THE USE OF PRETRAINED LANGUAGE MODELS IN SENTIMENT ANALYSIS

ÖMER YİĞİT YÜRÜTÜCÜ

MEF UNIVERSITY

APRIL 2022

MEF UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

MASTER’S in INFORMATION TECHNOLOGIES

M.Sc. THESIS

THE USE OF PRETRAINED LANGUAGE MODELS IN SENTIMENT ANALYSIS

Ömer Yiğit YÜRÜTÜCÜ

ORCID: 0000-0002-7745-0515

Thesis Advisor

Assist. Prof. Şeniz DEMİR

APRIL 2022

ACADEMIC HONESTY PLEDGE

I declare that all the information in this study, is collected and presented in accordance with academic rules and ethical principles, and that all information and documents that are not original in the study are referenced in accordance with the citation standards, within the framework required by the rules and principles.

Name and Surname: Ömer Yiğit YÜRÜTÜCÜ

Signature:

ABSTRACT

THE USE OF PRETRAINED LANGUAGE MODELS IN SENTIMENT ANALYSIS

Ömer Yiğit YÜRÜTÜCÜ

M.Sc. in Information Technologies

Thesis Advisor: Assist. Prof. Şeniz DEMİR

April 2022, 69 Pages

Natural language processing is one of the sub-topics of linguistic science and artificial intelligence. Sentiment analysis classifies a text on any topic according to its subjective content. It is one of the methods used to examine, analyze and interpret data such as the thoughts, feelings or attitudes of individuals about a subject on various platforms. The increase in social media shares has also increased the sentiment analysis studies conducted on these platforms. Different methods are used during sentiment analysis. Classification is carried out by sentiment analysis using machine learning and natural language processing algorithms. In recent years, pre-trained language models have been used together with machine learning methods or alone.

The aim of this thesis is to test the hypothetical advantages of sentiment analysis in social media comments with pre-trained language models. For this purpose, sentiment analysis was performed for Covid-19 related tweets on Twitter. Emotion intensities were determined using pre-trained language models and the results were compared. BERT, RoBERTa and BERTweet were used in the analysis. The results show that NLP techniques for sentiment analysis are as successful as other techniques.

Keywords: Sentiment analysis; Natural language Processing (NLP); BERT; RoBERTa; BERTweet; Covid-19

Numeric Code of the Field: 92416

ÖZET

DUYGU ANALİZİNDE ÖN EĞİTİMLİ DİL MODELLERİNİN KULLANIMI

Ömer Yiğit YÜRÜTÜCÜ

Bilişim Teknolojileri Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Şeniz DEMİR

Nisan 2022, 69 Sayfa

Doğal dil işleme, dil bilim ve yapay zekânın alt konularından biridir. Duygu analizi herhangi bir konuda bir metni öznel içeriğine göre sınıflandırma yapar. Genellikle bireylerin çeşitli platformlarda bir konu hakkında düşünce, duygu ya da tutumu gibi verileri irdelemek, analiz etmek ve yorumlamak amacıyla kullanılan yöntemlerden biridir. Sosyal medya paylaşımlarındaki artış bu platformlarda yapılan duygu analizi çalışmalarını da artırmıştır. Duygu analizi yapılırken farklı yöntemlerden yararlanılır. Makine öğrenmesi ve doğal dil işleme algoritmaları ile duygu tespiti ile sınıflandırma yapılır. Son yıllarda önceden eğitilmiş dil modelleri makine öğrenmesi metotlarıyla birlikte ya da tek başına kullanılmaya başlamıştır.

Bu tezin amacı önceden eğitilmiş dil modelleri ile sosyal medya yorumlarında duygu analizinin varsayımsal avantajlarını test etmektir. Bu amaçla Twitterdaki Covid-19 ile ilgili tweetler için duygu analizi yapılmıştır. Önceden eğitilmiş dil modelleri kullanılarak duygu yoğunlukları tespit edilmiş ve sonuçları karşılaştırılmıştır. Analizlerde BERT, RoBERTa ve BERTweet'ten yararlanılmıştır. Sonuçlar, duygu analizi için NLP tekniklerinin diğer teknikler kadar başarılı olduğunu göstermektedir.

Anahtar Sözcükler: Duygu analizi; Doğal dil işleme (NLP); BERT; RoBERTa; BERTweet; Covid-19

Bilim Dalı Sayısal Kodu: 92416

FOREWORD

I would like to forward gratitude to my advisor Assist. Prof. Şeniz Demir, shared her valuable knowledge and experience with dedication, guiding me, and contributing to my work. In addition, I would also like to thank my dear jury members Assist. Prof. Tuna Çakar and Assist. Prof. Didem Abidin, contributed to my education. I am very thankful to my family for their unlimited support during the thesis and entire of my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ABSA : Aspect-Based Sentiment Analysis

API : Application Programming Interface

BERT : Bidirectional Encoder Representations from Transformers

BOW : Bag of Words

K-NN : K-Nearest Neighbors

LR : Logistic Regression

ME : Maximum Entropy

ML : Machine Learning

MLM : Masked Language Modeling

MPQA : Multi-Perspective Question Answering

NB : Naïve Bayes

NER : Named Entity Recognition

NLP : Natural Language Processing

NSP : Next Sentence Prediction

OCR : Optical Character Recognition

POS : Part of Speech

RNN : Recurrent Neural Networks

SVM : Support Vector Machines

VADER : Valence Aware Dictionary and Sentiment Reasoner

INTRODUCTION

One of the distinguishing features of human intelligence is language. That's why research such as natural language processing is carried out to understand human language. With the development of the world, language is also developing, and new techniques are being developed to support this development of the language.

The rapid change on the internet opens new areas to the world. The increase in the information flow enables people to quickly reach the topics that they are researching and wondering about. As a result, the internet is one of the most frequently used platforms for sharing feelings and thoughts. This enables feedback collection for further analysis. Businesses are setting up different platforms such as blogs, surveys, and forums where people interact and provide feedback. As the amount of participation increases, it takes a long time to analyze the data. For this reason, the data are classified into groups automatically and interpreted from different perspectives analytically. This is one particular way of allowing companies get the opportunity to overcome their shortcomings. In addition, users also get opportunity to express their positive, negative or neutral feelings. Finally, newly developed infrastructures and interfaces make it easier to reach and interact with people.

In recent years, people have started to share more with the development of individual technological devices. Data such as shared photos, location, feelings and thoughts can be easily obtained. Large volumes of datasets are obtained from intermediaries such as social networking sites, blogs, news sites, forums, and complaint pages. However, it takes a long time and is not easy to determine how positive and negative the sharing are. Thus, several alternative solutions are provided for this, one of which is automatically identifying the sentiment of a given text.

Social media is a pool of information containing data belonging to customers in terms of organizations. Companies have adopted social media for purposes such as to increase customer loyalty, sales, revenue, customer satisfaction, brand awareness, customer traffic and gaining reputation (Culnan, et al., 2010; He et al., 2013; Kietzmann et al., 2011; Sinderen & Almeida, 2011; Weber, 2009). This mutual interaction has caused social media to be more effective. Insights obtained from social media can be analyzed with text mining techniques. Methods such as sentiment analysis, text classification, association rule learning or word clouds are used to obtain meaningful information from texts (Kapucugil & Özdağoğlu, 2015).

Natural language is the main need for ensuring the continuity of everyday life on all the platforms in question. It is a feature of human survival and communication. Also, it is required in everywhere such as writing texts, using signs, SMS, internet within the scope of communication. However, learning the language is quite difficult and time consuming. Although speaking in a native language may seem easy, it is not spontaneous. Words or situations that change over time in a language keep languages alive. The brain's thinking process in natural language is quite complex. It is more difficult to transfer this structure to computer environment. Natural language processing (NLP) has reached today's possibilities thanks to artificial intelligence modules.

Linguistics is a field that explores human use of language and aims to accurately describe the nature and functioning of language. The main problem is how grammar is deployed in the production and understanding of language (Stubb, 1996). NLP works on statistical natural language processing probabilities. Therefore, it differs from classical linguistic methods. NLP explores big data in text format using different text processing and machine learning methods. In this context, "*Information Retrieval*", "Text Categorization" and "Machine Learning" are important domains. The acquisition and coding of all information contributes to the development of effective and strong language systems and to the advancement of verbal machine learning methods.

Many examples of NLP fields can be given. For example, machine translation is one of the most widely used systems. When searching on Google, the question 'Did you mean this' uses the same information. Also, actions such as voice messaging, text correction or automatic word completion, spam filters, chatbots such as Microsoft Messenger and voice assistants such as Siri, Google Assistant and Alexa have been prepared with NLP. Apart from these, it is possible to see it in different applications in the business world. HR (Human Resources) also helps job seekers diversify their backgrounds, attract different candidates and recruit more qualified employee. Applications such as Outlook and Gmail detect text and split messages from specific people into folders you create. Another study is that sensitivity analysis is a tool. For example, it can be determined that tweets about companies are good or bad. The findings here have an important place in customer satisfaction data (Maksymenko, 2020; Shruthi et al., 2019).

Using neural networks before the education of the language is one of the ways to save additional description source in downstream tasks. In the NLP field, major progress has been recorded through independent language models without any supervision. Pretrained language models serve a lot of purpose today. One of them is sentiment analysis.

The Sentiment analysis is a subfield of Natural Language Processing (NLP). Sentiment analysis has several application purposes such as extracting and identifying opinions in social media, analyzing market sensitivity, and analyzing opinions about products (Bakliwal et al., 2013; Ghiassi et al., 2013; Zhang & Skiena 2010). Research on sentiment analysis started in the late (1990s). However, with the widespread use of the internet in recent years, there has been an increase in research on this subject. Thanks to the capable of communicating at a global level, people's social networks are also expanding. Especially, the increase in the number of users on platforms such as social media, blogs and forums increases social interaction. The common point of these platforms is the freedom of users to express their knowledge and feelings comfortably. This way, quality comments for a product brand, opinions about a movie, or customer complaints about a restaurant can be seen. For example, the Periscope application, which is Twitter's product, has NLP paired with visual recognition to create Trump-Emoticoaster, a data engine that processes language and facial expressions to monitor former US President Donald Trump's emotional state. From the perspective of researchers, these shares contain an important dataset. Sentiment analysis has enabled to determine the idea of determining the relevant emotion by examining it with different methods applied in the dataset on the subject.

In general, sentiment analysis can be defined as examining, analyzing and interpreting data such as thoughts, feelings or attitudes of individuals on a subject on various platforms. Sentiment analysis examines texts written on a particular topic. It analyzes positive, negative or neutral content in this text. The purpose of the analysis is to determine the opinion of the authors. It has become common to use mathematical models in natural language processing. In order to predict a model that describe the connections and relationships in the data are created and the language is processed and analyzed in this way (Goodfellow et al., 2016; Krizhevsky et al, 2012).

A sentiment analysis tool can be used in many domains such as social media platforms. Feelings and thoughts shared by users on platforms such as Instagram, Twitter, Facebook, LinkedIn can be identified via sentiment analysis (Abid, et al., 2019; Ghiassi et al., 2013). Sentiment analysis focuses on finding individuals' opinions on certain topics (Ghiassi & Lee, 2018). Its main objective is to categorize texts according to their poles. In another words, the task can be defined as determining whether the overall feeling in a given text is positive, negative or neutral (Azzouza et al., 2020). Syntax and vocabulary of the texts are considered during the analysis of emotions (Abid, et al., 2019). In this context, word clouds help us to understand the frequency of used words in a text and which words are used together. In addition, the data can be converted into a structural form and used as input in other research studies (Akın & Şimşek, 2018, p. 246). For example, classification, syntax analysis, detection of entities in the text and sentiment analysis can be performed (according to the data content) through the Google Cloud Natural Language API service. First of all, score and name are given to each sentence in the text and the value that measures the sentiment level is created. In the analysis, values varying between -1 and 1 are formed. The text becomes negative as the value decreases, and the text becomes positive as the value increases. It detects the intensity of sentiment in the text using the Google NLP library (Google CLN, 2018).

The purpose of this thesis is to make sentiment analysis using pre-trained language models. For this purpose, a study has been conducted on Twitter which is one of the social media networks. In the study, Twitter data related to Covid-19, which is one of the frequently researched topics was used. In the sentiment analysis, BERT, RoBERTa and BERTweet, which is a pre-trained language models were used. The results of the models are compared with each other.

Covid-19, which started in Wuhan in 2019 and spread almost all over the world in 2020, continues its effect in 2022. For this reason, a topic that is up to date was preferred. Covid-19 data on Twitter platform will be used as training and test sets.

In the first chapter, the sentiment analysis levels and the methods used in sentiment analysis are explained. Machine learning techniques and dictionary-based techniques used in sentiment analysis are mentioned. In addition, the learning methods of classification are explained. In the second chapter, the studies on sentiment analysis in the literature are included. These studies are grouped into classification-based, dictionary-based, and pre-trained language models. In the third chapter, there is Covid-19 sentiment analysis. Sentiment analysis was performed on Covid-19 tweet texts by using pre-trained language models. Naive Bayes, Logistic Regression and Random Forest classifiers and BERT, RoBERTa and BERTweet language models were used to train different classifiers. Finally, the last section contains discussion, and conclusion about this analysis.

1. SENTIMENT ANALYSIS THEORETICAL BACKGROUND

Sentiment analysis, or otherwise known as opinion mining, is the process of detecting and interpreting an emotion in a sentence, document or statement. The first studies in the literature work under the information retrieval area. It is understood that later on, opinion mining works were started. In general, sentimental polarity is the most studied application in the literature. The main purpose is to try to analyze the emotional words in the texts by dividing them into two groups as positive or negative. In this context, the chronological development of sentiment analysis can be shown as follows:

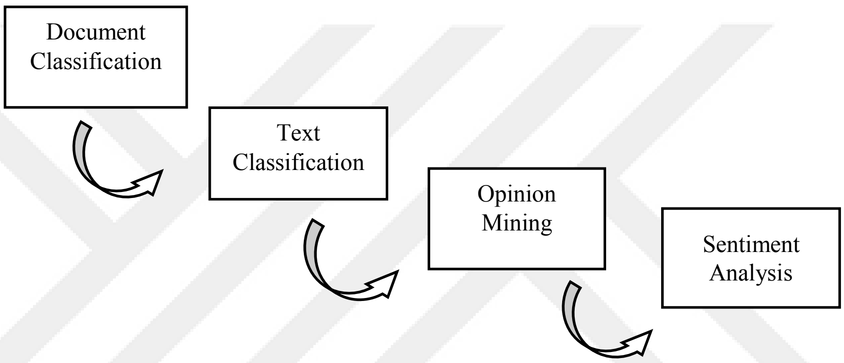

Figure 1.1 Sentiment analysis evolution (Uçan, 2014: 3)

Sentiment analysis or opinion analysis is the computational examination of the opinion, emotion and subjectivity of the text. Sentiment analysis and opinion analysis is not just natural language processing. While doing sentiment analysis, it is understood that there are research topics such as data mining, web mining and text processing (Mundalik, 2018). Sentiment analysis process is as follows:

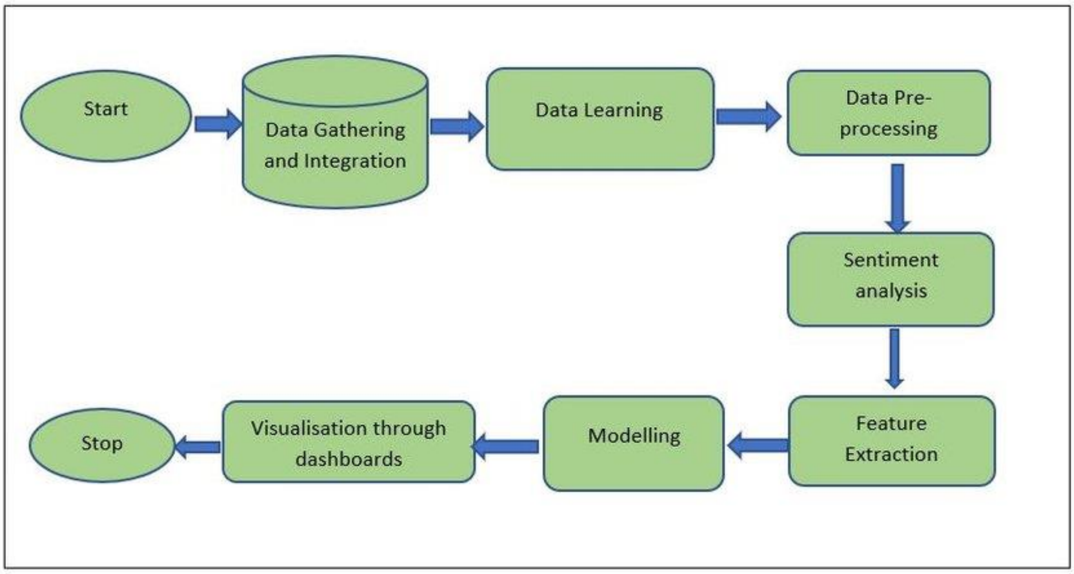

Figure 1.2 Aspect based sentiment analysis process flow diagram (Mundalik, 2018: 11)

When the studies in the literature are examined, it is seen that sentiment analysis focuses on three approaches. These approaches are statistical methods, word vectors and lexical work. Statistical methods are usually machine learning methods such as Naïve Bayes, Support Vector Machine or different algorithms. In a lexical study, emotions are scored with the help of a ready-made sentiment dictionary. Word vectors is the process of classifying text into words. In the studies, it is seen that sometimes single method and sometimes mixed methods are used.

