, there are some cases in respect of which even the human annotator is in doubt. This happens especially when an idiom is used, which may not always be known for the annotator, and during the use of irony and sarcasms, which can be differently interpreted by the evaluators. In the following sections all of the tasks are conducted in an automatic manner.

Aspect Term Extraction

Because we want to realize this task in an unsupervised manner, we have used the topic modeling techniques. The two most-used topic modeling techniques are: Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF). LDA uses a probabilistic approach whereas NMF is a linear-algebraic model. In this section we use both of the mentioned models and we compare the accuracy of the results in the terms of top-selected aspects and evaluate their F1 scores. In both models (LDA and NMF) the reviews and the list of stop-words are provided as inputs, and in both of them the number of topics is specified.

The models associate the words to each topic by an iterative procedure, and in this case we have done 50 iterations. In the following tables a sample of the results of aspect extraction is shown. In both models 346 words were extracted which were actually considered as aspect terms. In Table 1 the top 10 aspects for three different topics are displayed, generated by LDA model, while in Table 2 the generated aspects are from NMF model.

	Topic 1	Topic 2	Topic 3
1	ushqim	shkëlqyeshëm	kamarier
2	kajmak	shërbimi	shumë
3	respekt	restoran	pastër
4	fantastik	ambient	përvojë
5	yje	qytet	mënyrë
6	keqe	mrekullueshme	aromë
7	butë	patate	këndshëm
8	restoran	gjatë	çmime
9	kamarier	tmerrshëm	mire
10	speca	kohë	suksesi

Table 14. Aspects extracted with LDA model with 50 iterations

	Topic 1	Topic 2	Topic 3
1	ambient	viç	darkë
2	atmosfera	ushqimi	efikas
3	zhurmë	mish	staf
4	bukur	pica	jashtëzakonshme
5	dekor	pikërisht	menaxher
6	çmim	pulë	menu
7	dysheme	qepë	shërbim

8	elegant	shkëlqyeshëm	njerëz
9	freskët	thatë	pasjellshëm
10	jashtë	tryezë	pronar

Table 15. Aspects extracted with NMF model with 50 iterations

Since in both models we haven't used additional algorithms for separation of aspect and opinion (sentiment) it is obvious that there is a considerable amount of words that are wrongly categorized as aspects (noted in red color). However, despite this fact the precision in the first top 10 is around 60-70% in both models. We had a proportionally similar situation as well when we extracted 20 or more aspects. As can be noticed, the number of wrongly assigned words is equal or less in the NMF model. But, this is not a sufficient indicator to compare and choose which is more appropriate and with higher accuracy. That is why we have evaluated the results (in the next section) by the Precision, Recall and F1 measures. However, it is easily noticed that the classification of aspects into topics is better in NMF model, since in LDA we have a group of words that should correspond to one topic and here they are classified in more than one (e.g. restoran, kamarier, etc.)

In order to calculate the precision, recall and F1 score of both models we use the following formulas:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

and

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

True positives are words classified as aspect terms by the models that actually are aspects (meaning they are correct), false positives are cases the model incorrectly labels as aspects that actually are not so (they may be sentiments or other words), and false negatives are words that the model does not identify as aspects, though actually they are. The results of the three measures (Precision, Recall and F1 score) for both models are shown in the following table.

	Precision %	**Recall %**	**F1 score %**
LDA	47.46%	40.46%	43.68%
NMF	50.79%	43.73%	46.99%

Table 16. Results of the evaluation of LDA and NMF

Considering all three measures it can be noticed that both of the models provided by and large satisfactory results for extracting the aspects, however Non-negative Matrix Factorization model has a better accuracy in comparison to Latent Dirichlet Allocation. This is valid also for the aspect-terms' categorization into aspect topics. In the following sections this work will be extended by combining these models with other algorithms and techniques in order to achieve the most accurate separation of aspects and sentiments.

As it is stated in the section of dataset, for the evaluation of this phase we are considering the corpus with restaurant reviews. We focused on three aspects, such as:

ambiance, food and service. By including the Word2Vec model in this process, the word lists that are provided in input are extended and used for document indexing and document classification. In the following table, there is a listing of some terms obtained by extending the seed list with Word2Vec. And as can be noticed in the scope of the extension there are added both aspect-terms and sentiment-terms for each category.

	Ambiance	**Food**	**Service**
Aspects obtained with Word2Vec	elegant	shkëlqyeshme	mahnitshëm
	dekor	pjekur	minuta
	këndshëm	qepë	lire
	shkëlqyeshëm	tryezë	rregull
	keq	pulë	nxehtë
	madh	kompliment	veçantë
	vend	tymosur	rekomandoj
	zbrazët	ngrohtë	shërbehen
	autentik	mish	çmim
	ndriçim	shije	miqësor
	europian	ëmbëlsira	qetë

Table 17. Additional terms obtained with Word2Vec

In our model there is no need of language-dependent resources for improving the identification of aspect. Only a list of seed words that define the aspects based on domain knowledge is needed. Below is a list of seed words for all the aspects that we take into consideration; namely for ambiance, food and service. The chosen words affect directly the next phase and the final results. In the cases where the usage is aimed to be for more than one domain then the seeds are chosen to be more general, so that, they can be valid for all domains. On the other hand, when only one domain is aimed, the more specific the seeds are, the better results are expected.

