

JANUARY 2021

M.Sc. in Electronics and Computer Engineering

HATICE KÜBRA KILINÇ

**HASAN KALYONCU UNIVERSITY
GRADUATE SCHOOL OF
NATURAL AND APPLIED SCIENCES**

**FAKE NEWS DETECTION WITH DEEP LEARNING
METHODS**

**M.Sc. THESIS
IN
ELECTRONICS AND COMPUTER ENGINEERING**

**BY
HATICE KÜBRA KILINÇ
JANUARY 2021**

**Fake News Detection With Deep Learning and
Machine Learning Methods**

M.Sc. Thesis

In

Electronics and Computer Engineering

Hasan Kalyoncu University

Supervisor

Asst. Prof. Dr. Bülent HAZNEDAR

Hatice Kübra KILINÇ

2021



© 2021 [Hatice Kübra KILINÇ]



GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
INSTITUTE
M.Sc. ACCEPTANCE AND APPROVAL FROM

Electronics and Computer Engineering Department, Electronics and Computer Engineering Master program student **Hatice Kübra KILINÇ** prepared and submitted the thesis titled **Fake News Detection With Deep Learning and Machine Learning Methods** defended successfully on the date of 26/01/2021 and accepted by the jury as an M.Sc. thesis.

<u>Position</u>	<u>Title, Name and Surname</u> <u>Department/University</u>	<u>Signature</u>
Supervisor	Asst. Prof. Dr. Bülent HAZNEDAR Computer Engineering Department/ Hasan Kalyoncu University	
Jury Member	Prof. Dr. Muhammet Fatih Hasoğlu Computer Engineering Department/ Hasan Kalyoncu University	
Jury Member	Asst. Prof. Dr. Muhammet ÇINAR Civil Engineering Department/ Kahramanmaraş Sütçü İmam University	

This thesis is accepted by the jury members selected by the institute management board and approved by the institute management board.

Prof. Dr. Muhammet Fatih HASOĞLU
Director

Document no: ENS.FR.32 Yayın Tarihi: 26.03.2018 Rev no/ Date: 00/

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Hatice Kübra KILINÇ



ABSTRACT

FAKE NEWS DETECTION WITH DEEP LEARNING AND
MACHINE LEARNING METHODS

KILINÇ, Hatice Kübra

M.Sc. in Electronic - Computer Engineering

Supervisor: Asst. Prof. Dr. Bülent HAZNEDAR

January 2021

76 pages

In our age, especially since the rise of technology and social media, fake news has become a problem in society. News has had an impact on society throughout history. Today, when online platforms are one of the biggest communication tools, we access the news over the internet. When we cannot access the right source of information, serious problems arise for the life of society. Anymore we see fake news spreading more and faster in some situations than real information. The availability of technology has brought us many of its problems too alongside the advantages of it. Fake news has been used for directing society and for social engineering rather than an innocent information mistake. Today, we are faced with a false fraud that guides social movements, makes wrong decisions, and affects political and economic events. The events experienced show us that work such as big data processing and social media data analysis needs to be done. In this study, fake news data is used which is a more general topic than fake fabricated or fictional content that can change this economy, sociology, politics. I evaluate this different artificial intelligence method, which aims at using machine learning and deep learning methods of fake news, with fake news detection. I compare NLP, LSTMs, and traditional machine method learning. LSTMs has had higher classification success in most machine method learning methods. The fact that a lot of work has to be done related to big data processing, internet news analyzing, social media data analysis encouraged me to this field. This thesis is expected to be useful for producing more effective solutions to fake news, which is an important problem.

Keywords: Fake News Detection, Classification, Deep Learning, Machine Learning

ÖZET

SAHTE HABERLERİN DERİN ÖĞRENME VE MAKİNE ÖĞRENMESİ İLE TESPİTİ

KILINÇ, Hatice Kübra

Yüksek Lisans, Elektronik Bilgisayar Mühendisliği

Tez Danışmanı: Dr. Öğr. Üyesi. Bülent HAZNEDAR

OCAK 2022

76 sayfa

Yaşadığımız çağda, özellikle teknoloji ve sosyal medyanın yükselişinden bu yana, sahte haberler bir toplum sorunu haline geldi. Haberlerin tarih boyunca toplum üzerinde etkisi olmuştur. Online platformların en büyük iletişim araçlarından biri olduğu günümüzde haberlerde çoğunlukla internet üzerinden erişiyoruz. Doğru bilgi kaynağına erişemediğimiz zaman ise birey, toplum hatta ülkeler için ciddi sorunlara yol açan problemler oluşuyor. Artık sahte haberlerin bazı durumlarda gerçek bilgilerden daha fazla ve daha hızlı yayıldığını görüyoruz. Teknolojinin gelişmesiyle bilginin kolay erişebilirliği birçok problemi de birlikte getirdi. Sahte haberler masum bir bilgi yanlışlarından çok toplumu yönlendirmek, toplum mühendisliği yapmak için kullanılabilir oldu. Günümüzde toplumsal hareketlere yön veren, insanları yanlış kararlar almaya yönlendiren, siyasi ve ekonomik olaylara etki eden bir sahte haber kavramı ile karşı karşıyayız. Yaşanılan olaylar bize büyük veri işleme, sosyal medya veri analizi gibi konularda çalışmalar yapılması gerektiğini gösteriyor. Bu çalışmada ekonomiyi, sosyolojiyi, siyaseti değiştirebilen sahte uydurma veya kurgulanmış içerikler daha genel bir başlık olan sahte haberler kullanıldı. Çeşitli kaynaklardan oluşan farklı veri setleri kullanıldı. Sahte haberlerin tespitini amaçlayan bu çalışmada hem machine learning hem de deep learning yöntemleri kullanıldı. Kullanılan farklı yapay zeka yöntemlerinin sahte haber tespiti üzerindeki performansını değerlendiriyorum. NLP, LSTM ve geleneksel makine öğrenimi yöntemleriyle karşılaştırıyorum. Yapılan deneylerde NLP, LSTM, Derin öğrenme tabanlı sınıflayıcılar ise geleneksel sınıflayıcılardan olan Naive Bayes, Random Forest classifier, SGD classifier, K-neighborhood classifier, Logistic Regression ile kıyaslanmış sonuçlarda derin öğrenme yöntemleri başarı ölçütlerinin çoğunda daha yüksek sınıflama başarısına sahip olmuştur. Büyük veri işleme, internet haberleri, sosyal medya veri analizi gibi konularda yapılması çok fazla çalışma yapılması gerektiği gerçeği ile bu tezin önemli bir sorun olan sahte haberler konusunda daha etkili çözümler üretilebilmesi için yarar sağlaması beklenmektedir.

Anahtar Kelimeler: Sahte Haber Tespiti, Sınıflandırma, Derin Öğrenme, Makine Öğrenmesi

ACKNOWLEDGEMENTS

I want to thank my supervisor Asst. Prof. Dr. Bülent HAZNEDAR supported and encouraged me during the Master's process. It was a big proud to study with him due to his advice, guidance, and motivation. A special thanks to my mother, my father, and my siblings who taught me how to stand up in difficult and thanks to everyone who taught me how to stand up in difficult moments. Due to existing of them and their prayers, time of finding the right way became too short.

TABLE OF CONTENTS

	Pages
ABSTRACT	i
ÖZET.....	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES.....	v
LIST OF SYMBOLS AND ABBREVIATIONS	vi
CHAPTER 1.....	1
1. INTRODUCTION	1
CHAPTER 2.....	2
2. BACKGROUND AND LITERATURE REVIEW	2
2.1. FAKE NEWS CONTENT.....	2
2.1.1 WHAT IS FAKE NEWS.....	4
2.1.2 MIS-DISINFORMATION TYPES.....	6
2.2. LITERATURE REVIEW.....	9
CHAPTER 3.....	11
3. DATA PRE-PROCESSING	11
3.1. DATA PROCESSING.....	11
3.1.1. NEED OF DATA PREPROCESSING.....	12
3.2. DATA CLEANING.....	13
3.2.1. REMOVAL OF UNWANTED OBSERVATIONS.....	14
3.2.2. FIXING STRUCTURAL ERRORS.....	15
3.2.3. MANAGING UNWANTED OUTLIERS.....	15
3.3 TEXT PROCESSING.....	15
3.3.1. WHAT IS TEXT PREPROCESSING	15
3.3.2. WHAT IS NATURAL LANGUAGE PROCESSING.....	15
3.3.3. STOP WORDS REMOVAL.....	18
3.3.4 STEMMING AND LEMMATIZATION.....	18
3.3.5 TOKENIZATION.....	19
3.3.5.1. SENTENCE TOKENIZATION.....	20
3.3.5.2. WORD TOKENIZATION.....	20

3.3.6. LOWER CASING.....	21
3.3.7. WORD EMBEDDINGS AND TEXT VECTORIZATION.....	22
3.3.7.1 WORD2VEC.....	22
3.3.8 ONE HOT ENCODING.....	23
CHAPTER 4.....	25
4. METHODS AND MATERIALS.....	25
4.1. FEATURE SELECTION METHODS	25
4.1.1 CLASSIFICATION.....	25
4.1.2. CLASSIFICATION PROCESS.....	25
4.2. MACHINE LEARNING.....	26
4.2.1. THE MULTINOMIAL NAIVE BAYES CLASSIFIER.....	26
4.2.2.RANDOM FORESTS.....	27
4.2.3. K-NEAREST NEIGHBOR	28
4.2.4. STOCHASTIC GRADIENT DESCENT.....	28
4.2.5. LOGISTIC REGRESSION	29
4.3. DEEP LEARNING.....	30
4.3.1. NEURAL NETWORK.....	31
4.3.1.1. ACTIVATION FUNCTIONS.....	35
4.3.2. SIGMOID FUNCTION.....	36
4.3.3.RECURRENT NEURAL NETWORKS.....	36
4.3.4 VANISHING GRADIENT PROBLEM.....	38
4.3.5 LONG SHORT-TERM MEMORY	38
4.3.6. CONFUSION MATRIX	39
4.3.7. RECEIVER OPERATING CHARACTERISTIC CURVE	40
4.3.8. CROSS-VALIDATION.....	40
4.3.9. K-FOLD CROSS VALIDATION	41
CHAPTER 5.....	42
5. RESULTS AND DISCUSSIONS.....	42
5.1.1 FAKE NEWS DETECTION.....	42
5.2 FUTURE WORK	71
CHAPTER6.....	72
6.CONCLUSION.....	73
REFERENCES.....	73

LIST OF FIGURES

List of Figures

	Pages
Figure 2.1: Fake news on social media: from characterization to detection.....	3
Figure 2.2: Misleading image.....	3
Figure 2.3: Rating by Category of Fake News.....	5
Figure 2.4: Rating by Page of Fake News.....	5
Figure 2.5: Types of Fake News.....	7
Figure 3.1 Raw data processing steps.....	11
Figure 3.2: Rate of time spent by data scientists.....	12
Figure 3.3: Data processing steps.....	13
Figure 3.4: Data Cleaning.....	14
Figure 3.5: Vector Space Model.....	16
Figure 3.6: Stemming.....	19
Figure 3.7: Sentence Tokenization.....	20
Figure 3.8: Word Tokenization.....	20
Figure 3.9: Encoding.....	21
Figure 3.10: Lower Casing.....	21
Figure 3.11: CBOW and Skip-gram.....	23
Figure 3.12: One Hot Encoding Schema.....	24
Figure 4.1: Classification Process Schema of Fake News Detection.....	26
Figure 4.1. Stochastic gradient descent compared with gradient descent.....	29
Figure 4.1. Sigmoid function.....	30
Figure 4.1 : Neural networks.....	32
Figure 4.2 :Neural Networks as a Composition of Piece.....	33
Figure 4.3: Artificial Neural Network (ANN).....	33
Figure 4.4: Weight and Biases.....	34
Figure 4.5: Activation Function, Softmax.....	34
Figure 4.6: Convolutional Neural Networks for Visual Recognition.....	35
Figure 4.7: Sigmoid Function.....	36
Figure 4.7:RNN process.....	37
Figure 4.7:LSTM gates.....	39
Figure 4.7: Confusion Matrix.....	39
Figure 4.7: Diagram of k-fold cross validation.....	41

Figure 5.1: Rate of Liar Dataset.....	44
Figure 5.2: Number of Liar Dataset.....	44
Figure 5.3: Word Cloud of REAL News.....	45
Figure 5.4: Word Cloud of FAKE News.....	46
Figure 5.5: Confusion Matrix KNN-Liar dataset.	52
Figure 5.6: Confusion Matrix LR -Liar dataset.....	52
Figure 5.7 Confusion Matrix MNB -Liar dataset.....	53
Figure 5.8 Confusion Matrix SGD -Liar dataset.....	53
Figure 5.9 Confusion Matrix RF-Liar dataset.....	54
Figure 5.10. ROC -Liar dataset.....	54
Figure 5.11. Confusion Matrix LSTM -Liar dataset.....	56
Figure 5.12. Accuracy of LSTM model.....	57
Figure 5.13. Model Loss/Cost-Entropy Loss of LSTM model.....	57
Figure 5.14: Confusion Matrix Logistic regression- Reuters dataset.....	64
Figure 5.15: Confusion Matrix MNB - Reuters dataset.....	64
Figure 5.16: Confusion Matrix SGD Classifier - Reuters dataset.....	65
Figure 5.17: Confusion Matrix RF- Reuters dataset.....	65
Figure 5.18: Confusion Matrix KNN- Reuters dataset.....	66
Figure 5.19: ROC Curve -Reuters dataset.....	66
Figure 5.20: Accuracy and Model Loss/Cost-Entropy Loss of LSTM model.....	69
Figure 5.21: LSTM Confusion Matrix.....	70

List of Tables

Table 5.1 : Subject of Liar Dataset.....	43
Table 5.2: Model and Scores of Liar Dataset.....	47
Table 5.3: Accuracy for MultinomialNB - Liar dataset.....	47
Table 5.4: Accuracy for Random Forest Classifier-Liar dataset.....	48
Table 5.5: Accuracy for K-Neighbors Classifier-Liar dataset.....	49
Table 5.6: Accuracy for SGDClassifier-Liar dataset.....	50
Table 5.7: Accuracy for Logistic Regression -Liar dataset.....	51
Table 5.8. Sequential model -Liar dataset.....	55
Table 5.9. : Model and Scores.....	58
Table 5.10: Accuracy for MultinomialNB - Reuters dataset.....	59
Table 5.11: Accuracy for Random Forest- Reuters dataset.....	60
Table 5.12: Accuracy for KNN - Reuters dataset.....	61
Table 5.13: Accuracy for SGD Classifier - Reuters dataset.....	62
Table 5.14: Accuracy for Logistic regression- Reuters dataset.....	63
Table 5.15. : K-Fold Validation - Model and Score Reuters dataset.....	67
Table 5.16: Sequential model Reuters dataset.....	68

LIST OF SYMBOLS AND ABBREVIATIONS

AI	Artificial intelligence
ANN	Artificial Neural Network
ANN	Artificial neural network
CNN	Convolutional Neural Network
CNN	Convolutional neural network
CV	Computer Vision
DBN	Deep Belief Networks
DL	Deep learning
DT	Decision trees
FPR	False positive rate
KNN	K-Nearest Neighbor
KNN	K-nearest neighbor
LR	Logistic Regression
LSTM	Long Short-Term Memory
LSTMs	Long Short-Term Memory Networks
ML	Machine learning
MNB	Multinomial Naive Bayes
NB	Naive Bayes
NLP	Natural language processing
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NN	Neural Network
RF	Random forest
RNN	Recurrent neural networks

RNNs Recurrent Neural Networks

ROC Curve Receiver operating characteristic curve

SGD Stochastic Gradient Descent

TPR True positive rate



CHAPTER 1

1. INTRODUCTION

A huge amount of information is generated on the internet at any moment, and the data shared on social media and various news portals also take up a large amount of space. The rapid increase in social media posts on the web enables huge amounts of data to be generated. When an event occurs, we are quite used to most people discussing and sharing it on the web. As a news source, we can come across thousands of sources at once. Unfortunately, the accuracy of the information shared about an incident can be shared on the internet without being confirmed by an authorized institution. Over the past few years, the transmission of fake news has become a huge problem. As a result of this, we are exposed to a lot of inaccurate information in every event.