1.1 Sentiment Analysis Levels

Natural Language Processing (NLP) and Information Extraction (IE) aimed at separating and classifying the feelings expressed by the author in positive or negative comments by analyzing a large number of documents in general. Computational linguistics and Information retrieval (IR) techniques are combined for this (Sharma, Nigam & Jain, 2014).

Sentiment analysis studies are carried out at three different study levels. These are separated according to the information obtained from the inferences of the texts in the dataset and their scope.

1.1.1 Document-Level Sentiment Analysis

Document-level is a sentiment analysis based on uncontrolled approach that determines the sensitivity aspect of documents. Text with comments at this level is classified as positive or negative (Araque et al., 2017, Pang et al., 2002; Sharma et al., 2014).

Sometimes positive and negative opinions can be mixed up in the text. However, details are not necessary at this level. The text as a whole is classified as positive or negative.

1.1.2 Sentence-Level Sentiment Analysis

In the sentence-level sentiment analysis, each sentence appears as a separate unit. It is assumed that the sentence should contain only one opinion. This sentiment analysis has two tasks which are subjectivity classification of sensitivity analysis and the sensitivity classification.

The goal of feature level classification is generating a feature-based opinion summary. Both object properties and polarity of view is determined at this level (Jagtap & Pawar, 2013). Sentences with consensus are used to detect subjective sentences. The presence of word elements or semantic frames are used (Chaturvedi, Cambria, Welsch & Herrera, 2018: 68).

1.1.3 Aspect-Level Sentiment Analysis

Aspect-based sentiment analysis is a text analysis technique that separates text in different directions. It allocates to understand that the components of the product or service are accepted (Patil, Phansalkar & Kryssanov, 2018). Aspect-level sentiment analysis generally has two tasks. One of them is aspect extraction and the other one is aspect level sensitivity analysis. Aspect extraction tags certain words in a sentence of "n" words. The words tagged represent the beginning and non-beginning words of aspect terms, respectively. According to the information about words tagged obtained, the task is classified into appropriate emotion topics (Majumder, et al., 2020).

1.2 Approach Used in Sentiment Analysis

Natural Language Processing (NLP) is a commonly used method for preparing text for machine learning. Lots of text documents in various formats have missing characters, typos or words to filter. For this purpose, techniques such as noise removal, lexicon normalization and object standardization are used. Then, relevant topic is determined from the text with entity extraction.

Sentiment analysis is mostly used in two techniques which are statistical methods and natural language processing. One of the techniques developed specifically for natural language processing is sentiment analysis based on semantic results (Schuetze & Christopher Manning, 1999). The other one is to obtain and interpret numerical values of texts with statistical inferences (Jurafsky & Martin, 2008). The methods can be shown as follows:

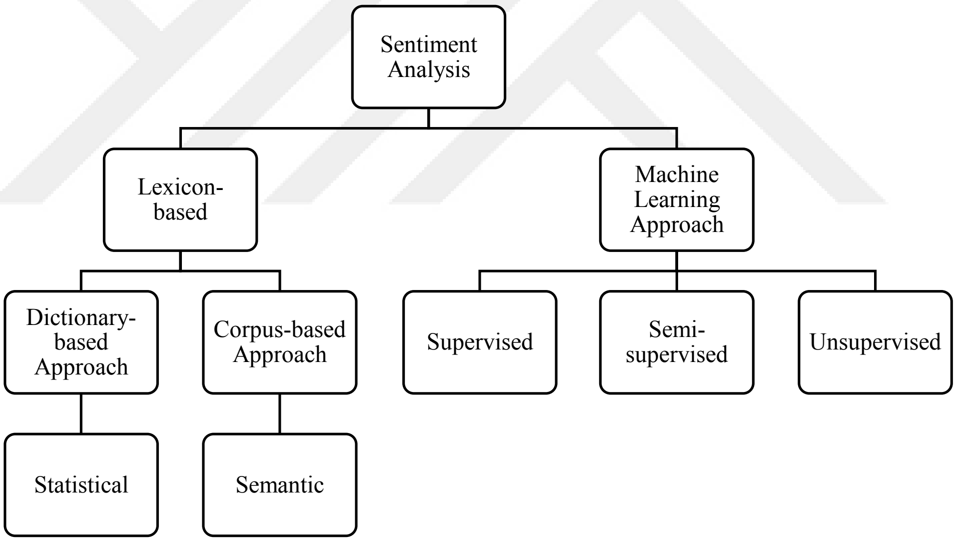

Figure 1.3 Sentiment classification techniques (Medhat et al., 2014)

One or more of the analyzes shown in the figure can be used in the same research. Sometimes both Lexicon-based and machine learning approach methods can be used to increase the accuracy of the study. The most commonly used analysis methods in sentiment analysis are as follows:

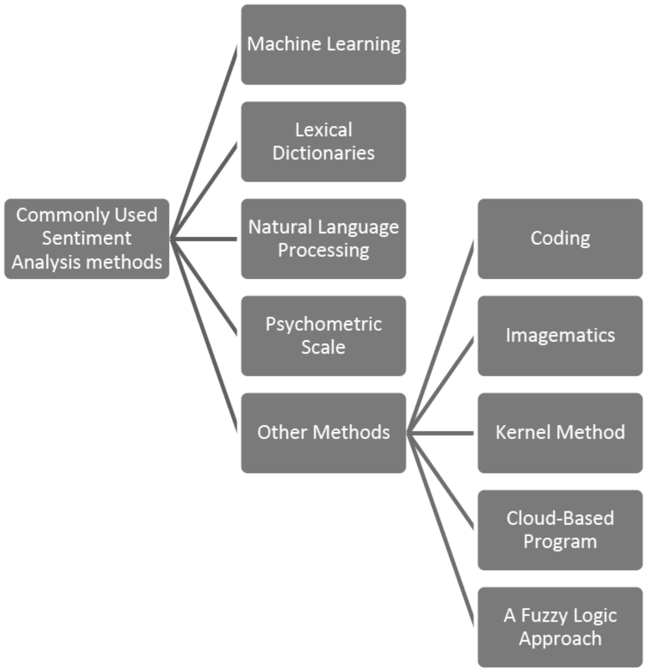

Figure 1.4 Commonly used sentiment analysis methods (Yuan, 2017)

1.2.1 Machine Learning Approach (ML)

Machine learning is a guide for algorithms and applications. "*Field of study that gives computers the ability to learn without being explicitly programmed*" (Samuel, 1959). The word meaning is sometimes confused with artificial intelligence, but it is predictive analytics or predictive modeling. As new data is sent to the algorithms, the computers learn and optimize the operation to develop intelligence.

The texts are grouped as positive, negative and neutral for sentiment analysis. Then, training dataset is created from these data. After that, operations such as cleaning symbols, punctuation marks in messages, separating the text into words, finding word roots, creating term frequencies are performed. Some of these operations are done by machine learning methods.

The methods used in machine learning are divided into supervised, semi-supervised and unsupervised. Firstly, in the supervised classification, a model is prepared and corrected when the estimates made are wrong. Correction continues up to the level of accuracy. Secondly, in semi-supervised classification, there is an expectation of a prediction model, but the model must learn to organize the data. The algorithms prepared are extensions of other flexible methods that make assumptions about how untagged data will be modeled. Thirdly, in unsupervised classification, entries in the dataset are not labeled. A model is prepared from the data. Mathematical operations are applied to reduce redundancy. After this processing, algorithms are organized by clustering, dimensional reduction, and association. (Brownlee, 2019).

The most commonly used machine learning algorithms; Regression Algorithms, Instance-based Algorithms, Regularization Algorithms, Decision Tree Algorithms, Bayesian Algorithms, Clustering Algorithms, Association Rule Learning, Artificial Neural Network Algorithms, Deep Learning Algorithms, Dimensionality Reduction Algorithms and Ensemble Algorithms (Pang, et al., 2002; Medhat, et al., 2014). Unsupervised machine learning approaches are PMI – Pointwise Mutual Information, LSI – Latent Semantic Indexing and LDA - Latent Dirichlet Allocation. Some of the algorithms frequently used in sentiment analysis are explained as follows.

1.2.1.1 Naïve Bayes (NB)

Naïve Bayes is one of the supervised learning-classification methods. The Naïve Bayes classifier is based on Bayes' theorem. It is used frequently because it is a classification method with high accuracy. Each value is classified independently from other values. It allows to predict a class based on a specific set of properties (Berrar, 2018: 403). It is a statistical method that calculates the probability of a new data entering any of the existing classes using data classified in the Bayesian Network. The method uses the Bayesian rule in probability theory (Heckerman, et al., 1995). Naïve Bayes method is used for sentiment analysis (Jurafsky & Martin, 2019; Suppala & Rao, 2019; Matharasi & Senthilrajan, 2017). In these studies, datasets are generally tested, and Bayesian method is used for the accuracy of classification.

1.2.1.2 k-Nearest Neighbors (k-NN)

This method is one of the methods used to categorize ungrouped data. The algorithm is run to find the groups in the dataset with the number of groups represented by the variable ‘k’. Then the detected data point is repeated to assign it to one of the ‘k’ groups (Myatt, 2007a: 179). Choosing a small value of ‘k’ causes the data to be assigned to the class of its closest neighbors and selecting it large causes the data to be evaluated. Therefore, the efficiency of the algorithm may vary according to the k value (Karaöz, 2018).

Three components are selected while applying Weighted k-NN algorithm. These are the weight vector, nearest neighbor number and distance metric. The general purpose of k-NN studies is to find the most appropriate dataset value that overlooks the characteristics of the data points that we want to predict the specific structure and labels of the dataset (Anava & Levy, 2016).

While doing sentiment analysis, the k-Nearest Neighbors method is used (Huq, Ali & Rahman, 2017; Bayhaqy et al., 2018; Toçoğlu et al., 2019). It is seen that different algorithms such as Random Forest, SVM, Artificial Neural Networks are included in the study together with k-Nearest Neighbors algorithms in studies where sentiment analysis was performed.

1.2.1.3 Support Vector Machines (SVM)

Boser et al. (1992) developed the basic learning theory proposed by Vapnik and Chervonenkis in previous years as a supervised learning method used in classification and regression analysis. SVM is one of the types of supervised linear classifier learning used to categorize the dataset and analyze the response. It is widely used in analyzing data and classifying text. Features such as unigrams or bigrams are used in research for text classification. SVM classifiers are created on the dataset without tags or tags using these features (Pang, Lee & Vaithyanathan, 2002).

The selected text is evaluated with the training data. The SVM algorithm creates a model with points on the linear representing the data. When mapped with test data, the area where the lower plane falls is classified (Taboada, Brooke, Tofiloski, Voll & Stede, 2011). For example, Brooke (2009) worked with a movie dataset where he extracted 100 unigrams of negative / positive features from the SVM classifier reaching 85.1% accuracy. Characteristics such as waste and disorganization were negative, and laughter and pleasure were selected as positive. Other features cannot be explained.

1.2.1.4 Maximum Entropy (ME)

Maximum Entropy models are one of the probabilistic methods that provide machine learning for classification and prediction, which are frequently used in areas such as econometrics and computer vision. This method was originally used in astronomical image restoration, was later used in different research areas such as linguistics, meteorology, and data processing (Yonamato, 2013: 1663). In natural language processing, the ME method can be used in sentence boundary detection, speech tagging, parsing and uncertainty solving (Latha, Varma & Govardhan, 2013).

In sentiment analysis, this method is applied with mixing models for classification. The mix model assumes that each class is a component of the mix. Each admixture component shows a generative model that provides the sampling probability of the specified term (Kiprono & Abade, 2016).

1.2.1.5 Random Forest

Random Forest is one of the machine learning methods that aims to store and develop batch classification trees. It creates a set of classifiers instead of a classifier. It classifies new data points according to estimates. Random forest is used for the same purpose as the Bagging and Boosting method (Bahrawi, 2019). The most important advantage of Random Forest is to be used for both classification and regression problems that make up machine learning systems.

In a very short time, it has become a data analysis tool used among other standard methods from fields such as microbiology and epidemiology. In estimators, each tree is dependent on the individual pattern values of random vectors and they are evenly distributed. It is a classifier with the structure {h (x, Ok), k = 1…}, which is one of the tree structured classifiers with {Ok} vectors that are independently distributed (Breiman, 2001).

It is seen that the Random Forest method is frequently used in studies where sentiment analysis is conducted (Bahrawi, 2019; Parmar, Bhanderi & Shah, 2014; Munshi, Sapra, & Arvindhan, 2020).

1.2.2 Lexicon-based Approach

In traditional structures, people ask their surroundings for help when they need information or ideas. Also, companies rely on surveys, consultants or public information. Big data is growing rapidly today, and it is not possible to follow millions of people every day from here. Therefore, different methods are needed to analyze sensitivity using unstructured texts on social media.

NLP is perceiving texts and sound waves in natural languages by computer systems and transferring them to the computer environment. It refers to a process consisting of certain steps. These are lexical analysis, syntactic analysis, disclosure integration, and pragmatic analysis (Data Science Earth, 2020). Firstly, lexical analysis means the collection of words and expressions in a language. In this way, all the chunks of a text are divided into paragraphs, sentences and words. Secondly, syntactic analysis allows words to be analyzed and edited to show the relationship between words. Thirdly, disclosure integration is the integration of meanings between sentences. Finally, pragmatic analysis undertakes the task of deriving linguistic aspects of synthesis that require real world knowledge (Şeker, 2015: 15).

Some methods are used to convert texts into numerical vectors. It is not common to use these character or word-based methods alone. However, they are the methods used in the analysis processes. Some of these methods are effective when analyzing sentiments, while others increase the success of the analysis. The most important of these methods can be explained as follows.

1.2.2.1 Bag-of-Words (BOW)

This model is a simplifying representation used in natural language processing. Bag-of-Words enables multiple words in a sentence to be evaluated as a single word. For this purpose, it is necessary to teach the computer phrases such as nouns and adjectives. As seen below in this way, it is easier to classify words in sentence queries.

Figure 1.5 Bag of word, bigrams and unigrams (Majahan, 2021)

1.2.2.2 Named Entity Recognition (NER) (Syntactic analysis)

NER is the process of removing predefined categories such as location and organization from text documents (Nadeau & Sekine, 2007). Enamex, numex, and timex are used to define the person, percentage, number or time. In named entity recognition studies, operations such as extracting the targeted word groups from the text, counting, determining their density and labeling are performed. For example, the word "Michael" could be a male name in America or the name of one of the archangels in the Bible. Different data labeling formats such as Raw, IOB, IOB2, BILOU are used to make this distinction. Although it does not create a positive or negative effect in sentiment analysis, it increases the success rate.

1.2.2.3 Optical Character Recognition (OCR) (Semantic analysis)

It is the stage of converting documents such as digital photographs, pdf files or paper documents into digital format. The main purpose is to detect text in images using different algorithms (İlkbahar, 2019). Many documents are difficult and time consuming to read and process (Ilango, 2019). This method increases productivity.

