THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY

APPLYING NLP MACHINE LEARNING FOR NEWS ANALYSIS AND CLASSIFICATION

Master's Thesis

MAHER ALREFAAI

ISTANBUL, 2021

THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY

APPLYING NLP MACHINE LEARNING FOR NEWS ANALYSIS AND CLASSIFICATION

Master's Thesis

MAHER ALREFAAI

Thesis Advisor: ASSIST. PROF. ÖZGE YÜCEL KASAP

ISTANBUL, 2021

THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY
GRADUATE SCHOOL

...../....../.....

MASTER THESIS APPROVAL FORM

Program Name:	
Student's Name and Surname:	MAHER ALREFAAI
Name of The Thesis:	APPLYING NLP MACHINE LEARNING FOR NEWS ANALYSIS AND CLASSIFICATION
Thesis Defense Date:	

This thesis has been approved by the Graduate School, which has fulfilled the necessary conditions as a Master thesis

Assoc. Prof. Dr. Burak KÜNTAY
Institute Director

This thesis was read by us, quality, and content as a Master's thesis has been seen and accepted as sufficient.

	Title/Name	**Signature**
Thesis Advisor's		
Member		
Member		

ACKNOWLEDGEMENT

I want to start by expressing my deep gratitude and thanks to my supervisor, Assist. Prof. Özge Yücel Kasap for her support during my research work. She was managing the thesis professionally and gave me lots of helpful advice and tutelage.

Secondly, I am extremely grateful to my company AJANS PRESS GROUP, especially the CEO, Mr. Doruk Özkan, the CTO, Mr. Saeid Babel and the Project Owner, Mr. Erman Koyunoğlu.Who were always pushing me to develop my skills and build my experiences. I also want to thank my co-workers for their technical support, the permanent sharing of experiences, and the fact that they have made life in Turkey better for me.

Last but not least, I would like to thank my family for their support and prayers that encourage success. My father Dr. Khaled Alrefaai, My mother Dr. Eman Fares, my brother Muhammed Osama Alrefaai, my aunt Mayson Fares, my uncle Naser Fares, and the rest of the family.

İstanbul, 2021 Maher ALREFAAI

ABSTRACT

APPLYING NLP MACHINE LEARNING FOR NEWS ANALYSIS AND CLASSIFICATION

Maher ALREFAAI

Computer Engineering Master Program

Thesis Supervisor: Assist. Prof. ÖzgeYücel Kasap

May 2021, 56 Pages

A massive amount of news is daily published worldwide. Today, the news is published in various structures, and it is easily accessible to all people via phones or tablets. However, what is written in newspapers, electronic magazines, or social media reflects the people or entities' perspectives and opinions that wrote and published it. With this vast daily news, it is impossible even for media monitoring agencies to manually classify news for positive and negative news based on text and sentimental analysis, or even do other operations like news summaries. That's why the demand for using Natural Language Processing - a section of Artificial intelligence and Machine learning - comes on. In this study, NLP traditional and modern algorithms and models to automate news classification based on sentimental analysis is used. Algorithms and models such as TF-IDF Vectorization, Word Embedding, Random Forest, fastText and Bidirectional Encoder Representations from Transformers (BERT) are implemented on a novel data set and an accuracy of 95% was obtained.

Keywords: NLP Natural Language Processing, Machine learning, Sentimental Analysis, fastText Transfer learning, BERT Transfer learning.

ÖZET

HABER ANALİZİ VE SINIFLANDIRMASI İÇİN NLP MAKİNE ÖĞRENMESİNİN UYGULANMASI

Maher ALREFAAI

Bilgisayar Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi ÖzgeYücel Kasap

Mayıs 2021, 51 Sayfa

Dünya çapında günlük olarak çok sayıda haber yayınlanmaktadır. Günümüzde haberler birçok farklı biçimde yayınlanmakta olup, telefon veya tabletler aracılığıyla tüm insanlar tarafından kolaylıkla erişilebilir hale getirilmektedir. Bununla birlikte, gazetelerde, elektronik dergilerde veya sosyal medyada yazılanlar, onu yazan ve yayınlayan kişi veya kuruluşların bakış açılarını ve görüşlerini yansıtmaktadır. Büyük çaptaki bu günlük haberlerle, medya izleme ajanslarının bile metin ve duygusal analize dayalı olarak olumlu ve olumsuz haberler için haberleri manuel olarak sınıflandırması veya haber özetleri gibi diğer işlemleri yapması imkansız hale gelmektedir. Bu nedenle Yapay Zeka ve Makine öğreniminin bir bölümü olan Doğal Dil İşleme'yi kullanma talebi ortaya çıkmaktadır. Bu çalışmada duygusal analize dayalı haber sınıflandırmasını otomatikleştirmek için Makine Öğreniminin bir bölümü olan Doğal Dil İşleme'nin geleneksel ve modern algoritmaları ve modelleri kullanılmıştır. TF-IDF, Kelime Temsil, Random Forest, fastText ve BERT algoritmaları özgün bir very seti ile denenmiş ve yaklaşık % 95 doğruluk elde etmiştir.

Anahtar Kelimeler: Doğal Dil İşleme, Makine öğrenimi, Duygusal Analiz, fastText Transfer öğrenmesi, BERT Transfer öğrenmesi.

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AI	:	Artificial Intelligence
ML	:	Machine Learning
SA	:	Sentimental Analysis
NLP	:	Natural Language Processing
TF-IDF	:	Term Frequency - Inverse Document Frequency Vectorization
NN	:	Neural Network
MLM	:	Masked Language Model
NSP	:	Next Sentence Prediction
BERT	:	Bidirectional Encoder Representations from Transformer

1. INTRODUCTION

The news fieldwork has evolved from just tracking the event, writing about it and publishing it. To the process of analyzing news, comprehending what lies behind it, and presenting it in the form of a quick report containing its contents, opinions expressed, sentimental analysis, and information.

The traditional way of news analysis in newspapers and media monitoring agencies was for employees to read the news and report about it. This report usually gives answers to questions such as: is it positive or negative news? Is there racist or hate speech? What is the newspaper's or reporter's opinion while writing the news? And how does The reader will receive it? After highlighting the critical words, events, and names that appear in the news and will be shown under the light spot, the employee summarizes the news to publish it under the title. But with the increased number of daily news and the ability for anyone to start a website or social media accounts for news publishing, Traditional work has become very difficult. Think about it, is it possible for a limited number of employees to work on massive news every day? Nielsen et al. (2016).

As a result, media monitoring companies and major newspapers, such as the AJANS PRESS company have adopted a more modern approach. They attempt to fully automate the process by utilizing artificial intelligence techniques and Natural Language Processing. Perhaps this is why Mark Cuban (Entrepreneur and Investor) stated, "learn AI or become a dinosaur!".

Modern Machine learning and Natural language processing solutions will be used in this study for sentimental analysis and text classification. Machine learning is a collection of algorithms that learn and improve by analyzing data. The process of machine learning has four main tasks: Data gathering, data preparation, model training and model testing or evaluation. Of course, The process will be explained in the upcoming sections. Hussain et al. (2019).

However, the Natural language process (NLP) is a part of artificial intelligence that focuses on the interaction between computer and human languages. Building NLP algorithms entails extracting numerical features from the text for the model to classify it. The News Dataset used in this study was provided by the AJANS PRESS GROUP company and is being used for the first time as a contribution to Machine Learning and the NLP field.

1.1. OBJECTIVE

Through the integration of Machine Learning and Natural Language Processing, this study aims to introduce a Model that automates classifying Turkish News as positive or negative based on Sentimental analysis. This will reduce cost and time for media monitoring and news agencies, will transfer opinions into useful knowledge and build a competitive advantage in its services and websites.

While working on news text analysis, some subtasks, such as news summarization, will also be included in this research.

2. LITERATURE REVIEW

2.1. SENTIMENTAL ANALYSIS

Many analytical procedures are performed on news to have a better understanding of it. Article headlines, keywords, Search Engine Optimization, and, in the study case, Sentimental Analysis are all examples of operations.
Sentiment analysis, often known as opinion mining, is a field of study that focuses on people's attitudes, sentiments, and feelings. It will analyze them and attempt to glean important information or knowledge about how people feel about a brand, news, or company.

Opinions define and reflect behaviors and attitudes in human actions and activities. The aim of the action taking process is to know how others react to it. Will they like it or dislike it? All these reactions will help in predicting their reactions based on their opinions. Businesses, too, want to know how their customers feel about their products and brands, as well as who will buy new products in the future. People in politics also aim to understand societal feelings through what they say or write.
Sentimental analysis, an NLP technique, delivers extremely appropriate answers to these queries by evaluating people's discussions, postings, blogs, tweets, news, and so on. Agarwal et al. (2011).

Sentiment analysis is becoming a key tool for monitoring and comprehending this sensation as individuals communicate their ideas and emotions more openly than ever before. Sentimental analysis is the contextual extraction of texts that identifies and extracts subjective information from source materials and assists media monitoring, news agencies, and trading companies, for example, in understanding the social feeling of its brand, product, or service during online conversations. Sensitivity analysis is the contextual extraction of texts that identifies and extracts subjective information from source materials (Liu 2010, pp. 627-666).

2.1.1. Applications Of Sentimental Analysis Approach

2.1.1.1. Customer feedback analysis

Customer feedback may be a significant search marketing tool, improving bio-scoring, encouraging more user clicks, and potentially increasing conversions. Customer reviews promote corporate accountability, which boosts the company's online reputation. Customer feedback can be used to validate third-party users who build trust in their brand and online campaigns. The sentiment analysis approach assigns a score based on these terms and the related emotion in the user's response (Giachanou and Crestani 2016, pp. 1-41). As a result, good input receives a higher score, while negative input receives a lower value.

2.1.1.2. Sentimental news classification

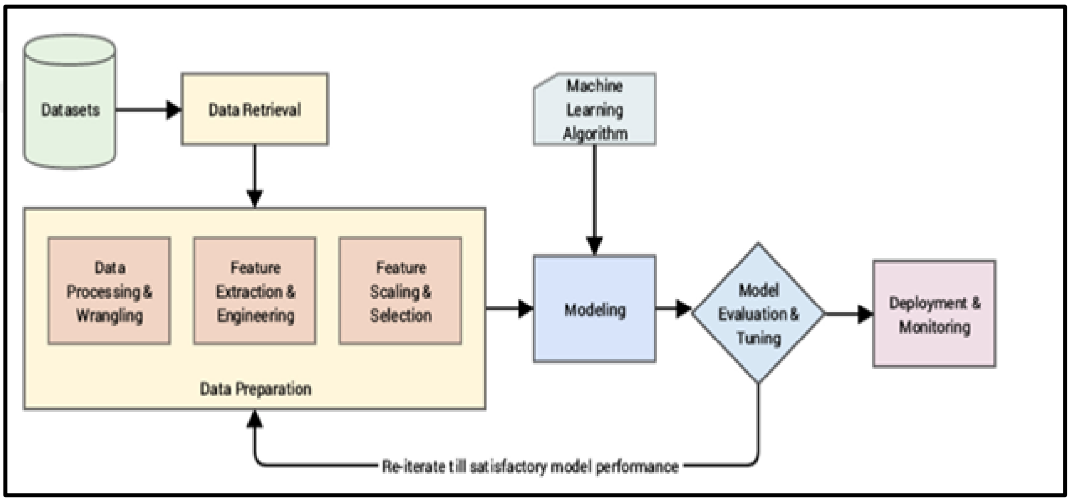
Today, news channels and newspapers are the centers of attention. Because the entire globe has become more accessible to one another and people enjoy keeping up with what is going on, especially now that the Internet and social media have created a great area for everyone to voice their thoughts and write what they want to say. As a result, it is critical to have a process in place that categorizes this content for people and extracts the emotions of journalists, newspapers, and satellite stations when they write and publish this or that piece. Knowing that these publications and television stations utilize a sophisticated and cutting-edge manner of targeting and disseminating news.

In any case, many news and media tracking websites have used sentiment analysis to better understand their viewers, which is the major purpose of this research.

2.2. MACHINE LEARNING

Machine learning is a branch of Artificial Intelligence (AI) that focuses on creating systems that can learn from data and improve their accuracy over time without being taught how to do so. Figure 2.1 depicts the machine learning workflow diagram Noguerol et al. (2019).

Figure 2.1: Workflow diagram of machine learning (Sarkar et al 2018)



Over the previous two decades, machine learning has evolved into one of the most important stages of information technology, and it has thus become an integral part of the daily lives. It combines concepts from neuroscience and biology, statistics, math, and physics to train computers. The learning process begins with observations or data, including direct expertise and guidance to find data trends and to make the best decisions in the future based on the examples given (Marsland 2015).