While the right news should be researched, criticized, and shared if it reflects the truth, for now, no system can provide this completely. However, the opportunity to view huge amounts of news or posts leaves us with the problem of information overload. We find ourselves in a situation that was taken us away from the ability of healthy analysis. Unreliable sources of information expose people to a variety of fake news; scams, manipulated content, satire, conspiracy theory, and misleading news to name a few. It can be easily observed that the false news that spread unexpectedly on issues threatening the dynamics of society is not in good faith. In many countries, fake news about presidential candidates during the election can direct politics and even the fate of the country.

These are just a few examples whose consequences are very dangerous and detrimental. Another example, even if the harms of climate change are known, it may not always be correct to give precise information about some natural events and climatic conditions. We can observe a perception that the popular is valuable. Fast spreading and unconfirmed information is often popular trend information, and in this case, it can cause fake news to spread. Fake news, false information, misunderstandings, and biased information fictionalized by individuals or institutions pose a threat to our lives.

CHAPTER 2

2. BACKGROUND AND LITERATURE REVIEW

2.1. FAKE NEWS CONTENT

The credibility of information has been defined in many words with reliability, credibility, truthfulness, fairness, objectivity, and other same concepts and definitions. Several studies use the machine learning approach to calculate the credibility of the news, as well as some of them, have tried different artificial intelligence models in their studies. Fake news is content that makes people believe inaccurate information. When news is received, they quickly distribute it to others through the internet and social media. This rapidly spreading information did not only affect a limited group.

The spread of fake news in today's digital world has affected a certain group. It is very normal to encounter interesting information that surprises us on social media platforms that have a significant place in our lives, however, it can sometimes confuse us. It is very difficult to classify not real but believable news that appears as fake news because sometimes fake news is close to the truth and gives information that makes sense. Some distributors believe it to be true and some distributors know that this is not real, but is deliberately deceiving and spreading. Those are different types of reasons for fake news.

Even though fake news has always been around throughout history, it has become a term we encounter more, especially after the 2016 presidential election in America. Nowadays, social media tools are used much more and the number of information is scattered via unpredictable sources. In addition to organic shares, it appeared in followers and broadcasters that occurred with bot accounts purchases and some situations took the subject to another dimension anymore. While fake news was not very common in our lives until a few years ago, it has now become a concept that we can hear almost everywhere and not be surprised when we hear it.

News manipulations also increased the distrust of the media. Moreover, these are the problems that will cause a big problem in society. To better understand what fake news is, it is necessary to solve the problem. To get it right to the definition is important because the use of fake news can be used for many genres. Only when it is correctly defined we can make the correct response. Machine learning (ML), natural language processing (NLP), and artificial intelligence (AI) are looked to different fields and techniques.

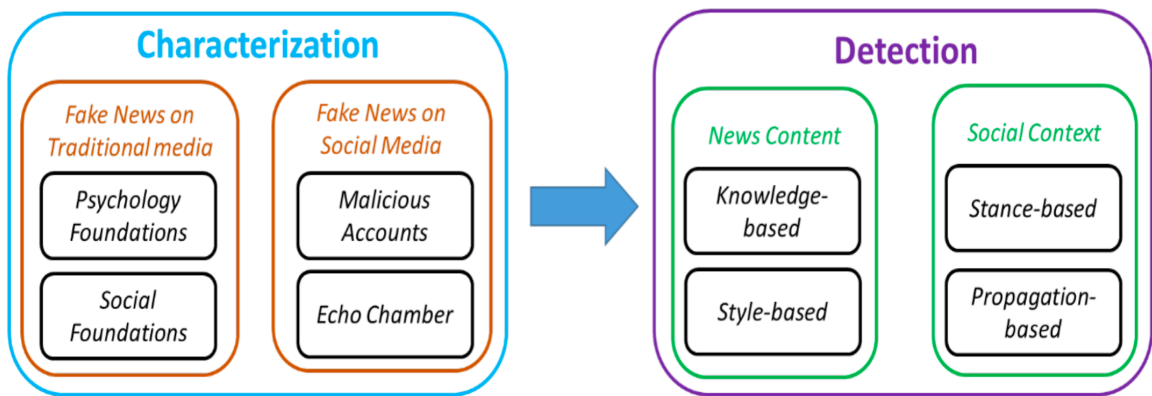


Figure 2.1: Fake news on social media: from characterization to detection (Kai Shu et al, 2017).



Figure 2.2: Misleading image

Figure 2.2 shows two images between two dates provided to illustrate deforestation, but images are given at the same time and as the same image (Wwf, 10yearschallenge, 2019.) Identifying reliable news sources has become more difficult, and this image fits the subject. The post was shared by the World Wildlife Fund (WWF) - Pakistan on multiple social media platforms with hundreds of thousands of followers, plus the same post was shared by more than 750,000 followers and attracted thousands of comments and reactions. On the contrary, with a reverse search on Google and Yandex search engines, we can find the original image attributed to photographer Richard Carey in Shutterstock images. This situation shows us that the problem situation is dangerous. This event is also a growing need for analysis tools that can provide artificial intelligence solutions that can provide insight into the reliability of online content in the show.

2.1.1 What is fake news

First of all, fake news is handled in two different parts. Those that are deliberately invented, albeit difficult to prove, and fictitious news that spread unnoticed, whether intentionally or not. Some problems make fake news detection difficult. A few examples of these; partial lies, biased news, incomplete information resulting from negligence. In addition to the difficulty of detecting fake news at present, there is also the difficulty of distinguishing and identifying. Although many factors make it difficult to work, many people and institutions have produced studies in this field. In many studies, we can see that the news is divided into different classification categories or they touch the subject from different angles.

In addition to scientific studies in this field, various news organizations shared information about the subject. NYT newspaper drew attention to this issue and these news reports “a made-up story to deceive” because of the importance of it. defined as. A study shows us that nowadays people easily believe more fake news than real news. This is a terrible rate (Sabrina Tavernise, 2016).

Writes that the proportion of fake and misleading news according to BuzzFeed News, which supports and works on the fight against fake news and provides datasets to developers as a news organization, gives the alarm. In the below part I showed a graphic originally prepared by BuzzFeed News, the rates of fake news published during the elections in America were shared. The data show us that these rates are alarming (Craig Silverman et al., 2016).

Rating by Category

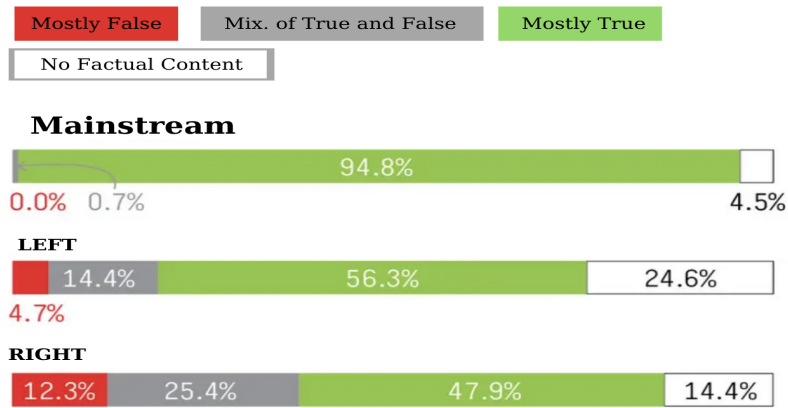


Figure 2.3: Rating by Category of Fake News

Rating by Page

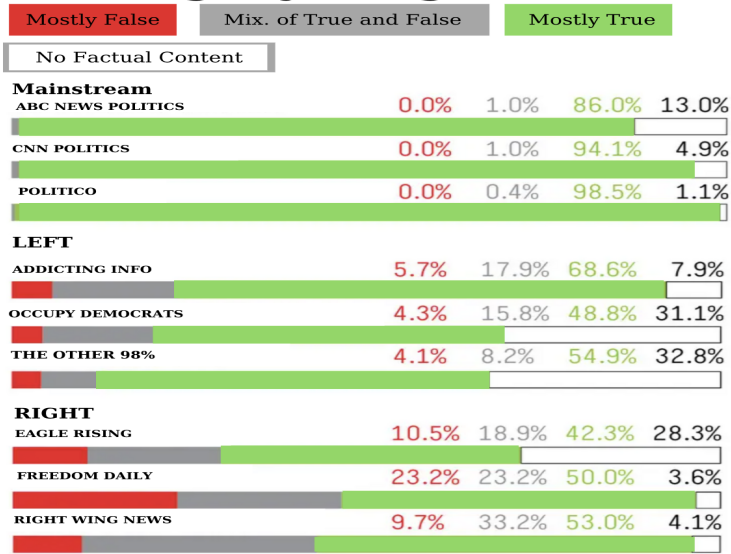


Figure 2.4: Rating by Page of Fake News

2.1.2 Mis-Disinformation Types

The information disorders can be broadly divided into three main categories: disinformation, misinformation, and wrong information (Claire Wardle et al, 2017). Disinformation, mislead, and/or are deliberately spreading false or incorrect information to deceive. Incorrect information is incorrect, content shared by people who are not aware that false or misleading (Liang Wu et al., 2019). Malinformatio's, to define the shared information in real harm. Also, there are other species caused by the deficiency or distortion related information. The rumor goes from person to person and is an unconfirmed story is true. (Xinyu Zhou et al, 2018) are often vague rumors or reveal the presence of the peace of the people threatening events. Despite all the difficulties many institutions and organizations struggle in this area. There have been institutions that divide fake and misleading news into different types and define them. The 7 identified types offer us the opportunity to distinguish more clearly on fake news. These types are listed as follows (Ingrid Brodnig,2017).

- Incorrect Link: Titles or subtitles do not support content
- Incorrect Context: Real content is shared with incorrect contextual information
- Modified content: Original information or images have been modified
- Satire or Parody: Potential to mislead but no intention to harm
- Misleading Description: An issue/use of misleading information to the individual frames
- Counterfeiting Content: When the impersonation of natural resources
- Fitting content: mostly wrong and deceive and new content designed to harm

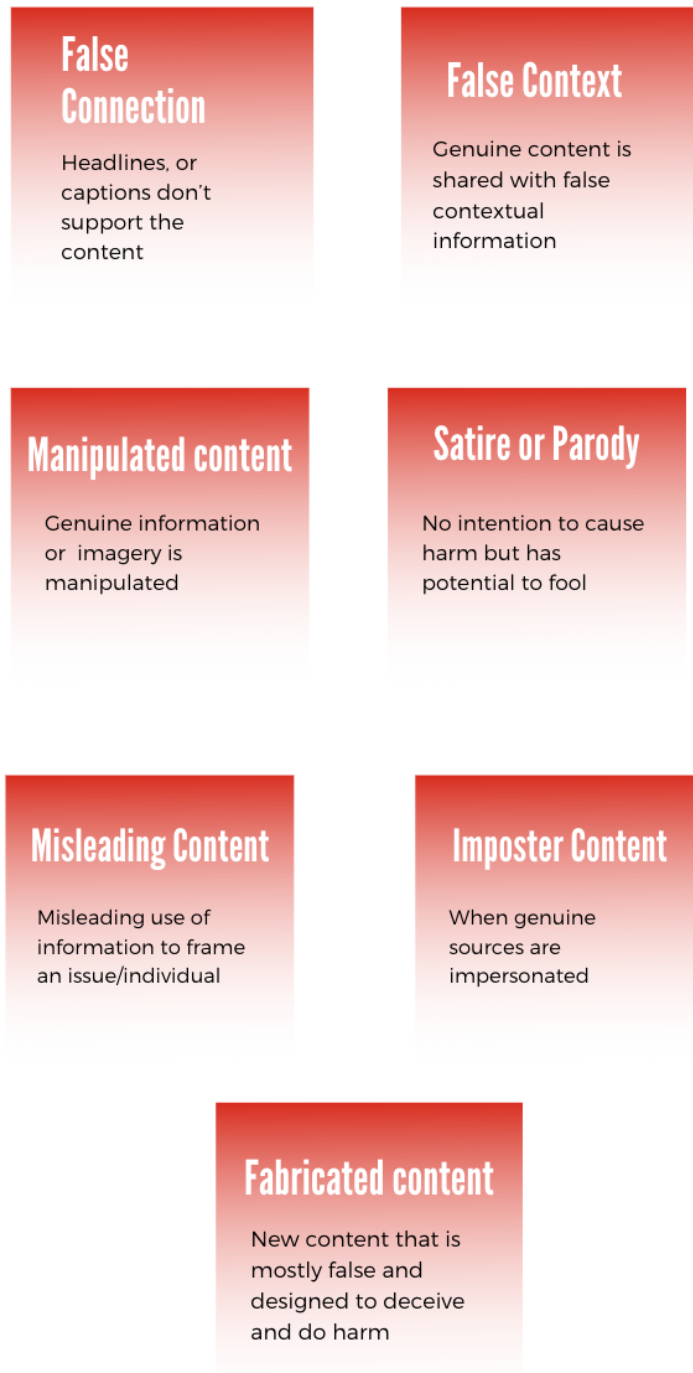


Figure 2.5: Types of Fake News

ML and DL offer us the means to overcome these challenges. Before moving on to this part, I would like to include information about why fake news is so prominent. Fake news doesn't just give political advantage to politicians, it's not the only reason. People can earn money with these methods. The reasons behind fake news include media manipulation and propaganda, political and social influence, provocation and social unrest, and financial profit. It is understood that before the US Presidential elections there may be a group of Macedonian youth behind the most important fake stories and that their aim is nothing but financial. In Figure 2.5, it is shown in detail in the visual prepared by First Draft News using the source (Miguel Malvarez, 2017).