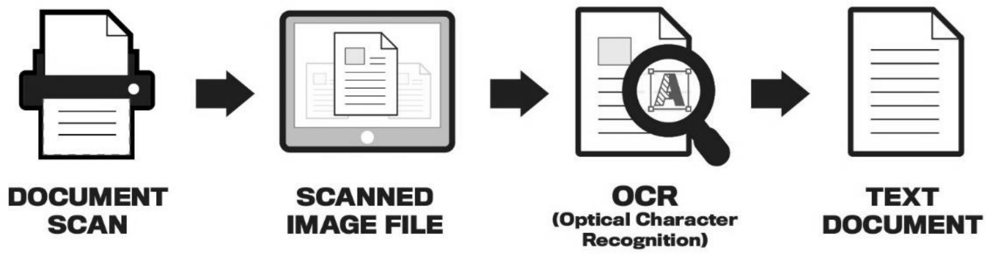

Figure 1.6 Paper document scanned and converted into a document using OCR (Ilango, 2019)

1.2.2.4 Seed Words

The Seed Words method is the words added to the dataset after the emotional expressions to increase the success of the analysis. It is established to increase the success of prepared systems. Seed words enable individuals to correct the variable that should be evaluated, by external intervention, instead of changing the data in the training set.

Controlled approaches require labeled polar words that are manually selected (seed words). However, the subjective nature of the manually selected word may cause poor performance. For example, the word bullish is a neutral word, but the bullish in the stock market has a positive meaning. Therefore, it is important to use a large number of common seed words when using such algorithms (Yu, Deng & li, 2013, pp. 855).

1.2.2.5 Part-of-Speech (POS) Tagging

POS-Tag is tagged according to the language structure of the words obtained from documents. Then a dataset is created from the tagged words. Classification is made from sentences queried according to tags. Rule-based POS taggers use contextual information to assign tags to unknown or ambiguous words. Emotion Tags are used for labeling emotions expressed in sign language. Emotion tags used in social media platforms are sometimes successful in expressing an emotion.

In this method, words are tagged one by one with the post-tag (POS) method and scored separately (Figure 1.7) (Patel, 2020). These scores are called prior polarity. In this way, words are associated with other words and sentiment scores are created. After the scores are calculated, it is determined what emotion it contains (Yu & Li, 2017).

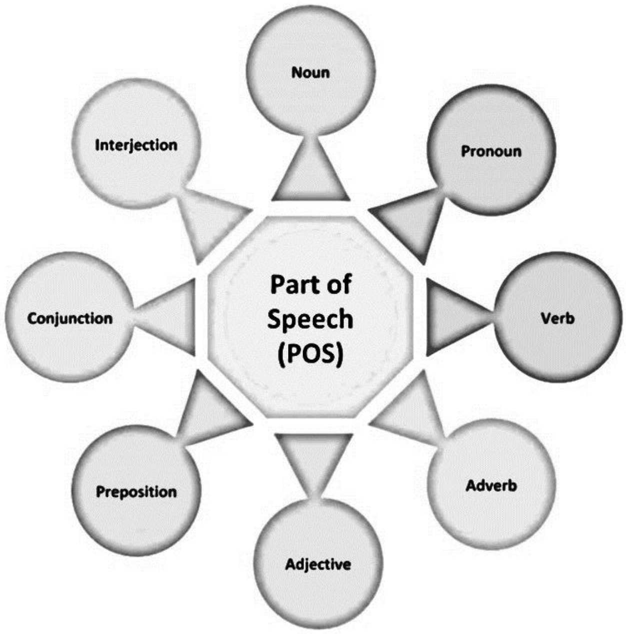

Figure 1.7 POS Tagging and chucking process in NLP using NLTK (Patel, 2020)

There are multiple elements in a sentence, such as subject, preposition, adverb, etc. For this reason, it is difficult to tag speech fragments. In addition, the usage of POS tags according to languages (English, German, Turkish etc.) is also different. POS tags also differ due to language. Some researchers have proposed a universal set of labels (Petrov, et al., 2012). However, universal addicts (Nivre, et al., 2016) preferred to use existing labels (Ruder, 2019: 39).

Syntactic parsing or sentiment analysis is frequently used in Natural Language Processing (NLP) methods. Because the correct (POS) tag information must be prepared for a specific word string. This need can be met with POS labeling. (Neunerdt et al., 2013).

2 PREVIOUS SENTIMENT ANALYSIS WORKS

Sentiment analysis, also called idea mining, is a field of study that analyzes people's views. Simple expressions are generally used to analyze the comments of the product or service in the studies in this field. However, due to differences such as language, culture, and other meanings of words, there may be problems in sentiment analysis. According to Liu (2012) on the issue, sentiment analysis does not show exactly what people like or dislike at the document or sentence level. To learn this, an aspect-based sentiment analysis (ABSA) should be done.

The main purpose is to separate subjective judgments from the text. Subjective judgment indicates the positive / negative mood class and a goal to which the sentiment is related. However, it shows that the methods (including new and different methods) used in sentiment analysis studies, especially in recent years, lead to correct results. In the historical process, there are many studies related to sentiment analysis in the literature. Part of them analyzed by machine learning techniques and dictionary-based models.

2.1 Classification-based Approach at Sentiment Analysis

Considering the sentiment analysis literature, Pang, Lee & Shivakumar (2002) can be given as an example for the first studies. In this study, a classification is made with controlled machine learning algorithms using unigram, bigram, POS tag and their combinations. As their dataset, they used the movie comments from IMDb. Comments are marked as positive and negative. They then concluded that sentiment classification is much more difficult than traditional subject classification. According to them, the best classification performance is provided by support vector machine algorithm.

In the same period, Turney & Littman (2002) listed positive terms as good, beautiful, excellent in the study, while they listed negative words as poor, negative, and false. PMI is the Pointwise Mutual Information degree for word t with each seed word t_i as a measure of their semantic association. The semantic orientation 't' here is calculated as follows:

$$O(t) = \sum_{t_i \in S_p} \mathrm{PMI}(t, t_i) - \sum_{t_i \in S_n} \mathrm{PMI}(t, t_i)$$

Nasukawa & Yi (2003) demonstrated a sentiment analysis approach with positive and negative poles for specific issues in the document, rather than categorizing the document as positive and negative. According to the distribution of the sensitivity words, 969 positive, 1.495 negative and only 1 neutral pole were determined. Thus, the polar analysis of 2465 documents were completed.

Hu & Liu (2004) proposed a system for product characteristics, which were evaluated by customer opinions. Using NLP techniques, they classified sentences as positive and negative. Its accuracy has also been proven thanks to the sentence polarity in the analyzes.

After these studies, it can be said that interest in sentiment analysis research has started to increase. Blitzer, Dredze & Pereira (2007) developed a different system for field adaptation in sentiment classification. They created a dataset consisting of four different product categories with the Structural Correspondence Learning (SCL) algorithm and analyzed this dataset. Prabowo & Thelwall (2009) suggested a hybrid system in the form of Rule-Based Classifier (RBC), Statistics Based Classifier (SBC), General Inquirer Based Classifier (GIBC) and SVM classifier.

Go, Huang & Bhayani (2009) applied sentiment analysis on Twitter. Remote controlled learning algorithm was used. The results were obtained by obtaining mutual information of the Naive Bayes algorithm and the unigram.

Martineau & Finin (2009) proposed a different version of the frequency (TF) * inverse document frequency (IDF) in the Delta TF * IDF weighting system to validate the classifications made in sentiment analysis.

Çelikyılmaz, Tür & Feng (2010) also conducted a sentiment analysis study on Twitter. In the study, they divided tweets into polar and non-polar forms. Accordingly, positive and negative tweets were determined. In this study, it was also argued that a broad polar lexicon is not needed to understand people's emotions.

Zhang, Ye, Zhang & Li (2011) studied thoughts written in Cantonese about restaurants. Naive Bayes and SVM machine learning techniques were used to distinguish positive and negative writings of the users. In the study, it was determined that Bayes classifier is better than SVM. The rate of negative thinking was 16.8%, negative thinking was 12.9% and 70.3 not for restaurant policy thoughts. It was also found that women (62.5%) shared more opinions than men (37.5%).

Zhang, Fuehres & Gloor (2011) analyzed the positive and negative moods of the masses on Twitter by comparing them with stock market indices such as Dow Jones, S&P 500, NASDAQ. It has been analyzed whether the tweets about the stock market contain fear or hope. Then, the correlations of these values with stock market indicators were examined. After the findings, a model has been developed that allows forecasting for stockbrokers.

Choy, Cheong, Laik & Shung (2011) study was also conducted on Twitter. This data was analyzed in the study conducted to predict the Presidential election result. They used an econometric model to model the 2011 Presidential Election. Based on the findings, it was estimated that there would be a small difference in votes between the two candidates, but the candidate could not be determined. The accuracy of the information obtained has been demonstrated.

Bollen, Mao & Zeng (2011) conducted sentiment analysis to examine individual behavior in terms of behavioral economics. The Dow Jones (DJIA) comments on the subject on Twitter have examined its relationship with the average value in the industrial area. Classified in 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). Granger causality analysis was performed, and Self-Organizing Fuzzy Neural Network was used. In the study, it was concluded that DJIA estimates can be improved with mood dimensions.

Kouloumpis, Wilson & Moore (2011) compared microblog-based, n-gram and word-based features on three different tweet datasets. Unigram and bigram were used. In addition, hashtags and POS characteristics representing sentiment were also taken into account. In addition, features have been added to capture some of the more domain-specific languages of micro-blogging. According to the findings of the study, it was understood that lexicon-based and microblog-based qualities were effective in determining emotion. Word type information did not show the same effect.

Pennacchiotti & Popescu (2011) tried to identify political relations, ethnic origins and interest in business areas by using Twitter users' behavior, network structures and linguistic knowledge. As a result of the analysis of machine learning methods, rich linguistic features are also valuable in different tasks.

Mittal & Goel (2012) examined the relationship between market data and general sentiments of the society on the dataset of user comments and financial sector information on Twitter. Self-Organizing Fuzzy Neural Network was used in the study. It has been observed that there is a positive process in the market when there are tweets containing happiness and calmness.

Wang, Chen, Thirunarayan & Sheth (2012) classified tweets into seven different emotions to analyze sentiments on Twitter. An average of 2.5 million tweets was used as a dataset. Some of them are positive (joy, surprise, love, etc.) and some are negative (anger, sadness, fear, etc.) emotions. Two different machinery learning algorithms were applied for sentiment identification. In the study, it was observed that emotion-carrying words are effective in confusing emotions.

Becker, Erhart, Skiba & Matula (2013) tested the polarity lexicon methods produced by Polarity Bag-of-Word, Pos Tag methods with SVM on 475 thousand untagged SMS and tweet data belonging to the years 2012-2013. High success was achieved in binary classifications, while results remained lower in neutral words.

Ghiassi, Skinner & Zimbra (2013) created a dataset of customer comments about a brand on Twitter. They made sentiment analysis with algorithms they developed themselves. A controlled feature reduction approach has been developed using n-grams and statistical analysis to create a Twitter specific dictionary.

He, Zha & Li (2013) tried to determine customer satisfaction with the data collected from Facebook and Twitter of three companies in the pizza industry. Analyses were performed with the SPSS Clementine text mining tool and NVivo mining. With these analyses, it was possible to analyze the sentiment in messages along with numerical values such as followers, comments, and shares. In this way, the competition analysis made provides benefits for companies.

Aston, Liddle & Hu (2014) decided to make sentiment analysis about tweets on a specific subject with Twitter becoming popular. Several algorithms have been proposed that classify the sensitivity of tweets in the data stream. An error rate as low as 0.23 and an F score as high as 0.78 were obtained for the detection of positive or negative sentiments in subjective tweets.

Nikfarjam et al., (2015) investigated the negative opinions of patients about the side effects of drugs by applying sentiment analysis on Twitter and comments on health forums. Machine learning based ADRMine method was used. High success has been achieved in classification.

Katz et al., (2015) proposed the ConSent (Context-Based Sentiment Analysis) model, which is a different sentiment analysis model. In their work, they compared this model with Naïve Bayes and SVM methods. However, the ConSent model has been less successful than the others.

Upadhyay & Singh (2016) conducted research on Twitter expressing opinions about electronic products. In the study, after separating the unnecessary words such as names and symbols, each thoughtful, thoughtless tweet was compared with the database of positive, negative and neutral words. Naive Bayes, SVM and maximum entropy were used as methods.

Dehkharghani et al., (2016) conducted a sentiment analysis on criticism of Turkish films. Based on the findings obtained, the accuracy varies between 60% and 79% at different analysis levels.

In the study of Kang, Wang, Zhang & Zhou (2017), opinions about school meals for the prevention of obesity in childhood were investigated. In the dataset consisting of 14.317 tweets, machine learning techniques were used to classify tweets as positive, negative and neutral. It was found that the opinions about the school food policy were mostly negative. The findings obtained created an opportunity for schools to cooperate with the society. It has also enabled gender or regional differences to be identified.

Albayrak et al. (2017), made sentiment analysis on tweets about paid military service. In addition, the dataset was compared with the SentiTurkNet dataset. The results showed that people were neither positive nor negative about it.

Onan (2017) utilized Naïve Bayes, SVM and logistic regression methods in the classification of one month of Turkish Twitter messages. The study also received support from the Zemberek library. The highest achievement was achieved in the Naïve Bayes method.

Ruhrberg et al., (2018), the purpose of their study is to analyze the tweets on the Islamic State on Twitter to categorize them. A python tool has been developed that interacts with the Twitter Streaming API to receive Islamic State related Tweets. Sentiment analysis of tweets was made by a tool invented by Janina Nikolic. The findings showed that most of the tweets have a negative attitude towards the Islamic State. It is understood that the number of neutral or positive tweets is in a minority.

In study of Ghag & Shah (2018), Bag-of-Words approach was used for sentiment analysis. The study focused on the classification of emotions, taking into account the syntactic and semantic structure of the sentences in the review. The performance of Conceptual Sentiment Analysis Model, one of the methods applied in the study, showed the highest performance compared to other techniques.

Yılmaz & Orman (2021) has conducted a sentiment analysis of tweets related to Covid-19 on Twitter. The LTSM system, which is one of the deep learning methods, has been utilized in the study. This study achieved a 97% success rate.

2.2 Lexicon-Based Approach at Sentiment Analysis

It is assumed that the words in the meaning of a text are associated with the fact that they are negative and positive poles in lexicon-based approaches. This relationship is related to adverb, noun, verbs and adjectives (Catelli, et al., 2022). In sentiment analysis, the lexicon is like a prerequisite. These lexicons which are defined as the emotion lexicon are constantly being renewed or created over time. Sentiment lexicon is the analysis of emotion in words containing expressions of emotion. When creating a sentiment lexicon, the words created before are usually examples. The first lexicon is "*Harvard General Inquirer*" founded in 1966 (Stone, et al., 1966). It could be called the first comprehensive lexicon at that time. Then, Hu & Liu, a different emotion lexicon was created in 2004. In addition, a lexicon of 8000 words has been created (Wiebe, et al., 2005).

Since these lexicons are not sufficient, research has increased for more comprehensive lexicons. The most commonly used lexicons are WordNet, SentiWordNet, MPQA Subjectivity Lexicon, SenticNet, and SentiFul.

There are many advantages to using lexicon-based techniques (Taboada, et al., 2011). For example, linguistic content can be classified through mechanisms that take into account both intensifiers (such as very sad) and negativities (not very happy) (Polanyi & Zaenen, 2006). Also, lexical entities can be separated according to their sensitivity dimension and language-dependent features can be included in these approaches.

However, lexicon-based approaches also have disadvantages. For example, the variability of opinion words between domains (Turney, 2002), contexts (Ding, Liu & Yu, 2008), and languages (Perez-Rosas, Banea & Mihalcea, 2012) call for a consistent and reliable lexicon (Taboada et al., 2011). However, dependencies make it difficult to maintain domain-independent lexicons (Qiu, Liu, Bu & Chen, 2009). Lexicon-based approaches are inherently unsupervised. Therefore, an unsupervised learning algorithm is used to classify (Wang & Araki, 2007). In this way, the relationship between words is used. These data are used while analyzing the sentiment.

The dissemination of this method has also led to sentiment analysis in different languages. For example, Kanayama & Nasukawa (2006) and Abbasi et al., (2009) conducted a sentiment analysis of movie criticism that different languages (Chiavetta et al., 2016).