The main goal is to automatically enable computers to learn and adjust actions without human intervention or assistance. Big data is time-consuming and hard to handle according to human standards, but the best food for training an algorithm for a computer. An algorithm is a mathematical processing sequence in the field of data research. Algorithms are trained in machine learning to identify patterns and characteristics in large quantities of data to make decisions and predictions based on new data. The smarter, more accessible and easier to read data is in a large dataset, the more efficient the machine learning algorithm is to be trained (Jordan and Mitchell 2015, pp. 255-260).

Machine learning has proved to be one of the technological advances that have changed most in the last decade. In the increasingly competitive business world, companies can quickly track digital transformation and enter an automation era. It turns out that this is precisely what has created a neural network, but even their dad does not recognize them now, after the innovations that have reinterpreted them as statistical students. Data mining, which looks at extracting useful knowledge from large databases, has led to the shift in machine learning research. The learning model is a component of artificial intelligence and creates programs that learn from the data presented and improve without human interference. The possible adoption of machine learning algorithms and their prevalence in businesses is well documented, with various companies taking machine learning on a vertical scale (Li and Zhang 2017, pp. 416-420).

Machine learning enables computers to learn and experience and perform tasks similar to people's, the latter to observe large volumes of data and to predict using a statistical algorithm, to make it clearer some examples have been shown below:

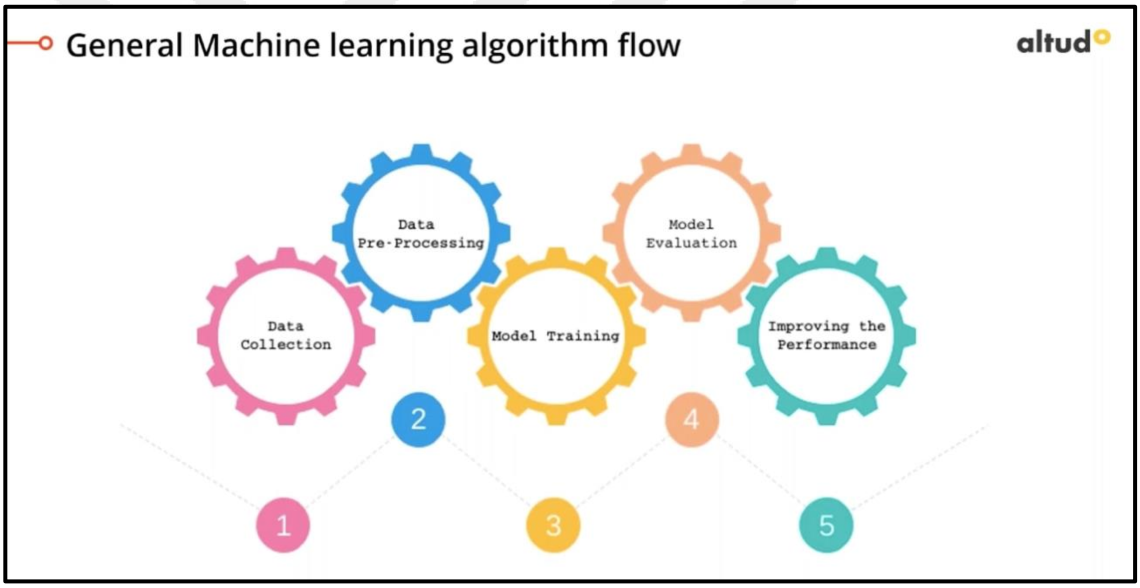
a. Automatic Running Cars: some companies, like Waymo, use machine learning methods to make self-driving cars that can be driven without a human captain. They use machine learning to observe what is going on around them and automatically generate a sense of how to behave. It shows the real importance of machine learning in helping human beings and predicts its importance and role in the future.

b. Virtual personal assistants: Siri, Alexa, Google Now are some of the most popular examples of virtual personal assistants. They help to find information on requests by voice, as the name suggests. a personal assistant is looking for the details, reminds customers of relevant questions or sends a command to other tools (such as telephone apps) to gather information. Machine education is a vital aspect of these personal helpers as they gather and refine the knowledge based on their previous participation. Later on, this data collection is used to produce results tailored to your requirements (Tulshan and Dhage 2018, pp. 190-201).

2.2.1. Machine Learning Process

An algorithm system works in accordance with its building model. The better the system, the more accurate and precise its work will be while processing. The main question that arrives to mind is, how does it work? The answer is not as complex as it may appear because anything that works has some form of procedure that must be followed. While considering machine learning, it may be obvious that it should be software there that collects data and distributes predictions and observations. Similarly, there are five steps shown in Figure 2.2. to build a machine learning model.

Figure 2.2: Steps of machine learning modelling (Kapoor 2021)



2.2.1.1. Data gathering

Data gathering is the process of gathering and measuring information from countless different sources. Algorithms require enormous amounts of data. Not to mention that the quality of the data is as essential as its size to train accurate models. Data collection enables you to record past events so that the data analysis can be used to find recurring patterns. Then predictive models can be built using machine learning algorithms to find trends and predict future changes from these patterns. There are many techniques and software for collecting data, such as scraping website tools. Note that some of these methodologies are not legal.

Mountains of data are available for machine learning, and certain companies (like Google) are willing to distribute it. The News Datasets in this study has been gathered and provided by AJANS PRESS GROUP Company.

2.2.1.2. Data preparation

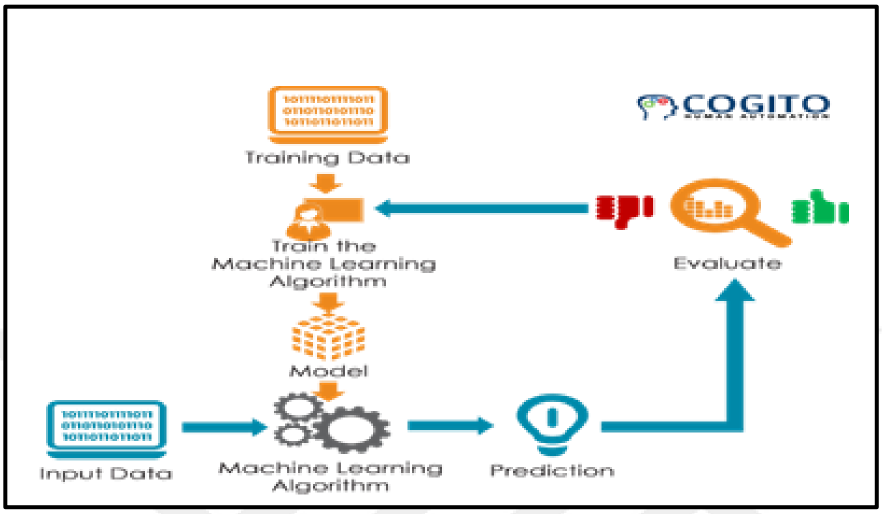
Before training the machine learning model or another, it is required to ensure that the data is clear, consistent and reliable. Data preparation technology is now in use for clean and annotated foundations necessary for modern machine learning. To make these forecasts reliable, the data must be accurately set and transformed (Swamynathan 2019). Formalized adverbial When the data is incorrect, the algorithm produces less precise or even misleading conclusions. The Data preparation step is the time to ensure that the data which is collected for the model is fully correct and will cause any problem that can twist the models' findings.

2.2.1.3. Model training

The training process of the ML model consists of providing an ML algorithm (i.e. the learning algorithm) with training data. Model machine language training is the process by which an ML algorithm with information is provided, so that good values are identified and learnt for all the attributes. In the training data, the learning algorithm finds samples that map the ingredients to a target (Brownlee 2013).

It involves executing algorithm variables and comparing the output to the results that it should produce, weight and bias in the algorithm to be adjusted, and the variables are executed until most of the time the algorithm returns the correct result. (IBM) Data teaches algorithms. You find connections, develop understanding, make decisions and assess your trust with the training data you receive. The better the training data, the more effective the model is and that's what is shown in Figure 2.3.

Figure 2.3: The complete workflow of model training (*Machine Learning* 2017)



2.2.1.4. Model testing

The last step is testing the complete model. It is the process in which the performance of a trained model is fully evaluated on a testing set. The model in this step will try to predict new testing data. The models' predictions are then compared to the actual outcomes to determine the proportion of right predictions. Model accuracy can be measured and expressed in a variety of ways, including F1 Score, Recall, and precision (Ghoneim 2019).

2.2.2. Machine Learning Methods

There are different methods for machine learning, which way to follow depends entirely on the problem statement. Mainly, machine learning tasks are classified into three different categories and they depend on the given data to the learning system as shown below:

a. Supervised Learning

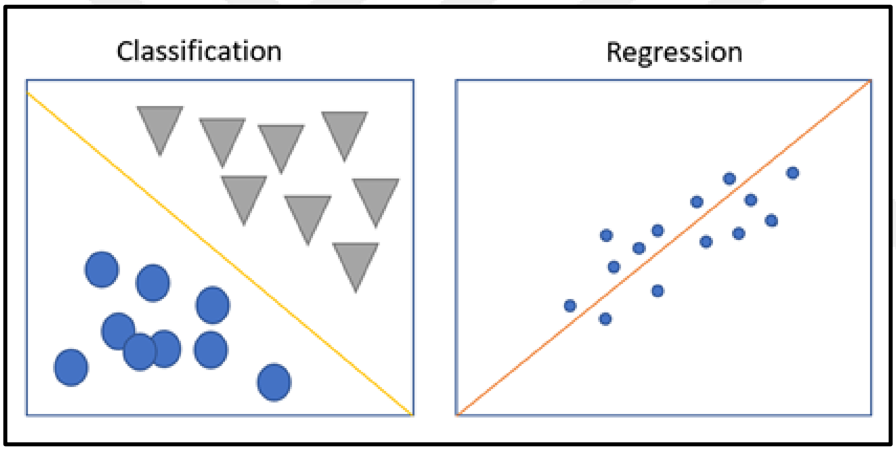
b. Unsupervised Learning

c. Transfer Learning

2.2.2.1. Supervised learning

When a student is learning a task under supervision like a teacher, the teacher in the end is going to judge his learning whether it is right or wrong. similarly, supervised machine learning has a full labelled data set the model to use while training. This trained model will eventually be used to make predictions on new unknown data. Xu et al (2019). The two famous supervised machine learning techniques are classification and regression.

Figure 2.4: Supervised learning classification and regression

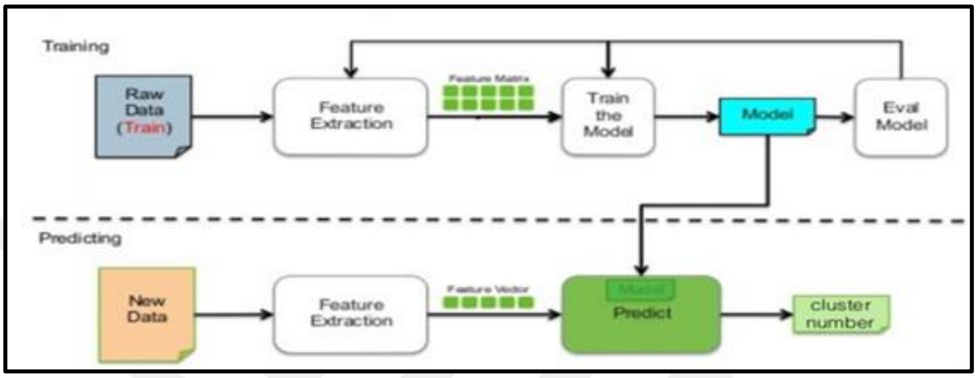


2.2.2.2. Unsupervised learning

Unsupervised learning is the process of training a machine with data that has not been classified or labelled, allowing the algorithm to act autonomously. Without prior data training, the machine's mission is to collect unsorted information based on similarities, patterns, and differences. Such learning can help transform data into groups based exclusively on statistical features. Goodfellow et al(2016).

Unsupervised learning does not require large-scale training. There is less to automate decisions and predictions, and more to identify patterns and connections in data that people might not know about. Clustering is a very famous unsupervised learning technique. Figure 2.5 explains how the unsupervised learning technique is working.