ASPECT	SEEDS
Ambient (Ambiance)	bukuri, koncept, dekor, elegancë, interior, ndriçim, zhurmë, natyrë, vend, madhësi, temperature, ambient.
Ushqim (Food)	mish, viç, pulë, vezë, pizza, brumë, pasta, ëmbëlsirë, pije, lëng, qingj, ushqim.
Sherbim (Service)	Staf, kamarrier, punëtor, kuzhinier, shërbim, profesionist.

Table 18. Aspect seeds

With these lists we contribute a lot to increasing the accuracy of aspect extraction. That is why in the future evaluations we may experiment with other seed words which may be more general or specific depending on the aimed results. The following table demonstrates the output that is generated by the system for classifying the aspect-terms into aspect classes. The classification measure is denoted from 0 to 1.

REVIEWS	Ambiance	Food	Service
ketu qenka mire me shku.	1	9.73E-16	9.73E-16
vend ideal per te gjithe!	1	5.02E-16	5.02E-16
a do ma jepni be vec nje kajmak prej ne saraj do vij :)	0.95644	0.04356	2.13E-15
a do shkojm ta hame nje darke se shume mire e paskan be	1.22E-15	1	1.22E-15
ai moment kur te shkon mendja per kajmak	1.85E-15	1	1.85E-15
ai qe e viziton nje her nuk e harron per jet	0.333333	0.333333	0.333333
aty asnjehere ska qene problem cfare do hame, po kush do paguaje.	0.844708	0.155292	2.22E-15
aty ku freskia eshte specialitet ,provoeni skare te perzier	0.949443	0.050557	2.52E-15
aty ku natyra e bukur qe te rrethon,sherbimi profesional e mbi te gjitha kualiteti dhe shija qe kane ushqimet e pergatitura,te bejne t? ndihesh me mire se kudo tjeter.	1	1.20E-15	1.20E-15
aty ku ne cdo koh ke shije dhe kualitet	5.59E-15	0.932557	5.59E-15
beni gati nja 10kg kajmak	0.552613	0.447387	1.30E-15
bravo i qofte pronarit	5.07E-15	5.07E-15	1
bravo mrekulli	1	8.91E-16	8.91E-16
bukur shum ...klm e pershendes gjellbersin	1	7.44E-16	7.44E-16
bukuri e jashtzakonshme	1	1.66E-15	1.66E-15
bukuri e rall respekt per pronarin	1	3.58E-15	3.58E-15
bukuri e ralle me ngjyre te bardhe	1	1.68E-15	1.68E-15
bukuri natyrore e papare	0.898291	0.101709	3.41E-15
cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	0.399139	0.600861	5.90E-16

Table 19. Automatic aspect-term classification

As may be noted, it is a high percentage of proper classification, especially in the ambiance category. What is very important to be mentioned is the case when a review was not with an exact aspect, and we categorized it in all three aspects. In the automatic classification is done the same evaluation. This can be seen in the following example, where the classification is done evenly in all three aspects.

ai qe e viziton nje here nuk e harron per jete	0.33333	0.333333	0.333333

Sentiment Extraction

Sentiment Extraction is the next phase after identifying the appropriate aspects in each sentence. In this phase we calculate the value of sentiment polarity. Namely, we need to classify the opinions as positive or negative. In order to achieve the classification we should follow the below-stated steps:

- We should manually build two seed terms' lists. For each aspect we should create one list for positive sentiment terms and one for negative sentiments.
- Similarly to one of the previous sections where we extended the aspect terms lists, at this stage we should extend the sentiment terms lists by using a Word2Vec model.
- Each document of the corpus should pass in a pre-processing step which includes the stemming and the stop-words removal. This is done on each extended sentiment terms lists.
- Finally, in each document we search for sentiment terms which are relative to the most related topics which are identified therein.

By finding the average value of the number of positive and negative terms, which are found in relation to a specific topic, we calculate the polarity for each aspect. A positive sentiment is valued with 1 and the negative with -1. The final label that will be assigned to the opinion depends on the sign of the average score. As expected, the negative sentiments are harder to detect. So, taking into account the structure of the negation in Albanian language, at this stage we must consider involving some additional negation rules. We solve this issue by adding a short list of negation markers which will switch into opposite polarity of a sentiment. In the following table we list the seeds for sentiment polarity classification, where for each aspect we should list positive and negative seeds.

Aspect	**Polarity**	**Seeds**
Ambient (Ambiance)	Positive	bukur, elegante, komode, moderne, mirë, e madhe
	Negative	ngushtë, i errët, shtrenjtë, i zhurmshëm, keq
Ushqim (Food)	Positive	i pjekur, i tymosur, i butë, i shijshëm, i lëngshëm
	Negative	i thatë, pa gatuar, i vjetër, pa shije, i djegur, i gjallë
Shërbim (Service)	Positive	i vëmendshëm, efikas, i shpejtë, i sjellshëm, i qeshur
	Negative	i pavëmendshëm, i vrazhdë, i ngadalshëm, i vonuar, nevrik

Table 20. Sentiment polarity seeds

In the phase of sentiment polarity classification we considered only the reviews which have been correctly classified in the phase of aspect term classification. We have considered as seed words those that are listed in the above table. After providing the seed words as an input, the analysis of the sentiment polarity proceeds. Table 20 shows a small part of the results that the system gives us as one of the outputs.