WordNet is the first thing that comes to mind for lexicon used in natural language processing. WordNet is a database developed for many years with the contributions of volunteer employees. This database was created with the contribution of psychology, natural language processing and linguists (Fellbaum, 1990). Nouns, adjectives, adverbs and verbs are grouped with cognitive, most meaningful clusters that express a different concept. These groups are called synset. Synset is embedded in a large and hierarchical ontology. WordNet assigns words to sets of synonyms that have equal sense. The meaning in the lexicon is the meaning of a word defined in WordNet. In short, the meaning of each word is in a different match (Hanks &

Pustejovsky, 2005). SentiWordNet is another developed lexicon. SentiWordNet is a dictionary that uses the lexicon database (WordNet).

The most common relationship in the lexicon is the hyponym such as milk and cheese. There is a part-whole relationship like the eye is part of the head (meronymy) and sub-action relationship such as crying and sadness, in addition there is a cause-effect relationship such as thinking and talking. There are also derivative adjectives such as eye-observer. SentiWordNet describes all synsets of WordNet according to positive, neutral and negative concepts. It contains 117.659 synonym sets (Baccianella, et al., 2010). The sentiment analysis score is calculated by subtracting the percentage of negative words from the percentage of positive words (Singh & Gupta, 2016). Synset scores are determined by 8 triple classifiers.

```
great  0.08973680845374968
good  0.3883563601675836
better  0.5480991330178288
chinese  0.0
such  -0.041666666666666664
common  -0.021350496905303964
low  -0.07680417032349182
big  0.08006682443206818
worth  0.15306122448979592
excellent  1.0
good  0.3883563601675836
big  0.08006682443206818
good  0.3883563601675836
easily  0.1715328467153285
less  -0.06802721088435375
possible  0.21532846715328466
okay-type  0.0
much  0.1157468243539203
friendly  0.1398071625344353
```

Figure 2.1 Sample polarity score from SentiWordNet dictionary (Singh & Gupta, 2016)

The sentiment analysis pipeline process performed through SentiWordNet from the document selected as the dataset is as follows:

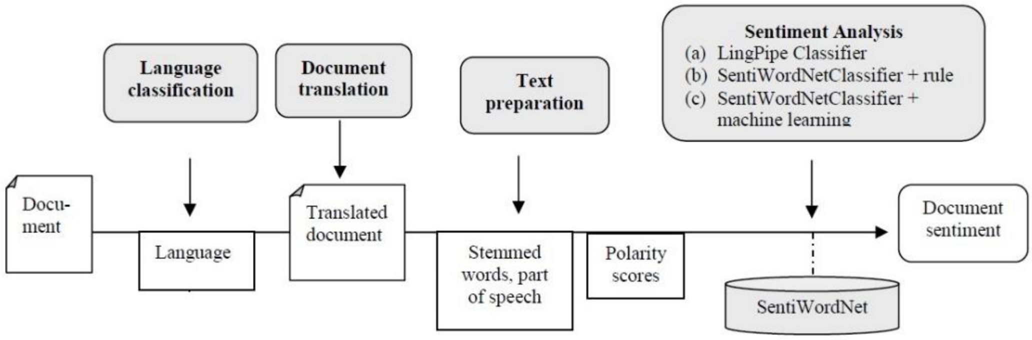

Figure 2.2 Processing pipeline for sentiment analysis (Denecke, 2008)

The disadvantage of the SentiWordNet lexicon is that it is in English. It has been translated into some similar language structures such as Italian, French and German. However, it is difficult to translate for different language structures such as Turkish (Esuli & Sebastiani, 2006; Basile, 2013). New lexicons written for use in Turkish are being developed. However, similar methods are used in all lexicons used in natural language processing.

Another lexicon similar to SentiWordNet is SentiWords. It actually derives from SentiWordNet. The most important difference is that it assigns scores directly to words instead of synthesis. In this way, a lexicon can be used without clarifying the document text. (Francesco, 2020).

One of the dictionaries generally used on texts on social media is VADER (Valence Aware Dictionary and Sentiment Reasoner) dictionary. It is a lexicon with less scope (almost 7000 words) than others that are created manually. However, the difference from the others includes emojis and abbreviations that are not in lexicons (Francesco, 2020).

Developed by TÜBİTAK (2005), Zemberek won the best free software award. It is a Turkish natural language processing lexicon. Thanks to its Open Office feature, it can control Turkish language errors (Özker, 2019).

Tumasjan et al. (2011) tried to predict 2009 German elections using Google-Profile of Moods through tweets. They also used OpinionFinder to understand positive and negative emotions.

2.3 Pre-Trained Language Based Sentiment Analysis

Sentiment analysis of pretrained language models also give similar results with the methods of machine learning. Sometimes there are studies that implement mixed methods. For example, Pak & Paroubek (2010) automatically collected tweets using the Twitter API, and then developed a corpus that explains the sentiments. Tree Tagger is used for POS tagging. A sentiment analysis was made with the Multinomial Naive Bayes classifier. As a result of this analysis, it was determined that they have positive, negative or neutral emotions.

Ghag & Shah (2013) compared the techniques used for sentiment analysis in their study. In the study, a multilingual approach was emphasized, and they showed that no available technique is language independent and not a generalized sentiment analyzer.

Mostafa (2013) tried to identify the most frequently used words next to the company names. The words in the Twitter about companies from different sectors have been analyzed. Tweets from well-known brands such as Nokia, T-Mobile, IBM, KLM and DHL were used as datasets. A predefined dictionary containing around 6800 seed adjectives was used in the analysis. It has been argued that the results obtained reflect the correct information about attitudes towards brands.

In a study in which 120 million clinical texts at the University of Tokyo Hospital were datasets, the BERT base was used. It is concluded that the advantage of training on domain-specific texts can become apparent in more complex tasks on the actual clinical text (Kawazoe, 2020).

Sentiment analysis was performed on Turkish texts with a dictionary-based approach. 63 different human sentences are classified. The SentiTurkNet dictionary was used in the research. In order to find the roots of words, the Zemberek module was used (Yoldaş, 2021).

Sel & Hanbay (2021) were used BERT, DistilBERT and Electra which are machine learning methods and a pre-trained language model. In the analysis made on Turkish tweets, the highest performance was obtained in the BERT model.

Masarifoglu et al. (2021) analyzed the comments collected from customers using banking services using NLP methods. BERT and XLM-ROBERTA were used.

With the rapid increase of public opinion data, Weibo text sentiment analysis technology has also gained importance. For this purpose, Li et al. (2021) proposed a new model based on BERT and deep learning for Weibo text sentiment analysis. It is understood that the performance of a comparative experiment conducted on the Weibo text dataset collected during the COVID-19 pandemic was higher.

In addition to these studies, there are also studies on creating the new dictionary or different dictionary training. For example, Cruz et al. (2016) created a sentiment word from seed words using AcroDefMI3 to create a sentiment dictionary. Then it is compared with SentiWordNet.

3 COVID-19 SENTIMENT ANALYZER

3.1 Motivation

The basic structure of the current development of social media is the invention of the internet. The world met the internet with the invention of Tim Berners-Lee (1989). Originally, the purpose of web technologies was only to deliver content. Later, it was possible to link data such as text, images, videos with these pages. The active participation of its users on the web page made social media start. There is a constant change in the technologies used on the internet with the rapid development of the internet. For example, the concept of semantic network has developed with the participation of search engines (Lehman, 1992). It has been possible to integrate millions of independent data on the Internet. After the algorithm developed by Google (1998) to rank and classify related pages, it became easier to search for content on the web. In this way, social networks on the network have become more meaningful.

The result of these developments is the use of social media, which is one of the most used areas in daily life. The most basic features of social media are communication, collaboration and content sharing (Anttiroiko & Savolainen, 2011). Individuals create their own profile and participate in online platforms that allow instant communication. Feedback and participation are welcome on social media services. Voting, commenting and sharing information is easy. Creating content (if there is no barrier) is possible. Two-way communication can be established. People who are active on social media act as members of a group. In addition, a sense of loyalty also develops in people who are active (Mayfield, 2008).

With the increase in the use of social media, innovations began to be added to the concepts of socialization and communication. The most effective area of socialization and communication from past to present has been in the media faced with. However, it has been understood that it has been effective in virtual media or digital platforms to communicate or share feelings and thoughts in recent years.

The most used social media platforms today are platforms such as Twitter, Facebook, YouTube, LinkedIn, Instagram, Snapchat. The data shared by individuals are from social networks (Facebook, LinkedIn et al.), Blogs (Blogger, Live Journal et al.), Microblogs (Twitter, Tumblr et al.), Wiki applications (Wikipedia et al.), Social news (Digg) or YouTube, it can be obtained from multimedia shares such as Flickr, Instagram (Sinha, et al., 2012).

Twitter is a social network and microblog site founded in 2006 by Jack Dorsey. It is one of the most visited platforms in recent years. Messages called tweets consist of 140 characters. At the same time, tweets can be tagged with a # (hashtag) sign. The main complaint of users is character limitation in tweets because this rule is an obstacle to emotions and thoughts. Despite this, Twitter is currently one of the most used platforms.

Participants who tweet communicate at different levels. These include micro-level interpersonal messaging, establishing relationships at the level of visibility of the share by followers, and interacting with all twitter participants using hashtags. Tweet sharing takes place in different dimensions. Sometimes, thanks to the retweet feature of Twitter, the opinions of the participants can reach a wide audience. Thanks to the reply feature, other participants can increase the interaction by commenting (Bruns & Moe, 2016). Tweets are in different sizes. Sometimes there are posts containing a feeling and a thought. Sometimes it is promoted or criticized through a popular topic hashtag. Apart from this, compliments or complaints about products and services are shared.

Some features of Twitter make it stand out from other platform. Tweet messages are maximum 140 characters. For this reason, feelings and thoughts have to be told with 140 characters. Because it is a short message, abbreviations or words with special meanings are frequently used. Frequent updating of messages shows the users' first reaction to events. In terms of data, the amount of data is very large. Twitter API, on the other hand, facilitates data collection (Kumar & Sebastian, 2012; Maharini, 2013). Some users use Twitter to track friends (what ate in the evening, where is it on the weekend). However, these do not make sense to most users. The number of people who use it as news and event tracking is quite high. Recently, there has been a significant increase in the number of users for emergency and communication purposes (Rogers, 2016).

The spread of the Covid-19 outbreaks to the whole world, starting from China, has caused life to change on a global scale. The obligation of people to stay away from each other has led to the transition to stay home, work from home, hybrid work and education. In many countries, the fact that the lockdown came at sometimes, taking measures in places such as educational institutions, shopping malls, markets, or being closed sometimes, led people to be alone. This process has also led to the use of technological equipment, mobile conversations and video calls, and social media tools.

As we mentioned above, social media is an important resource for our research. That's why I did a sentiment analysis on Twitter. As a research topic, I chose Covid-19 tweets, which are one of the current topics, as a dataset. There are many different methods in NLP for analysis. We decided to use pre-trained models in this study and compared the performance of the models. We chose Bert-base-uncased, RoBERTa and BERTweet, which are among the alternative pre-trained models. We compared and interpreted the results.

3.2 Task Definition

Sentiment analysis is based on two concepts commonly referred to as emotion polarity and emotion score. The polarity or category of emotion is a positive or negative binary value. Emotion data dictionary is a list of statements prepared beforehand. Sentiment analysis draws a conclusion with the words that people who correspond to this phrase list convey their ideas and feelings. In this way, it enables to determine the positive and negative majority of a sentence or text (Palanisamy, Yadav, Elchuri, 2013).

Language modeling is the task of guessing the next word according to a piece of text. One of the most important recent developments in natural language processing is that a model trained for language modeling can be successfully finetuned for the flow task with small changes. The pretrained representation of words obtained by language models is often used in NLP research. In this context, we decided to apply pre-trained language models for sentiment analysis. Pretrained language models will be applied in the experiments. BERT base model, BERTweet and RoBERTa base model is a large-scale transformer model based on the pre-training and fine-tuning process. RoBERTa is an improvement over the original BERT transformer model.

BERTweet is the first model of its kind presented to other researchers for further improvements and new applications (Baker, 2021). The task of such a system is to analyze the words so that they are classified according to the most dominant sentiment or polarity contained in that text. Pretrained models are widely available in many languages. This approach eliminates the time-consuming and resource-intensive model training directly on tweets from scratch, allowing you to focus only on their fine tuning. Also produces better performance (Pota et al., 2021). Sentiment analysis is usually done in the following format:

- Obtaining the dataset
- Pre-processing the text and making it suitable for analysis,
- Selecting the attributes to be used for analysis
- Classification
- Determining the sentiment score of the words according to dictionaries,
- Calculation of total sentiment score according to each emotion category,
- Classified according to the category with the highest sentiment score.

In order to produce a successful model, the data preprocessing steps were performed as follows:

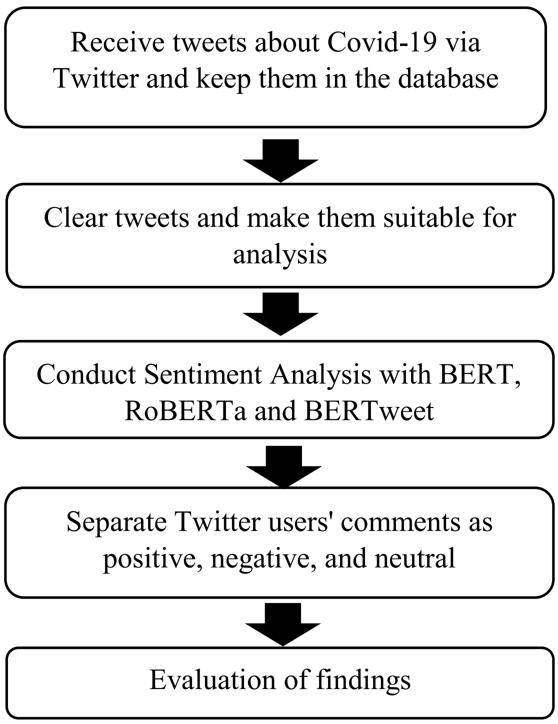

Figure 3.1 The data preprocessing steps

3.3 BERT, RoBERTa, BERTweet Pre-Trained Model

In recent years, many pretrained language models such as ELMo, ULMFIT, OpenAI transform have been developed. BERT is the latest of these models. BERT (Bidirectional Encoder Representations from Transformers) was developed by researchers at Google AI Language. Its most important feature is that it brings a new approach to NLP tasks, especially SQuAD v1.1 (question answering) and MNLI (Natural Language Inference) (Horev, 2018).

The main task of the dictionary is to find the target word whose explanation is given. BERT pretrained language model was used (Devlin, et al., 2016). Natural Language Processing (NLP) also codes information from the text sequences using a model, such as Bidirectional Encoder Representations Transformers (BERT). In this way, the result is obtained for text recognition and classification (Pota, et al., 2021, Cesconi, 2020).

Apart from being bidirectional, BERT is trained with two techniques called Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) (Uçar, 2020). In the masked language model, 15% of the words in sentences are masked randomly and it is ensured that these words are guessed correctly in a bidirectional way. The prediction of the next sentence is whether it follows the sentence before it (Çelikten & Bulut, 2021). BERT is based on byte-pair-encoding (BPE) sub word encoding. Data sparseness between words can also be reduced because it divides the sentence into sub words. Thanks to this sorting, BERT can successfully perform the appearance-based sensitivity analysis task and other tasks such as summarization (Sun, et al., 2019). For instance,

- If you keep walking like this, I think you'll drop your phone.
- I enjoyed every last drop of my coffee.

As can be seen, the same word is used in both sentences, but their meanings are different from each other. To achieve a strong understanding of language like this one, pretrained (GPT) transformers are needed. If some words are masked or hidden in a sentence, then it is important to read the sentence both from left to right and from right to left. It is one of the NLP techniques developed for BERT bidirectional encoder representatives developed by Google (Cesconi, 2020). BERT (Fine Tuning) is used for a specific task. In classifications such as sentiment analysis, a classification section is added above the transformer output.