After learning the information shared by the former KGB (Komitet Gosudarstvennoy Bezopasnosti) agent Yuri Bezmenov in his interview, we can understand more easily how serious the fake news used in psychological warfare methods is. Agent Yuri Bezmenov lists psychological warfare in 4 stages. He stated in his interview that these are demoralization, destabilization, crisis, and normalization. He also stated that this method aims to change people's real perceptions and thus prevent the mind from seeing reality. In this technique called Ideological Destruction, people cannot see the truth despite the abundance of information



2.2. LITERATURE REVIEW

On social media platforms, the volume of the data sent by the users grows very large, so it is very difficult to determine the subject to which the content is related (Shu et al., 2017). Also, news platforms, which are considered to be a reliable source, can produce fake news that does not suit them and people do not expect. It is to share accurate, precise, and real news that suits news organizations. Unfortunately, news organizations can share misleading and inaccurate news in a way that does not suit them, and in today's age, with the possibilities of technology, fake news can spread at an incredible speed.

This situation shows that the research is very important. It is a positive step towards a critical effort. There are many studies in this field, and one of them is an article by Marco L. Delia Vedov et al. An easy method was discovered to detect fake news using the Naïve Bayes classifier. Facebook data was used at this stage. Instead of news articles on the internet. As a result, it achieved 74 % accuracy in the paper grading process. Aside from this low accuracy rate, using social media data alone is not enough for analysis. For this reason, in my thesis, I used data sets consisting of news from more than ten sources.

The article was written by Rubin et al. It deals with the fake news area of satire news. Satirical stories provide clues that deliberately reveal their deception. While fake news requires readers to believe in a false truth, satire news should be understood as a gesture at the end. This article provides an in-depth view of the characteristics of humor and satire news, along with the reporting style of the authors. The newspaper evaluated news articles from twelve current news topics in four different fields: citizenship, science, business, and social news. The article proposes a Support Vector Machine-based algorithm that can detect satire news based on features such as Nonsense, Humor, Grammar, and Punctuation. The models achieved 90% accuracy and 84% recalls. The goal is to reduce the negative impact of satire news on readers.

We cannot separate each news piece as true or false with a definite difference, because this news has typed. For example, satire is one of and it is not news to deceive or harm people. Some researchers have done studies in this direction. Rubin et al. Examples of this subject. As a purpose, the satire news deliberately provided clues revealing its deception. As the first thought, fake news should ask readers to believe in a false truth, while satire news should be understood as a gesture. So there is a joke and humor in satire. In this study, it gains importance as it is different from the general stories. As a result of their work, the models achieved an accuracy of 90 % and a recall of 84 %.

Some researchers have based the news on different sources in their studies. For example, Namwon Kim et al. detects fake news based on unified key sentence information. Also, instead of the most preferred English datasets, the Korean dataset is used. The reason why English datasets are used more is that it is the most used language in the world and this language is since there are more resources in social media. He performs his studies by using an efficient matching process for word vectors obtained using bidirectional LSTM. As a result, they achieved an accuracy varying between 64 % and 69 %.



CHAPTER 3

3. DATA PRE-PROCESSING

3.1. DATA PROCESSING

Converting data from given or raw form to the desired form makes it more usable and more meaningful. Data Processing is a stage of cleaning the data to be used and preparing it for training. Mathematical modeling, algorithms, such as this process can be automated. For taking meaningful results from data, data processing is an acceptable beginning process. The given form data cannot yield a successful result. The reason of it, raw data has noisy inconsistent parts in it. In this way, the program cannot understand relations between words and sentences, as a result of this program is not able to show a successful inference. As is raw data called dirty data (Erhard Rahm et al.,2020). This method is a method used to transform the obtained raw data into a clean data set. If missing or raw data are used in a model, it can cause various errors and consequently reduce its accuracy.

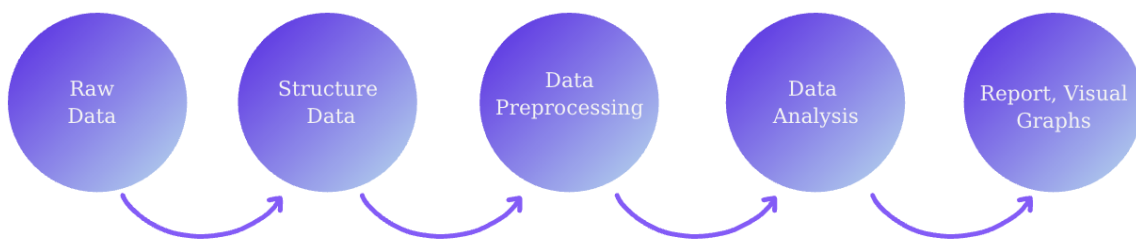


Figure 3.1 Raw data processing steps

An article published by Forbes shared the information that data scientists spend 60% of their time cleaning and organizing data. This ratio shows that the process is very important. The same research shares that they spend 19% of their time collecting data. This shows us that data scientists spend about 80% of their time preparing and managing data for analysis. The next sections describe the data pre-processing stage (Gil Press, 2016).

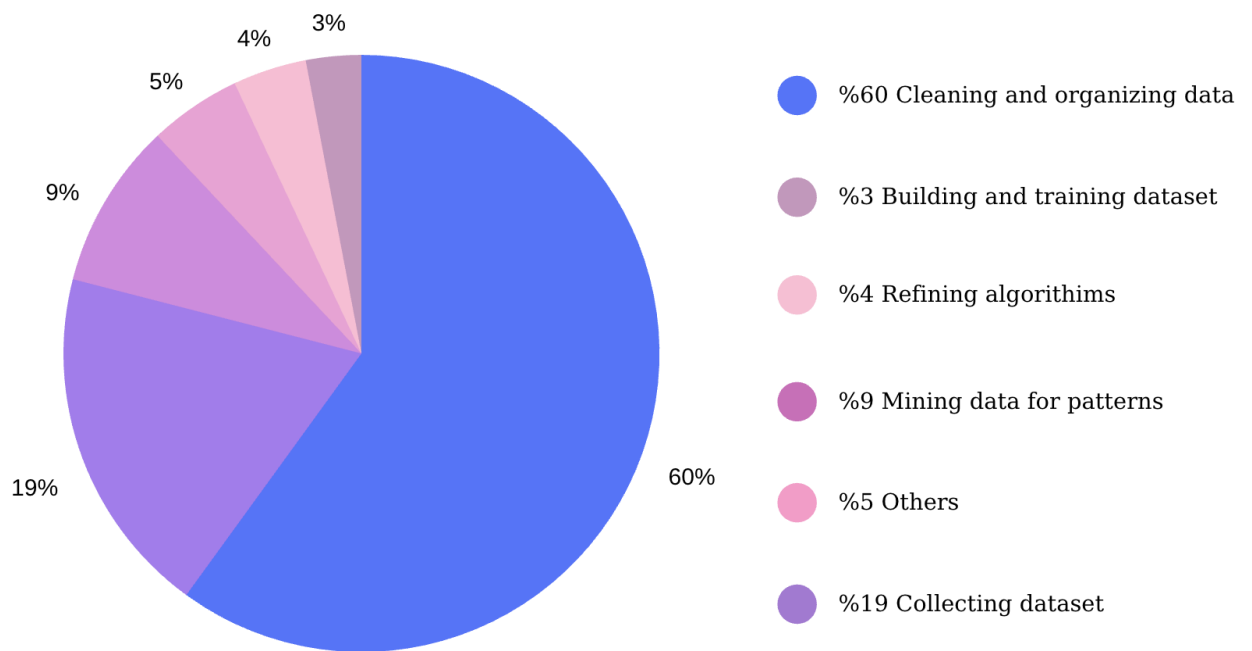


Figure 3.2: Rate of time spent by data scientists

3.1.1. Need for Data Preprocessing

In Machine Learning (ML) and Deep Learning (DL) projects, data is formatted to get better results from the model. The importance of optimizing data is better understood when we examine algorithms. DL and ML models must have information in a specific format. For example, the Random Forest algorithm does not support null values, so random forest algorithms need to be managed from the original raw data set for the operation of the null value. Therefore, before feeding data into an algorithm, it is a necessary step to apply different types of transformations to our data, and this is called preprocessing (Sireesha Moturi, 2020). Another aspect of the data set on which to run multiple machine learning and deep learning algorithm on a data set and the best of them should be formatted so that selection.

Data Processing steps to perform are as follows:

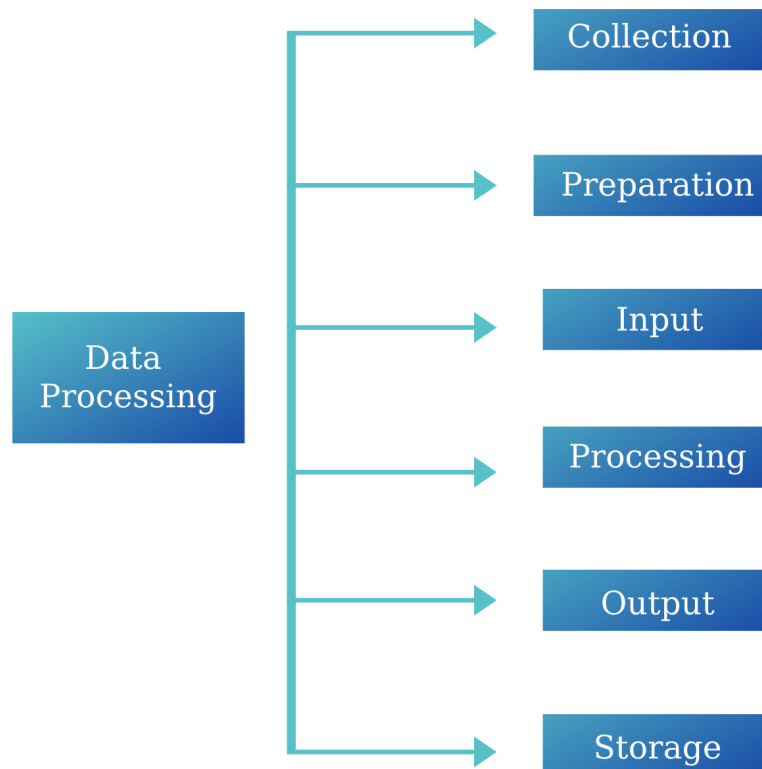


Figure 3.3: Data processing steps

3.2. DATA CLEANING

The data cleaning phase is one of the important parts. It plays the most important role in creating a model. Data cleansing is the process of preparing incorrect, incomplete, irrelevant, duplicated data to analyze data by removing or replacing malformed data. Data cleaning is not just about cleaning unnecessary information. While the first things that come to mind for data cleanup are steps such as correcting typos, standardizing data sets, and removing blank spaces, it includes more actions. The data cleaning can make better or decrease the quality of a project.

When it comes to analyzing data, raw data is often not necessary or useful because it can hamper the process or provide false results. Maximize the accuracy of data sets before deleting information. It is a phase in which professional data scientists often spend a great deal of their time. The desired result of algorithms depends largely on a well-cleaned dataset. Depending on the answers being sought there are several methods for data cleaning. Different types of data require different cleanups.

I explained what the general stages are, the types required and used in this study in the following section.

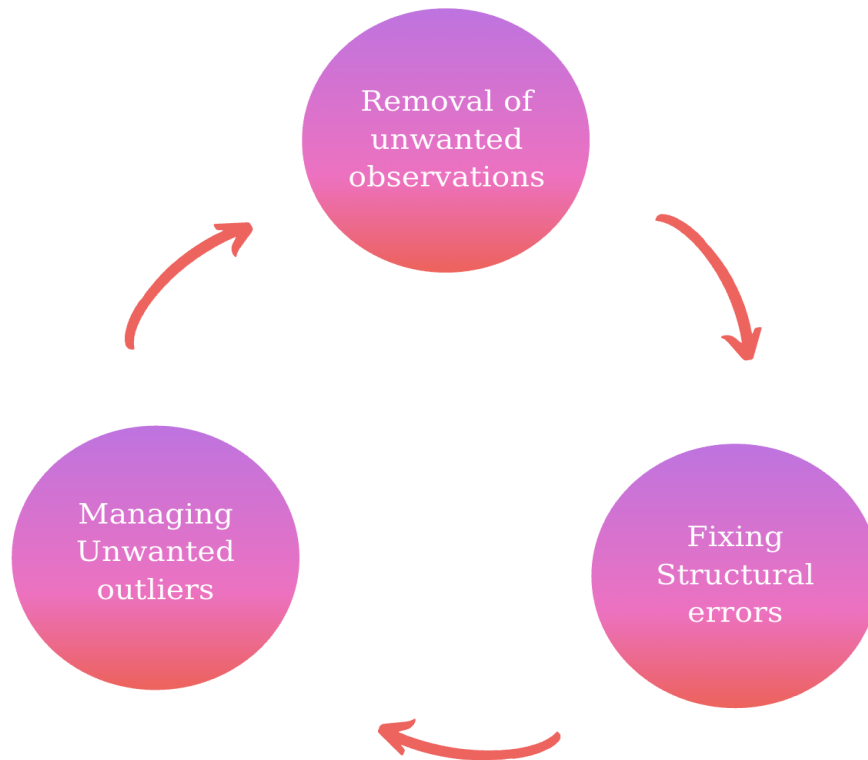


Figure 3.4: Data Cleaning

3.2.1. Removal of unwanted observations

There are 2 types of this; duplicates and irrelevances. Duplicate Observations are data repeated a lot of time in a dataset. Also in some situations, during error, this type can happen. Cause of the error participant couldn't send more than one submission. Irrelevant Observations, not related to the subject can be inside a dataset. It is possible for a data set to damage integrity in irrelevant content outside of its purpose. If my news dataset contained market information, it would be an example.

3.2.2. Fixing Structural errors

Structural errors, errors that occur in measurement, data transfer, or other similar situations. Spelling errors in the name of properties, the same attribute with a different name, incorrectly labeled classes, separate classes that really should be the same, or inconsistent capitalization include structural errors. To prevent this, lower case processing is applied to the data set.

3.2.3. Managing Unwanted outliers

Linear regression models are less resistant to irregular and irrelevant values than decision tree models and can be given as an example of these outliers. For each case, we can not say that removing them improves performance or not (GeeksforGeeks, 2018).

3.3 TEXT PROCESSING

3.3.1. What is text preprocessing

Text pre-processing, text can be estimated for a task can be analyzed and means to bring a form. cleaning and preparation of the text data, the data is an approach that makes available in a specific context, and the context of my thesis is to make the data available for the detection of false news. "Text data is usually number formats, date formats, and prepositions, articles and pronouns are not likely to help, such as text mining involve some special formats such as the most common words." (Dr. S. Kannan et al., 2014). I used natural language processing (NLP) pipelines before model training.