REVIEWS	**Ambiance**	**Food**	**Service**
ketu qenka mire me shku.	1	0	0
vend ideal per te gjithe!	1	0	0
a do ma jepni be vec nje kajmak prej ne saraj do vij :)	0	0	0
a do shkojm ta hame nje darke se shume mire e paskan be	0	1	0
ai moment kur te shkon mendja per kajmak	0	0	0
ai qe e viziton nje her nuk e harron per jet	0	0	0
aty asnjehere s'ka qene problem cfare do hame, po kush do paguaje.	0	-1	0
aty ku freskia eshte specialitet ,provoeni skare te perzier	0	1	0
aty ku natyra e bukur qe te rrethon,sherbimi profesional e mbi te gjitha kualiteti dhe shija qe kane ushqimet e pergatitura,te bejne t? ndihesh me mire se kudo tjeter.	1	1	0
aty ku ne cdo koh ke shije dhe kualitet	0	1	0
beni gati nja 10kg kajmak	0	0	0
bravo i qofte pronarit	0	0	1
bravo mrekulli	1	0	0
bukur shum ...klm e pershendes gjellbersin	1	0	0
bukuri e jashtzakonshme	1	0	0
bukuri e rall respekt per pronarin	1	0	0

bukuri e ralle me ngjyre te bardhe	0	0	0
bukuri natyrore e papare	-1	0	0
cdo ushqim qe ju pergadisni e beni ne menyre perfekte.	0	1	0

Table 21. Automatic sentiment polarity classification

The accuracy of these results for the classification is presented in the following table, and as can be noticed, the performance of negative polarity is lower than the positive one. The results of our evaluations in this task may be slightly lower than other state of the art models (Pontiki, et al., 2015); however our model is noticeable for the independence from extra linguistic resources that takes up a lot of time and human resources.

	Reviews
Accuracy	0.81
Precision (positive polarity)	0.82
Recall (positive polarity)	0.85
Precision (negative polarity)	0.80
Recall (negative polarity)	0.76

Table 22. Sentiment Polarity Results

The greatest advantages of the model are the domain and language flexibility. In order to achieve the domain independence we should carefully choose the seed words in the stage of aspect extraction. They should be seeds of a more general nature so that they can cover more domains. Another point which may lead to lower performance is the frequency of used aspect terms in each language. On the other hand, to achieve language flexibility, we should just translate the seed words that are used in aspect extraction and sentiment polarity classification. This may be done by automatic translation; however the more secure way to avoid underperformance is the manual translation.

Another thing that we should pay attention to is the set of grammar rules for a specific language. These terms would not be easy to translate automatically from one language to another. So, the negation terms for Albanian language where compiled manually. Considering all of these, we assume that the right tuning of seed words and grammar rules can achieve a good performance.

Comparison of Aspect Extraction and Sentiment Classification of Albanian and English Texts

In this section our goal is to compare the accuracy of the results from applying the same model to Albanian and English language reviews. Here, it is very important to consider the language differences and challenges that affect the tasks of aspect-extraction and sentiment polarity. In the section of Language Challenges we have discussed the main issues while comparing both languages. Some of them are: inflections, homonyms, dialect and regional speeches, irony, sarcasm, ambiguous sentence structures, etc. Among grammatical and structural differences, in the phase of testing the inputs of stop-words, aspect-seeds and sentiment seeds differ as well. We have since done the comparison with reviews from the same domain, namely from the restaurant reviews. And the size of the testing dataset is the same as well.

The list of stop-words for Albanian language may be found in the appendix,. As may be noted in the lists, the negative markers are not included, since if they are ignored they would switch the sentiment polarity.

In both analyses we have considered three aspects, such as: ambiance, food and service. Regarding the aspect –terms seeds for Albanian language are already mentioned and for English language they are presented in the following table.

Aspect	Seeds
Ambiance	Beauty, concept, decor, elegance, interior, lighting, noise, nature, place, size, temperature, environment, ambiance.
Food	Meat, beef, chicken, eggs, pizza, dough, pastries, dessert, drink, juice, lamb, food.
Service	Staff, waiter, worker, cook, service, professionals.

Table 23. Aspect-term seeds for English language

In order for the comparison to be more even the same aspect seeds that were used in Albanian are translated in the English language. With this, we can easily note that the adaption of the model is easy for different languages. That is why the used model has the characteristic of language-independent systems. Beside the aspect-seeds the model takes as an input the list of sentiment seeds. Here, same as we did in the aspect seeds, we will translate the same sentiment seeds that are used for Albanian language. The considered seeds are presented in the table below.

The translation of the seed words can be done manually or automatically, but since the automatic translation is not always 100% true, manual translation is preferred. However, as we saw in the tables, the number of seeds is very small, so just a minimal supervision is needed. Below are the results of the aspect-term extraction which are evaluated in three measures: Precision, Recall, F1 score.

Aspect	Polarity	Seeds
Ambiance	Positive	beautiful, elegant, cozy, modern, good, great
	Negative	narrow, dark, expensive, noisy, bad
Food	Positive	ripe, smoked, soft, tasty, juicy
	Negative	dry, uncooked, old, tasteless, burnt, lively, stale
Service	Positive	attentive, efficient, fast, polite, smiling
	Negative	inattentive, rude, slow, delayed, irritable

Table 24. Sentiment seeds for English Language

	Precision %	Recall %	F1 score %
English	52.16%	74.3%	61.72%
Albanian	50.79%	43.73%	46.99%

Table 25. Accuracy of aspect-term extraction

The accuracy in all parameters in the Albanian language is lower, compared to the English language. We consider that the main reasons are the processes of stemming and lemmatization. Since we don't have a lexicon for POS taggers in Albanian language which is needed in the lemmatization task, we proceeded only with the stemming task.