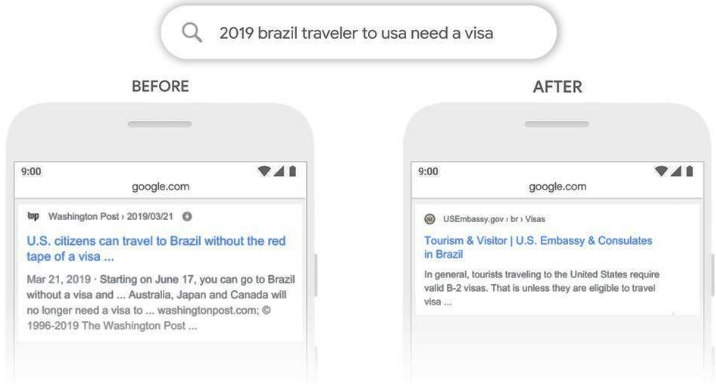

Figure 3.2 Use of BERT model in Google search engine

BERT is trained by the source cluster during pre-education. According to the frequency of use, it selects 30 thousand words and are divided into token according to the listing of the attachments (Muller, et al., 2019). The BERT language model uses a structure called a WordPiece to break down input sentences into word particles. The WordPiece structure divides each input word into sub-words called word fragments (tokens). Words with the characters as the middle point as the word particles protect their linguistic meanings. Even with a small dimensional vocabulary, it can also work in the case of fragments of words that are outside the vocabulary (Song, et al., 2020). In addition, the BERT Language model is also actively used to find the most associated query results between the words and the results set in the Google search engine (Sar, 2021). For instance, the results of a query before and after on the Google search engine developed using the BERT language model are as follows (Nayak, 2019).

According to the results of the researchers, a bidirectional trained language model can have a deeper language context than unidirectional language models. This shows that better results will be obtained in studies such as sentiment analysis (Hover, 2018). However, BERT can output contextualized representation for a word. Thus, the polysemy problem can be much relieved. In addition, the mBERT is suitable to tackle the cross-lingual reverse dictionary. Because BERT shares some sub words between different languages, there is no need to align different languages explicitly (Devlin, et al., 2016).

RoBERTa Model is a pretrained model which uses masked language modeling (MLM). This model is sensitive to lowercase and uppercase letters. So RoBERTa is a self-supervised and pretrained transformer model on English data. It allows one to learn a two-way representation of a sentence (Singh, 2021).

RoBERTa uses 160 GB of text, along with those used in BERT. Additional data includes the Common Crawl News dataset (76 GB), Web text corpus (38 GB), and Stories from the Common Crawl (31 GB). This, coupled with the fact that the 1024 V100 Tesla GPU worked for a whole day, led to ROBERTA's pretraining (Khan, 2021).

BERTweet Model is the first public large-scale language model that is pretrained for tweets in English. BERTweet is trained using the RoBERTa pre-training procedure with the same model configuration as BERT-base (Liu, et al., 2019). In the experiments, it is understood that BERTweet is performed well (Conneau, et al., 2020). It is trained with goal of modeling masked language which is same structure as BERT base model. BERTweet has achieved the most advanced performance in many Tweets NLP tasks, such as in BioBERT's Bio NLP field and SciBERT's Scientific NLP field (Nguyen, 2020).

The corpus used to pre-train BERTweet consists of 850 million English tweets (16-billion-word tokens ~ 80GB), containing 845M Tweets streamed from 01/2012 to 08/2019 and 5 million tweets related the COVID-19 pandemic (Nguyen, Vu & Nguyen, 2020). It’s pretraining procedure is based on RoBERTa (Liu et al., 2019), which optimizes the BERT pre-training approach for more robust performance. (Devlin et al., 2019).

3.4 Our Dataset

Tweets about Covid-19 have been used on Twitter. Covid-19 Sentiment Analysis Data shared on Kaggle (https://www.kaggle.com/datatattle/covid-19-nlp-text-classification) were used in the study. The data was created from tweets with the about “covid-19” for an average of 1.5 months between 03.02.2020 and 04.14.2020.

In order to improve the overall performance of all models, the dataset has been edited. There are many meaningless characters in Twitter messages such as hashtag, links, and mention. For the training set to complete the correct classification process, these characters were removed from the dataset. In addition, retweets and repetitive tweets and punctuation marks at the end of sentences have been deleted. Non-English characters have been cleared. In addition, all lowercase and uppercase letters have been corrected to lowercase. Tweets with less than five words have also been cleared. All models used tokenizer to implement pre-trained BERT language models. In this way, the BERT transformer has been fine-tuned.

The most important data for sentiment analysis inference is text data. First of all, the preprocessing step ensures that text data is cleaned of noisy or unnecessary objects. Tweets are usually written in everyday colloquial speech, so they have a noisy structure. Usually, words are misspelled or written with lengthening/shortening. In addition, it can be written in social media jargon or special emojis can be used. For this reason, in order to make attribute extraction on tweets, it is necessary to clean the texts first. In this way, a successful classifier can also be obtained. After all the preprocessing steps, clear text data was obtained as follows. The columns and their numbers in the dataset before cleaning are as follows (Figure 3.3).