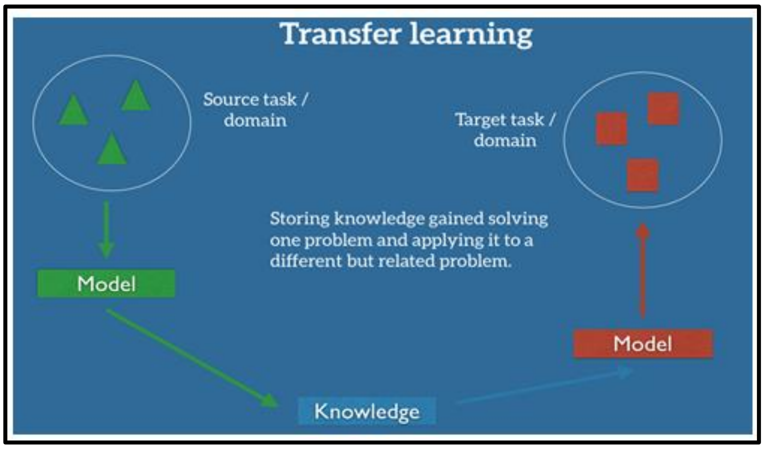
Figure 2.5: Unsupervised learning workflow diagram (Anon 2020)



2.3. TRANSFER LEARNING

The reuse of a pre-trained model for a new problem means learning from the transfer. These pre-trained models can have very sophisticated structures, require a large quantity of training data, and can be expensive for high-quality hardware settings. It is difficult for a single developer or researcher to train these models on their own. He can, instead, use them as a starting point for machine and deep learning projects. The mechanism of the transfer learning technique has been shown in Figure 2.6.

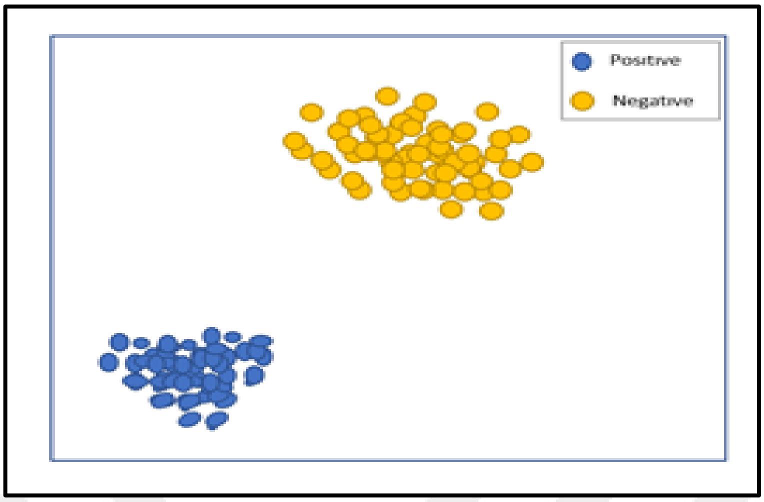
Figure 2.6: Chart model of transfer learning (Algorithm-X 2019)



2.4. CLASSIFICATIONS AS SUPERVISED LEARNING METHODOLOGY

Classification is defined as the process of recognizing, comprehending, and categorizing objects and concepts. Classification is a subset of supervised learning that aims to predict future instance class labels based on previous observations. A variety of algorithms are used in machine learning programs to classify future data sets into the matching and appropriate categories using these pre-categorized training data sets. In the case of a discrete value, classification is the task of predicting which class a data point belongs to. Linares-Vásquez et al (2014). Figure 2.7 shows how the positive and negative responses may appear according to the data classification system.

Figure 2.7: Data classification for positive and negative (neptune 2021)



2.4.1. How Classification Works

The training model will walk through every row in the training data and tries to understand the features and attributes of each class. It tries to figure out what makes this data row of class A or B. Later on, in the prediction phase of new data, the model will look at the new row attributes and see if it is closer to class A or B. For example, data from cancer patients who have already been classified as "needing surgery" or "does not need surgery" could be used as input for model training. Later, based on features learnt from the training data, such as tumor size and location, the model can generate predictions for future patients Banegas-Luna et al (2021).

2.4.1.1. Binary classification

In this type of classification, the classifying task of the input variable set is divided into two class/group labels, i.e., predicting which of the two classes corresponds to each element. Binary classification tasks usually involve a normal class and an abnormal class. For example, whether an image is of a cat or a dog or if an e-mail is spam or is not spam are binary problems. La Grassa et al (2020) "Not spam," for example is the normal and "spam" the abnormal state. Class 0 is assigned to a standard status class and class 1 is assigned to a class with an irregular condition.

Popular binary classification algorithms include:

a. Naive Bayes
b. Decision Trees
c. Support Vector Machine
d. Logistic Regression
e. k-Nearest Neighbors

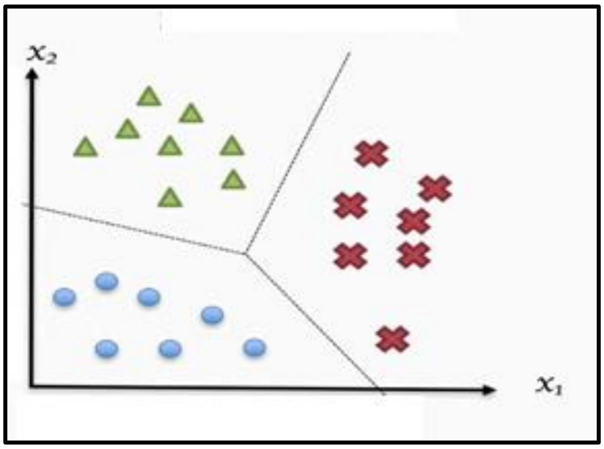
Some algorithms are designed specifically for binary classification and support no longer than two classes, such as logistic regressions and vector support machines.

2.4.1.2. Multi-class classification

The classifying task of the input variable set is separated into three or more class/group labels in this sort of classification. Unlike binary classification, multi-class classification has no concept of normal or abnormal findings. Rather, samples are drawn from one of several well-known categories (Ayodele 2010, pp. 19-48).

Some cases of this classification can be used: Some news is classified into different categories (political, sports/entertainment). In some cases, the number of class may be very high. For example, in a face recognition system, a model may predict a photograph as being one of the thousands or tens of thousands of faces. The data results according to multi-class classification have been shown in Figure 2.8.

Figure 2.8: multi-class classification (neptune 2021)



Popular classification algorithms that may be used include the following:

a. k-Nearest Neighbors.
b. Decision Trees.
c. Naive Bayes.
d. Random Forest.
e. Gradient Boosting

Some binary classification algorithms can also be adapted in multi-class classification.

2.4.1.3. Text classification

Text classification for supervised machine learning is one of the important and typical tasks (ML). Documents that can be a website, library books, media articles, galleries, etc. are assigned categories with many applications including spam filters, email routing, feeling analytics etc. Text Classification is an example of a supervised learning machine since a data set that contains text documents and their labels will be used for the formation of a classifier. Text classification also provides a good framework for getting acquainted with the processing of textual data without a lack of attention. Text classification systems can be used to organize, structure and categorize almost every type of text, from documents, medical research and files to the entire web Khan et al (2010).

2.5. NATURAL LANGUAGE PROCESSING

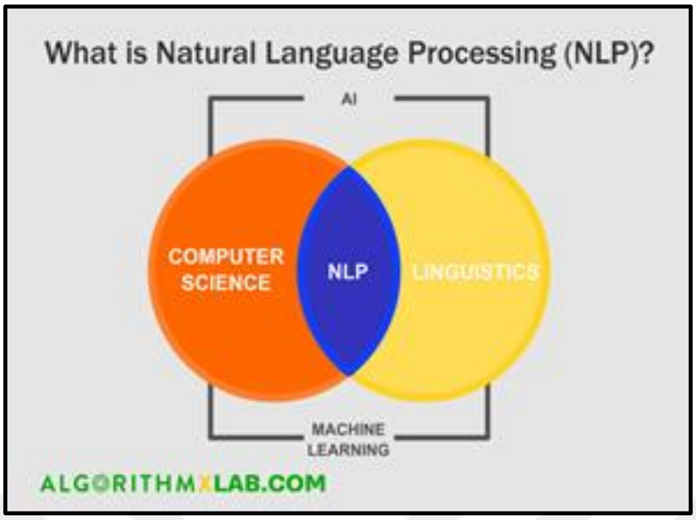
Natural Language Processing (NLP) is a part of AI that provides computers with the ability to comprehend words as people perceive them. In several fields, including computing and linguistics, NLP focuses on the difference between human communication and machine comprehension. The field of artificial intelligence is always visualized as giving machines the abilities of the human mind. However, language is one of the most important human mind creations that has contributed to civilization improvement along with history. Without a doubt, researchers are seeking around the world to integrate language into the field of artificial intelligence.

NLP offers computers a way of complying with humans in their language. For example, NLP enables computers to read, hear, understand, measure feelings and decide which parts are important.

Technology is progressing rapidly and big data, powerful computing and improved algorithms are available in addition. Machines can now examine more language data than humans, indefinitely and objectively. Developers can utilize NLP to organize and structure information in order to do tasks such as automated synthesis, translation, identification, relationship extraction, and emotion analysis.

NLP involves the use of algorithms for the identification and extraction of natural language rules to transform the unstructured language data into a computer-understandable form. If the text is given, the computer uses algorithms to extract meaning for each sentence and collect essential data from it (Indurkhya and Damerau 2010, pp. 4-20). Figure 2.9 shows the mechanism of natural language processing.

Figure 2.9: Diagram natural processing ideology (Algorithmxlab 2019)



2.5.1. Benefits Of NLP

Making a widespread analysis: Computers automatically interpret and evaluate vast volumes of text information such as social media commentaries, customer service tickets, online review, news reports, and more, using natural language processing.

Real-time automation of processes: Machines can learn to sort and travel information with little or no human interaction with a natural language processing tool fast, effective, accurately and round the clock.

Natural Language Processing Famous Use Cases

a. Recognition of Named Entity
b. Sentiment Analysis
c. Text Summarization
d. Aspect Mining
e. Topic Modeling

2.5.2. Applications Of NLP Approach

NLP is widely used in daily life applications and many examples can be seen around us. Here Are two important examples of natural language processing.

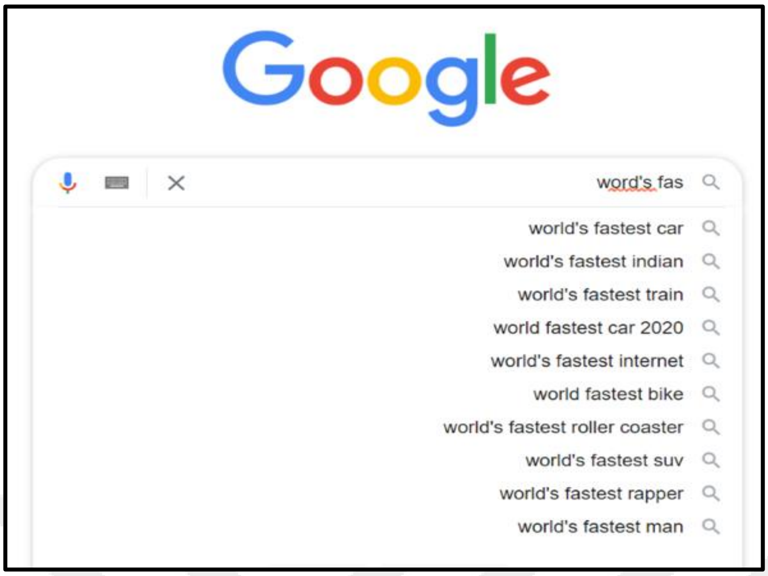
2.6. SEARCH AUTOCORRECT AND AUTOCOMPLETE

Google will start giving you recommendations for your search words after typing just two to three characters in the search input area. Or it corrects you if you look for something with the typos and still find the results that are important to you. Isn't it incredible? Another type of NLP that many people use every day and have grown to expect when searching for something is autocomplete and autocorrect search. This is largely due to pioneers like Google, who used this feature for many years on their search engine. Everybody uses it every day but never takes much care of it.

It is a fantastic implementation of the production of natural languages and a good example of how it affects millions around the world, you and me. Self-fulfilling and autocorrecting help both of us find precise answers quite effectively.

Now, several other businesses, including Facebook and Quora, have started using the feature on their websites. Figure 2.10 explains the process clearly.

Figure 2.10: Search autocorrect and autocomplete



2.7. SOCIAL MEDIA MONITORING

Social networking has become a major tool for e-retailers to consider consumer buying patterns, forecast demand for products or track marketing trends. If companies know what consumers feel about a brand on social media, they will continue to have a better product, service or customer experience. NLP facilitates surveillance and feedback response. Sentiment analysis enables companies to analyze this data and gather the feelings behind social media discussions to understand how and why people speak about a certain product or subject. The most common social media industries, restaurants, retail stores, internet providers and airlines can easily identify disgruntled clients, categorize issues as a matter of urgency and give priority to answers by means of sentiment analyses.