3.3.2. What is Natural Language Processing

Natural Language Processing (NLP) is an application area that studies how computer language can be done to understand and process text in natural languages. (Dr. S. Vijayarani ve diğerleri 2015). NLP researchers, how to understand people's language and aims to collect information about the use and in this way to fulfill preferred roles of the computer system to understand natural language and appropriate tools and techniques can be developed to ensure that manipulated.

Also, my goal is to make the text understandable when processing it. As previous research has shown, NLP is one of the main methods used for this. NLP is a subfield of computer science and

artificial intelligence and deals with the interactions between computers and human (natural) languages. NLP is a field of Artificial Intelligence that gives machines the ability to read, understand, and infer.

Some examples of NLP implementation are as follows: Speech recognition, document summarization, machine translation, spam detection, named entity recognition, sentiment analysis, language translation, question answering, autocomplete, and fake news detection. Used to apply machine learning and deep learning algorithms to text and speech.

NLP is actually in our lives in more areas than we think. The speech recognition feature on the smartphones we use serves by using the NLP infrastructure. Natural language processing takes text as input. These texts are collected from many data sources. Before analyzing the data, many steps and cleansing are required.

In the vector space model, each word or term is represented as an axis/dimension. Text is represented as a vector in a multidimensional space. Unique means the number of words and dimensions (Harshith, 2019).

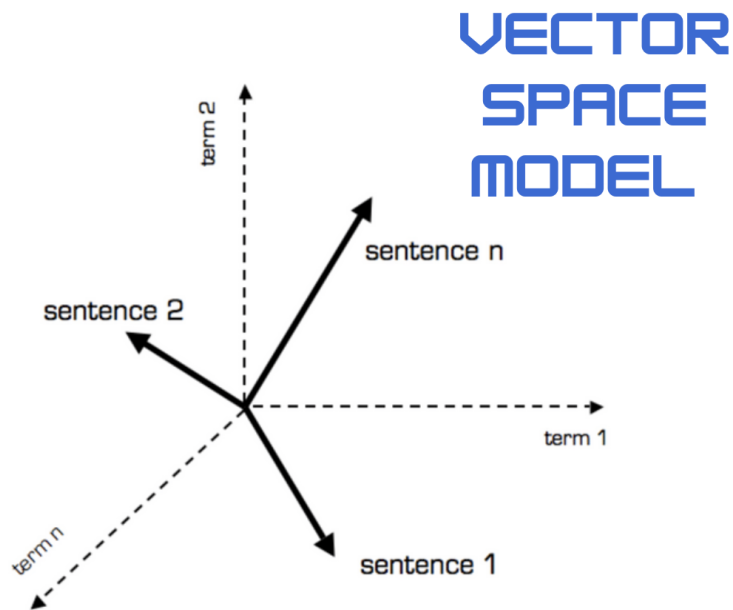


Figure 3.5: Vector Space Model

Examples of NLP can be explained as follows;

- NLP records Patients' statements and conversations, as well as reviewing the patient's old electronic health records, allowing the disease to be identified and predicted. This ability to depression, cardiovascular disease, and even cancer diagnoses are being investigated in the changing health conditions.
- With NLP, organizations can determine the opinions, ideas, and trends of their customers, and what they say about a service or product. NLP draws the desired result by identifying information in sources such as social media.
- An employee at IBM developed a personalized search engine. This tool can be called a cognitive assistant who knows a lot of information about you.
- Companies like Yahoo and Google, by analyzing the text in the e-mail servers and spam from your e-mail with NLP stopping from entering your inbox by e-mail filter and categorize.
- To help identify the fake news, the NLP Group at MIT has developed a new system to determine the source of a correctly or whether politically biased. a system like this is a serious example of a false news set and determines whether they trusted a trusted news source in this way.

The various text preprocessing steps that I used in this study are:

- Tokenization
- Lower Casing
- Stop words removal
- Stemming
- Lemmatization

3.3.3. Stop words removal

Sometimes, some extremely common words that have little value in helping to select documents that fit a user's need are completely excluded from the vocabulary. These words are called stop words. A stop list overall strategy for determining is to sort the frequency of collection of terms (in the collection of documents each term the total number appear) and then take the most common terms used, usually the domain of indexed documents for semantic content, as a stop list, members then indexing discarded during (nlp.stanford.edu, 2009).

A stop word is a commonly used word (such as "the", "a", "an", "in") and these words do not have any significance as they do not help distinguish the two documents. These words can turn out to be too many problems. Therefore, it is necessary to delete these unrelated words.

"Or, and as boolean common features, however, and pronouns should be removed because the influence and these words do not take very least adds value or no value to the classification process" (Amman Ismael I Kadhim, 2018). The model is programmed to ignore stop words as a result.

Stop words, is a set of commonly used words in a language. stop words in English, "a", "the", "is", "are", etc. is. Nevertheless, there is no one universal list of words to be blocked, so it may vary. The list of stop words and practices vary by language. Stop using intuition behind the words, instead of removing words from the text that contains little information is focused on the important word.

Analyzing instead of all words from the word lists, word Removal can stop applying heavy Preventors this stage. The popular search area systems are applying fields of stop words, text classification applications, topic modeling, topic extraction, and others.

The tool has a predefined list of StopWords NLTK that refers to the most common words. By storing a list of the words that you consider to stop words can remove them easily. NLTK (Natural Language Toolkit) in python has a list of stopwords stored in 16 different languages.

3.3.4 Stemming and lemmatization

Stemming Stanford University NLP according to the source, stemming often indicates a rough intuitive process often degrade the ends of the target word in hopes of achieving accuracy and usually involves the removal of attachment differentiation.

Once again, according to Stanford University NLP sources, Lemmatization defines it as follows: "Lemmatization refers to doing things neatly using a vocabulary and morphological analysis of words, usually aimed at removing inflectional endings and restoring the basis of a word or verbal form. This system is known as lemma (nlp.stanford.edu, 2009). For grammatical and linguistic reasons, documents will use different forms of a word. For example, to edit them, organize and edit

comes with different forms of the same word database. Also, families of these words, another example of democracy, and democracy and democratization have both derive the purpose of explanation, the shot is a basic form of a word and sometimes reducing common forms of related derivative form. Also, families of these words, another example of democracy, and democracy and democratization have both derive the purpose of explanation, the shot is a basic form of a word and sometimes reducing common forms of related derivative form. For instance:

am, are, is ⇒ be

car, cars, car's, cars' ⇒ car

The result of this mapping of text will be something like:

the boy's cars are different colors ⇒

the boy car be a different color

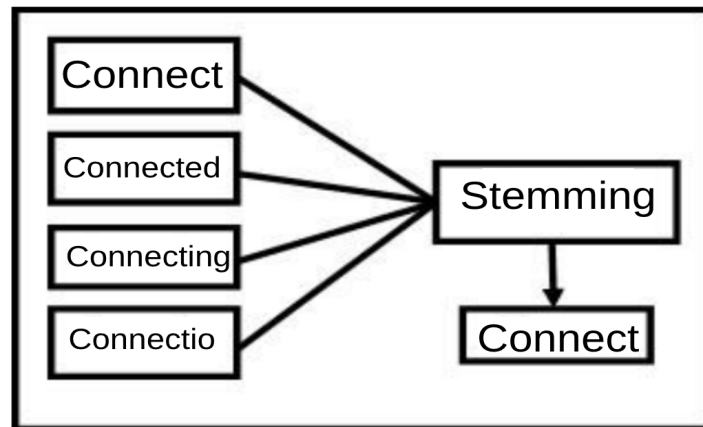


Figure 3.6: Stemming

3.3.5 Tokenization

The tokenization is called text segmentation or lexical analysis. The basic definition of tokenization is splitting data into small chunks of words. Tokenization is a popular process. Tokenization is one of the types of natural language text preprocessing. In this step, all the basic information about the data without compromising the security of sensitive data modification and having a unique identification symbol is preserved. As a result, raw text is pre-treated and divided into textual units (I Kadhim Ismael Ammar, 2018). There are 2 types of tokenization.

1. Sentence Tokenization
2. Word Tokenization

3.3.5.1. Sentence Tokenization

The sentence on simgeselleştir is divided into a series of written language component of a sentence. The idea here seems very simple. English and several other languages, the separate sentence is a punctuation mark, we see a common need.

```
Its history can be traced back nearly 5,000 years to archeological discoveries in the Middle East.
```

```
It is a two player game where each player has fifteen checkers which move between twenty-four points according to the roll of two dice.
```

Figure 3.7: Sentence Tokenization

3.3.5.2. Word Tokenization

Word icon making, is divided into component words of a series of written language. gaps in many other languages that use a form of English and Latin, the word separator is a better approach.

```
['Backgammon', 'is', 'one', 'of', 'the', 'oldest', 'known', 'board', 'games', '.']
```

```
['Its', 'history', 'can', 'be', 'traced', 'back', 'nearly', '5,000', 'years', 'to', 'archeological', 'discoveries', 'in', 'the', 'Middle', 'East', '.']
```

```
['It', 'is', 'a', 'two', 'player', 'game', 'where', 'each', 'player', 'has', 'fifteen', 'checkers', 'which', 'move', 'between', 'twenty-four', 'points', 'according', 'to', 'the', 'roll', 'of', 'two', 'dice', '.']
```

Figure 3.8: Word Tokenization

```
In [32]: print("The encoding for document\n",df.combined[0],"\n is : ",train_sequences[0])
```

```
The encoding for document
budget fight looms republicans flip fiscal script washington reuters head conservative republican faction congress v
oted month huge expansion national debt cuts called fiscal conservative sunday urged budget restraint keeping sharp p
ivot republicans representative mark meadows speaking face nation drew hard line federal spending lawmakers bracing b
attle january return holidays wednesday lawmakers begin trying pass federal budget fight likely linked issues immigra
tion policy november congressional election campaigns approach republicans seek control congress president donald tru
mp republicans want budget increase military spending democrats want proportional increases defense discretionary spe
nding programs support education scientific research infrastructure public health environmental protection trump admi
nistration willing going increase defense discretionary spending percent meadows chairman small influential house fre
edom caucus said program democrats saying need government raise percent fiscal conservative rationale eventually peop
le money said meadows republicans voted late december party debt financed overhaul expected balloon federal budget de
ficit trillion years trillion national debt interesting hear mark talk fiscal responsibility democratic representativ
e joseph crowley said crowley said republican require united states borrow trillion paid future generations finance c
uts corporations rich fiscally responsible bills seen passed history house representatives think going paying years c
ome crowley said republicans insist package biggest overhaul years boost economy growth house speaker paul ryan suppo
rted recently went meadows making clear radio interview welfare entitlement reform party calls republican priority re
publican parlance entitlement programs mean food stamps housing assistance medicare medicaid health insurance elderly
poor disabled programs created washington assist needy democrats seized ryan early december remarks saying showed rep
ublicans overhaul seeking spending cuts social programs goals house republicans seat senate votes democrats needed ap
prove budget prevent government shutdown democrats leverage senate republicans narrowly control defend discretionary
defense programs social spending tackling issue dreamers people brought illegally country children trump september ma
rch expiration date deferred action childhood arrivals daca program protects young immigrants deportation provides wo
rk permits president said recent twitter messages wants funding proposed mexican border wall immigration changes exch
ange agreeing help dreamers representative debbie dingell told favor linking issue policy objectives wall funding nee
d daca clean said wednesday trump aides meet congressional leaders discuss issues followed weekend strategy sessions
trump republican leaders white house said trump scheduled meet sunday florida republican governor rick scott wants em
ergency house passed billion package hurricanes florida texas puerto rico wildfires california package exceeded billi
on requested trump administration senate voted
is : [460, 3207, 1, 402, 284, 2187, 2347, 8, 16, 270, 726, 154, 147, 412, 1898, 53, 1, 45, 292, 256, 2065, 1379, 4,
2621, 4315, 69, 2834, 6206, 2934, 3, 10, 1, 81, 115, 398, 1177, 284, 16, 12, 2755, 3081, 3581, 116, 142, 231, 126, 6
9, 1982, 98, 494, 85, 483, 187, 420, 284, 995, 3383, 1473, 2834, 170, 297, 61, 2347, 181, 460, 2, 956, 67, 1, 1368, 2
187, 3581, 116, 270, 286, 105, 2330, 1092, 2574, 1019, 143, 69, 45, 509, 21, 2427, 9317, 4043, 6970, 284, 414, 3744,
6719, 2, 1715, 808, 191, 45, 1267, 1202, 3894, 812, 1377, 2834, 149, 13135, 1730, 173, 284, 1404, 2226, 1264, 918, 14
77, 2922, 2834, 2135, 1332, 6835, 280, 2, 2834, 1676, 903, 26, 2666, 284, 1338, 289, 806, 1264, 1178, 26, 1379, 2934,
460, 2, 126, 6206, 2940, 3573, 6719, 460, 2, 436, 256, 442, 637, 334, 94, 3459, 1839, 2621, 665, 4315, 284, 7312, 16,
12, 410, 14, 6719, 6206, 2666, 154, 2, 81, 446, 64, 11, 460, 491, 3574, 1557, 2310, 1242, 6719, 786, 49398, 38141, 1
7, 1124, 56, 1006, 284, 361, 16, 12, 7312, 5877, 818, 1380, 16, 12, 2, 260, 1898, 215, 812, 1377, 57, 6750, 126, 207,
7248, 28632, 2, 436, 143, 1757, 21, 422, 816, 5247, 28632, 17, 8, 824, 769, 2018, 232, 4, 426, 1236, 2188, 379, 94, 9
4, 293, 598, 143, 4, 1043, 1443, 1753, 28]
```

Figure 3.9: Encoding

3.3.6. Lower Casing

Text pre-processing is one of the simplest and most effective ways. This incredibly simple process is a very easy step in the processing of words in the entire dataset. Most of the issues are valid for text mining and NLP problems are not too big and your data sets that can help where help significantly to the consistency of the expected output. Lower Case Conversion rarity thanks to a fantastic way to solve the problem, many people are using it, and this explains why the use of lowercase letters. Here is an example of how to solve the problem of the small letters below rarity;

RAW	LOWERCASED
Canada Canada CANADA	canada
TOMCAT Tomcat toMcat	tomcat

Figure 3.10: Lower Casing

3.3.7. Word Embeddings and Text Vectorization

The process of converting words to numbers to manipulate natural language text and extract useful information from a particular word is called vectorization and word embedding. A sentence or text used in machine learning and deep learning techniques must be transformed into a set of real numbers, a vector. Word Embedding, in other words, is known as word vectorization. Each word is represented by a vector of true value, usually tens or hundreds of dimensions, where words are compared to thousands or even millions of dimensions. A word embedding is a learned representation for text where words with the same meaning have a similar representation. Natural language processing is realized through this approach to represent words and documents that can be considered as one of the fundamental stages of deep learning in challenging problems. Most neural network toolkits do not work well with very high dimensional, sparse vectors, and therefore the use of computational low dimensional vectors is one of the advantages (Yoav Goldberg, 2017). Each word is mapped to a vector and the vector values are learned like a neural network, so technically deep learning is often clustered as one-hot encoding. Likewise, I applied the model to one hot coding. Burial words, substantially at a predefined vector space of the individual words are an engineering grade being represented by real-valued vectors. Word2vec developed by Google, as examples thereof Glove developed by Stanford and has issued only tools for this process. Both of the two N-dimensional vectors mathematically as reflected in linear space (item1, item2) is a measure of the cosine of the angle between. The advantage of the cosine similarity is the Euclidean distance even if predicting the document similarity. Word embeddings are used in the following use cases (Prabhu, 2019).