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 41157 entries, 0 to 41156
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   UserName       41157 non-null  int64
 1   ScreenName     41157 non-null  int64
 2   Location       32567 non-null  object
 3   TweetAt        41157 non-null  object
 4   OriginalTweet  41157 non-null  object
 5   Sentiment      41157 non-null  object
dtypes: int64(2), object(4)
memory usage: 1.9+ MB
```

Figure 3.3 Dataframe information

As the following example shows, location, date, originaltweet and sentiment status of tweets are created in the dataset. According to the emotion, the tweets were shown as positive, negative, neutral, extremely positive and extremely negative (Figure 3.4).

UserName	ScreenName	Location	TweetAt	OriginalTweet	Sentiment
3807	48759	Atlanta, GA USA	16-03-2020	Due to COVID-19 our retail store and classroom...	Positive
3808	48760	BHAVNAGAR,GUJRAT	16-03-2020	For corona prevention,we should stop to buy th...	Negative
3809	48761	Makati, Manila	16-03-2020	All month there hasn't been crowding in the su...	Neutral
3810	48762	Pitt Meadows, BC, Canada	16-03-2020	Due to the Covid-19 situation, we have increas...	Extremely Positive
3811	48763	Horningsea	16-03-2020	#horningsea is a caring community. LetÂ s ALL ...	Extremely Positive

Figure 3.4 Examples of covid-19 tweets from dataset

Examples of tweets that have been cleaned in order to get more efficient results are as follows:

	OriginalTweet	Sentiment	text_clean
3782	In light of all that s happening with the #cor...	0	in light of all thats happening with the coron...
3783	Need medical advice: What's safer, 1) Going t...	2	need medical advice whats safer 1 going to a p...
3784	#while hamsters are buying canned food, I deci...	0	while hamsters are buying canned food i decide...
3785	I m going to create a new battle Royale game b...	2	im going to create a new battle royale game ba...

Figure 3.5 Examples of tweets after cleaning

A dataset containing 41,157 pieces of data has been created for the training process. 60% of the compiled data is divided into a training dataset, 20% is a test dataset, and the other 20% is a validation dataset. After data preprocessing, 40,923 tweets were obtained. There are 11381 positive, 9889 negative and 7560 neutral comments in the compiled data.

Table 3.1 Sentiment column analysis

Positive	11381
Negative	9889
Neutral	7560
Extremely Positive	6618
Extremely Negative	5475

Training and Test data were combined with negative values of 0, neutral values of 1 and positive values of 2 in a matrix layout and made ready for the training stage. The total number of tweets for each class:

Table 3.2 Sentiment values

2	17999
0	15364
1	7560

With the train_test_split function of the Sklearn library, a validation part will be pulled from the training part. In this way, it is aimed to monitor the accuracy and thus prevent overfitting.

3.5 Experiment and Results

3.5.1 Evaluation Criteria

The confusion matrix measures the document's accuracy of belonging to the relevant class. The measurement of the success rate of the training stage was provided by the confusion matrix method. A confusion matrix is a machine learning concept that contains information about the success of real and predicted classifications made by a classification system (Deng et al., 2016).

An actual value that is positively labeled in a dataset is called True Positive (TP) if it is also classified as positive because of classification. On the contrary, an actual value that is negatively labeled in a dataset is called True Negative (TN) if it is also classified as negative because of classification. Also, the actual value is negative, but the model predicted a positive value is called False Positive (FP) and the predicted value labeled falsely negative is called False Negative (FN) (Sarıman & Mutaf, 2020).

Accuracy is used to measure the success of a model. However, it is a metric that is not sufficient by itself. The accuracy value is to be calculated by the ratio of the fields that we estimate correctly in the model to the total dataset.

$$\frac{TP + TN}{TP + FP + TN + FN}$$

Precision shows how many of the values we estimate as positive are actually positive. A high accuracy value is an important criterion for choosing a model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall is a metric that shows how much of the transactions we need to estimate positively. The sensitivity value is also a metric that will help us in situations where the cost of estimating False Negative is high. It is necessary that it be as high as possible.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1 Score value shows us the harmonic mean of the Precision and Recall values. The reason why it is a harmonic mean instead of a simple mean is that we should not ignore extreme cases either.

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

The most basic reason why the F1 Score value is used instead of Accuracy is not to make an incorrect model selection in datasets that are not evenly distributed. There is also a calculation that contains all errors not just False Negative or False Positive. Therefore, the F1 score is very important.

3.5.2 Experimental Models

Three different classifiers and pre-trained language models were used in the experiments. The mentioned ML classifiers are Naïve Bayes, Logistic Regression and Random Forest. BERT, RoBERTa and BERTweet were also used as pre-trained language models.

3.5.2.1 Naïve Bayes

Datasets are tested for sentiment analysis and Bayesian method is used for classification accuracy. The Naive Bayes algorithm is based on the Bayes rule, which can be represented as follows, and the process is as follows until the lexicon is formed.

P(X|Y) = P(Y)P(Y|X)P(X)

The probability of the word appearing can be estimated by the ratio of positive, neutral and negative tweets. To estimate the sensitivity of a tweet, we just need to summarize the possibility that the words in the tweet are logical together with the log prior (Mashalkar, 2020).

logprior: - log($P(Dpos)$)−log($P(Dneg)$)=log($Dpos$)−log($Dneg$)

I also made a classification with Naive Bayes in this study. Before using the pretrained BERT models, the baseline models' Naïve Bayes Classifier model was also trained to perform sentiment classification. The results are as follows:

Table 3.3 Classification scores for Naïve Bayes

	Precision	Recall	F1-score	Support
Negative	0.75	0.78	0.74	1629
Neutral	0.57	0.43	0.49	614
Positive	0.73	0.73	0.73	1544
Accuracy			0.70	3787
Macro avg.	0.67	0.64	0.65	3787
Weighted avg.	0.70	0.70	0.70	3787

The classification score is a performance evaluation metric in machine learning. It is used to show the accuracy, recall, F1 Score, and support of trained classification model. This report is illustrated with the classification report object of the metrics method at the Sklearn library. The algorithm performance is acceptable. As a result, the accuracy and F1 are about 70%.

3.5.2.2 Logistic Regression Algorithm

Logistic regression (LR) is like a regression problem in which the dependent variable is a categorical variable. It is widely used in linear classification problems. Although it is called regression, a classification is made. The logistic regression algorithm determines the classes based on the samples used in the dataset. Another feature is that it captures the data logarithmically on the curve, not linearly (Sarıman & Mutaf, 2020). In this analysis, the word values calculated by TF-IDF method represent the independent variables, and the positive, negative and neutral values of tweets represent the nominal dependent variable. The classification results with Logistic Regression are as follows (Table 3.4):

Table 3.4 Classification score for logistic regression algorithm

	Precision	Recall	F1-score	Support
Negative	0.82	0.78	0.80	1629
Neutral	0.59	0.75	0.66	614
Positive	0.83	0.79	0.81	1544
Accuracy			0.78	3787
Macro avg.	0.75	0.77	0.76	3787
Weighted avg.	0.79	0.78	0.78	3787

The results showed that the logistic regression algorithm performed relatively well compared to the Naive Bayes algorithm. The algorithm performance is acceptable. As a result, the accuracy and F1 are about 78%. This result is related to the fact that the logistic regression algorithm does not make as many assumptions as the pure Bayesian algorithm.

3.5.2.3 Random Forest

Random Forest is a method used to model predictions in sentiment analysis. It is an algorithm used to classify large data. A series of regression decision trees combine the results of the results to make an output prediction. Each tree is independent and all trees have the same distribution. The trees depend on the random vector sampled from the input data (Williams et al., 2020). The classification results with Random Forest are as follows (Table 3.5):

Table 3.5 Classification score for Random Forest

	Precision	Recall	F1-score	Support
Negative	0.71	0.66	0.68	1629
Neutral	0.56	0.60	0.58	614
Positive	0.69	0.71	0.70	1544
Accuracy			0.67	3787
Macro avg.	0.65	0.66	0.65	3787
Weighted avg.	0.67	0.67	0.67	3787

Results were lower than other classifications. Algorithm performance is acceptable. According to the results, the score of F1 is 67%.

3.5.2.4 Pretrained Model Classification

Finally, the pre-trained BERT model has been imported from the Huggingface from the models library. In order to be able to fine-tune the BERT transformer, special function was implemented to accommodate the pre-trained BERT model. Therefore, an output layer with 3 neurons has been added, which is necessary to perform the classification of 3 different classes (3 emotions) of the dataset.

The success values of the designed model were calculated. The Classification Scores for BERT is as follows:

Table 3.6 Classification score for BERT-Base-Uncased

	Precision	Recall	F1-score	Support
Negative	0.88	0.92	0.90	1629
Neutral	0.85	0.79	0.82	614
Positive	0.92	0.89	0.90	1544
Micro avg.	0.89	0.89	0.89	3787
Macro avg.	0.88	0.87	0.87	3787
Weighted avg.	0.89	0.89	0.89	3787
Samples avg.	0.89	0.89	0.89	3787

As a result of the training, the total F1 score is 90% (positive), 90% (negative) and 82% (neutral). The highest F1-score class is the positive and negative sentiment class, while the success rate in the neutral sentiment class is lower than in other classes.

Data training was carried out by dividing into 4 parts. While the model is being trained, not all data participate in the training at the same time. They take part in training in a certain number of parts. The first part is trained, the success of the model is tested, the weights are updated with backpropagation according to success. Then, with the new training set, the model is retrained and the weights are updated again. This process is repeated at each training step to try to calculate the optimal weight values for the model.

Each of these training steps is called an epoch. Success will be low in the first epochs, and success will increase as the number of epochs increases. However, after a certain step, the learning status of the model will decrease significantly. Epoch values are a hyper parameter that defines the number of times the learning algorithm will run through the entire training dataset. The success criteria of the training are as follows.

Table 3.7 Valid loss for transformer BERT

	Training loss	Valid. loss	Valid. Accur.
Epoch			
1	0.5630	0.3511	0.8741
2	0.2890	0.2573	0.9043
3	0.1927	0.2149	0.9222
4	0.1297	0.2045	0.9348
5	0.0889	0.2278	0.9302
6	0.0579	0.2231	0.9422
7	0.0423	0.2425	0.9428
8	0.0282	0.2960	0.9302
9	0.0248	0.2569	0.9419
10	0.0195	0.2848	0.9446

The change in values specified in Table 3.7 indicates that the losses incurred for each training round (epoch) change during the rounds. For BERT-Base-Uncased model, accuracy is indicated by the train (blue line) and test (orange line) on 10 epochs. The design model shows that validation during the epoch, and a line graph has been created that shows learning curves and test sets for the model accuracy on the train throughout each training period (Figure 3.6).

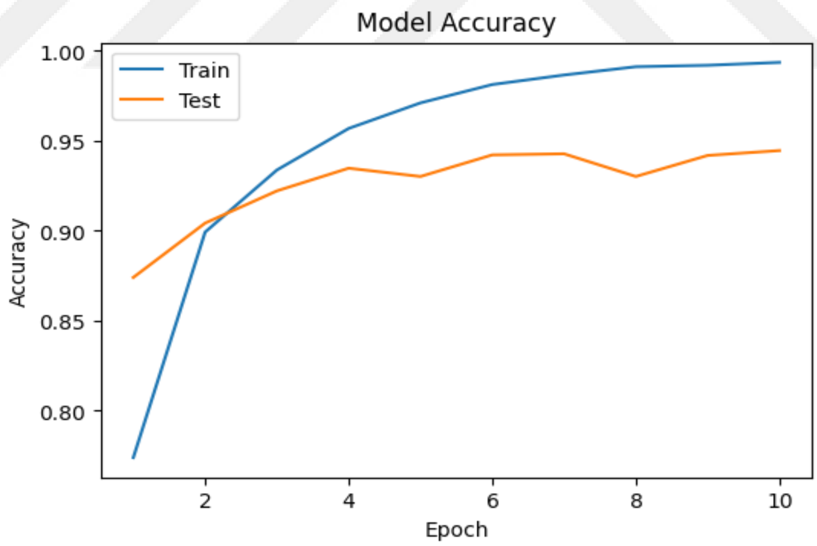

Figure 3.6 Accuracy over the epochs for the BERT model.

Figure 3.6 gives the accuracy values of the training and test sets. Accuracy has not increased since Epoch 7. The curve graph of the loss rate produced by the model is shown in Figure 3.7.

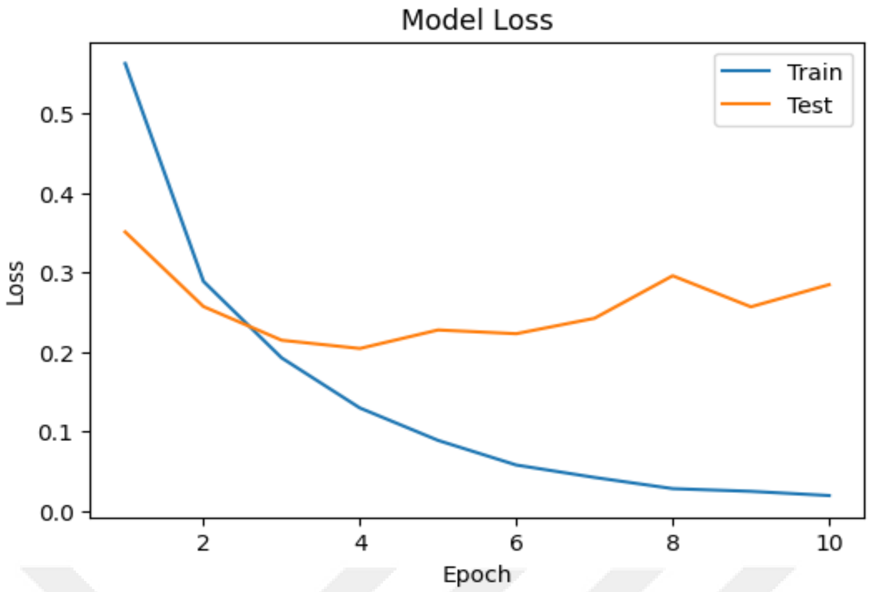

Figure 3.7 BERT loss function curve

The function curve shows the variation of the loss that occurs for each epoch during the tours. The reason for the fluctuations is that text is transmitted to the model by small parts. Train loss is seen throughout the epochs. When we looked at the change in Loss values, it was observed that this value was gradually reduced. The following shows the confusion matrix created by estimating the designed model.

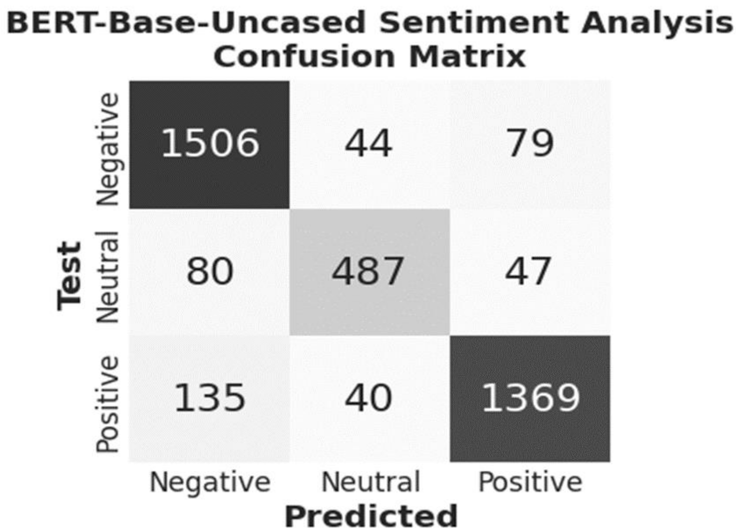

Figure 3.8 BERT sentiment analysis confusion matrix

RoBERTa Modelling;

Table 3.8 Valid loss for transformer RoBERTa

	Training loss	Valid. loss	Valid. Accur.
Epoch			
1	0.5769	0.3656	0.8687
2	0.3420	0.2775	0.8994
3	0.2520	0.2683	0.9106
4	0.1872	0.2228	0.9259
5	0.1435	0.2219	0.9313
6	0.1047	0.2428	0.9298
7	0.0786	0.2740	0.9357
8	0.0604	0.2180	0.9426
9	0.0475	0.2614	0.9409
10	0.0381	0.2743	0.9365

For the RoBERTa model, accuracy is indicated by the train (blue line) and test (orange line) on 10 epochs. A line graph has also been created for model accuracy on Train, showing learning curves and test sets throughout each training period (Figure 3.9).

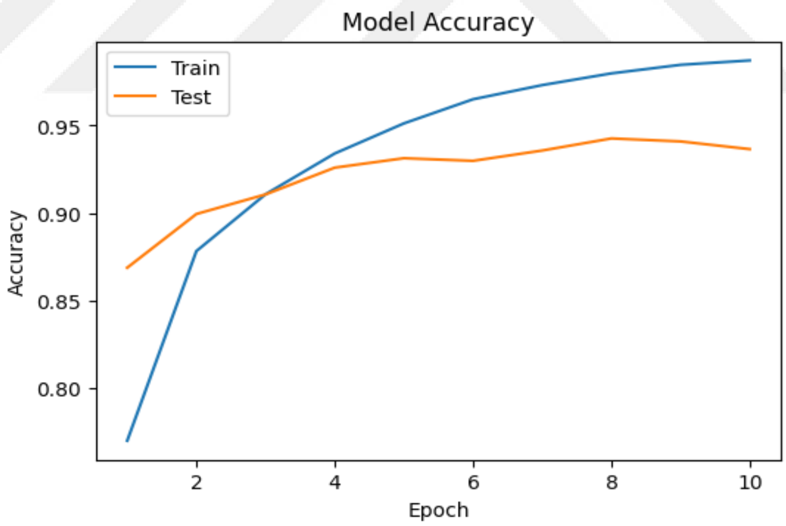

Figure 3.9 Accuracy over the epochs for the RoBERTa model

Figure 3.9 gives the accuracy values of the training and test sets. Accuracy has not increased since epoch 8. A comparable consistent accuracy has been found in both the train and the test set. The curve graph of the error rate generated by the model is given in Figure 3.10.

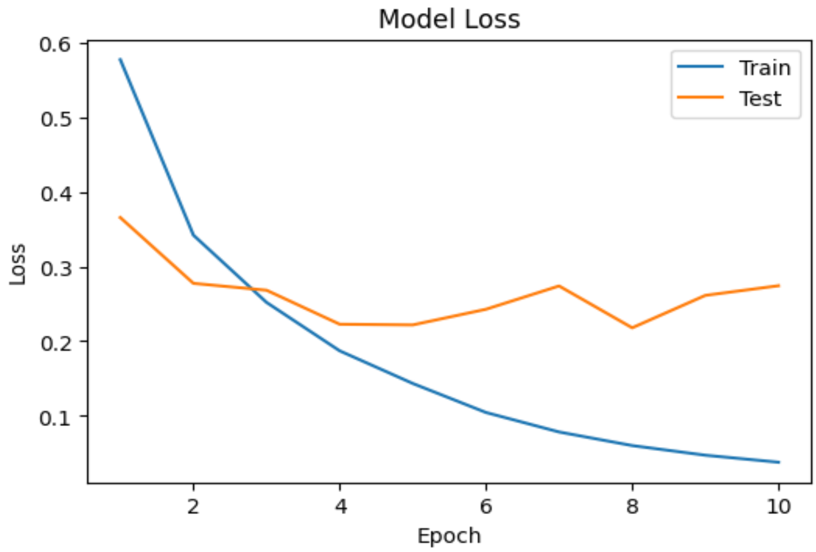

Figure 3.10 RoBERTa loss function curve

The loss value change shows the variation of the loss that occurs for each epoch during the tours. Train loss is seen throughout the epochs. When we looked at the change in loss values, it was observed that this value was gradually reduced. The following shows the confusion matrix created by estimating the designed model.

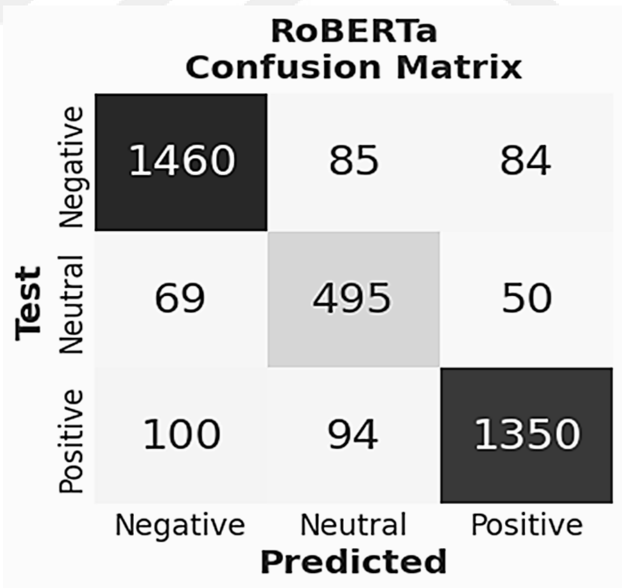

Figure 3.11 RoBERTa sentiment analysis confusion matrix

Table 3.9 Classification score for RoBERTa

	Precision	Recall	F1-score	Support
Negative	0.90	0.90	0.90	1629
Neutral	0.73	0.81	0.77	614
Positive	0.91	0.87	0.89	1544
Micro avg	0.87	0.87	0.87	3787
Macro avg	0.85	0.86	0.85	3787
Weighted avg	0.88	0.87	0.87	3787
Samples avg	0.87	0.87	0.87	3787

As a result of the training, the total F1 score is 89% (positive), 90% (negative) and 77% (neutral). The highest F1-score class is the positive sentiment class, while the success rate in the neutral sentiment class is lower than in other classes.

BERTweet Modelling;

Table 3.10 Valid loss for transformer BERTweet

	Training loss	Valid. loss	Valid. Accur.
Epoch			
1	0.5406	0.3642	0.8724
2	0.2957	0.2693	0.9057
3	0.2129	0.2431	0.9174
4	0.1643	0.2219	0.9207
5	0.1250	0.2195	0.9365
6	0.0953	0.2440	0.9294
7	0.0743	0.2207	0.9376
8	0.0559	0.2344	0.9428
9	0.0461	0.2601	0.9454
10	0.0361	0.2405	0.9441

For the BERTweet model, accuracy is indicated by train (blue line) and test (orange line) on 10 epochs. The design model shows that validation increases during epoch (Figure 3.12). A line graph has also been created for model accuracy on train, showing learning curves and test sets throughout each training period.

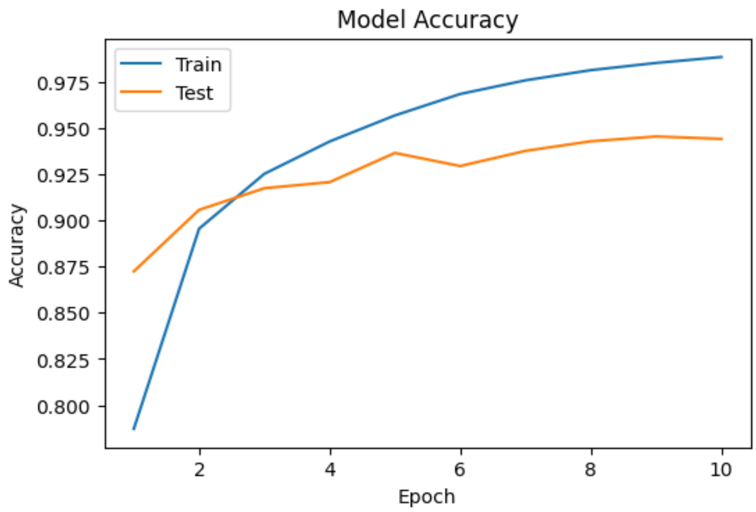

Figure 3.12 Accuracy over the epochs for the BERTweet model

Accuracy has not increased since Epoch 9. The error values of the designed model are shown in the change table (Table 3.13). Training and testing set and near accuracy have been seen to increase. The curve graph of the loss rate produced by the model is shown in the figure.

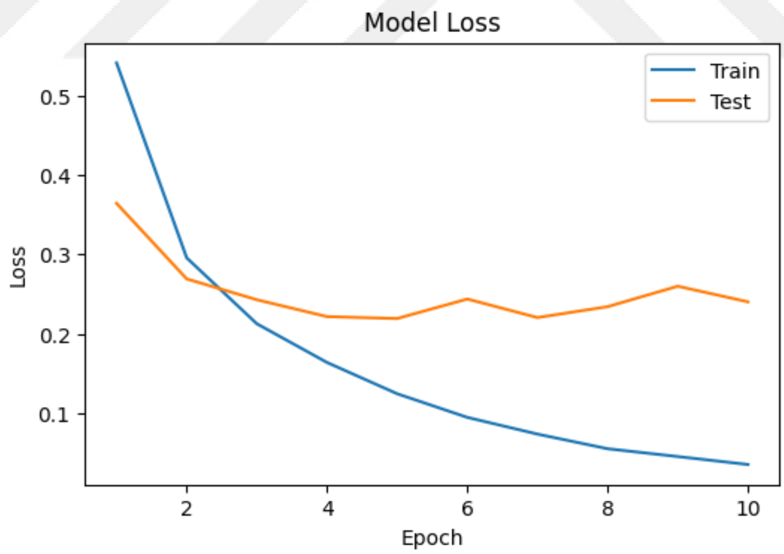

Figure 3.13 BERTweet loss function curve

The loss value change shows the variation of the loss that occurs for each epoch during the tours. Train loss is seen throughout the epochs. When we look at the change in loss values, we see that this value gradually decreases. This means that after a specific step value, the accuracy and loss values in the test dataset remain constant and the difference in value between the training dataset increases. In this case, our models of these datasets may say that there is an overlearning situation. The following shows the confusion matrix created by estimating the designed model.

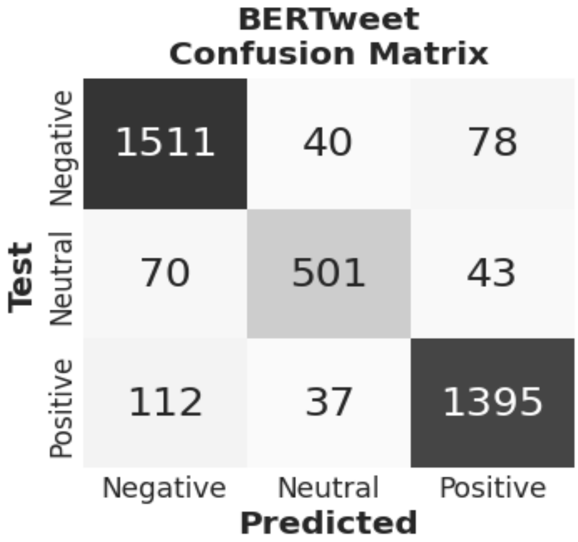

Figure 3.14 BERTweet sentiment analysis confusion matrix

Table 3.11 Classification score for BERTweet

	Precision	Recall	F1-score	Support
Negative	0.89	0.93	0.91	1629
Neutral	0.87	0.82	0.84	614
Positive	0.92	0.90	0.91	1544