3. DATASET

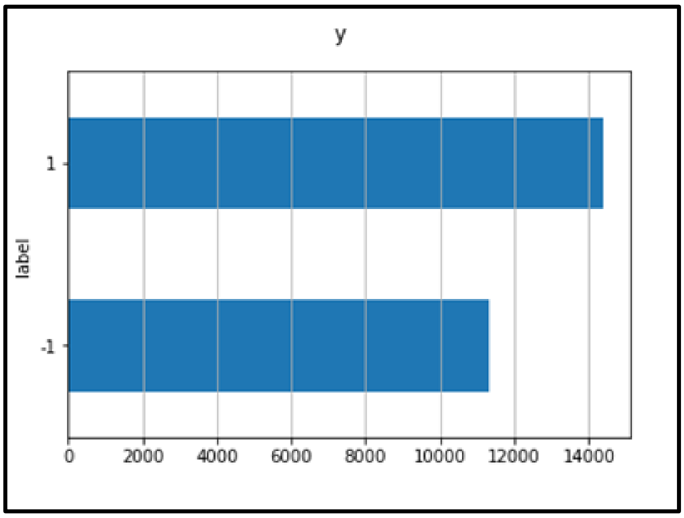
Only one database was used in this thesis to complete the classification of news into bad and good news. It is worth noting that this dataset is a novel dataset. it is never used before in Natural Language Processing or Machine learning researches. It was thought of using it for the first time as a contribution and addition to the researches of these fields.

3.1. DATASET

AJANS PRESS Company collects and provides the selected data. This text data comprises around 26,000 rows of manually categorized news as "Negative" or "Positive." As previously stated, there is a necessity to automate this classification process.

The histogram for the database is shown below, and it indicates the number of bad and good news. The dataset is semi-balanced, with 14500 positive and 11500 negative news items as mentioned in Figure 3.1.

Figure 3.1: News dataset positive and negative classes histogram



Since the computer cannot understand human languages directly. Researchers need to change dataset categories to numerical values; this operation is referred to as Label Encoder. where each label is assigned with an integer representative.

In Figure 3.1 it can be seen that the labels have changed to -1 and 1 where the "positive" news category is assigned to the value "1" and "negative" news is assigned to "-1". Now, computer algorithms can understand and work with these classes. This is one of the first very basic steps of the data preparation phase that will be explained in the upcoming sections.

4. METHODOLOGY

The work done on the news dataset to train ML classification models is described in this section. These models' structure and learning methodologies are detailed more in the following parts, but first, the data preparation, which is the initial stage will be discussed.

4.1. DATA PREPARATION

As it can be seen from Table 4.1, the first row contains positive news, that why it is assigned with 1. the second row is an example of negative news, which's why it is assigned with -1. Note that the original news is Turkish but it is here translated to English just for demonstration.

Table 4.1: A few rows of the news dataset

Turkish News	News Translated to English for Demo.	Category
100. yıl anısına 100 zeylin ağacı Büyükçekmece Belediyesi, Türkiye Büyük Millet Medisi'nin kuruluşunun 100'üncü yılında "23 Nisan Ulusal Egemenlik 100.	Büyükçekmece Municipality planted 100 trees.	1 ("Positive news")
Silahlı bir soygunda üç kişi öldü. Polis, geçen Pazartesi günü meydana gelen silahlı soygun sonrasında üç kişinin öldürülmesiyle ilgili	Three people were killed in an armed robbery. Police said they have launched an investigation into the killing of three	-1 ("Negative news")

soruşturma başlattığını söyledi.	people after an armed robbery last Monday	
…… +26000	…… +26000	…… +26000

The question that now arises is how textual news will be translated into characteristics. the answer to this question will be provided in this section as several operations have been informed on the text to convert it into classifiable material, these operations involve:

4.1.1. Data Cleaning

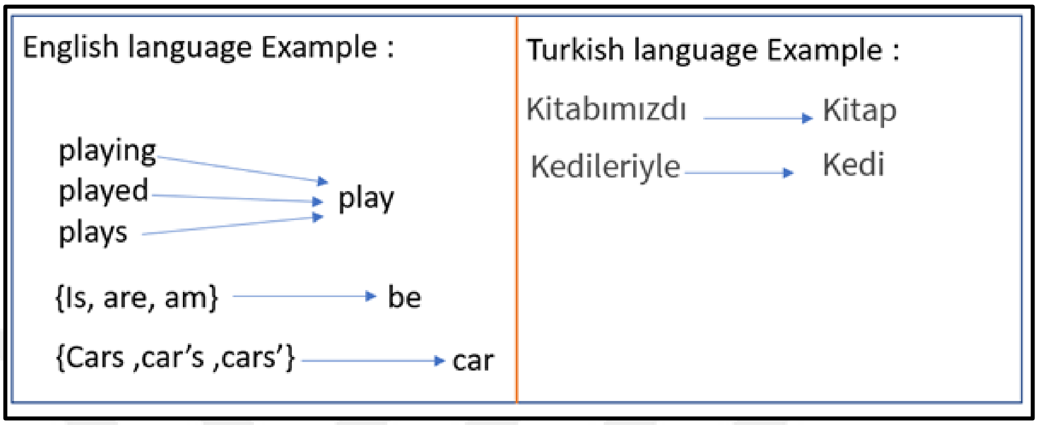
Using Python programming language could be achieved by going through the news text and remove things that increase the text size and do not constitute or provide useful information that can be relied on in classification. Typically, these useless additions occur during data collection and online scraping. The python program will consider removing numbers, e-mails, links, specials characters, single characters, Punctuation, multiple spaces, convert text to lowercase and prefixed. All these extras and useless information will make the classification less accurate and the classification process more difficult. As it increases the size of the data and does not add to the context and meaning.

4.1.2. Stemming And Lemmatization

The model shouldn’t treat one word as a group of different words because of the chronological conjugation, plural and singular, or possession and other language's grammar that may change the shape of the word and some of its letters. Therefore, during the data preparation phase, using some free open-source libraries such as the Turkish Stemmer python library may be effective to do what is called Stemming and Lemmatization.

These libraries are responsible for returning the word to its root as it is in the dictionary. fortunately, some of these libraries like the Turkish Stemmer library that used in this research support the Turkish language otherwise it can be difficult to use it since Turkish is an agglutinative language, some examples are shown in Figure 4.1.

Figure 4.1: Stemming and lemmatization example



4.1.3. Remove Stop Words

The analyst should create a list of the useless words and seek to remove them from the text as well. These words are called stop words. Of course, it differs from one language to another, and in most cases removing these words will help to raise the classification efficiency and show better results because the model will focus on the important words that make a difference. But on the other hand, this may also negatively affect the flow of the sentence and its context. There is a group of words agreed that removing it will not affect much and generally this issue depends on the opinion of the analyst or programmer. in this research, the analyst is highly encouraged to read some of the texts in his possession and understand the important and useless words, then he may add or remove them from the stop word list whenever he wants as shown in Figure 4.2.

Figure 4.2: Stop words list example

English stop words Example :		Turkish stop words Example:	
an	anybody	þey	þu
also	anyhow	þeyden	þuna
and	anyone	þeyi	þunda
another	anything	þeyler	þundan
any			þunu
....		....	

4.1.4. Handel Incomplete Data

Some data rows do not contain news, for example, there may be no text in them, or it may be a text that does not represent news, inside it contained the newspaper's name. There is also news that is classified as null. These situations must be treated by feeding the classification model complete and consistent training data. In this research, when a situation like this is encountered, the database provider -AJANS PRESS Company- would be asked to replace the incomplete or null news.

4.1.5. Split Dataset to Training and Testing

To complete the machine learning process, two separate datasets will be required. a dataset for training the model and another dataset to measure the performance of the model. In this search, only one dataset is available. Therefore, the available Turkish News Dataset will be divided into two parts. Usually, 80 per cent for training and 20 per cent for testing. The model will learn from the training dataset and will predict classes for testing.

4.1.6. Term Frequency - Inverse Document Frequency Vectorization

ML models consist of computer algorithms and computer algorithms can only deal with numbers. That's why they represent news texts must by numbers to feed in the training model. There are many ways to achieve that, in this section, an introduction of TF-IDF Vectorization will be discussed to train the first model. Later on, other methods such as Word Embeddings will be discussed Tirtha (2020) TF-IDF Vectorization consists of two parts Term Frequency and Inverse Document Frequency; where:

a. **Term Frequency (TF):** it shows how frequently a term or a word appears in a document.

$$Term\ Frequency = \frac{Number\ of\ term\ t\ appears\ in\ a\ document}{Total\ number\ of\ term\ in\ the\ dcoument} \tag{4.1}$$

b. **Inverse Document Frequency (IDF):** the weight of the rare words, where that occurs rarely in the document have high IDF score.

$$Inverse\ Document\ Frequency = \frac{\log(total\ number\ of\ document)}{(number\ of\ documents\ with\ term\ t\ in\ it)} \tag{4.2}$$

Finally, TF-IDF Vectorization is equal to the multiplication between TF and IDF. At the end of these calculations, each word in each news text row will be presented as a number making it a feature that can be inputted in the training model for classification.

$$TF - IDF\ Vectorization = TF * IDF \tag{4.3}$$

4.1.7. Term Frequency Other Applications News Summaries

A summary is a smaller set of the most important sentences in the original text. In NLP there are many ways to identify important sentences and Term Frequency is one example of that.

Journalists emphasize concepts and keywords in the story by repeating them often throughout the news text. By calculating TF for each word, the most important words that are repeated often excluding stop words will be identified. Thus, sentences that contain many important words are more important than other sentences as well. In the end, the sentences will be arranged according to their importance and take a quarter or a third of the number of sentences in the original text and put them together to be a summary of the news. The extracted important words can also be used as keywords, search words, and hashtags to help search engines and algorithms across different sites to reach the intended news Alqaryouti et al. (2018).

4.2. TRAINING AND TESTING MODELS EXPLANATION

By reaching this step, the data preprocessing and preparation has been finished, moving on to model training to explain the three models that have been trained which are Random Forest, Fast Text and Bert.

4.2.1. Random Forest Model Training and Testing

Random Forest is a very popular supervised Machine learning model, the main concept of it is ensemble learning where multiple classifiers are used to classify the training data and do predictions. The building block of Random Forest is Decision Tree, and It can be concluded from the word forest in the model's name that it contains a large number of decision Trees, a thousand trees, for example. Each tree takes a portion of the data and classifies it, and at the end, the Random Forest classifies each data row with a class that the majority votes. But what is the Decision Trees first? (Machine Learning Random Forest Algorithm – Javatpoint).

4.2.1.1. Decision trees

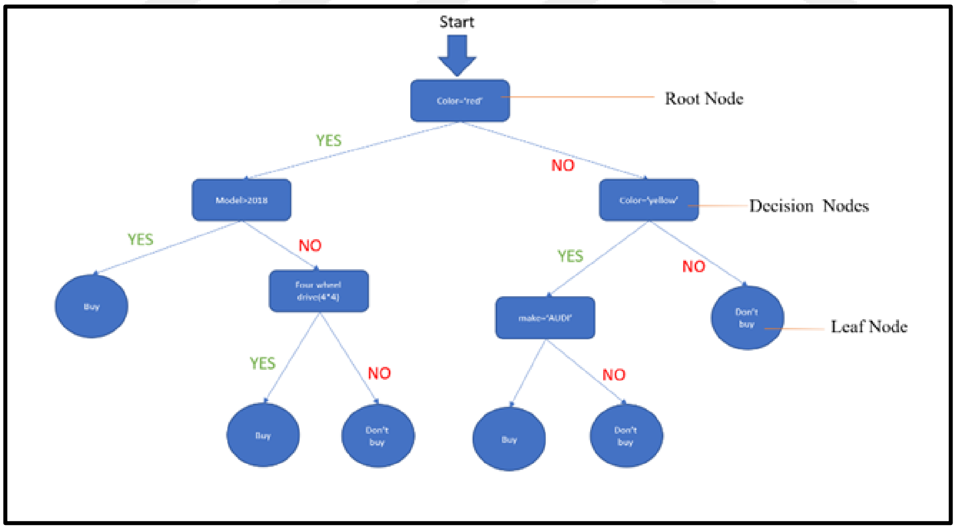
It is a tree-structured classifier used in supervised Machine learning to do predictive analysis based on historical data and experiences. Figure 4.3 explains the workflow of the decision tree. The tree is formed by a group of different types of nodes:

I. Root Node: the starting node of the tree, where the whole dataset features have been inputted and the features have been divided into many Decision Nodes.

II. Decision Nodes: each decision node represents a feature. This node helps in dividing the original big tree into subtrees. the decision node has another decision node under it as a child.

III. Leaf Node: is the final output of the tree. leaf nodes do not have any further child and it shows to which class the data has been finally labelled.