- Compute similar words
- Text classifications
- Document clustering/grouping
- Feature extraction for text classifications
- Natural language processing.

3.3.7.1 Word2Vec

For efficient learning one of the methods is Word2Vec that converts word embedding from a text corpus. For explaining the Word2Vec giving example makes more clear, if you give a string like [King, Queen, Prince], the Word2Vec method will show "Prince" as the words closest to them. Word2Vec allows for vector-oriented reasoning based on distances between words. Learned automatic relationship of male and female and with induced vector representations, reasoning based on distances between words

became more clear. For example, "King - Male + Female" results in a vector very close to the "Queen" and there are similarities in the field. Generally speaking, when you hear the news that "a woman was killed by a neighbor", the neighbor is thought to be a man. This shows exactly what perception management is. The mainstream media feeds our minds as "Men are usually murderers", as a result, we have a certain perception. (Mücahit Gültekin,2016)

There are two different words to find the embed learning model that can be used as part of the word2vec approach; Continuous Bag-of-Words, in other words, CBOW Skip-gram model and Continuous Model. These are two based on different things. Briefly, the CBOW model learns to place guessing based on the current word in context, and other care to continuously jump-gram model, learn by guessing words surrounding an existing word constantly jump-gram model, the words surrounding the current word.

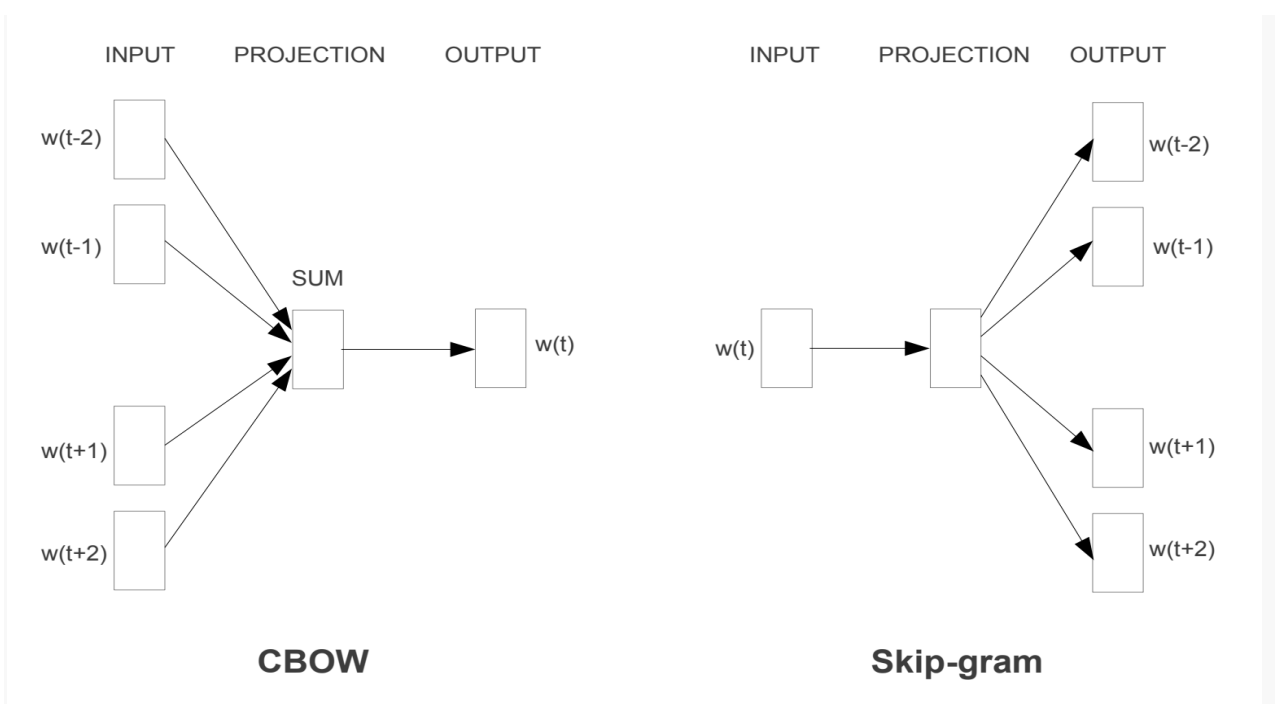


Figure 3.11: CBOW and Skip-gram

3.3.8 One Hot Encoding

One hot encoding, such as one of the multi-class classifications is still the most common procedure to fulfill their duties. This procedure modeled a $f : R^p \rightarrow Z_2^n$ the function that maps image or audio samples to a probability distribution on a separate set of n tags of target categories. This process requires categorical values to be mapped to integer values first. Then, for each integer values, all the values except for the one marked integer index is represented as a binary vector of zero.

Unfortunately, when the output of magic, class labels, the current label does not cover the full label area mainly due to the cross-correlations properly. As a result, the output of higher-dimensional (orthogonal) projection of a field, because the calculation of the parameter space of the model significantly increases, one hot coding can fail in fine-grained classification tasks. Also, the sampling rate per track for a plurality of data sets with a tag is typically reduced. Therefore, it is used in a hot encoding and embedding the integrity of my work. It can be thought of as a kind of "binary encoding".

OneHotEncoder class, which is also under the preprocessing library, can be used to perform this transformation as follows. (Pau Rodríguez, 2018)

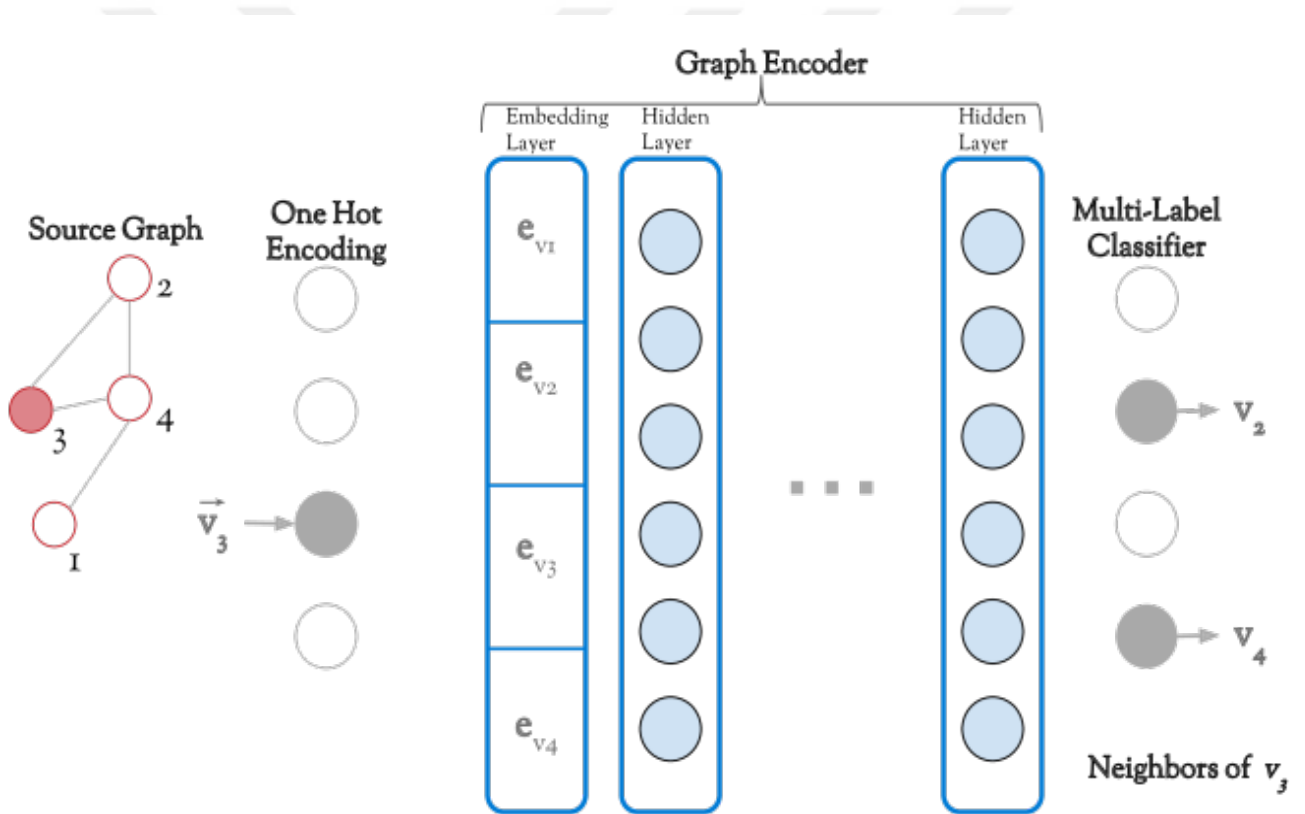


Figure 3.12: One Hot Encoding Schema

CHAPTER 4

4. METHODS AND MATERIALS

4.1. FEATURE SELECTION METHODS

4.1.1 Classification

Supervised machine learning algorithms, after analyzing a series of observations, input and try to express as a mathematical dependence between the output and this defines descriptive models that capture the relationships between the data. Classification is an area of supervised machine learning. Be the labels or categories according to the characteristics of an asset and tries to predict. It deals with many important machine learning problems. Classification problems have classes or categories called discrete and finite output. For example, if an article at the right is a classification problem. The two main types of problem of classification, classification, and binomial and multinomial classification.

1. Bilateral or binomial classification: exactly two classes you can choose from (usually 0 and 1, or true and false positive and negative).

2. Very class or polynomial classification: three or more classes of output selection can be made.

“If there’s only one input variable, then it’s usually denoted with x . For more than one input, you’ll commonly see the vector notation $\mathbf{x} = (x_1, \dots, x_r)$, where r is the number of the predictors (or independent features). The output variable is often denoted with y and takes the values 0 or 1.” (Mirko Stojiljković, 2020).

4.1.2. Classification Process

Below is a schematic representation of the classification process. The first step of the process is to pre-process the data which includes stopping word removal, lower casing, tokenization, word embedding of the entire characters. Also, special characters are removed from the corpus. The One-Hot encoding process is used, and a vector is formed representing the documents involved. The ultimate step of the classifier is to predict the class of the given news article into real or fake in the label column. Inside the label column, real and fake news is represented by 1 and 0. There are a total of 5 machine learning models and deep learning models being tested here to arrive at the best choice. The data is partitioned into train and test.

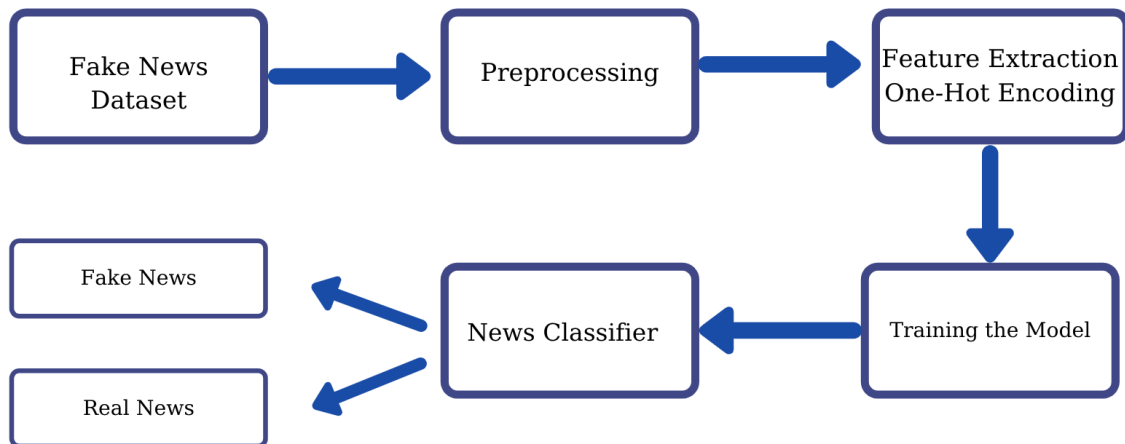


Figure 4.1: Classification Process Schema of Fake News Detection

4.2. Machine Learning

Machine Learning is an area that the ability to learn computers. This ability to learn is taught without explicitly programmed. Machine learning is no longer suitable for many areas and provides exciting results. This is a human talent is the work area where computer learning is taught. Machine learning has various types of classification. In this section, I will explain some types of use.

4.2.1. The Multinomial Naive Bayes classifier

Scikit-learn.org, which can be accepted as the Main Source, has defined this classification method as follows; “The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution normally requires integer feature counts. However, in practice, fractional counts such as tf-idf may also work.”

Naïve Bayes, which is computationally very efficient and easy to implement, is a learning algorithm frequently used in text classification problems and the multinomial Naive Bayes classifier is suitable for classification with discrete features as word counts for text classification and fake news classification.

Multinomial Naive Bayes classifier is used for distributed data polynomial. The emergence of a feature does not affect the possibility of the other features and also performed better than multinomial Naive Bayes's most powerful alternative to small sample sizes. Application is easy, fast, and accurate is used in many different areas.

To understand how Naïve Bayes works, first, have to make sure to understand the concept of Bayes' rule. This model was formulated by Thomas Bayes (1701-1761) :

$$P(A|B) = \frac{(P(B|A)P(A))}{P(B)}$$

where,

$P(A)$ = the prior probability of occurring A

$P(B|A)$ = the conditional probability of B given that A occurs

$P(A|B)$ = the conditional probability of A given that B occurs

$P(B)$ = the probability of occurring B

4.2.2. Random Forests

During training time, Random forests by building multiple decision trees. Random forests generally outperform decision trees. However, data characteristics can affect their performance, cause of this, cleaning data is also important.

An RF step is to install before, and this ensures that each unique random decision tree in the forest, which reduces the overall variance of the RF classifier. For the final decision, the RF classifier collects the decisions of the individual trees. In the end, we can see that the RF classifier shows a good generalization. RF classifiers tend to perform better than most other classification methods in terms of accuracy. RF classifier does not need the feature scaling; This feature is available in DT. RF classifier, the training is a more robust selection of samples and noise in the training data set. Although more difficult to interpret the RF classifier, the advantages are easier to set hyperparameters according to DT classifiers.