Micro avg.	0.90	0.90	0.90	3787
Macro avg.	0.89	0.88	0.89	3787
Weighted avg.	0.90	0.90	0.90	3787
Samples avg.	0.90	0.90	0.90	3787

As a result of the training, the total F1 score is 91% (positive), 91% (negative) and 84% (neutral). The highest F1-score class is the positive sentiment class, while the success rate in the neutral sentiment class is lower than in other classes.

The classification success of all three algorithms is close to each other. We can see that all three algorithms performed well in the classification task with performance scores around 90%. In addition, classification success rates are also sufficient. Bert models gave higher accuracy rates because Naive Bayes and logistic regression neglects semantic and structural information of the context, focuses on causality between characteristic words and categories, while the pre-trained BERT models takes into account semantic and contextual information of the text. The comparison table is as follows.

Table 3.12 Comparison BERT, RoBERTa and BERTweet

Epoch		Negative			Neutral			Positive			Weighted avg.		
	Model	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Covid-19 Dataset	BERT-base	0.88	0.92	0.90	0.85	0.79	0.82	0.92	0.89	0.90	0.89	0.89	0.89
	RoBERTa	0.90	0.90	0.90	0.73	0.81	0.77	0.91	0.87	0.89	0.88	0.87	0.87
	BERTweet	0.89	0.93	0.91	0.87	0.82	0.84	0.92	0.90	0.91	0.90	0.90	0.90
	Naïve Bayes	0.75	0.78	0.74	0.57	0.43	0.49	0.73	0.73	0.73	0.70	0.70	0.70
	Logistic Regression	0.82	0.78	0.80	0.59	0.75	0.66	0.83	0.79	0.81	0.79	0.78	0.78
	Random Forest	0.71	0.66	0.68	0.56	0.60	0.58	0.69	0.71	0.70	0.67	0.67	0.67

According to the results of the experiment, the Logistic Regression from the classification algorithms has yielded the highest F1 scores. The second-high algorithm is Naïve Bayes. The lowest algorithm is Random Forest results. As can be seen from the table, the same result was not always obtained in all three models in the identification of sentiments. It seems that the results of the three models are close to each other. In addition, we also see that BERTweet is a little more successful in positive and negative scores. However, we cannot say that one model is more successful than others with values close to each other. Examples of what sentiments are assigned to some tweets according to the models can be given. For instance, table 3.12 lists some of the emotions the three models assigned to the tweets.

Table 3.13 Tweet Data Sample

Tweet	Correct Sentiment	BERT	RoBERTa	BERTweet
Just thinking if we have to close schools, how many children may lose access to main source of meals. And what will happen to demand at food banks.	Negative	Negative	Negative	Negative
If you have booked a ticket to an event as part of a package holiday you will be offered an alternative or a refund by your travel provider, if it has been cancelled due to Coronavirus. Check ABTA's consumer QA	Positive	Negative	Neutral	Neutral
I don't understand the run on toilet paper. You would think that you stock up on food first. Without food, you won't need the toilet paper.	Negative	Neutral	Neutral	Negative
Does everyone really need to stock up on toilet paper and paper towel? Two weeks of food I can understand, but you don't need a year worth of toilet paper and paper towel. It's not like the world is ending	Negative	Neutral	Negative	Negative
Breaking: New Jersey officials urge residents to stock up for a two week coronavirus quarantine Just in case COVID19	Neutral	Negative	Neutral	Neutral

CONCLUSION AND FUTURE WORK

Advances in information technologies enable the analysis of big data and the acquisition of information. Important information is revealed from the analysis of unstructured data. NLP techniques make texts understandable within the framework of the rules of the language. These techniques also provide interpretation of thoughts in media such as Twitter.

In this thesis, three pre-trained language models were used to conduct sentiment analysis. As a result of the analysis, the intensity of the emotions of tweets about Covid-19 on Twitter was revealed and currently only describes tweets in English about the COVID-19 pandemic. According to the findings obtained after the experiments, it was observed that all three language models performed closely to each other. It seems that the BERTweet model achieved the highest F1 score (91%) and Precision (93%), but others are also close to success. Neutral comments in the comments have the lowest intensity of emotions. In order to achieve higher accuracy values in the success rates of the models, more data may be required. However, quality of data is reduced due to the deficiencies in grammar and spelling rules in the processing of data. However, since Twitter is an informal platform, users may not pay attention to the grammar and spelling. Although the tweets are not very regular, the meaning in the sentence is not distorted.

There are sentiment analysis studies on Covid-19 tweets in the literature and on Kaggle. There are also different methods applied to the dataset used in this study. In the classification made in the same dataset, the highest success rate of 97% was achieved (Bakrey, 2022). The results of an analysis in which the BERT model was applied are also similar to this study (Naduvin, 2022). In addition, it is seen that 80% and above performance is achieved in studies conducted with different Covid-19 tweet datasets (Chakraborty et al., 2021, Karaca & Aslan, 2021; Malla & Alphonse, 2021; Topbaş et al., 2021). A high number of data is used in pre-trained language models. These data usually consist of regular and meaningful sentences. In language models, the meaning of words in sentences is encoded according to their relationship with other words. For this reason, analyses with pre-trained language models can be more successful than classical machine learning models. The results obtained contribute both to the development of dictionaries and to the evaluation of researchers' analyses in future studies.

In summary, all the necessary information and inferences can be obtained by using big data sources such as social media in information technologies. Using these tools, new algorithms and lexicons can be developed with the desired flexibility. The results of this thesis are a reference to other studies. However, it was not possible to apply the methods and experiments applied in the study in different datasets. In future studies, new studies can be done on different datasets with different pre-trained language models. In future studies, it may also be considered to determine the relationship between positive case and death rates of the Covid-19 virus with new analyzes of tweets by country and region. Pre-trained language models can also be applied to predict future tweets about similar disease types. In addition, it is expected that studies will be conducted with other existing datasets and methods proposed in this thesis, as well as methods such as Emotion Tags and Bag-of-words, will improve the results.

REFERENCES

Abid, F, Alam, M. Yasir, M., & Li, C. (2019). Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. *Future Generation Computer Systems*, 95, 292-308. https://doi.org/10.1016/j.future.2018.12.018.

Akın, B. K., & Şimşek, U. T. G. (2018). Adaptif öğrenme sözlüğü temelli duygu analiz algoritması önerisi. *Bilişim Teknolojileri Dergisi*, *11*(3), 245-253.

Albayrak, M., Topal, K., & Altıntaş, V. (2017). Data analysis on social media: twitter. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22, 1991-1998.

Anava, O., & Levy, K. F. (2017). k-Nearest neighbors: from global to local. https://www.semanticscholar.org/paper/k*-Nearest-Neighbors%3A-From-Global-to-Local-Anava-Levy/a5d121a5d262c99612afcc2a6eb05baf49ec514d (accessed: 31.08.2020).

Anttiroiko, A. V., & Savolainen, R. (2011). Towards library 2.0: the adoption of web 2.0 technologies in public libraries. *Libri,61*(2), 87-99.

Araque, O., Corcuera, I., Sánchez, F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications, *Expert Systems with Applications*, 77, 236-246. http://dx.doi.org/10.1016/j.eswa.2017.02.002

Aston, N.; Liddle, J., & Hu, W. (2014), Twitter sentiment in data streams with perceptron. *Journal of Computer and Communications*, *2*(3), 11-16. http://dx.doi.org/10.4236/jcc.2014.23002

Azzouza, N.; Aklii, K., & İbrahim, R. (2020). TwitterBERT: Framework for twitter sentiment analysis based on pre-trained language model representations. *International conference of Reliable Information and Communication Technology* (428–437). https://doi.org/10.1007/978-3-030-33582-3_41.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the International Conference on Language Resources and Evaluation, LREC Valletta, Malta.*

Bahrawi, B. (2019). Sentiment analysis using random forest algorithm-online social media based. *Journal of Information Technology and Its Utilization*, *2*(2), 29-33.

Baker, W. (2021). *Using large pre-trained language models to track emotions of cancer patients on twitter.* (Undergraduate Honors Theses). University of Arkansas. Retrieved from https://scholarworks.uark.edu/csceuht/92

Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L., & Hughes, M. (2013). Sentiment analysis of political tweets: Towards an accurate classifier, *Association for Computational Linguistics,* 49-58.

Bakrey, M. (2022). Make a classification for tis tweets coved-19. Retrieved from https://www.kaggle.com/code/mohamedbakrey/make-a-classification-for-tis-tweets-coved-19 (accessed: 05.05.2022).

Basile, V. (2013). Sentiment analysis on Italian tweets, 4th Workshop on Computational Approaches to Subjectivity, *Sentiment and Social Media Analysis*, Georgia, USA.

Bayhaqy, A., Sfenrianto, S., Nainggolan, K., & Kaburruan, E. R. (2018). Sentiment analysis about e-commerce from tweets using decision tree, k-nearest neighbor, and naïve bayes. *International Conference on Orange Technologies (ICOT).* https://doi.org/10.1109/ICOT.2018.8705796

Becker, L., Erhart, G., Skiba, D., & Matula, V. (2013). AVAYA: Sentiment analysis on twitter with self-training and polarity lexicon expansion. *Second Joint Conference on Lexical and Computational Semantics (*SEM).*

Berrar, D. (2018). Bayes' theorem and Naive Bayes classifier. *Encyclopedia of Bioinfor-matics and Computational Biology,* 403-412. https://doi.org/10.1016/B978-0-12-809633-8.20473-1

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, Bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, 440-447.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Boser, B.E., Guyon, I.M., & Vapnik, V.N. (1992). A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92),* Pittsburgh, 144-152.

Breiman, L. (2001). Random forests. *Machine. Learning*, *45*(1), 5–32.

Brooke, J. (2009). *A semantic approach to automatic text sentiment analysis*. (M.A. thesis). Simon Fraser University, Burnaby, B.C., Canada.

Brownlee, J. (2019). A tour of machine learning algorithms. machine learning mastery. https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/ (accessed: 10.01.2022).

Bruns, A., & Moe, H. (2016). Twitter'da iletişimin yapısal katmanları. In K. Weller, A. Bruns, J. Burgess, M. Mahrt & C. Puschmann (Eds.), *Twitter ve toplum.* (pp. 62 - 78). İstanbul: Kafka, Epsilon Yayıncılık.

Catelli, R., Pelosi, S., & Esposito, M. (2022). Lexicon-Based vs. Bert-based sentiment analysis: A comparative study in Italian. *Electronics*, *11* (3), 374. https://doi.org/10.3390/electronics11030374

Cesconi F. (2020). Natural language processing: Explaining BERT to business people. https://hackernoon.com/natural-language-processing-explaining-bert-to-business-people-obz3uno (accessed: 18.12.2020).

Chakraborty, A. K., Das, S., & Kolya, A. K. (2021). Sentiment analysis of Covid-19 Tweets using evolutionary classification-based LSTM model. *arXiv*. https://doi.org/10.48550/arXiv.2106.06910

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, 44, 65-77. https://doi.org/10.1016/j.inffus.2017.12.006

Chiavetta, F.; Lo Bosco, G., & Pilato, G. (2016). A Lexicon-based approach for sentiment classification of Amazon books reviews in Italian language. *12th International Conference on Web Information Systems and Technologies.* http://dx.doi.org/10.5220/0005915301590170

Choy M., Cheong M.L.F.; Laik Ma N., & Shung K. P. (2011). A sentiment analysis of Singapore presidential election 2011 using Twitter data with census correction. 1093-1113. *Research Collection School of Computing and Information Systems.* https://ink.library.smu.edu.sg/sis_research/1436

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzm ́an, F., Grave, E., Ott, M., Luke Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *In Proceedings of ACL, page to appear.*

Cruz, L., Ochoa, J., Roche, M., & Poncelet, P. (2016). Dictionary-based sentiment analysis applied to specific domain using a web mining approach. *SIMBig: Symposium on Information Management and Big Data,* Cusco, Peru. 57-68

Culnan, M., McHugh, P., & Zubillaga, J. (2010). How large U.S. companies can use twitter and other social media to gain business value MIS. *Quarterly Executive*, *9*(4), 243-259.

Çelikyilmaz A., Hakkani, D., & Feng, F. (2010). Probabilistic model-based sentiment analysis of Twitter messages. *IEEE Workshop on Spoken Language Technology.*

Data Science Earth (2020). Doğal dil işleme (NLP) teknikleri. https://www.datasciencearth.com/dogal-dil-isleme-nlp-teknikleri/ (accessed: 28.12.2020).

Dehkharghani, R., Yanıkoğlu, B., Saygın, Y., & Oflazer, K. (2016). Sentiment analysis in Turkish at different granularity levels. *Natural Language Engineering, 23*(4), 535-559. https://doi.org/10.1017/S1351324916000309

Denecke, K. (2008). Using SentiWordNet for multilingual sentiment analysis. *IEEE 24th International Conference on Data Engineering Workshop.*

Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*. 340-341, 250-261.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: human language technologies. 1. Minneapolis, Minnesota*, 4171-4186. https://www.aclweb.org/anthology/N19-1423

Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining.* https://doi.org/10.1145/1341531.1341561

Esuli, A., & Sebastiani, F. (2006). A publicly available lexical resource for opinion mining. *5th International Conference on Language Resources and Evaluation, Genoa, Italy,* 417-422.

Fellbaum, C. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235-244.

Francesco, E. (2020). Sentiment analysis dictionaries. https://www.baeldung.com/cs/sentiment-analysis-dictionaries (accessed: 07.01.2020).

Ghag, K., & Shah, K. (2013). Comparative analysis of the techniques for sentiment analysis. *ICATE 2013 paper.*

Ghag, K. V., & Shah, K. (2018). Conceptual sentiment analysis model. *International Journal of Electrical and Computer Engineering (IJECE),* 8(4), 2358-2366. https://doi.org/10.11591/ijece.v8i4.pp2358-2366

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications*, *40*(16), 6266-6282. https://doi.org/10.1016/j.eswa.2013.05.057

Ghiassi, M., & Lee, S. (2018). A domain transferable lexicon set for twitter sentiment analysis using a supervised machine learning approach. *Expert Systems with Applications*, *106*(15), 197-216. https://doi.org/10.1016/j.eswa.2018.04.006

Go, A., Bhayani, R., & Huang, L. (2009). *Twitter sentiment classification using distant supervision.* (Project Report No. CS224N), Stanford, 1-12.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. 1. MIT press Cambridge.

Google CLN (2018). Cloud natural language web site. https://cloud.google.com/natural-language (accessed: 19.01.2021).

Hanks, P., & Pustejovsky, J. (2005). A pattern dictionary for natural language processing. *Revue Française De Linguistique Appliquée*, *2*(x), 63-82.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464-472.

Heckerman, D., Mamdani, A., & Wellman, M. P. (1995). Real-world applications of Bayesian networks. *Communications of the ACM*, *38*(3), 24-26.

Horev, R. (2018). BERT Explained: State of the art language model for NLP. Retrieved from https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270 (accessed: 18.12.2020).

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 168-177.

Huq, M. R., Ali, A., & Rahman, A. (2017). Sentiment analysis on Twitter data using KNN and SVM. *(IJACSA) International Journal of Advanced Computer Science and Applications*, *8*(6), 19-25.

Ilango, R. (2019). Using NLP (BERT) to improve OCR accuracy. https://medium.com/states-title/using-nlp-bert-to-improve-ocr-accuracy 385c98ae174c (accessed: 28.12.2020).

İlkbahar, N. (2019). OCR (Optical Character Recognition) tanımı. https://medium.com/@ilkbaharnaz/ocr-optical-character-recognition-optik-karakter-tan%C4%B1ma-268593b6e284 (accessed: 29.12.2020).

Jagtap, V. S., & Pawar, K. (2013). Sentence-level analysis of sentiment classification. *National Conference on Emerging Trends in Engineering, Technology & Architecture*.

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing*. New Jersey: Prentice Hall.