After taking into account only the correct aspects, we proceeded with the sentiment classification. The same steps of the task of sentiment classification, which are mentioned and described in the previous section, are conducted for English language as well. These tasks are also done by Pupura (2018) with the similar model. The following table shows the results of accuracy for this classification, considering the positive and negative polarities.

	English Reviews	**Albanian Reviews**
Accuracy	0.87	0.80
Precision (positive polarity)	0.89	0.82
Recall (positive polarity)	0.91	0.85
Precision (negative polarity)	0.85	0.80
Recall (negative polarity)	0.80	0.76

Table 26. Accuracy of sentiment polarity classification

As in the aspect term extraction, the same trend appears in the process of sentiment classification. Again the accuracy of applying the model to English language reviews is higher. It is noticeable that in both languages the accuracy of negative polarities is lower. This is because of the negation markers and rules that we have discussed in the section of Language Challenges. From this comparison it is concluded that each difference in the languages (semantic or structural) may affect the accuracy of both aspect-term extraction and sentiment polarity classification. The used language in the reviews, which may include formal and informal language, slangs, abbreviations, irony, sarcasm, etc., plays the biggest role in these analyses. Another factor is the input that we give to the system, which contains the seed words and stop-words. However, with this double application, we can see that the model is easily adaptable for any language and for any domain, since it depends solely on the seed-words.

Conclusions

After introducing the idea and the main objectives we want to achieve through this thesis in the above sections, first we have done an elaboration of Aspect-based Sentiment Analysis and its tasks and then conducted them with our adapted model. While trying to compute all the tasks of Aspect-based Sentiment Analysis in an almost unsupervised manner for textual data in Albanian language, we have encountered a lot of challenges that were mostly a result of the language structure and the lack of data and research in this field for our target language. Since our goal was to adapt a model that will fulfill the tasks of ABSA with minimal supervision, we focused on the unsupervised approaches and on language- and domain-flexible models.

The first challenge of the process was to compare and decide with which machine learning approach we were going to work. After considering the existing research and our experiments, we managed to compare the following approaches: rule-based, topic-based, seed-based, FURIA (Fuzzy Unordered Rule Induction Algorithm)-based and seed-based, combined with topic modeling approaches. According to the results obtained and according to all detected advantages and disadvantages of each model, we concluded that in comparison with rule-based model, all the other machine learning models have had same or higher accuracy results. Currently most approaches that are being tested go along the lines of obviating human labor. So, this is one of the greatest advantages of other models over rule-based technique, which works on very strict pre-defined rules. With this, we have confirmed the first stated hypothesis.

Regarding the corpus, as we have explained, currently we own more than 30.000 comments, and by analyzing the Albanian ecosystem, which was one of our research questions, we have found out that the top topics are: politics, sports, restaurant and showbiz. That is why we have used for experimental needs within our work only the comments/reviews from the domain of restaurants, and later showed the easy adaption for other domains. In the phases of preprocessing data and data normalization, more precisely in the process of stemming and lemmatization, it is noticed that absence of POS taggers for Albanian language adds one more challenge to the implementation of an ABSA system for this language. That is why we have conducted only the stemming process. Besides the corpus, other inputs that are needed to the system are the list of stop-words, which we have provided in the Appendix 1, the list of aspect seed words and sentiment polarity seed words and, if needed, some additional grammar rules which depend on the used language.

In the task of aspect extraction a comparison of topic modelling algorithms (LDA and NMF) was done, which has been one of our research questions as well, and the result was that e NMF works slightly better for Albanian language than LDA algorithm. So, the other stages are completed with the aspect terms which were generated by NMF algorithm. In the second task, we have extracted the sentiment corresponding to the aspect-terms that we gained in the first task, and then we have classified the sentiment polarities. Regarding the polarities we have had cases with positive, negative, neutral and conflicting polarities. There have been a lot of special cases that were included in the dataset, however in order for the results to be more realistic, we did not exclude them. Some of these situations sometimes were a result of other language challenges such as: homonyms, dialects, irony, sarcasm and the ambiguity of opinions. Here, the negation rules have been a great challenge as well, since they differ in the Albanian and the English language. The Albanian language includes, in addition to a single negation, double and triple negations as well, and the number of negation markers in the Albanian language is greater than in English, which affects the correlation, and as a result, the accuracy. The rate of accuracy in the negative sentiment polarity was lower than in positive polarity.

Due to the lack of research and implementation of ABSA in Albanian language through supervised approaches as well, we can only serve the results of ABSA through weakly supervised method, which were satisfactory and acceptable for future applications. On the other hand, we compared the gained accuracy with the implementation of the same or similar models for English language and we concluded that the accuracy is higher for English language, and this is due to the above mentioned language differences in structure, grammar and semantics and other language limitations.

We have also tested the application of the system for different domains and depending on the seed words that we provide to the system as inputs, the model is easily adaptable, after giving the proper seeds words. With all these experiments in the conduct of ABSA tasks for Albanian texts and comparisons among languages we have achieved our goal to start the application of ABSA for Albanian language and make use of all unused data in the online world. This model, since it is adaptable for different domains, can be used in different areas, such as: tourism, public and private institutions, restaurants, hotels, politics, arts, etc. In our future work we are aiming to implement it in educational institutions for the inner-evaluations, with a focus on teacher evaluations.