The steps for creating a random forest classifier (Breiman, 2001):

- “Give the variable named 'N' a proper value, which is the number of elements of each subset of features.
- Depending on the value of N, select a new feature subset m from the whole feature set at random.
- Train the data set with the feature sub-set to build a decision tree for each training set group.

- Select a new m that is different from others and do the whole thing again until all the subsets of the features are used. A random classifier of the forest is completed.
- Give the test data and determine the class name of the study according to the ranking results of each classification.” (Simsek,2020)

4.2.3. K-Nearest Neighbor

K-NN (K-Nearest Neighbor) algorithm is one of the simplest and most widely used classification algorithms. K-NN nonparametric (nonparametric), lazy (lazy) is a learning algorithm. If we try to understand the concept of lazy, unlike eagerness to learn, there is a training phase of lazy learning. Training data can not be learned, rather than the training data set that "memorizes". We want to make a prediction, call the nearest neighbors in the entire data set.

A K value is determined for the algorithm to work. This K value is the number of elements to look at. When a value arrives, the distance between the value is calculated by taking the closest K element. The Euclidean function is often used in the distance calculation process. Manhattan, Minkowski, and Hamming functions can also be used in addition to the Euclidean function. It is sorted after calculating a distance and a corresponding value is assigned to the corresponding class (Hall, 2018).

4.2.4. Stochastic Gradient Descent

SGD has been successfully applied to large-scale and sparse machine learning problems often encountered in text classification and natural language processing and the reason for this, it performs operations on 1 random data at each step. It processes 1 data randomly at each step, which makes it more effective. Since the points are constantly changing, it aims to reach the optimum point. In my thesis, the reason for using the SGD algorithm instead of Batch Gradient Descent is SGD is faster than Batch Gradient Descent. Fast and effective processes are of great importance in the model development process.

Just a way of training, SGD is essentially an optimization technique. In contrast. SGD does not correspond to a family of machine learning models. SGD is an efficient and ease of implementable optimization technique. All these advantages have some disadvantages too. In SGD feature scaling steps are sensitive and require several hyperparameters such as the regularization parameter and the number of iterations (scikit-learn.org,2020).

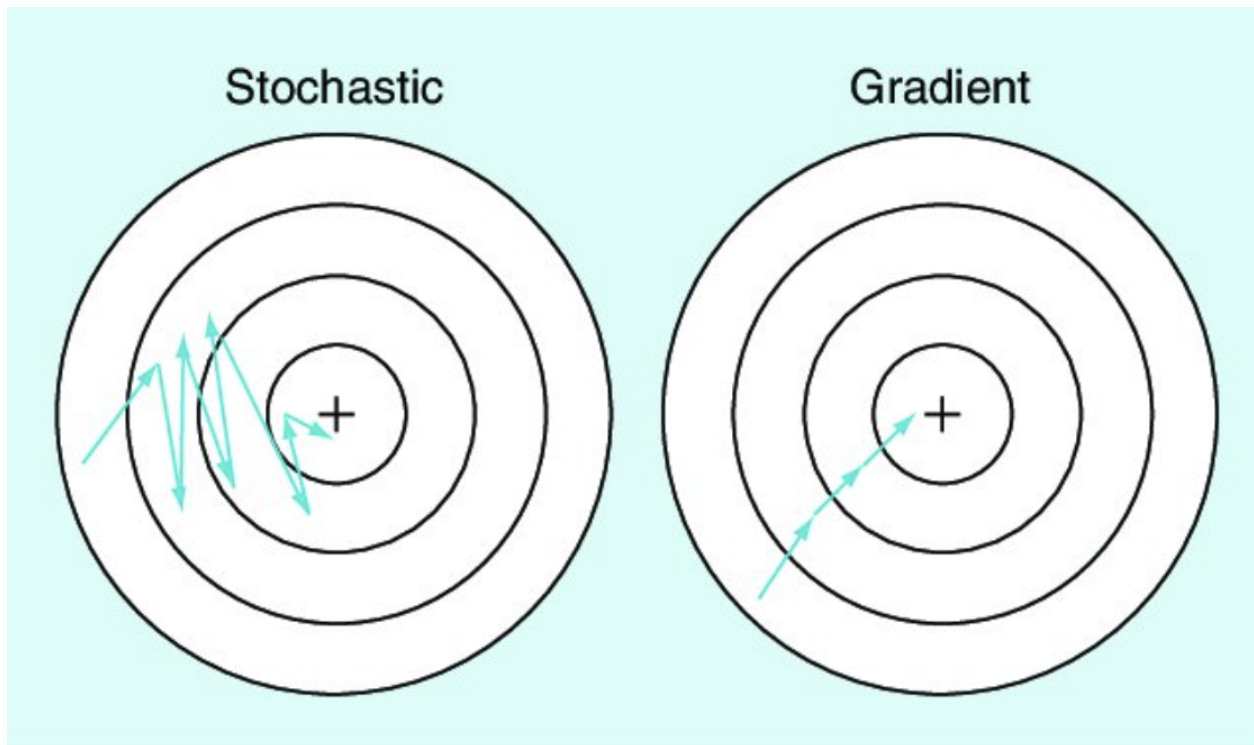


Figure 4.1. Stochastic gradient descent compared with gradient descent.

Source: https://www.researchgate.net/figure/Stochastic-gradient-descent-compared-with-gradient-descent_fig3_328106221

4.2.5. Logistic Regression

Logistic Regression (LR), one Machine Learning classification algorithms. LR is used to estimate the probability of a categorical dependent variable. In logistic regression, the dependent variable 1 (yes, successful, etc.) Or 0 (no failure etc.) Is encoded as a binary variable that contains the data. In other words, the logistic regression model predicts $P(Y = 1)$ as a function of X . "(wikipedia.org).

There are certain Logistic Regression requirements. Some of the steps required to make the Logistic Regression process feasible are as follows; First of all, for Binary regression, the 1st factor level of the dependent variable should represent the desired result and only important variables should be included. Without these, it is difficult to get a healthy result. Moreover, the arguments must be independent of each other. In other words, the model should have little or no multi-link capability. Basic classification technique of logistic regression, linear classifiers belong to the group and are said to be somewhat similar to the polynomial and linear regression.

To understand LR we need to understand the sigmoid function. The sigmoid function is defined in the below part. Although in a short, the sigmoid function has values very close to either 0 or 1 across most of its domain and so this one makes it suitable for application in classification methods. This image shows the sigmoid function of some variable x :

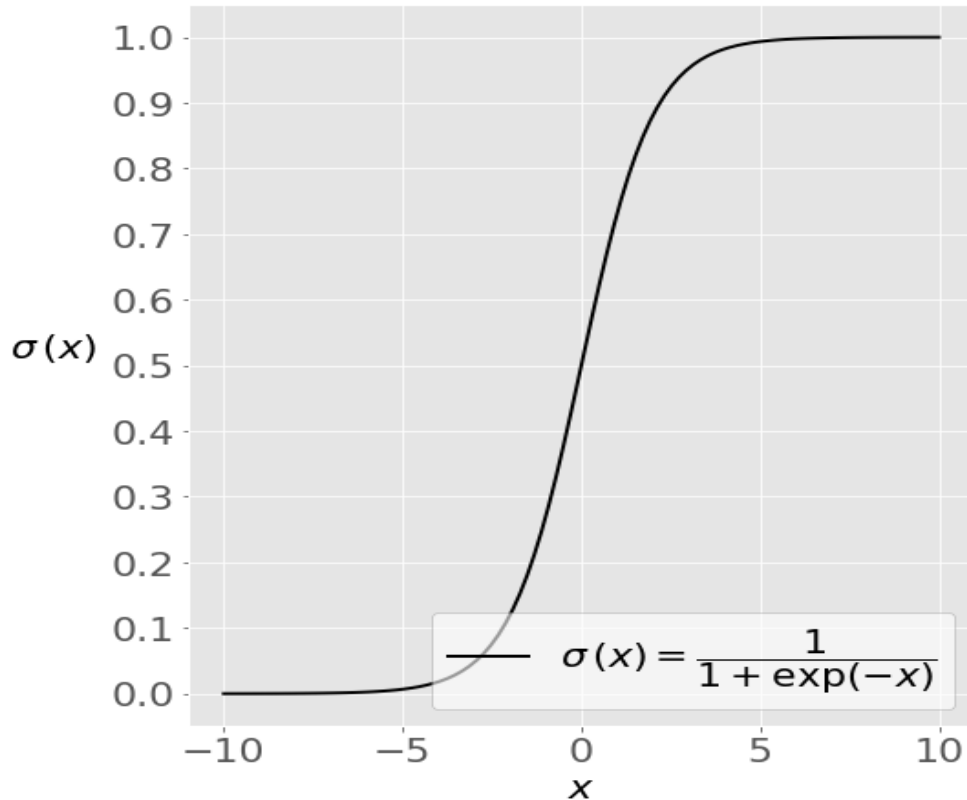


Figure 4.1. Sigmoid function

4.3. Deep Learning

Deep learning exists as a subfield of machine learning. There are certain similarities between machine learning and deep learning. One of the ways to implement machine learning is DL. Deep Learning methods are a modern update of Artificial Neural Networks that are concerned with building much larger and more complex neural networks. There are some types of deep learning, such as supervised and unsupervised, and also the most popular deep learning algorithms are:

- Convolutional Neural Network (CNN)
- Recurrent Neural Networks (RNNs)
- Long Short-Term Memory Networks (LSTMs)
- Deep Belief Networks (DBN)

In my thesis, I used RNN and LSTMs. There are mainly important reasons for the increasing use of deep learning today; the significantly lowered cost of computing hardware is one of those reasons (L. Deng, 2014).

Thanks, kind of these advantages deep learning applies in many fields;

- Medical: For cancer cell detection, brain, gene printing, etc.
- Banks: Stock prediction, financial decisions.
- Information Retrieval: search engines, both text search, and image
- Computer Vision (CV)
- Natural Language Processing (NLP)
- Recommender Systems
- Reinforcement Learning
- Graphical Models

Machine Learning Deep Learning the difference between, on the one machine learning workflow begins with the removal of the images manually specifications, and deep learning in a workflow is automatically removed from the display-related features. Also, deep learning, a network of raw data and given the task to be carried out "by learning from the end u" performs. Typically, the data size will continue to grow in development. The use of small-sized raw data makes it possible to get bad results.

Because deep learning generally uses neural networks, deep learning models are also called deep neural networks. The word "deep" is used to describe the number of hidden layers in the neural network. One of the main differences with Machine Learning is the difference in the number of hidden layers. While this number may be even 150 deep learning in traditional neural networks 2-3 hidden layers. This is one of the causes of profound learning generally work better and are the more highly preferred method. Deep learning models, it does not require manual feature extraction. It learns properties from data, is trained using large label datasets and neural network architectures (mathworks.com).

4.3.1. Neural Network

Improved technology allows computer programs can reflect the behavior of the human brain. Although the neural network is part of the human brain, nowadays science applies and reflects this system on programs. NN solves common problems in the fields of AI. (www.ibm.com)

Learning is a human brain feature. New information is generated after raw information is learned. NN is an information system that can automatically generate this information. The learning method of this system depends on the modeling of it mathematically. In artificial neural networks, it experiences information first in the learning process. Structurally, NN consists of many layers. These layers are three as input, output, and hidden layers. The middle part is called a hidden layer. Layers contain a certain number of neurons and these neurons are connected by Synapses. There is a coefficient in synapses. These coefficients tell how important the information in the neuron to which it is connected is.

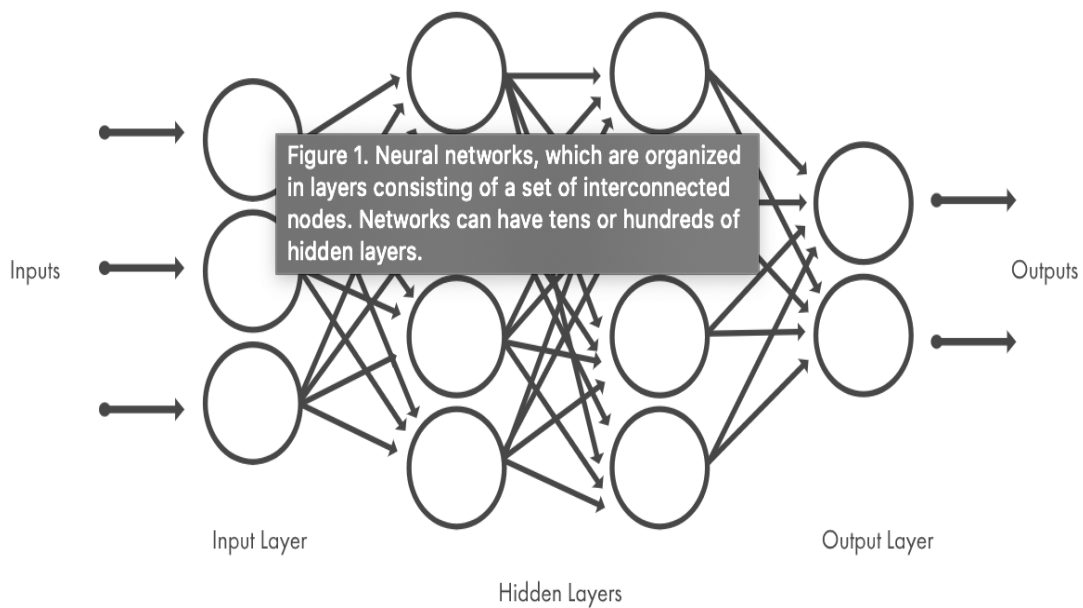


Figure 4.1: Neural networks

This figure shows us Neural networks, which are organized in layers consisting of a set of interconnected nodes, and also networks can have tens or hundreds of hidden layers.

Advantages of Artificial Neural Networks

- Artificial Neural Networks consist of many cells and this process performs complex tasks by working equally.
- Learning ability can be learned with different learning algorithms.
- Results (information) can be produced for unseen outputs. Unsupervised learning is involved.
- They can make pattern recognition and classification. They can complete missing patterns.
- They have fault tolerance. They can work with incomplete or ambiguous information. They show graceful degradation.
- They can work in parallel and process real information (www.veribilimiokulu.com,2020).