Jurafsky, D., & Martin, J. H. (2019). Naive Bayes and sentiment classification in speech and language processing. https://web.stanford.edu/~jurafsky/slp3/

Kang, Y., Wang, Y., Zhang, D., & Zhou, L. (2017), The public's opinions on a new school meals policy for childhood obesity prevention in the U.S.: A social media analytics approach. *International Journal of Medical Informatics*, 103, 83-88, https://doi.org/10.1016/j.ijmedinf.2017.04.013

Katz, G., Ofek, N., & Bracha Shapira, B. (2015). ConSent: Context-based sentiment analysis. *Knowledge-Based Systems* 84, 162-178.

Khan, S. (2021). BERT, RoBERTa, DistilBERT, XLNet — which one to use? https://towardsdatascience.com/bert-roberta-distilbert-xlnet-which-one-to-use-3d5ab82ba5f8 (accessed: 07.01.2022).

Kapucugil, A., & Özdağoğlu, G. (2015). Text mining as a supporting process for VoC clarification. *Alphanumeric Journal*, *3*(1), 25-40.

Karaca,Y. E., & Aslan, S. (2021). Sentiment analysis of Covid-19 Tweets by using LSTM Learning Model. *Journal of Computer Science,* IDAP-2021(special issue), 366-374. https://doi.org/10.53070/bbd.990421

Karaöz, B. (2018). *Büyük veri ve işletme analizi: Sosyal medya ve duygu analizi ile bir öngörü modeli.* (Doktora Tezi). İstanbul Üniversitesi, İstanbul.

Kietzmann, J.H., Hermkens, K., I.P., & McCarthy, B.S. (2011). Silvestre social media? get serious! understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241-251.

Kiprono, K. W., & Abade, E. O. (2016). Comparative Twitter sentiment analysis based on linear and probabilistic models. *International Journal on Data Science and Technology*, *2*(4), 41-45.

Kouloumpis, E., Wilson, T., & Moore, J. (2021). Twitter sentiment analysis: The good the bad and the OMG! *Proceedings of the International AAAI Conference on Web and Social Media,* 5(1), 538-541. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14185

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *NIPS Advances in Neural Information Processing Systems Conference.* 1-9.

Kumar, A., & Sebastian, T. M. (2012). Sentiment analysis on Twitter. *International Journal of Computer Science*, *9*(4), 372-378.

Latha, I. H., Saradhi Varma, G. P., & Govardhan, A. (2013). Sentiment analysis tool using machine learning algorithms. *Elixir International Journal*, 58, 14791-14794.

Lehman, F. (1992). Semantic networks. *Computers Mathematic Application*, *23*(2-5). 1-50.

Li, H.; Ma, Y.; Ma, Z., & Zhu, H. (2021). Weibo text sentiment analysis based on bert and deep learning. *Applied Sciences*, 11, 10774. https://doi.org/10.3390/app1122107

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Liu, Y.; Ott, M.; Goyal, N. Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: ARobustly optimized BERT pretraining approach. *Computer Science Computation and Language,* arXiv:1907.11692

Maharani, W. (2013). Microblogging sentiment analysis with lexical based and machine learning approaches. *International Conference of Information and Communication Technology*, 439-443.

Mahaljan, P. (2021). Bag of words in natural language processing. https://aipoint.tech/blog/post/Bag-of-words-in-NLP/ (accessed: 03.11.2021).

Majumder, N., Bhardwaj, R., Poria, S., Zadeh, A., Gelbuklh, A., Hussain, A., & Morency, L. (2020). Improving aspect-level sentiment analysis with aspect extraction. https://arxiv.org/abs/2005.06607 (accessed: 05.07.2020).

Maksymenko, S. (2020). 5 use cases for natural language processing application in marketing. https://customerthink.com/5-use-cases-for-natural-language-processing-application-in-marketing/ (accessed: 28.12.2020).

Malla, S., & Alphonse, P.J.A. (2021). COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets. *Applied Soft Computing*, 107.

Masarifoğlu, M., Tigrak, U., Hakyemez, S.; Gul, G.; Bozan, E.; Buyuklu, A. H., & Özgür, A. (2021). Sentiment Analysis of Customer Comments in Banking using BERT-based Approaches. *Signal Processing and Communication Applications Conference (SIU).* https://doi.org/10.1109/SIU53274.2021.9477890

Matharasi, P. B., & Senthilrajan, A. (2017). Sentiment analysis of twitter data using naïve bayes with unigram approach. *International Journal of Scientific and Research Publications*, *7*(5), 337-341.

Martineau, J., & Finin, T. (2009). Delta TFIDF: An improved feature space for sentiment analysis. *Proceedings of the Third International Icwsm Conference*, 258-261.

Mashalkar, A. (2020). Sentiment analysis using logistic regression and naive bayes. https://towardsdatascience.com/sentiment-analysis-using-logistic-regression-and-naive-bayes-16b806eb4c4b (accessed: 07.02.2022).

Mayfield, A. (2008). *What is social media*? http://www.icrossing.com/uk/sites/default/files_uk/insight_pdf_files/What%20is%20Social%20Media_iCrossing_ebook.pdf

Medhat W, Hassan A., & Korashy H. (2014). Sentiment analysis algorithms and applications: a survey. *Ain Shams Engineering Journal*, *5*(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

Mittal, A., & Goel, A. (2011). Stock prediction using twitter sentiment analysis. Stanford University. Retrieved from https://www.semanticscholar.org/paper/Stock-Prediction-Using-Twitter-Sentiment-Analysis-Mittal/4ecc55e1c3ff1cee41f21e5b0a3b22c58d04c9d6

Mohamed, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments, *Expert Systems with Applications*, *40*(10), 4241-4251.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, *40*(10), 4241-4251.

Muller, B., Sagot, B., & Seddah, D. (2019). Enhancing BERT for lexical normalisation. *Proceeding of the 2019 EMNLP Workshop W-NUT: The 5th Workshop on Noisy User-generated Text, Hong Kong*, (297-306).

Mundalik, A. (2018). *Aspect based sentiment analysis using data mining techniques within Irish airline industry.* (MSc Research Project Data Analytics). National Collage of Ireland, Ireland.

Munshi, A., Sapra, S., & Arvindhan, M. (2020). A novel random forest implementation of sentiment analysis. *International Research Journal of Engineering and Technology (IRJET), 7*(6), 2821-2824.

Myatt, G. J. (2007). *Making sense of data: A practical guide to exploratory data analysis and data mining.* Wiley-Interscience, https://doi.org/10.1002/0470101024

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes*, *30*(1), 3-26. https://doi.org/10.1075/li.30.1.03nad

Naduvin, R. (2022). Covid-19 text classification using BERT (TPU). Retrieved from https://www.kaggle.com/code/ravikumarmn/covid-19-text-classification-using-bert-tpu (accessed: 05.05.2022).

Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. *Proceedings of the 2nd international conference on Knowledge capture, Sanibel Island, FL, USA*. https://doi.org/10.1145/945645.945658

Nayak, P. (2019). Understanding searches better than ever before. https://blog.google/products/search/search-language-understanding-bert/ (access date: 11.01.2021).

Nikfarjam, A., Sarker, A., O'Connor, K., Ginn, R., & Gonzalez, G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features, *Journal of the American Medical Informatics Association, 22*(3), 671-681.

Nguyen, D. Q.; Vu, T., & Nguyen, A. (2020). BERTweet: A pre-trained language model for English Tweets. *Computer Science.* https://doi.org/10.18653/v1/2020.emnlp-demos.2

Nguyen, A. T. (2020). TATLat WNUT-2020 Task 2: A transformer-based baseline system for identification of informative covid-19 English tweets. *Proceedings of the 2020 EMNLP Workshop W-NUT: The Sixth Workshop on Noisy User-generated Text,* 319-323.

Onan, A. (2017). Sentiment analysis on twitter messages based on machine learning methods. *Yönetim Bilişim Dergisi, 3*(2), 1-14.

Özker, U. (2019). Zemberek-doğal dil işleme. Retrieved from https://ugurozker.medium.com/zemberek-nlp-7add032881e9 (accessed: 20.01.2021).

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the 7th International Conference on Language Resources and Evaluation*, 320-1326.

Pang, B., Lee, L., & Vaithyanathan S. (2002). Thumbs up? sentiment classification using machine learning techniques. *Conference on Empirical Methods in Natural Language Processing (EMNLP),* 79-86.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1-135. https://doi.org/10.1561/1500000011

Parmar, H. H., Bhanderi, S., & Shah, G. (2014). Sentiment mining of movie reviews using random forest with tuned hyperparameters. *International Conference on Information Science, Kerala.*

Patel, S. (2020). NLP: POS (Part of speech), Tagging & Chunking. https://medium.com/@suneelpatel.in/nlp-pos-part-of-speech-tagging-chunking-f72178cc7385 (accessed: 06.12.2021).

Patil, P. P., Phansalkar, S., & Kryssanov, V. V. (2018). topic modelling for aspect-level sentiment analysis, *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, 221-229.

Pennacchiotti, M., & Popescu, A. (2011). A machine learning approach to Twitter user classification, *ICWSM*, 281-288.

Perez-Rosas, V., Banea, C., & Mihalcea, R. (2012). Learning sentiment lexicons in Spanish. *LREC Conferences*, 12, 73.

Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. *Computing Attitude and Affect in Text: Theory and Applications*, 1-10. https://doi.org/10.1007/1-4020-4102-0_1

Pota, M., Ventura, M., Catelli, R., & Esposito, E. (2021). An effective BERT-based pipeline for twitter sentiment analysis: A case study in Italian. *Sensor (Basil), 21*(1), 133. https://doi.org/10.3390/s21010133

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143-157.

Qiu, G., Liu, B., & Chen, C. L. P. (2009). Expanding domain sentiment lexicon through double propagation. IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA.

Rogers, R. (2016). Twitter'ı Sıradanlıktan Kurtarmak: Bir Çalışma Nesnesinin Dönüşümü. In Weller, K., Bruns, A. Burgess, J. Marth, M. & Puschmann, C. (Eds.), *Twitter ve Toplum*. 24-33. Epsilon Yayıncılık.

Ruder, S. (2019). Neural transfer learning for natural language processing. (Degree of School of Engineering and Informatics). Natıonal Unıversıty of Ireland, Galway.

Ruhrberg, S.D., Kirstein, G., Habermann, T., Nikolic, J., & Stock, W.G. (2018). #ISIS—A comparative analysis of country-specific sentiment on Twitter. *Open Journal of Social Sciences,* 6, 142-158. https://doi.org/10.4236/jss.2018.66014

Samuel, A. (1959). Arthur Samuel: Pioneer in machine learning. Retrieved from http://infolab.stanford.edu/pub/voy/museum/samuel.html (accessed: 13.02.2022).

Sar, K. T. (2021). *Yapay Sinir Ağlari ve Bert Dil Modeli Kullanilarak Zaman Bazlı Duygu Analizi: Whatsapp Yeni Gizlilik Sözleşmesine Yönelik Yorumların Araştırılması.* (Yüksek lisans tezi). Dokuz Eylül Üniversitesi, İzmir.

Sarıman, G., & Mutaf, E. (2020). Sentiment analysis of Twitter messages in Covid-19 process. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences,* https://doi.org/10.38065/euroasiaorg.149

Sel, İ., & Hanbay, D. (2021). Gender identification from Turkish Tweets using pre-trained language models. *Fırat Üniversitesi Müh. Bil. Dergisi*, *33*(2), 675-684. https://doi.org/10.35234/fumbd.929133

Schuetze, H., & Manning, C. (1999). Foundations of statistical natural language processing. USA: MIT Press.

Sharma, R., Nigam, S., & Jain, R. (2014). Opinion mining of movie reviews a document level. *International Journal on Information Theory (IJIT), 3*(3), 13-21. http://dx.doi.org/10.5121/ijit.2014.3302

Shruthi, J., & Swamy, S. (2020). A prior case study of natural language processing on different domain. *International Journal of Electrical and Computer Engineering (IJECE*), *10*(5), 4928-4936. http://dx.doi.org/10.11591/ijece.v10i5

Sinderen, M.V., & Almeida, J.P.A. (2011). Empowering enterprises through next-generation enterprise computing. *Enterprise Information Systems*, *5*(1), 1-8. https://doi.org/10.1080/17517575.2010.528802

Singh, A. (2021). Evolving with BERT: Introduction to RoBERTa. https://medium.com/analytics-vidhya/evolving-with-bert-introduction-to-roberta-5174ec0e7c82 (date of access: 07.01.2022).

Singh, P. K., & Gupta, N. (2016). Recommendation model for infrequent items. *First New Zealand Text Mining Workshop, Hamilton, New Zealand.*

Sinha, V., Subramanian, K. S., Bhattacharya, S., & Chaudhuri, K. (2012). The contemporary framework on social media analytics as an emerging tool for behavior informatics, Hr Analytics and Business Process. *Management, 7*(2), 65-84.

Song, X. Salcianu, A. Song, Y. Dopson, D., & Zhou, D. (2020). Linear-time word piece tokenization. *arXiv preprint*. arXiv:2012.15524v1. 1-13.

Springboard (2019). Natural language processing use case – How do personal assistant apps work? Retrieved from https://in.springboard.com/blog/natural-language-processing/ (accessed: 28.12.2020).

Sun, C., Huang, L., & Qui, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*

Suppala, K., & Rao, N. (2019). Sentiment analysis using naïve bayes classifier. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8*(8), 264-269.

Stone, P., Dunphy, D., & Smith, M. (1966). *The General Inquirer: A Computer Approach to Content Analysis.* MIT press.

Stubbs M. (1996). *Text and corpus analysis: Computer-assisted studies of language and culture.* Oxford: Wiley- Blackwell.

Şeker, S. E. (2015). *Doğal Dil İşleme (Natural Language Processing).* YBS Ansiklopedi, *2*(4), 14-22.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M., (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.

Toçoğlu, M. A., Çelikten, A., Aygün, İ., & Alpkoçak, A. (2019). Türkçe metinlerde duygu analizi için farklı makine öğrenmesi yöntemlerinin karşılaştırılması, *DEUFMD*, *21*(63), 719-725.

Topbaş, A., Jamil, A., Hameed, A. A., Ali, S. M., Bazai, S., & Shah, S. A. (2021). Sentiment analysis for COVID-19 Tweets using Recurrent Neural Network (RNN) and Bidirectional Encoder Representations (BERT) models. *Proceeding IEEE Xplore Conference*. 21503449, Pakistan: Quetta. http://dx.doi.org/10.1109/ICECube53880.2021.9628315

Turney, P. D. (2002). Thumbs up or Thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia*, 417-424.

Turney, P., & Littman, M. (2002). *Unsupervised learning of semantic orientation from a hundred-billion-word corpus.* (Technical Report ERB-1094). Technical Report, Institute for Information Technology, Council Canada.

Uçan, A. (2014). *Otomatik duygu sözlüğü çevirimi ve duygu analizinde kullanımı.* (Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.

Uçar, T. (2020). BERT modeli ile Türkçe metinlerde sınıflandırma yapmak. https://medium.com/@toprakucar/bert-modeli-ile-t%C3%BCrk%C3%A7e-metinlerde-s%C4%B1n%C4%B1fland%C4%B1rma-yapmak-260f15a65611 (accessed: 24.11.2021).

Upadhyay, N., & Singh, A. (2016). Sentiment analysis on Twitter by using machine learning technique. *International Journal for Research in Applied Science & Engineering Technolog (IJRASET*), *4*(5), 488-494.

Vu, L., & Li, T. L. (2017). A lexicon-based method for sentiment analysis using social network data. *International Conference. Information and Knowledge Engineering*, IKE'17.

Walaa, M., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093-1113.

Wang, G., & Araki, K. (2007). Modifying SO-PMI for Japanese Weblog Opinion Mining by using a balancing factor and detecting neutral expressions. *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion*, 189–192

Wenbo, W., Chen, L., & Thirunarayan, K. (2012). Harnessing Twitter "big data" for automatic emotion identification. Proceedings *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*.

Weber, L. (2009). *Marketing to The Social Web: How Digital Customer Communities Build Your Business*. (2nd ed.), Wiley, Hoboken, NJ.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39, 165-210.

Williams, B., C. Halloin, C., Löbel, W., Finklea, F., Lipke, E., Zweigerdt, R., & Cremaschi, S. (2020). Data-driven model development for cardiomyocyte production experimental failure prediction. *ESCAPE30,* 48, 1639-1644.Italy: Milano.

Yılmaz, M. C., & Orman, Z. (2021). Sentiment analysis from twitter data during the Covid-19 pandemic era with LSTM deep learning approach. *ACTA INFOLOGICA, 5*(2), 359-372. http://dx.doi.org/10.26650/acin.947747

Yoldaş, İ. N. (2021). Türkçe metinlerde duygu analizi: sözlük tabanlı yaklaşım ve insanların tepkilerinin karşılaştırılması. *Journal of ESTUDAM Information*, *2*(1).

Yonamato, Y. (2013). Application of maximum entropy method to semiconductor engineering.*, Entropy,* 15, 1663-1689.

Yu, H., Deng, Z., & li, S. (2013). Identifying sentiment words using an optimization-based model without seedwords. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 855-859.

Yuan, B. (2017). *Sentiment analytics: Lexicons construction and analysis.* (Master Theses). https://scholarsmine.mst.edu/masters_theses/7668

Zhang, W., & Skiena, S. (2010). Trading strategies to exploit blog and news sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.*

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter - I hope it is not as bad as I fear-. *Procedia-Social and Behavioral Sciences,* 26, 55-62.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications* *38*(6), 7674-7682.