References

Afzaal, M., Usman, M., Fong, A.C.M. et al. (2016). Fuzzy aspect based opinion classification system for mining tourist reviews. Advances in Fuzzy Systems. https://doi.org/10.1155/2016/6965725

Akhtar, M.S., Kumar, A., Ekbal, A., Biemann, C., and Bhattacharyya, P. (2019). Language-agnostic model for aspect-based sentiment analysis. In Proceedings of the 13th International Conference on Computational Semantics-Long Papers, pages 154–164.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J., & Iglesias, C. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Systems With Applications, 77, 236-246. doi: 10.1016/j.eswa.2017.02.002

Axhiu, M. (2019). Language Challenges in Aspect-Based Sentiment Analysis: A Review of Albanian Language. Knowledge International Journal, 31(6), 1709 - 1712

Axhiu, M., Aliu, A. (2021). Semi-Supervised Aspect-Based Sentiment Analysis On Albanian Restaurant Reviews. Knowledge International Journal, Volume 45.3 http://ikm.mk/ojs/index.php/KIJ/article/view/5012

Axhiu, M., Aliu, A., (2020). Aspect-Term Extraction from Albanian Reviews with Topic Modeling Techniques. Machine Learning and Applications. 12th International Conference, ICT Innovations 2020- ISSN 1857-7288

Axhiu, M., Freisleben, B. (2018). Aspect-Based Sentiment Analysis of Albanian Texts-Review. Information and Communication Technologies at DSC Proceeding p.89-97

Bagheri, A., Saraee, M., de Jong, F. (2014). ADM-LDA: An aspect detection model based on topic modelling using the structure of review sentences. Journal of Information Science 40:621-636. doi: 10.1177/0165551514538744.

Bekteshi, A. (2020). Syntactic Structures and Semantic issues of Negation in English and Albanian - A contrastive analysis (PhD). South East European University.

Biba, M., Mane, M. (2014). Sentiment Analysis through Machine Learning: An Experimental Evaluation for Albanian. Advances in Intelligent Systems and Computing 195-203. doi: 10.1007/978-3-319-01778-5_20

Brun C. et al. (2014). XRCE: Hybrid Classification for Aspect-based Sentiment Analysis, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) pp. 838 - 842, Dublin, Ireland, August 23-24, 2014.

Brychcin T. et al. (2014). UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval, 2014) pp. 817 - 822, Dublin, Ireland, August 23-24, 2014.

Caka, N., & Caka, A. (2011). Korpusi i Gjuhës Shqipe – Rezultatet e Para, Problemet dhe Detyrat. In Shqipja dhe gjuhët e Ballkanit – Albanian and Balkan Languages (pp. 643-656). Prishtinë.

Carenini, G., Ng, R. T., & Zwart, E. (2005). Extracting knowledge from evaluative text. In Proceedings of K-CAP, pp. 11–18.

Carenini, G., Ng, R., & Pauls, A. (2006). Multi-document summarization of evaluative text. In Proceedings of EACL, pp. 305–312.

Chawla, R. (2017). Topic Modeling with LDA and NMF on the ABC News Headlines dataset. In: Medium.https://medium.com/ml2vec/topic-modeling-is-an-unsupervised-learning-approach-to-clustering-documents-to-discover-topics-fdfbf30e27df.

Chen, Y., Zhang, H., Liu, R., et al. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowledge-Based Systems 163:1-13. doi: 10.1016/j.knosys.2018.08.011

Cichocki, A., Phan, A.H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. IEICE Trans. Fundam. Electron., Commun. Comput. Sci. 92(3), 708–721

Colhon, M., Badica, C. and Sxendre A. (2014). Relating the opinion holder and the review accuracy in sentiment analysis of tourist reviews. Proceedings of the international conference on knowledge science, engineering and management, Melbourne, VIC, Australia, 19–20 August 2014, pp. 246–257. New York: Springer. https://doi.org/10.1007/978-3-319-12096-6_22

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In Proceedings of WWW, pp. 519–528.

De Clercq, O. et al. (2015). LT3: Applying Hybrid Terminology Extraction to Aspect-Based Sentiment Analysis, Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) pp. 719 - 724, Denver, Colorado, June 4-5, 2015.

Deriu, J. (2016). Sentiment Analysis using Deep Convolutional Neural Networks with Distant Supervision (Graduate). ETH Zurich.

Devopedia. (2020). "Word2vec." Version 4, https://devopedia.org/word2vec

Doshi, S. (2019). Skip-Gram: NLP context words prediction algorithm. Retrieved from https://towardsdatascience.com/skip-gram-nlp-context-words-prediction-algorithm-5bbf34f84e0c

Giannakopoulos, A., Musat, C., Hossmann, A., & Baeriswyl, M. (2017). Unsupervised Aspect Term Extraction with B-LSTM and CRF using Automatically Labelled Datasets. Proceedings of The 8Th Workshop On Computational Approaches To Subjectivity, Sentiment And Social Media Analysis. doi: 10.18653/v1/w17-5224

Hamdan H. et al. (2015). Lsislif: CRF and Logistic Regression for Opinion Target Extraction and Sentiment Polarity Analysis, Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) pp. 719 - 724, Denver, Colorado, June 4-5, 2015.