Figure 4.3: Example of decision tree buying a car



The question that poses itself now is what feature should be at the top as root or parent Nodes and what feature should come under it as a child? well this depends on Selecting Measures such as:

I. Information Gain: it calculates the change in Entropy after dividing the dataset by an attribute. This will display the quantity of information provided by a feature about a class when a higher value of information gain was attempted to be reached while developing the decision tree.

$$Information\ Gain = Entropy\ (S) - [(Weghted\ Avg) * Entropy(each\ feature) \quad (4.4)$$

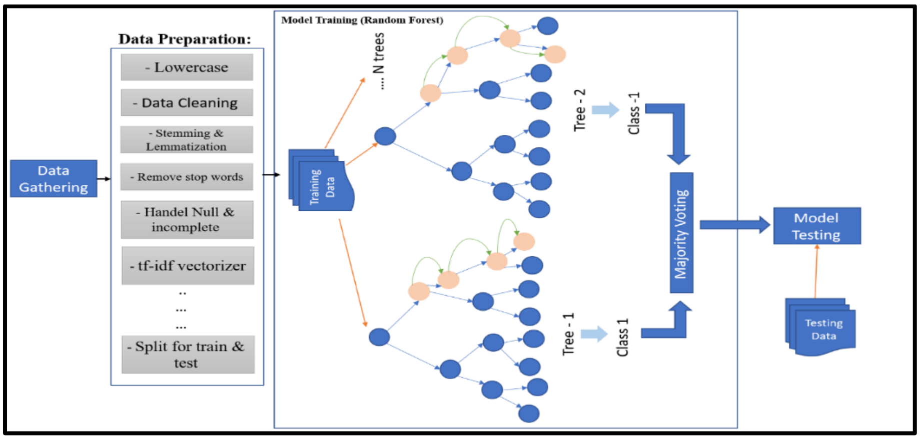
II. Entropy: is a measurement metric that shows the impurity of a given attribute.
$$Entropy(s) = -P(yes)log2\ P(yes) - P(no)log2\ P(no) \quad (4.5)$$
Where: S=total number of samples P(yes)=probability of yes P(no)=probability of no.

III. Gini Index: it is a measure of impurity and purity used in building a decision tree. attribute with the low Gini is better than attribute with high Gini.
$$Gini\ Index = 1 - \varepsilon_j\ P_j{}^2 \quad (4.6)$$

As mentioned before, Random Forest is a set of many Decision Trees explained above. Random Forest splits the data into many subsets and then for each subset it builds a Decision Trees. After that, it uses majority voting or averaging to get the final output.

The accuracy of the Random Forest model can be calculated by using the testing set. The model will be asked to make predictions for each row in the testing set and then compare the predicted class with the actual class for each row. Finally, the number of times the model properly predicted the class to obtain the accuracy value will be taken into consideration. A more detailed explanation will be written in the results section.
The work has been done so far for data preparation and Random Forest training and testing as simply represented by Figure 4.4.

Figure 4.4: Random forest training and evaluation architecture



4.2.2. Fasttext Model Training and Testing

FastText is a library for learning word embeddings and text classification created by Facebook's AI Research (FAIR) lab. Facebook trained this model on very massive data with 294 languages and made this pre-trained model available to use and fine-tune by transfer learning. NLP problem does need massive data because words in languages have infinite uses. three words with different order can form dozens of sentences.

Transfer learning and Word Embeddings will help us a lot in improving the accuracy as will be noticed in the results section. To build and train this model, Facebook relied on basic concepts found in machine training sciences and NLP such as Word Embeddings, Neural Networks and hierarchical SoftMax (Bird et al 2019).
To understand the mechanism and structure of the fastText model, these concepts should be explained:

a. Word embeddings
TF-IDF Vectorization has been discussed previously by saying that it helps us in converting text's words into numbers and features that the model understands when training. Word Embeddings is another way to convert text words into numbers and features. But it is a more advanced and effective method. What Word Embeddings presents is a vector for each word as it was by TF-IDF Vectorization, but it is also

distinguished by the fact that these vectors converge with each other when they carry a meaning that is close and diverging from each other when the meaning is opposite or different. Therefore, a text containing "this is awful news " will be close to a text containing "this is bad news" since awful and bad kind of Synonymous words Malafosse (2019). Here lies Word Embedding's strength and surpasses TF-IDF Vectorization who only gives a numeric weight to the word and may treat two words with the same meaning differently.

So, Word Embeddings starts creating a vector for every word and trains these vectors on divergence and convergence in the vector space using mathematical equations such as Cos Distance and COS Similarity. Of course, this is no stranger in machine training science and artificial intelligence science as these sciences are engineering applications of the math and physics fundamentals. Word Embeddings uses these equations is explained in figure 4.5. Once again, this will mark a major improvement in the accuracy and results of the model and here are the mathematical equations:

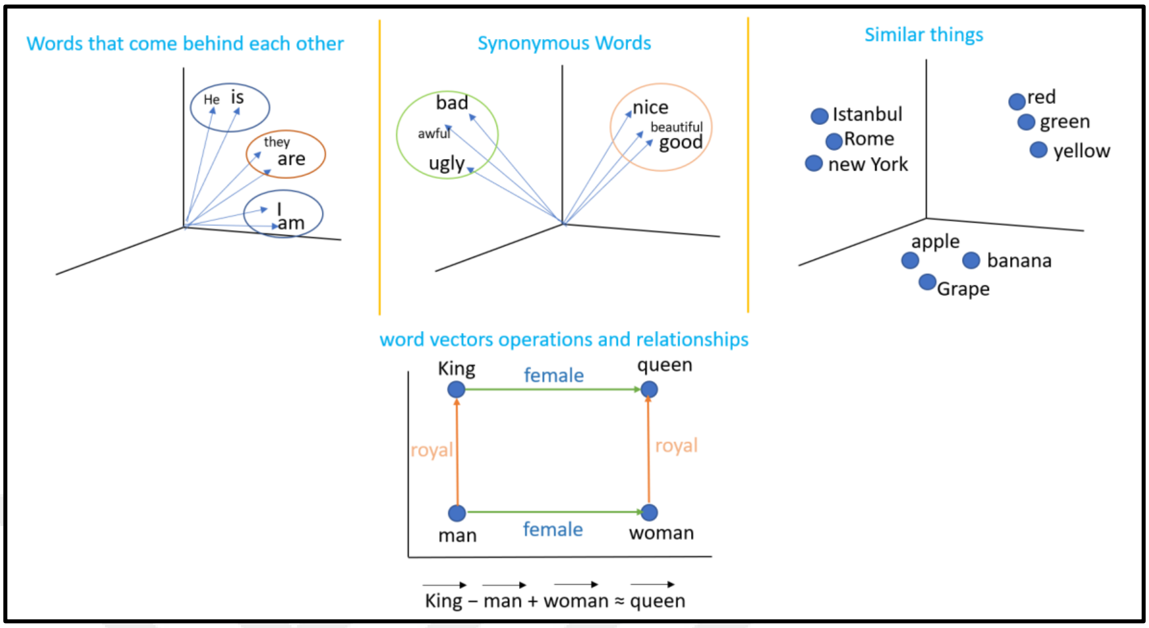
$$\cos\theta = \frac{\vec{a}.\vec{b}}{||\vec{a}|| \, ||\vec{b}||} \tag{4.7}$$

$$||\vec{a}|| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2} \tag{4.8}$$

$$||\vec{b}|| = \sqrt{b_1^2 + b_2^2 + b_3^2 + \cdots + b_n^2} \tag{4.9}$$

Where; a and b here represent a word vector

Figure 4.5: Examples of word embeddings



b. Neural networks

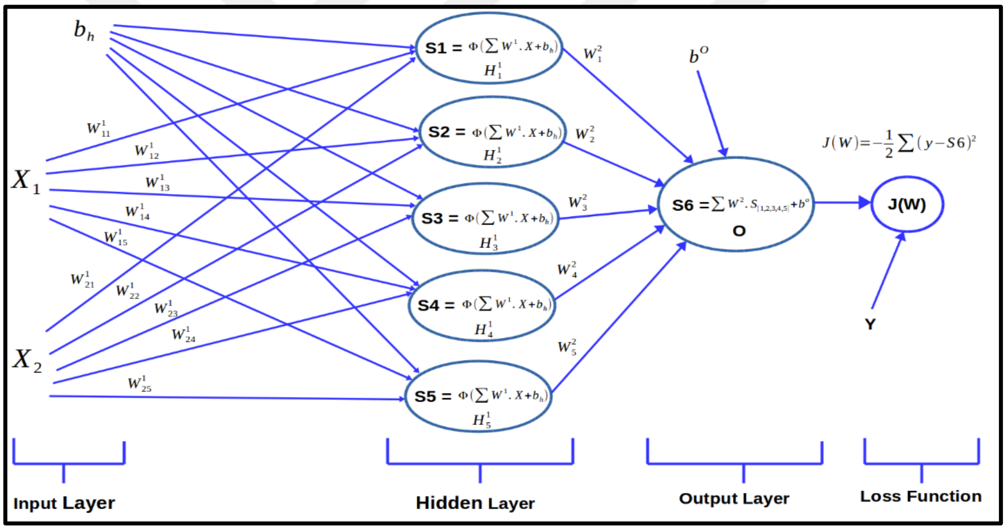
Neural networks are a method of machine learning. Scientists have long likened neural networks to the way neurons in the brain operate. As the science of artificial intelligence is nothing but an attempt to simulate the human brain. A neural network is formed from a neuron that takes an input value, a constant weight, and gives an output. A group of neurons creates a layer, and a group of layers creates Neural networks (Chouksey 2020). There are three basic layers of Neural networks, which are:

I. The input layer: this is the layer that starts the learning process. each neuron takes an input for example a word vector multiplied by weight matrix W.

II. The hidden layer: this is where all the magic happens, the hidden layer takes the output of the input layers, do extra calculations and map it to the next layer. There are different types of functions used within the hidden layers, and the analyst chooses what suits his learning purpose. But for the sake of explanation, an example of a function called Sigmoid Activation will be selected as it sums the output of the previous layer and adds a bias term b_h for it. the bias is just a constant that can be updated many times during each learning cycle. then outputs of the hidden layer will be multiplied with another weight vector, W^2.

III. The output layer: also has different types of functions used within it for example Linear Activation function that sums the inputs of the previous layer and adds another bias term, b^O. Finally, the learning process is evaluated by functions such as the loss function which calculates the error by measuring the difference between the actual and predicted output. Then it will update the weights and bias term and start all over again until the loss function error is minimum. Figure6.4 shows the structure and formulas of the explained Neural network. Where the hidden units use sigmoid activation and the output unit uses linear activation.

Figure 4.6: Neural network with a single hidden layer with 5 units(chouksey 2020)

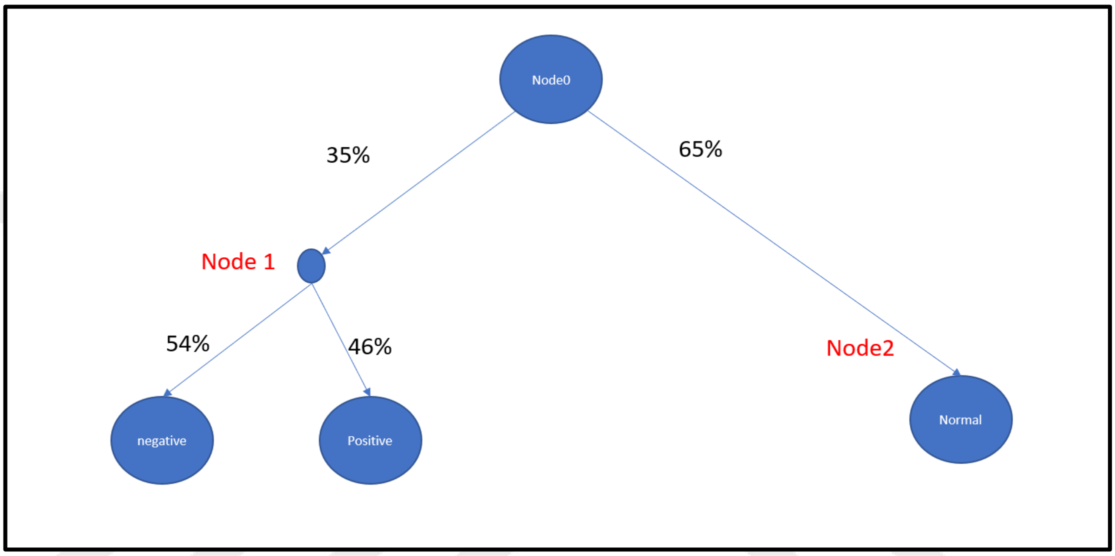


fastText uses Neural Networks for many purposes, such as creating Word Embeddings features or for training the model to classify these features into labels.it usually uses the Shallow Neural Network which is a Neural Network that has a small number of hidden layers one or two layers to make the learning process faster.

c. Hierarchical SoftMax

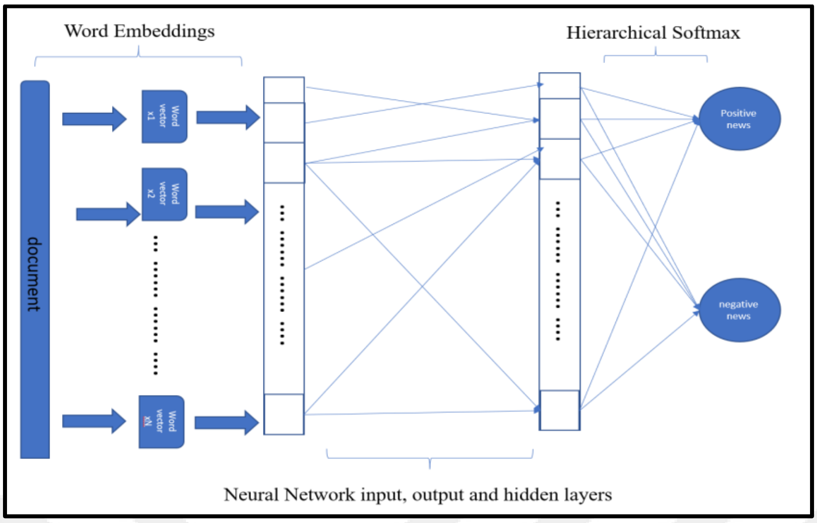
It is a binary tree used in the final stage of the fastText model architecture. Hierarchical SoftMax represents each node as a label and it is used to map the mathematical calculations of the fastText Neural Network to the output classes. Hierarchical SoftMax is also making the learning process faster by going throw a path based on first node probability that will lead to reaching the final class without needing to calculate other paths probabilities as shown in Figure 4.7.