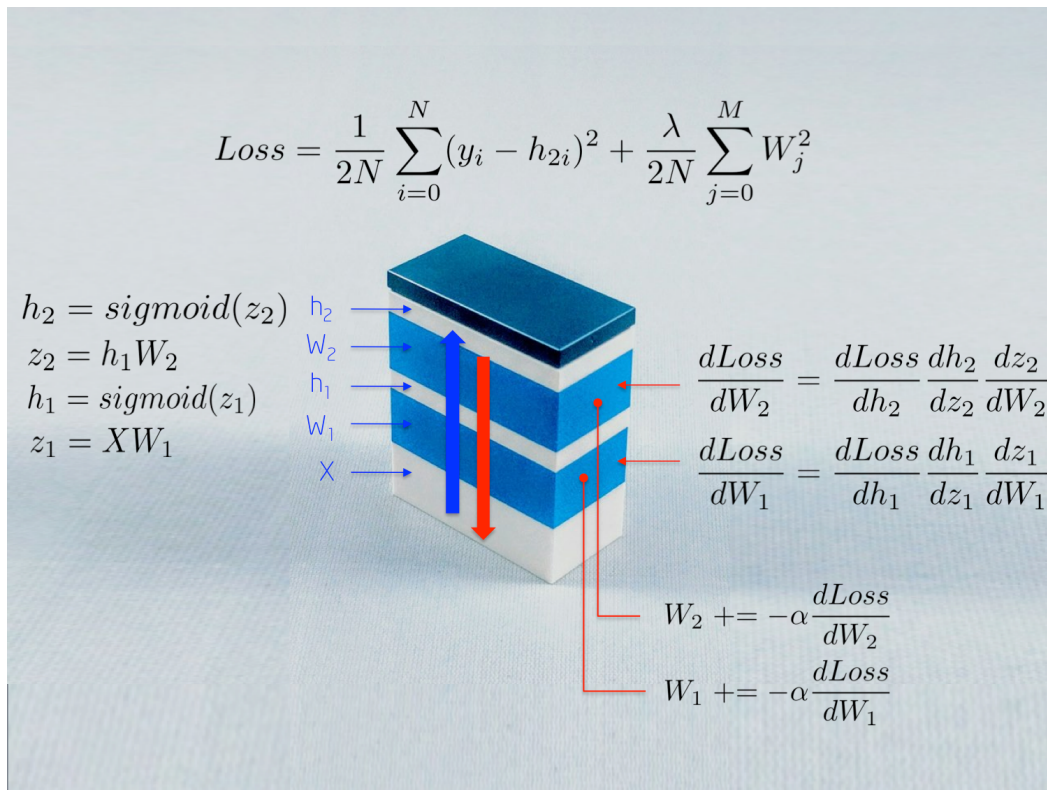


Figure 4.2 : Neural Networks as a Composition of Piece

Source: https://omar-florez.github.io/scratch_mlp/

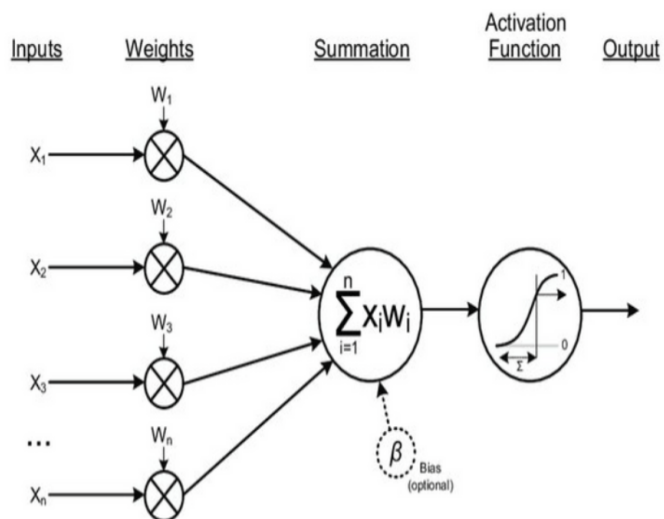


Figure 4.3: Artificial Neural Network (ANN)

As you see in Figure 4.3 this is a second-generation Artificial Neural Network (ANN), with a sigmoidal activation function. The output is continuously based on the inputs received. In a simple neural network, the input x is defined as the weight w and the value passed to the output of the network $f(x)$ or activation, we performed. Then there will be input to the final output or another layer.

Weights: A goal of weight is deciding how much influence the input will have on the output and weight control the signal between two neurons.

Biases: Biases that are fixed is an additional input to the next layer that will always get a value of 1, and is updated for it. Even if all inputs are zeros the bias unit guarantees here will still be activation in the neuron.

This equation is the mathematical operation we do when connecting neurons. The formula for this is given in the following section.

$$L = X \cdot W + b$$

Figure 4.4: Weight and Biases

The activation function is applied for each row. The formula for this is given in the following section.

$$Y = \text{Softmax}(X \cdot W + b)$$

Figure 4.5: Activation Function, Softmax

4.3.1.1. Activation Functions

The task of the Activation Function is to process the net input to the cell and determine the output that the cell will produce in response to this input. If the activation function is not applied, the output signal becomes a simple linear function. Linear functions are only one-order polynomials. A neural network without the activation function will behave like a linear regression with limited learning power. But we want our neural network to learn nonlinear situations as well. Therefore, generally, a nonlinear function is chosen as the activation function. Because text will be given which is real-world data for our neural network to learn. Multi-layer deep neural networks can thus make meaningful inferences from data and learn its properties.

Another point that should be considered while choosing the activation function is that the derivative of the function can be calculated easily. This is important in terms of not getting tired of the device/computer used in terms of hardware. Since the derivative of the activation function is also used in feedback networks, a function whose derivative is easily calculated is chosen so that the calculation does not slow down. I used the sigmoid function which is an activation function in the program I wrote. The Sigmoid function" is generally used as the activation function in the "Multilayer perceptron" model, which is widely used today (Nilgun Sengoz,2017).

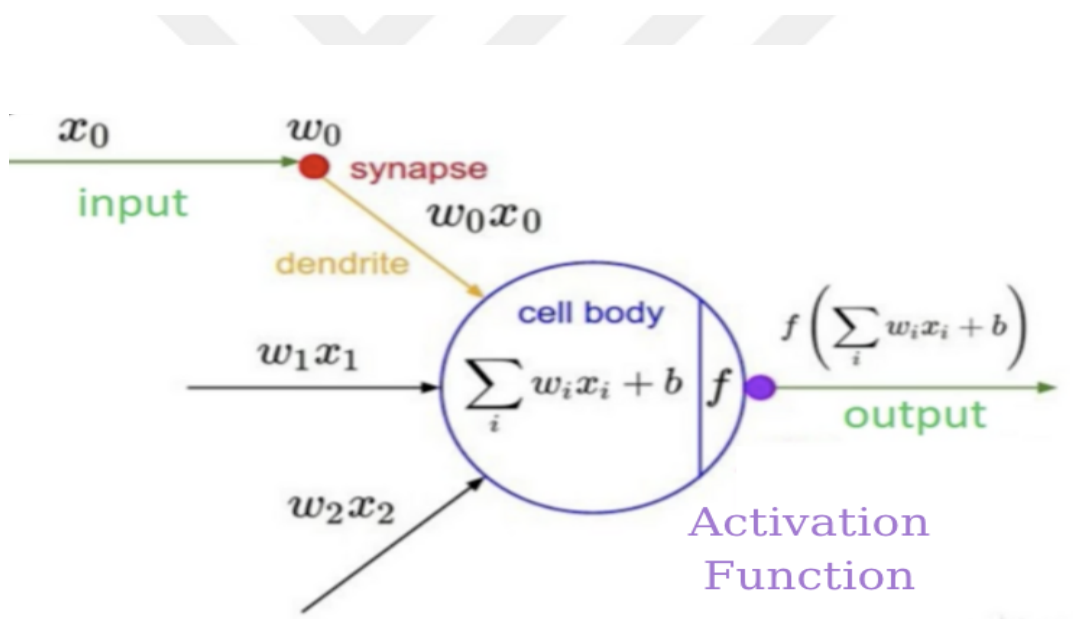


Figure 4.6: Convolutional Neural Networks for Visual Recognition

Source: <http://cs231n.stanford.edu/index.html>

4.3.2. Sigmoid Function

Also known as the transfer option. It is the function used to output the neural nodes in artificial neural networks. It allows the neural network to output a yes or no like. It compresses the obtained value between 0 or 1 or -1 or 1 depending on the function.

The function is defined for a connection with 0 and 1 and converges to 0 and 1 before the $x = 0$ axes, and always gets (0,1) worth knowing information when encountering (-infinite, + infinite). So the activation leap. The Sigmoid function, which is a continuous and derivative program, is also the most frequently used in unrealistic network applications.

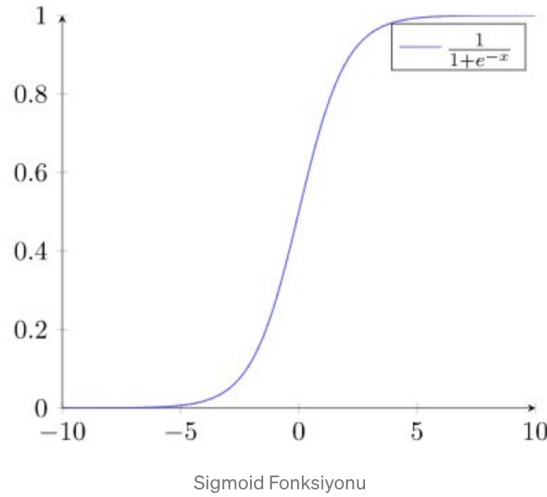


Figure 4.7: Sigmoid Function

4.3.3. Recurrent neural networks

Recurrent neural networks, also known as RNNs, are a class of neural networks that allow previous outputs to be used as input while having hidden states. In the transaction cycle of the RNN, the result from the hidden layer is written to both output and content units. In this way, its new input is processed along with the content units resulting from the processing of previous entries. If there is a correlation between different memorized data, it is called a "long term" dependency. RNN is a network that can calculate the relationship between these long-term dependencies.

$$h_t = \varphi(Wx_t + Uh_{(t-1)})$$

h_t is the result of the hidden layer at time t , x_t input is multiplied by the weight of W . Then, at time $t-1$, the value of $h(t-1)$ kept in the content unit is multiplied by the weight U and added by Wx_t . The W and U values are the weights between the input and the layer. Here, the weight matrix takes values according to which of the previous and current data has more or less effect on the result. The error resulting from these operations is calculated and new weight values are rearranged by backpropagation. The backpropagation continues until the error is sufficiently minimized. The sum of $Wx_t + Uh_{(t-1)}$ is inserted into the activation function like sigmoid, tanh. Thus, very large or very small values are taken into a reasonable range (Hamza Ergüder,2018).

Unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as fake news detection. All the inputs are independent of each other in other neural networks. Feedforward networks are also one of the NN types but not enough for these kinds of data sets (Aditi Mittal 2019).

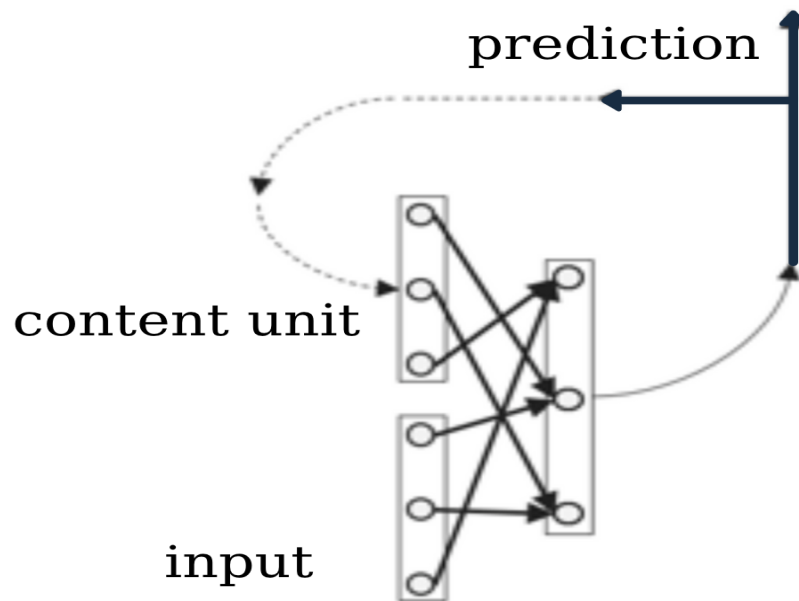


Figure 4.7:RNN process

There are some advantages and disadvantages of Recurrent Neural Network

Advantages of RNN;

- In the model sequence of data so that each sample dependent on previous ones
- RNN use convolutional layers to extend the effective neighborhood of the text vector

Disadvantages of RNN;

- Gradient vanishing and exploding problems.
- Training an RNN is a very difficult task and this is a computational challenge.
- It cannot process very long sequences depending on activation functions like *tanh*, *relu*.

4.3.4 Vanishing Gradient Problem

Recurrent neural networks during the update of the network involve the unrolling of the network for each input time step and create a very deep network that requires weight updates. and that makes a problem. This is called the vanishing gradients, also known as a particular problem with recurrent neural networks (Azer Çelikten,2020).

This creates a disadvantage. When a slow learning event occurs, the optimization that minimizes the error can be stuck to local minimum values and we cannot get the capability that can be taken from the artificial neural network model (Ayyüce Kızrak, 2019). Afterward, continued training does not result in anymore when we examine the changes in weights during training, we will see more learning in layers closer to the output layer and less change in layers closer to the input layer (Wang, 2019). As a solution to this problem, LSTM and GRU models, which are specialized versions of RNN, have been proposed. I used LSTM.

4.3.5 Long Short-Term Memory

Long Short-Term Memory (LSTM) networks, makes it easier to remember history data in memory. LSTM networks, and is a modified version of similar recurrent neural networks. Wherein the gradient problem was solved rnn'n disappeared. LSTM is very convenient to classify. The unknown period has given delays in the time-series forecasting and jobs and trains the model using the backpropagation. For these reasons, LSTM networks are the preferred technology.

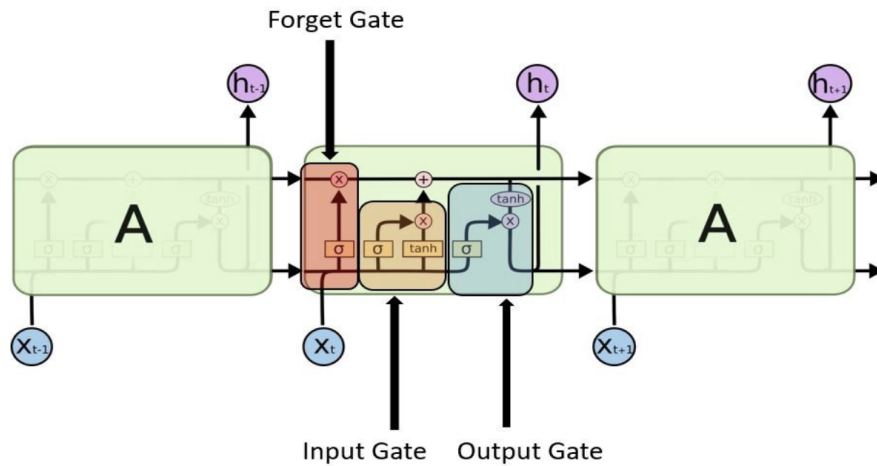


Figure 4.7:LSTM gates

Source: <https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e>

4.3.6. Confusion Matrix

A confusion matrix is known as the error matrix (Stephan, 1997). Each line represents the estimated range of the matrix and each column represents a real class, it can also be the opposite (David M. W., 2011).