Hamdan, H., Bellot, P., & Bechet, F. (2014). Supervised Methods for Aspect-Based Sentiment Analysis. In 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 596-600). Dublin, Ireland.

Hasanaj, B. (2012). A Part of Speech Tagging Model for Albanian. Lambert Academic Publishing, Saarbrücken.

Hazarika, D., Poria, S., Vij, P., Krishnamurthy, G., Cambria, E. and Zimmermann, R. (2018). Modeling inter-aspect dependencies for aspect-based sentiment analysis. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2, (pp. 266–270).

Hercig, T. et al. (2016). Unsupervised Methods to Improve Aspect-Based Sentiment Analysis in Czech. Computación y Sistemas, Vol. 20, No. 3, 2016, pp. 365–375

Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., & Kaymak, U. (2013). Exploiting emoticons in sentiment analysis. Proceedings of the 28Th Annual ACM Symposium on Applied Computing - SAC '13. doi: 10.1145/2480362.2480498

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.

Kabashi, B., & Proisl, T. (2016). A Proposal for a Part-of-Speech Tagset for the Albanian Language. In Tenth International Conference on Language Resources and Evaluation (LREC 2016). Paris: European Language Resources Association (ELRA).

Kadriu, A. (2013). NLTK tagger for Albanian using iterative approach. In Proceedings of the 35th International Conference on Information Technology Interfaces (ITI).

Kayaalp, N.F., Weckman, G.R., Millie, D. et al. (2017). Extracting customer opinions associated with an aspect by using a heuristic based sentence segmentation approach. International Journal of Business Information Systems (IJBIS), Vol. 26, No. 2. (pp.236–260). https://doi.org/10.1504/IJBIS.2017.086335

Kim, H., & Park, H. (2008). Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method. SIAM Journal on Matrix Analysis and Applications, 30(2), 713-730. doi: 10.1137/07069239x

Kim, S., & Hovy, E. (2005). Automatic detection of opinion bearing words and sentences. In Proceedings of IJCNLP, pp. 61–66.

Kim, S.-M., & Hovy, E. (2006). Automatic identification of pro and con reasons in online reviews. In Proceedings of COLING ACL, pp. 483–490.

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) pp. 17461751, Doha, Qatar, October 25-29, 2014.

Kiritchenko, S. et al. (2014) NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) pp. 437 - 442, Dublin, Ireland, August 23-24, 2014.

Laddha, A., & Mukherjee, A. (2016). Extracting Aspect Specific Opinion Expressions. In 2016 Conference on Empirical Methods in Natural Language Processing (pp. 627-637). Austin, Texas: Association for Computational Linguistics.

Laskari, N., & Sanampudi, S. (2016). Aspect Based Sentiment Analysis Survey. IOSR Journals. 18. 2278-661. 10.9790/0661-18212428. Retreived from: https://www.researchgate.net/publication/304252751_Aspect_Based_Sentiment_Analysis_Survey

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. In Proceedings of WWW, pp. 342–351.

Lu, Y., & Zhai, C. (2008). Opinion integration through semi-supervised topic modeling. In Proceedings of WWW, pp. 121–130.

Ma, J., Yu, M., Fong, S. et al. 2018. Using deep learning to model the hierarchical structure and function of a cell. Nat Methods 15, 290–298. https://doi.org/10.1038/nmeth.4627

Madhoushi, Z., Hamdan, A., & Zainudin, S. (2015). Sentiment Analysis Techniques in Recent Works. In Science and Information Conference. London.

Mani, I. (2001). Automatic summarization, Vol. 3. John Benjamins Pub Co

McDonald, R., Hannan, K., Neylon, T., Wells, M., & Reynar, J. (2007). Structured models for fine-to-coarse sentiment analysis. In Proceedings of ACL, pp. 432–439.

Mikolov, T., Yih, W.-t., and Zweig, G. (2013). Linguistic regularities in continuous space word representations. Proceedings of NAACL-HLT, pages 746-751.

Mozetič, I., Grčar, M., Smailović, J. (2016). Multilingual Twitter Sentiment Classification: The Role of Human Annotators. PLoS ONE 11(5): e0155036. https://doi.org/10.1371/journal.pone.0155036

Muangon, A., Thammaboosadee, S. and Haruechaiyasak, C. (2014). A lexiconizing framework of feature-based opinion mining in tourism industry. In: Proceedings of the fourth international conference on digital information and communication technology and it's applications (DICTAP), Bangkok, Thailand, 6–8 May 2014, pp. 169–173. New York: IEEE. https://doi.org/10.1109/DICTAP.2014.6821677

Mukherjee, A., and Liu, B. (2012). Aspect extraction through semi-supervised modeling. In: Proceedings of the 50th annual meeting of the association for computational linguistics: long papers, vol. 1, Jeju Island, Korea, 8–14 July 2012, pp. 339–348. New York:ACM.

Naskar, A. Continuous Bag of Words (CBOW) - Single Word Model - How It Works - ThinkInfi. Retrieved from https://thinkinfi.com/continuous-bag-of-words-cbow-single-word-model-how-it-works/

Nguyen, T., & Shirai, K. (2015). Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction. In 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (pp. 1354–1364). Beijing: Association for Computational Linguistics. Retrieved from https://aclanthology.org/P15-1131

Pablos, A., Cuadros, M., & Rigau, G. (2017). W2VLDA: Almost Unsupervised System for Aspect Based Sentiment Analysis. Arxiv.org. Retrieved from https://arxiv.org/abs/1705.07687

Pang, B. & Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of ACL, Barcelona, Spain.

Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP, pp. 79–86.

Patel, D., Saxena, S. and Verma,T. (2016). Sentiment Analysis using Maximum Entropy Algorithm in Big Data. International Journal of Innovative Research in Science, Engineering and Technology., vol. 5, pp. 8356-8361, available at: Google Scholar.

Pavlopoulos, I. (2014). Aspect Based Sentiment Analysis. PhD Thesis, Athens University of Economics and Business

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015). SemEval-2015 Task 12: Aspect Based Sentiment Analysis. Proceedings of the 9th International Workshop on Semantic Evaluation (Semeval 2015). doi: 10.18653/v1/s15-2082

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 27-35). Dublin: Association for Computational Linguistics.

Popescu, A.M., Nguyen, B., & Etzioni, O. (2005). OPINE: Extracting product features and opinions from reviews. In Proceedings of EMNLP/HLT, pp. 339–346.

Purpura, A., Masiero, C., Susto, G.A. (2018). WS4ABSA: An NMF-Based Weakly-Supervised Approach for Aspect-Based Sentiment Analysis with Application to Online reviews. International Conference on Discovery Science, (pp. 386-401).

Rana, T.A., and Cheah, Y.N. (2016). Aspect extraction in sentiment analysis: Comparative analysis and survey. Artif. Intell. Rev., vol. 46, no. 4, pp. 459-483, Dec. 2016.

Raychev, V., & Nakov, P. (2009). Language-Independent Sentiment Analysis Using Subjectivity and Positional Information. In International Conference RANLP (pp. 360-364).

Ruder, S., Ghaffari, P., and Breslin, J.G. (2016). A hierarchical model of reviews for aspect-based sentiment analysis. arXiv preprint arXiv:1609.02745

Ruppert, E., Kumar, A., & Biemann, C. (2017). LT-ABSA : An Extensible Open-Source System for Document-Level and Aspect-Based Sentiment Analysis.

Sarwan, N. (2017). An Intuitive Understanding of Word Embeddings. Retrieved 11 September 2021, from https://neerajsarwan.github.io/2017-06-04-an-intuitive-understanding-of-word-embeddings/

Sauper, C., & Barzilay, R. (2013). Automatic Aggregation by Joint Modeling of Aspects and Values. Journal of Artificial Intelligence Research, 46, 89-127. doi: 10.1613/jair.3647

Seki, Y., Eguchi, K., Kanodo, N., & Aono, M. (2005). Multi-document summarization with subjectivity analysis at DUC 2005. In Proceedings of DUC at EMNLP/HLT.

Seki, Y., Eguchi, K., Kanodo, N., & Aono, M. (2006). Opinion-focused summarization and its analysis at DUC 2006. In Proceedings of DUC at NAACL/HLT, pp. 122–130.

Semeval-2015 task 12: aspect based sentiment analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation.

Singh, A., and Ullah, E. (2015). Aspect based Sentiment Analysis. Retrieved from http://home.iitk.ac.in/~enayat/files/report_aip.pdf

Snyder, B., & Barzilay, R. (2007). Multiple aspects ranking using the good grief algorithm. Proceedings of NAACL/HLT, pp. 300–307.

Sun, C., Myers, A., Vondrick, C., Murphy, K., & Schmid, C. (2019). VideoBERT: A Joint Model for Video and Language Representation Learning. Retrieved from: https://arxiv.org/abs/1904.01766

Suri, P., Roy, N.: Comparison between LDA & NMF for event-detection from large text stream data. 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT). doi: 10.1109/ciact.2017.7977281 (2017).

Syamala, M. (2019). A Deep Analysis on Aspect based Sentiment Text Classification Approaches. International Journal of Advanced Trends in Computer Science And Engineering, 1795-1801. doi: 10.30534/ijatcse/2019/01852019

Taylor, E. M., Velasquez, J.D. and Marquez, F.B. (2013). Opinion zoom: a modular tool to explore tourism opinions on the web. In: Proceedings of the 2013 IEEE/WIC/ACM international joint conferences on web intelligence (WI) and intelligent agent technologies

(IAT), vol.3, Atlanta, GA, 17–20 November 2013, pp.261–264. New York: IEEE. https://doi.org/10.1109/WI-IAT.2013.193

Taylor, E. M., Velasquez, J.D. and Marquez, F.B. (2014). A novel deterministic approach for aspect-based opinion mining in tourism products reviews. Expert Sys Appl 2014; 41: 7764–7775. https://doi.org/10.1016/j.eswa.2014.05.045

Thet, T. T., Na, J. C., Khoo, C.S.G. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. Journal of Information Science archive Volume 36 Issue 6 pp. 823-848 Sage Publications, Inc. Thousand Oaks, CA, USA Optim. 20(3), 1364–1377

Titov, I. and McDonald, R. (2008). Modeling online reviews with multi-grain topic models. In Proceedings of the 17th international conference on World Wide Web, pages 111–120. ACM.

Toh, Z. & Su, J. (2015). NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction, Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) pp. 719 - 724, Denver, Colorado, June 4-5, 2015.

Toh, Z. & Wang, W. (2014). DLIREC: Aspect Term Extraction and Term Polarity Classification System, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) pp. 235 - 240, Dublin, Ireland, August 23-24, 2014.

Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In Proceedings of ACL, pp. 417–424.

Vashisht, G., & Jailia, M. (2020). Emoticons & Emojis Based Sentiment Analysis: The Last Two Decades!. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 9(03).

Vavasis, S.A. (2009). On the complexity of nonnegative matrix factorization. SIAM J.

Vidhya. (2017). Understanding Word Embeddings: From Word2Vec to Count Vectors. Retrieved 11 August 2020, from https://www.analyticsvidhya.com/blog/2017/06/word-embeddings-count-word2veec/

Vryniotis, V. (2013). Machine Learning Blog & Software Development News. Datumbox, 20 Nov. 2013, blog.datumbox.com/machine-learning-tutorial-the-max-entropy-text-classifier/.

Wagner, J. et al. (2014). DCU: Aspect-based Polarity Classification for SemEval Task 4, Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) pp. 223 - 229, Dublin, Ireland, August 23-24, 2014.

Wang, W., Wang, H., and Song, Y. (2017). Ranking product aspects through sentiment analysis of online reviews. Journal of Experimental & Theoretical Artificial Intelligence vol. 29 (pp.227–246). https://doi.org/10.1080/0952813X.2015.1132270

Wiebe, J.; Wilson, T.; Bruce, R.; Bell, M.; and Martin, M. (2004). Learning Subjective Language. Computational Linguistics 30(3):277–308.

Wilson, T., Wiebe, J., & Hoffmann, P. (2009). Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. Computational Linguistics, 35(3), 399-433. http://dx.doi.org/10.1162/coli.08-012-r1-06-90

Wright, A. (2009). Sentiment Analysis Takes the Pulse of the Internet. Nytimes.com. Retrieved, from http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html

Xiang, B., Zhou, L. (2014). Improving Twitter Sentiment Analysis with Topic-Based Mixture Modeling and Semi-Supervised Training. Proceedings of the 52nd Annual Meeting of the

Association for Computational Linguistics (Volume 2: Short Papers). doi: 10.3115/v1/p14-2071

Xu, J. (2020). Topic Modeling with LSA, PLSA, LDA & lda2Vec - KDnuggets. In: KDnuggets. https://www.kdnuggets.com/2018/08/topic-modeling-lsa-plsa-lda-lda2vec.html

Xu, Q., Zhu, L., Dai, T., et al. (2019). Non-negative matrix factorization for implicit aspect identification. Journal of Ambient Intelligence and Humanized Computing. doi: 10.1007/s12652-019-01328-9

Yoo, B., & Rayz, J. (2021). Understanding Emojis for Sentiment Analysis. The International FLAIRS Conference Proceedings, 34(1). doi: 10.32473/flairs.v34i1.128562

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of EMNLP, pp. 129–136. Association for Computational Linguistics.

Zafra, J.S.M., Valdivia, M.T.M., Camara, E.M., Lopez, L.A.U. (2016). Combining resources to improve unsupervised sentiment analysis at aspect-level. J Inf Sci 2016; 42: 213–229. https://doi.org/10.1177/0165551515593686

Zhao, L., Li, M., Meng, D., Li, X., Zhang, Z., & Zhuang, Y. et al. (2018). Deep Convolutional Neural Networks with Merge-and-Run Mappings. In Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18) (pp. 3170-3176). Retrieved from https://www.ijcai.org/proceedings/2018/440

APPENDIX – Albanian

- a
- ai
- ajo
- andaj
- apo
- asnjë / asnje
- ata
- ate
- ato
- be
- bëhet
- bëj
- bëje
- bën
- bëre
- bëri
- bie
- ca
- çdo/cdo
- çfarë/cfare
- deri
- derisa
- dhe
- dicka
- dickaje
- dikë
- dikujt
- dikush
- disa
- do
- drejt
- duke
- e
- edhe
- është/eshte
- etj
- fare
- gjatë
- gjë
- gjitha
- gjithë
- herë
- i
- ia
- ishte
- iu
- ja
- jam
- janë / jane
- jap
- jemi
- jeni
- ju
- juaj
- ka
- kam
- kanë

- kaq
- kë/ke
- kemi
- keq
- kështu/keshtu
- kete / këtë
- kishte
- ku
- kur
- kurse
- la
- le
- m'
- madhe
- marr
- mban
- mbi
- me
- më
- megjithatë
- megjithëse
- meje
- meqë/meqe
- merr
- mori
- mos
- mu
- mund
- na
- ndaj
- ndërsa/ndersa
- ndonëse/ndonese
- ndonje/ndonjë
- ne
- në
- nën
- nëpër
- nëse
- nga
- ngaqë/ngaqe
- një
- nuk
- o
- ose
- pa
- pak
- para
- pas
- pasi
- për / per
- përndryshe
- përpara
- po
- por
- porsa
- posa
- pra
- prandaj
- prej
- qe
- që
- qëkurse
- qenët

- rri
- s/ s’
- sa
- saj
- sapo
- saqë/saqe
- se
- së
- seç
- sepse
- sesa
- sh
- shih
- shumë
- si
- sipas
- t’
- ta
- tanë
- te
- të
- teje
- tek
- teksa
- ti
- tij
- tjera
- tjerët
- tjetër
- tonë
- tuaj
- ty
- tyre
- u
- ua
- unë
- vend
- veta
- vete
- vetë
- vetëm
- vjen
- zakonisht