Figure 4.7: Hierarchical softmax example



fastText puts all of the previously explained topics together to provide the pre-trained model. It is easy to download and import this model with python libraries. The news text dataset will be given as input in order for the model to do the rest for us. The next step as shown in Figure 4.8 will be using the model to do predictions on the test data set and calculate the accuracy.

Figure 4.8: Fast text pre-trained model architecture



4.2.3. Bidirectional Encoder Representations from Transformers (Bert) Model Training and Testing

Bert revolutionized NLP science in recent years. It is a model created by Google to help solve solving NLP problems such as Sentimental analysis, translating to different languages, and question answering. The original paper is published by Jacob Devlin in 2018. in the original paper, the model was trained on Wikipedia data with 800M words. This is a huge number of data that motivates all researchers to use this pre-trained Model in solving their problems. This is not the only motivation; Bert Encoder is also excellent at understanding language and context Devlin et al (2018).

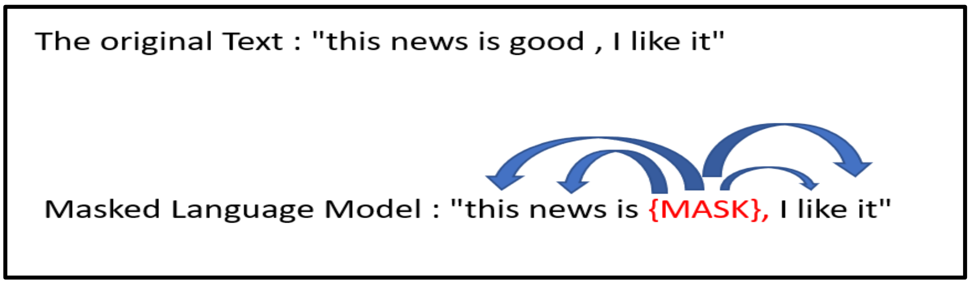
4.2.4. The Transformer's Encoder and Decoder

Transformers Encoder is a network that takes Word Embeddings as input and its main task is to understand the language, the grammar and most important the context of that language. The language has been used in this study is the Turkish language. Transformers Encoder does that by doing these two main tasks simultaneously:

a. Masked language model (MLM): here the model tries to mask some words in the given sentence and then it predicts the masked word tokens as shown in figure 4.14.

This is very similar to fill in the blanks. By repeating this process, the model will understand the relationship of each word with its neighbors.

Figure 4.9: Masked language model example



b. Next sentence prediction (NSP): In this phase, the model does a binary classification. it takes two sentences B and A and tries to determine if sentence B comes after sentence A. It outputs either 0 Or 1. That is how Transformer's Encoder Network learns about language and context, on the other hand, Transformer's Decoder uses the output of the Encoder to figure out how this learned language is related to the problem that has been solved. For example, if the model is trying to translate from Turkish to English. The Encoder will first learn the Turkish language and its context, then the Decoder will learn how does Turkish is related to the English language. Since Bert is an Encoder only, Decoder is considered as out of scope in this research.

4.2.4.1. Bert word embeddings

Although as explained above Word Embeddings previously, Word Embeddings in Bert differs slightly. So how Bert generates Word Embeddings must be explained. Bert generates Word Embeddings by adding three different values together. They are:

I. Token Embeddings: the first input in the Bert model is the Text and the text class. In this study, the news text and the news class was either "Positive class " or " negative class ". Token Embedding is a unique victor given to each word in this input. Again, vectors are closer when the words have a similar meaning as explained in fastText

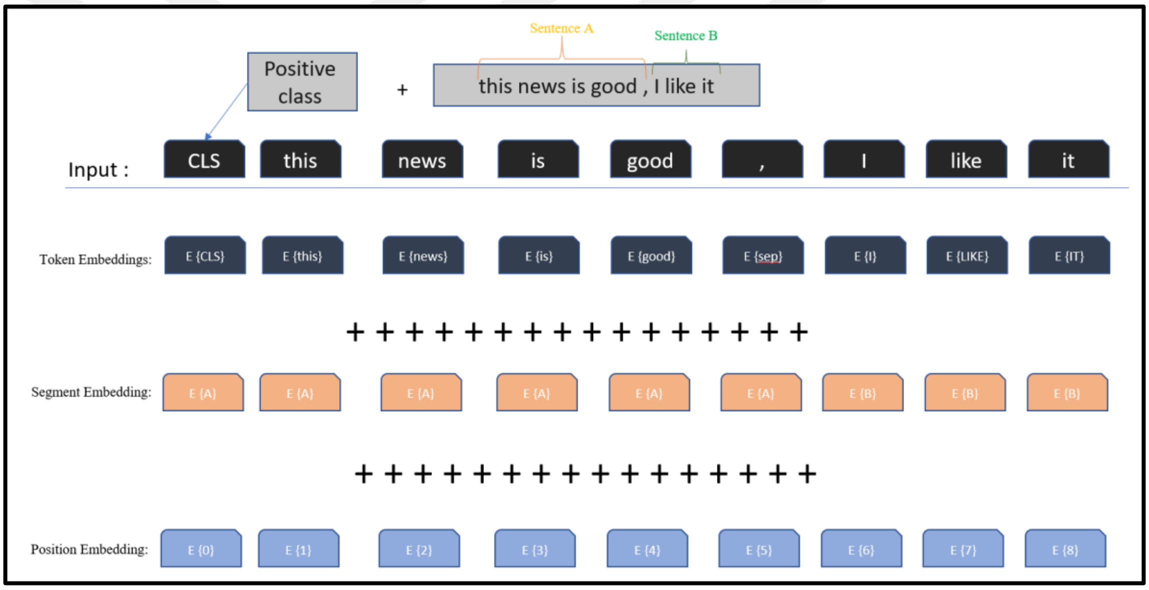
Embeddings. Token Embeddings is created automatically with Bert where the original paper uses 30K vocabulary tokens for this purpose.

II. Segment Embedding: it is the sentence number that is encoded into a vector.

III. Position Embedding: it is the Position of the word in that sentence.

These values are added to gather to generate the Bert Embedding as explained in figure 4.10.

Figure 4.10: Example of the Bert embedding



Bert Word Embeddings are the inputs of the Bert Encoder, and the outputs are the Next Sentence Prediction (NSP) and Masked Language Model (MLM). this is how Bert was pre-trained.

Now to use this pre-trained model for the NLP problem which is Sentiment analysis needed to download and import the pre-trained Bert model from libraries such as Transformers or in this study Hugging Face which has Turkish pre-trained Bert models to use in transfer learning. Then the model will be given the Turkish News training dataset as input. the model will create Bert Word Embeddings and start fine-tuning the Encoder Network. The rest of the process looks similar, SoftMax will map the encoder output to the final output layer which is the supervised problem classes "positive" or "negative " news in this study building.

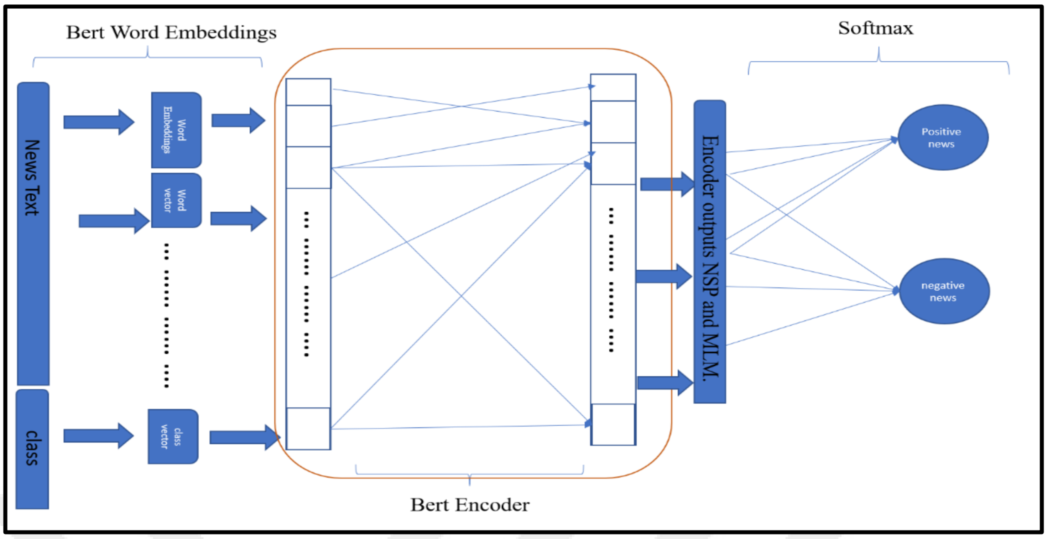
Finally, the errors in Bert is measured by the function Cross-Entropy Loss. Cross-Entropy loss is the distance between the predicted class vector and the actual class vector. (Gombru 2018)

$$Cross\ Entropy\ Loss = D(P, A) = -\sum_{i}^{c} A_i \log(P_i) \qquad (4.10)$$

Where; D is the distance, A is the actual class vector, P is the predicted class vector and C is the number of classes. One of the most important things noticed while training the models is that Bert takes a lot of time to be trained the architecture of the Bert model training has been explained in Figure 4.11.

In many cases, the researcher faces error, telling him that memory or RAM is full and Bert training cannot be completed. To solve this problem, a Google Colab platform that gives free GPU and RAM for developers OR analyst to work on has been used. It took us two hours to train Bert on Google Colab. on the other hand, training FastText faster, it takes only 20 minutes to train FastText on a private personal computer.

Figure 4.11: Bert model architecture



4.2.4.2. Bert training and testing summary:

I. Bert takes the News training dataset as input.

II. convert text news to Bert's Embedding with some word masked.

III. MLM and NSP will help Bert's Encoder network to better understand the language, word positions, word importance and sentences orders.

IV. the output of the Encoder will be mapped to positive and negative classes by SoftMax in the last layer of the network.

V. Cross-Entropy Loss will calculate the training error.

VI. the training process will be repeated until the error is minimum.

VII. Bert will use the News Testing dataset to do predictions for the classes.

VIII. in the end, the model's predictions will be compared to real values and the accuracy will be calculated.

5. RESULTS

After training and testing of classification models on the Turkish news dataset, A low accuracy for the manual trained model Random Forest about 53 per cent has been achieved and a very satisfying accuracy for the two transform learning models, 92 per cent for fastest and about 95 per cent for Bert. The results will be discussed in details in the discussion section.

First of all, the terms used to measure the model's performance has been defined as shown below:

a. True Positives (TP): refers to correctly predicted positive values for example if the actual news class is "positive news" and the model predicted class is "positive news" too.

b. True Negatives (TN): refers to correctly predicted negative values for example if the actual news class is "negative news" and the model predicted class is "negative news" too.

c. False Positives (FP): When the actual class is negative and the predicted class is positive.

d. False Negatives (FN): When the actual class is positive but the predicted class is negative.

e. Precision: it is the ratio of correctly predicted positive observations to the total predicted positive observations.

$$Precision = \frac{TP}{TP} + FP \tag{5.1}$$

f. Recall: it is the ratio of correctly predicted positive observations to all observations in actual class positive.

$$Recall = \frac{TP}{TP} + FN \tag{5.2}$$

g. F1 score: it is the weighted average of Precision and Recall.

$$F1Score = 2x(precision * Recall)/(precision + Recall) \tag{5.3}$$

h. Accuracy: it is a performance measure that is equal to the correctly predicted observation to the total observations.

$$Accuracy = TP + \left(\frac{TN}{TP}\right) + TN + FP + FN \qquad (5.4)$$

5.1. RANDOM FOREST RESULTS

For negative news the precision is 0.45, recall is 0.27 and the F1-Score is 0.33. where for positive news precision is 0.56, recall is 0.74 and the F1-Score is 0.63. The model overall accuracy is 0.53. The results according to the random forest model are written in Table 5.1.

Table 5.1: Random Forest results

	Precision	Recall	F1-Score	Support
Negative news	0.45	0.27	0.33	1146
Positive news	0.56	0.74	0.63	1424
Accuracy			0.53	2570
Macro avg	0.50	0.50	0.48	2570
Weighted avg	0.51	0.53	0.50	2570

5.2. FASTTEXT RESULTS

For negative news the precision is 0.92, recall is 0.93 and the F1-Score is 0.92. Also, for positive news precision is 0.93, recall is 0.92 and the F1-Score is 0.92. The model overall accuracy is 0.92. The results according to the fastText model are written in Table 5-2 below.

Table 5.2: fasttext results

	Precision	Recall	F1-Score	Support
Negative news	0.92	0.93	0.92	1250
Positive news	0.93	0.92	0.92	1250
Accuracy			0.92	2570
Macro avg	0.92	0.92	0.92	2570
Weighted avg	0.92	0.92	0.92	2570

5.3. BERT RESULTS

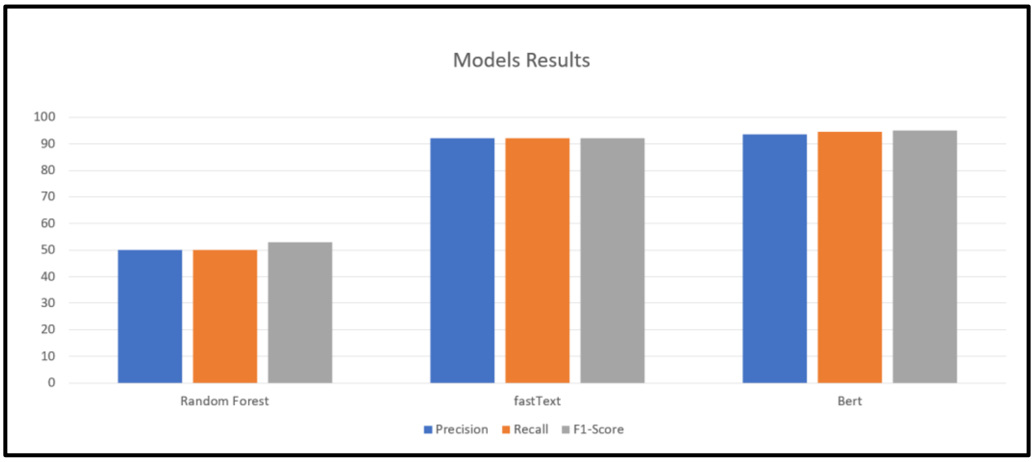
For negative news the precision is 0.90, recall is 0.94 and the F1-Score is 0. 920. Also, for positive news precision is 0.971, recall is 0.95 and the F1-Score is 0.96. The model overall accuracy is 0.947. The results according to the Bert model is written in Table 5-3 below.

Table 5.3: The results according to Bert

	Precision	Recall	F1-Score	Support
Negative news	0.900	0.940	0.920	1382
Positive news	0.971	0.950	0.960	2895
Accuracy			0.947	4277
Macro avg	0.935	0.945	0.940	4277
Weighted avg	0.948	0.947	0.947	4277

Finally, all model results are put together in one figure So that the difference between them is visible. Figure5.1 explains the differences in the results according to the different models by making a comparison between them.

Figure 5.1: Comparison of the different model's results



6. DISCUSSION

6.1. WHY BERT AND FASTTEXT ARE BETTER THAN THE RANDOM FOREST MODEL

Although the data has been prepared in an acceptable way before training Random Forest, the data has been re-cleaned in many steps to convert the text into features with TF-IDF Vectorization. But BERT and fastText have more sophisticated ways to do this. BERT for example has millions of millions of training data and thus they can understand the language, contextual texts, important and unimportant words of the sentence. They also generally use Word Embeddings, which gives the words vectors that converge and diverge a sense of similarity and difference in meanings. As for the TF-IDF Vectorization used in Random Forest, it gives only a numerical weight to the word, and the model has been trained only on the Turkish news dataset, meaning that it is trained on a limited number of words, sentences and limited understanding of the language. Of course, this is not enough in NLP problems that often require an enormous number of data. Anyway, throughout Random Forest training, a better grasp of text data preparation, supervised text classification, text summarization, and keyword extraction was gained, which was encouraging to begin with.

6.2. COMPARING FAST TEXT AND BERT MODELS

Although the results of these two models are close, Bert is still better than fast Text by accuracy measures such as Precision, Recall and F1-Score.
Once again, although the two models use Word Embeddings, Bert's Word Embeddings is generated differently, so that the word, its location and place in the sentence are preserved and then given to Bert Encoder. Also, the Bert Encoder has a more complex structure to understand and learn the language than fast Text. How Bert Encoder is concerned with

each word and its relationship to adjacent words by Masked Language Model (MLM) and even the relationship of each sentence to the next sentence with Next Sentence Prediction (NSP) has been explained in this research.

However, this structure makes the Bert model slow to train and is often time and resource-consuming such as RAM and CPU. This makes it more difficult to use in the servers and the clouds.

On the other hand, fastText, even if it is a little bit lower accuracy but is faster than Bert in training and it consumes less amount of time and RAM. In this study, Bert has been trained in two hours, but fastText only took twenty minutes. This is what makes some people prefer it over Bert at times.

7. CONCLUSION

In the current world, news spread on an endless number of platforms. With this expansion, large newspapers and media monitoring companies need to extract useful information from the text of this news and classify them according to their Sentimental analysis into positive and negative news. Thus, it will give analytical services to its clients and its users. Also, the normal newsreader will understand more deeply the news, its content and the opinion of its writer. Of course, this is no longer possible traditionally and manually. That’s why this research has been done by using Machine Learning and Natural language Processing to construct neutral models capable of completing the task of text Sentimental analysis classification automatically.

Detailed information about the traditional and non-traditional methods of preparing texts for input in the model such as TF-IDF Vectorization and Word Embeddings has been learned through the study as well as the ability to understand the supervised text classification with the Random Forest has been reached. However, looking at the results shows that NLP problems require an enormous number of data, and that is why the transfer learning to train BERT and fastText from Google and Facebook has been used and as a result, BERT was better in accuracy terms than fastText, but fastText was faster to train and learn from BERT.

Some other applications of NLP have been used during the study, such as summarizing news and extracting search words and the final results for the models are 53 per cent for Random Forest, fastText 92 per cent and BERT 95 per cent.

8. FUTURE WORKS AND CHALLENGES

What can be good news for a person can be bad news for another person. Therefore, the satisfaction shouldn't be happened with analyzing the news text, but the reader should be knowable as well. So that building a profile for him and based on his profile, models can predict what is good news and what is bad news for him.

The news sector is huge. And the data in this sector is very huge as well. Therefore, big data technologies like Hadoop are needed to save data in Clusters and Spark to pass data between these clusters and other technologies that help in managing and storing this massive data and keep models updating and learning.

9. TECHNOLOGIES AND LIBRARIES USED

I. Python language programming.

II. Pandas: open-source data analysis library, helps in creating and managing data frames.

III. Natural Language Toolkit (NLTK): for building Python programs to work with human language data.it provides a list of stop words for most of the languages, TfidfVectorizer, Stemming and Lemmatization.

IV. Turkish Stemmer: an open-source python library used for returning Turkish word to its root in the dictionary.

V. Sklearn: Machine learning library for feature extraction, pipelines and model selection such as Random Forest Classifier.

VI. fastText: this is the fastText model provided by Facebook.

VII. Hugging Face: it has the pre-trained Turkish Bert that can be used for transfer learning.

VIII. Anaconda: it is a data science platform that helps to download the previous libraries and solve version conflicts between them. using Anaconda, the developer can create many development environments that suit his task.

IX. Google Colab: it is a data science platform for Machine learning engineers, it offers a stable development environment with free GPUs, TPUs and RAMs to make work faster.

10. RELATED RESEARCHES AND PAPERS

In this section, the most important papers and research that this study has built on will be listed such as the original google Bert paper and fast text papers and under each paper title, a summary of the main concepts presented in the paper, the experiments, and the results of these experiments will be explained.

10.1. BERT, PRE-TRAINING OF DEEP BIDIRECTIONAL TRANSFORMERS FOR LANGUAGE UNDERSTANDING

Bidirectional Encoder Representations from Transformer (BERT) is a new language model designed to combine deep two-way unlabeled text representation both left and right in all layers. BERT representations can be completed with just one extra layer for a wide range of tasks for the production of state-of-the-art models. It includes evaluating tasks, such as natural language references and paraphrasing, to predict the full application of sentences to each other and how to classify token level so that modelling can be applied to fine-grained token results. Features-based and fine-tuning are two strategies where feature-based architectural features are used as additional functionality and fine-tuning introduces minimum task-specific parameters, which are trained in downstream tasks. These approaches serve the same purpose to learn the representation of general language by means of unidirectional language models. Through a masked language model, this BERT model improves these two approaches, providing only contextual fore comes of the original masked word vocabulary. In the pre-BERT world, a language model may have been examining this text series during training from left to right and from right or in combination. When BERT enters the world of language, the language context and flow can be understood more clearly than the one-management model.

The masked language model allows both the left and the correct contexts to be combined, which allows an advanced byway transformer, contrasting with the left and right pretraining model.

In contrast to previous models, the bidirectional encoder representation from Transformer is automatically trained in the uniformity of language and a slim linkage between left and right. Although initialized with the same pre-trained parameters, BERT is the first finished model to achieve this after performing many specific tasks, to ensure their state-of-the-art performance in a wide array of sentence-level and token-level jobs.
Bidirectional Encoder Transformer representatives achieve cutting-edge results for 11 natural languages processing tasks, including MultiNLI (86.7 per cent accuracy), SQuAD v2.0 (83.1) SQUAD V1.1 (93.2 per cent) and GLUE (80.5 per cent) etc. These results enable even low-resource tasks to benefit deep unidirectional architectures. The improvements resulting from the transfer of language models have shown that rich and unexpected pre-training is an integral part of many language understanding systems (devlin et al, 2018).

10.2. ENRICHING WORD VECTORS WITH SUBWORD INFORMATION

Consistent word representations, learned in large and unstuck companies, help many tasks with natural language treatment and have restrictions, especially in high vocabulary and rarely used languages. Popular models that learn such representations ignore word morphology by assigning a different vector to each word. This constraint, based on the grammar paradigm, is addressed by a modern solution in which any word is represented as an n-gram bag of character. Every n-gram is linked to a vector representation; the terms are the sum of those statements. These interpretations are usually derived from large unmarked businesses using statistics on joint events, yet a wide range of work called distribution semantics have examined the properties of these approaches. Collobert and Weston suggested learning word embedding using a neural feed network in their neutral network community by predicting two words on the left and two words on the right. Each word in the vocabulary is represented by different vectors, most of these technologies neglect the internal structure of words which, without parameters being shared, constitute an important limit for morphological languages such as Turkish or Finnish.

Because many word formations follow rules the use of knowledge at the character level can improve vector representations for morphologically rich languages. Learning representations of n-gram character and word representation as a sum of n-gram vectors should be set as a purpose. While the main contribution was to extend the skip-gram model continuously, which takes into account the subword knowledge, testing this model and show the advantages of the used approach in nine languages is an important point as well. Morphology is shaped by a sum of its characteristics N-grams through the use of subword units and words. It has decided to use n-grams which vary from the fact of n-grams covering a wide range of information and is based on the application of n-gram character. It is quickly trained because of the simplicity of the model and does not require preprocessing or monitoring, and it exceeds the basic performance, which takes no account of subword information, and methods based on moral analysis. This model will open up a platform for comparing future work on sub worth learning representations. Compared to recent morphological word representations noticing that these tasks are performed advanced by the vectors Bojanowski et al (2017).

10.3. BAG OF TRICKS FOR EFFICIENT TEXT CLASSIFICATION

In terms of precision and precision in deep learning, fast text classification is a very open source for quicker education and evaluation. Fast Text is a free, open-source, usable library for text education and text representation on various engines. The classification of text is a major task in natural language processing in many applications, including web-based search, classification of documents, ranking and information gathering. Text classification offers the right framework for the uninterest-free processing of textual data. There are many interesting applications for text classification, for example, spam detection and feeling analysis. A quick text classifier also includes different models for the completion of such tasks and achieves excellent performance, but both in train and in test time are relatively slow, which limit their use on large data sets in the results. In the meantime, linear classifiers are considered as solid foundations of text classification issues. Despite their simplicity, they achieve modern performance, often when using the right features.

Training fast Text in under 10 minutes with a standard multicore CPU with over one trillion words and classify half a million sentences into 312K classes in less than a minute could be done. A new way to scale the basic lines into a large corpus with large output areas in Text Classification should be found. Inspired by the recent work on effective speech use. This is why the algorithm of the text classification is analyzed and the principal benefit is as quick as possible the simplicity of text is extremely useful for formal analysis. While deep neural networks have far greater power of representation in theory than faultless models, it is not clear whether simple issues of text classification such as sentiment analysis are right to evaluate. One fundamental approach that can be adopted is to make the text classification model a far better way than other models such as word2vec to make good sentence representation. The model is trained in accordance with recent methods and much quicker. (Joulin)

10.4. TEXT CLASSIFICATION MODEL BASED ON FASTTEXT

For many applications, text classification is a major issue, such as spam, feel analysis or smart answering. Many text grading models based on traditional algorithms have dimensionality and performance problems. Apprentices' classification of text machine learning methods based on data observations. Therefore, fastText offers a great way to construct NLP templates and generate live development environments forecasts. This model examines important text information through feature engineering and obtains low-dimension, continuous and good text by means of the fastText algorithm. As technology develops daily and rapidly popularized, text information in this field has increased rapidly. When looking for the data from such a large dataset, the user has to face many problems. The issue is solved to a certain degree by the fastText classification by identifying the user's predictions by reducing massive text information causing the problem.

The most common methods for text classification are the Vector Space Model (VSM), the Nive Bayes (NB), the nearest neighbor (kNN), the Latent Dirichlet Assignment Model (LDA), the SVM, and the neural network method. All of this method is not sufficient to capture the principal semantics of the text, which lead to poor test performance. These models created the testelm of the text, which was solved in the creation of a quick text pattern. It represents text in a small, continuous space and feature engineering, and main semantics with high input text quality. It shows the positive or negative user reviews, which can help determine the product's quality. The research demonstrates that the model is effective in many ways and that it is higher than other classification algorithms. The preprocessing of text and the tokenization of the classifiers are important and have a major influence on the performance of the classifier. Therefore, training new algorithms for the machine is very beneficial and will be widely used in future to achieve better results.

10.5. TWITTER SENTIMENT ANALYSIS USING DEEP LEARNING MODELS

Twitter is one of the most famous social networking websites where users can read and post messages on a person, an event, a product and events worldwide. The data in its database is enormous and people show their instant feelings and perspectives, which can provide a great platform for people to understand what they think and understand. The classification of social media is analyzed by different tools, and the polarity of feelings is primarily assessed by textual content and other platforms are neglected. The text of social media like Twitter, Goggle BERT, Long Speaker (LSTM) and Convolutional Neural Network, can be classified using 3 sentimental classification models. A natural language processing and in-depth learning techniques should be used to gain subjective knowledge of the text and attempt to classify it as having a positive, neutral and negative polarity. These instruments are considered state of the art algorithms for most natural language processing tasks. This allows us to predict any company's stock markets and determine the general impression of a sales product.

Many people use social media to connect or keep up with news and keep an eye on what is going on all over the world. In products, film review, scholarships, government policy and other matters, Twitter has the highest consideration. People from around the world give Twitter their precious time and share their thoughts and comments on different subjects such as sports, films, policy debate etc. Sentimental analysis suggests that the evaluations should be predicted and that the quality of the source product should be determined to know whether or not this would succeed in the future. This is not an easy task, though, because the data contains different types of words that are usually used and that could be difficult to categorize the text from the entire data. These profound learning models are excellent when the data are filtered and the unnecessary text is removed from the system. The three models discussed are the best for sentimental twitter analysis and after their performance evaluation, one can say that BERT is superior and has all outperformed them (Roy and Ojha, 2020, p, 1-6).

10.6. BUSINESS REVIEWS CLASSIFICATION USING SENTIMENT ANALYSIS

The field of research in sentiment analysis in recent years has become more prominent. Business developers want to know not only about product and benefits marketing based on the number of sales but also about feedback and ideas of people who use these products. Reviews have gained a very important role in business strategies and determining the quality of the product. The analysis of sentiments is a field in which the research focuses on the drawing of insights and perspectives from textual or structured data on a particular issue. Millions of people express themselves on social media websites, blogs or popular services or products websites. The people's active feedback is valuable for companies not only to analyze their customers' satisfaction and to monitor their competition but also for consumers looking to buy a product. Yelp Challenge's dataset is a review provider for different companies, following a sentimental approach in various corporate reviews.

The proposal included two extraction methods and four machine learning models, several automated sentimental classification approaches. Yelp Challenge data package includes information on 61,000 local enterprises, 1,6 reviews, and 366,000 users in 10 cities in four countries worldwide rating from 1 to 5. These reviews are classified with the help of machine learning algorithms through sentimental analysis and various techniques of natural language processing. A way to classify sentimental business reviews automatically has been proposed based on either positive or negative answers. Analyzing the results at 1 to 5 levels is not, however, an easy task and the system will also show the steps that enable the system to achieve high accuracy in the classification task. Factors such as word connotation, environmental connotation, language, ambiguity in speech or the use of sarcasm are considered to be a difficult job in identifying the basic sentiment for corporate evaluation. The classification of Yelp business reviews by taking into consideration the user's star rating was successful through sentimental analytical methods. In my opinion, it may continue to increase the accuracy of the Systems to differentiate between the same word characteristics that serve as different POS, using bigger or trigrams, word chunks and even part-of-speech (Salinca, 2015, p, 147-250).

REFERENCES

Books

Alqaryouti, O., Khwileh, H., Farouk, T., Nabhan, A. and Shaalan, K., 2018. Graph-based keyword extraction. In Intelligent Natural Language Processing: Trends and Applications (pp. 159-172). Springer, Cham.

Ayodele, T.O., 2010. Types of machine learning algorithms. New advances in machine learning, 3, pp.19-48.

Banegas-Luna, A.J., Peña-García, J., Iftene, A., Guadagni, F., Ferroni, P., Scarpato, N., Zanzotto, F.M., Bueno-Crespo, A. and Pérez-Sánchez, H., 2021. Towards the Interpretability of Machine Learning Predictions for Medical Applications Targeting Personalised Therapies: A Cancer Case Survey. International Journal of Molecular Sciences, 22(9), p.4394.

Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T., 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, pp.135-146.

Giachanou, A. and Crestani, F., 2016. Like it or not: A survey of twitter sentiment analysis methods. ACM Computing Surveys (CSUR), 49(2), pp.1-41.

Goodfellow, I., Bengio, Y. and Courville, A., 2016. Machine learning basics. Deep learning, 1, pp.98-164.

Hussain, M., Zhu, W., Zhang, W., Abidi, S.M.R. and Ali, S., 2019. Using machine learning to predict student difficulties from learning session data. Artificial Intelligence Review, 52(1), pp.381-407.

Indurkhya, N. and Damerau, F.J. eds., 2010. Handbook of natural language processing (Vol. 2). CRC Press.

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

Linares-Vásquez, M., McMillan, C., Poshyvanyk, D. and Grechanik, M., 2014. On using machine learning to automatically classify software applications into domain categories. Empirical Software Engineering, 19(3), pp.582-618.

Liu, B., 2010. Sentiment analysis and subjectivity. Handbook of natural language processing, 2(2010), pp.627-666.

Marsland, S., 2015. Machine learning: an algorithmic perspective. CRC press.

Conference Papers

Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R.J., 2011, June. Sentiment analysis of Twitter data. In Proceedings of the workshop on language in social media (LSM 2011) (pp. 30-38).

Nielsen, R.K., Cornia, A. and Kalogeropoulos, A., 2016. Challenges and opportunities for news media and journalism in an increasingly digital, mobile, and social media environment. Mobile, and Social Media Environment (December 1, 2016).

Xu, H., Zhou, J., G Asteris, P., Jahed Armaghani, D. and Tahir, M.M., 2019. Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate. Applied sciences, 9(18), p.3715.

Salinca, A., 2015, September. Business reviews classification using sentiment analysis. In 2015 17th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 247-250). IEEE

Li, X. and Zhang, T., 2017, April. An exploration on artificial intelligence application: From security, privacy and ethic perspective. In 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 416-420). IEEE

Tulshan, A.S. and Dhage, S.N., 2018, September. Survey on virtual assistant: Google assistant, siri, cortana, alexa. In International symposium on signal processing and intelligent recognition systems (pp. 190-201). Springer, Singapore.

Roy, A. and Ojha, M., 2020, December. Twitter sentiment analysis using deep learning models. In 2020 IEEE 17th India Council International Conference (INDICON) (pp. 1-6). IEEE.

Journal Articles

Khan, A., Baharudin, B., Lee, L.H. and Khan, K., 2010. A review of machine learning algorithms for text-documents classification. Journal of advances in information technology, 1(1), pp.4-20.

Noguerol, T.M., Paulano-Godino, F., Martín-Valdivia, M.T., Menias, C.O. and Luna, A., 2019. Strengths, weaknesses, opportunities, and threats analysis of artificial intelligence and machine learning applications in radiology. Journal of the American College of Radiology, 16(9), pp.1239-1247.

Other Publications

Algorithm-X Lab. (2019). 10 Amazing Examples Of Natural Language Processing. [online] Available at: https://algorithmxlab.com/blog/natural-language-processing/ [Accessed 7 Jun. 2021].

Anon, (n.d.). Unsupervised Machine Learning – BinaryPlanet. [online] Available at: http://binaryplanet.org/category/unsupervised-machine-learning/.

Bird, S., Klein, E. and Loper, E., 2009. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.". https://towardsdatascience.com/fasttext-sentiment-analysis-for-tweets-a-straightforward-guide-9a8c070449a2

Brownlee, J., 2013. A tour of machine learning algorithms. Machine Learning Mastery, 25.

Chouksey, S. (2020, May 26). Simple Neural Network from Scratch. Medium. https://medium.com/swlh/simple-neural-network-from-scratch-130b175eb1e6.

Chouksey, S. (2020). Simple Neural Network from Scratch. [online] Medium. Available at: https://medium.com/swlh/simple-neural-network-from-scratch-130b175eb1e6 [Accessed 7 Jun. 2021].

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ghoneim, S., 2019. Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on. Obtenido de Towards data science: https://towardsdatascience. com/accuracy-recall-precision-f-score-specificity-whichto-optimize-on-867d3f11124.

Gomez, R., 2018. Understanding categorical cross-entropy loss, binary cross-entropy loss, softmax loss, logistic loss, focal loss and all those confusing names. URL: https://gombru. github. io/2018/05/23/cross_ entropy_loss/(visited on 29/03/2019).

Hands-On Transfer Learning with Python. [online] Available at: https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781788831307/1/ch01lvl1sec05/standard-ml-workflow [Accessed 7 Jun. 2021].

Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T., 2016. Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.

La Grassa, R., Gallo, I. and Landro, N., 2020. Ocmst: One-class novelty detection using convolutional neural network and minimum spanning trees. arXiv preprint arXiv:2003.13524.

Machine Learning | Artificial Intelligence | Data Annotation. (n.d.). What is a training data set in Machine Learning and rules to select them. [online] Available at: https://machinelearningasaservice.weebly.com/blog/what-is-a-training-data-set-in-machine-learning-and-rules-to-select-them.

neptune.ai. (2021a). Cross-Entropy Loss and Its Applications in Deep Learning. [online] Available at: https://neptune.ai/blog/cross-entropy-loss-and-its-applications-in-deep-learning [Accessed 7 Jun. 2021].

Swamynathan, M., 2019. Mastering machine learning with python in six steps: A practical implementation guide to predictive data analytics using python. Apress.

www.altudo.co. (n.d.). 5 Steps defining the Machine Algorithm Flow to deliver a customized user experience. [online] Available at: https://www.altudo.co/resources/blogs/5-steps-to-define-ml-flow-to-deliver-custom-user-experience [Accessed 7 Jun. 2021].