The name of the confusion matrix, the system is due to the ease of mixing seeing two classes mix. Your mistake made by the classifier, and more importantly, gives an idea of the types of errors made. Print may be two or more classes is a performance measure for classification problems. of predicted and actual values is a table containing the four different combinations.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 4.7: Confusion Matrix

Source: <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>

- A true positive is an outcome of a model that correctly predicts the positive class.
- A true negative is an outcome of the model that correctly predicts the negative class.
- A false positive is an outcome of the model that incorrectly predicts the positive class.
- A false negative is an outcome of the model that incorrectly predicts the negative class.

It is so useful for measuring Recall, Precision, F1 Score, Accuracy, and most importantly AUC-ROC Curve. (Ibrahim Mohamed Gad, 2020)

Accuracy: Classification accuracy is calculated as below:

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

Recall: Sensitivity value is also a metric that will help us in situations where the cost of estimating as False Negative is high. It should be as high as possible.

$$recall = \frac{TP}{TP + FN}$$

Precision: It shows how many of the values we guess as Positive are Positive.

$$precision = \frac{TP}{TP + FP}$$

F1 Score: The F1 Score value shows us the harmonic mean of the Precision and Recall values.

$$F-1 = \frac{2 * (Recall * Precision)}{Recall + Precision}$$

4.3.7. Receiver Operating Characteristic Curve

ROC receiver operating characteristic curve or curve is a graphical plot showing. The separation is binary classifier system diagnostic capabilities for which the threshold has changed. ROC curve, the false positive rate (FPR) against the true positive rate (TPR) is formed by drawing. ROC analysis is a tool that enables the best possible pattern of selecting and disposing of insufficient independent cost context or the class distribution. ROC analysis further, the diagnostic decision-making cost/benefit analysis is also directly related to the (wikipedia.org).

4.3.8. Cross-Validation

Cross-Validation is a technique used in model selection to better predict the test performed in a model. The idea behind cross-validation is to create sample observation sections as validation sets from the training resource set. Fitting the model on the training data and working accurately on the model for the real data is not easy. For this, assuring that our model got the correct patterns from the data is an obligation. Also, ensuring that our model is not getting up too much noise is one of the requirements. For this goal, I use the cross-validation technique. Place a model in the training place, then measure it against its new validation set, and then get a better assessment of how the

model will perform when new observations are requested. The number of segments depends on the observation and bias in the sample dataset that the sample judgment on variable compensation changes depending on the more splitting the smaller the bias and the more changes (veribilimci.com,2017).

4.3.9. K-Fold Cross-Validation

Splitting k number of subsets known as folds perform then perform training on all the subsets but leave one(k-1) subset for the evaluation of the trained model. Using a method, I iterate k times with a different subset reserved for testing purposes each time.

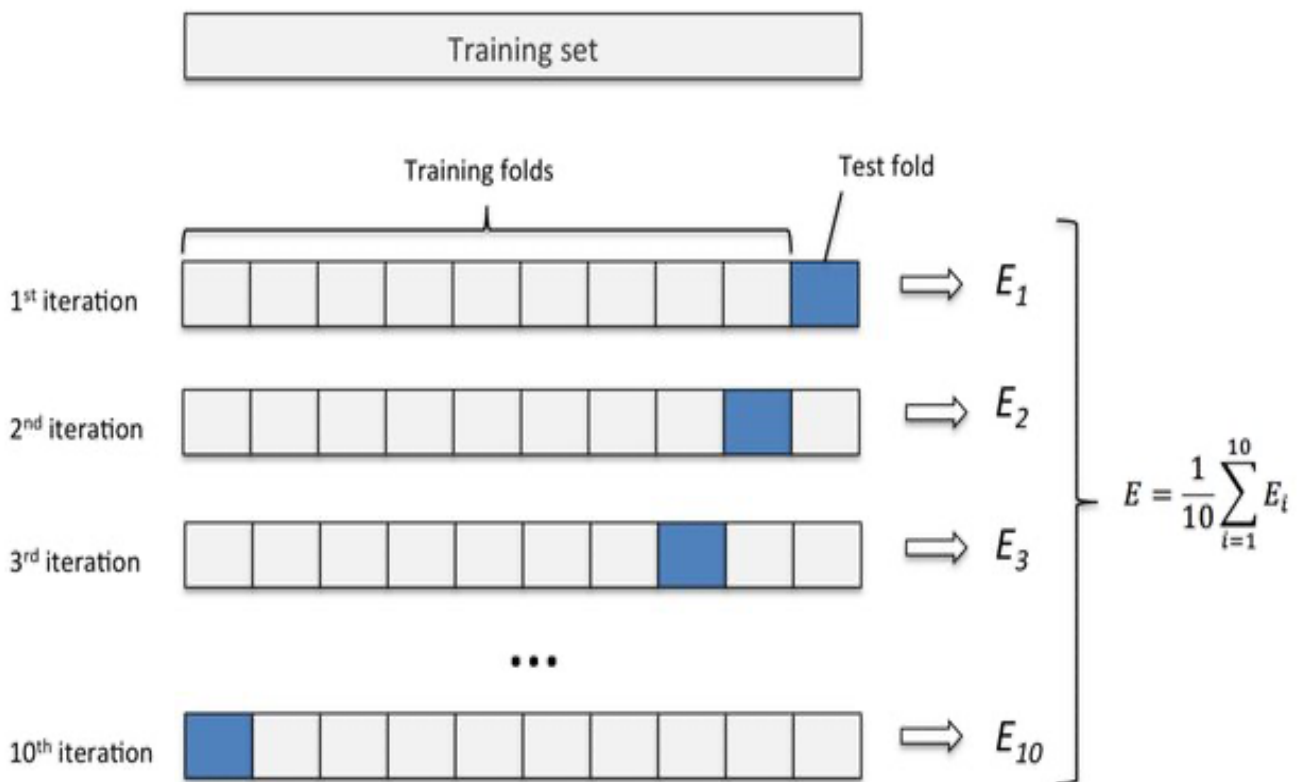


Figure 4.7: Diagram of k-fold cross-validation

Source: Karl Rosaen Log <http://karlrosaen.com/ml/learning-log/2016-06-20/>

CHAPTER 5

5. RESULTS AND DISCUSSIONS

In this chapter, experimental results for fake news detection with deep learning and classification of two different news datasets are completed. I have used pre-trained word embedding in training. The various machine learning algorithms, as well as deep learning algorithms, have been employed for classifications. Five different classification algorithms are used to classify detecting fake and real news. Also to these five classification algorithms I have used RNN LSTM methods. It gives much information about the news, and too many analysis steps are required to detect fake news.

5.1. EXPERIMENTAL DATASETS

Two different data were used in this thesis, named Liar and Reuters. Detailed information about these data sets can be found below.

Liar dataset: The first dataset is the Liar dataset. The LIAR dataset was published at first in July 2017. data is taken from PolitiFact's API. The main sources for deceptive detection of opinions are crowdsourced datasets (Ott et al., 2011; Pe´rez-Rosas & Mihalcea, 2015). Very useful datasets for examining deception detection, but positive training data are collected from a simulated environment. Because fake news on TVs and social media is much shorter than customer reviews, these datasets are not suitable for detecting fake statements. Vlachos and Riedel year at 2014 was the first to create datasets that check fake news and facts, and they received 221 statements by CHANNEL 42 and POLITIFACT.COM3, a Pulitzer award-winning website. PolitiFact specifically covers a wide range of political issues with detailed labels so that it makes detailed decisions. (William Yang Wang, 2017)

Reuters dataset: Reuters is a benchmark dataset for document classification. We can say more clearly that there are multiple classes in it. It has 90 classes, 7769 training documents, and 3019 test documents. It is a subtest of Reuters-21578 benchmark ModApte. The average number of words per document is between 93 and 1263 and they are grouped by class in the Training set. The word size of the training set is 35247 and even if you limit it to words that appear at least 5 times and a maximum of 12672 times in the training set, there are still 12017 words, which shows how useful the dataset is. (Thoma,2017)

5.1.1 Fake News Detection

Firstly, the classifications method was applied to the LIAR dataset. Dataset was split into a training set and a test set. %70 of the dataset used for training and %30 for testing. K-Fold cross-validation accepted as 5 for increasing the dataset variance. After preprocessing of datasets, classification

algorithms were applied on a liar and Reuters datasets. Firstly, classical algorithms were applied, and then the results were obtained with deep learning methods. Classical methods are K-Nearest Neighbors, Random Forest, Naive Bayes, Logistic Regression, Stochastic gradient descent, five different methods in Machine Learning and RNN, LSTM were used Artificial Neural Networks. Afterward, methods were applied, all these results were compared in the tables.

isfake	subject	text
0	Goverment News	1570
	Middle-east	778
	News	9050
	US_news	783
	left-news	4459
1	politics	6841
	politicsNews	11272
	worldnews	10145

Table 5.1 : Subject of Liar Dataset

This list is subject of Liar Datase and it displayed inside of source code of this fake news project. Subject labels was declared by main source of dataset. I have listed what I started with this topic in the software I wrote, and you see the topic list and the number of news in this table.

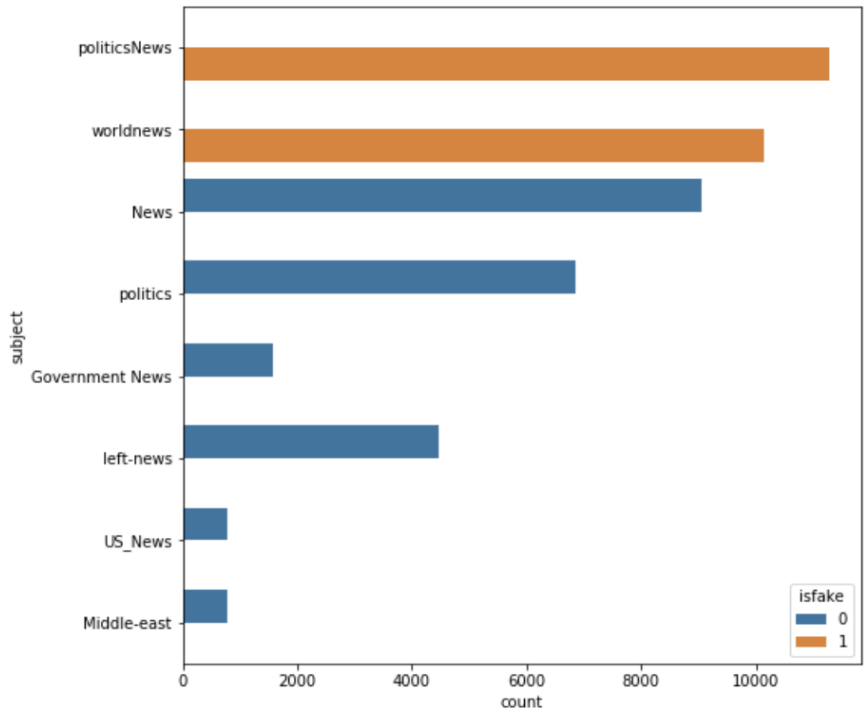


Figure 5.1: Rate of Liar Dataset

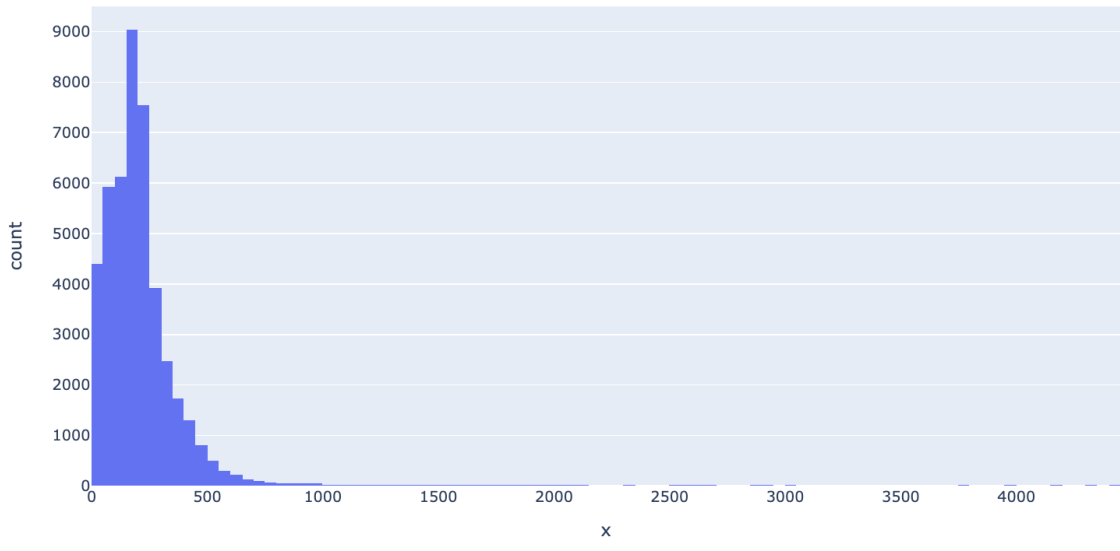


Figure 5.2: Number of Liar Dataset

Model	Score
LSTM	99.87506
Random Forest	99.042316
Naive Bayes	92.717149
SGD	90.801782
KNN	85.356347

Table 5.2: Model and Scores of Liar Dataset

First, the Naive Bayes classification algorithm was applied to the Liar dataset. 70% of the dataset is reserved for training and 30% for the test. After training, the algorithm was tested for 30% test data and classification was completed with an accuracy of Naive Bayes 92.71%. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated. Accuracy for MultinomialNB: 0.93

	Precision	recall	f1-score	support
Fake	0.93	0.93	0.93	4763
Real	0.92	0.93	0.92	4217
Accuracy			0.93	8980
Macro avg	0.93	0.93	0.93	8980
weighted avg	0.93	0.93	0.93	8980

Table 5.3: Accuracy for MultinomialNB - Liar dataset

After the Naive Bayes, the Random Forest algorithm was applied. Again 70% of the data set is reserved for training and 30% for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of Random Forest 99.04%. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated. This one is the closest one to LSTM results.

Accuracy for Random Forest: 0.99

	Precision	recall	f1-score	support
Fake	0.99	0.99	0.99	4763
Real	0.99	0.99	0.99	4217
Accuracy			0.99	8980
Macro avg	0.99	0.99	0.99	8980
weighted avg	0.99	0.99	0.99	8980

Table 5.4: Accuracy for Random Forest Classifier-Liar dataset

After the Random Forest, the KNN algorithm was applied. Again 70% of the data set is reserved for training and 30% for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of KNN 85.35%. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for K-Neighbors Classifier: 0.85

	Precision	recall	f1-score	support
Fake	0.91	0.80	0.85	4763
Real	0.80	0.91	0.85	4217
Accuracy			0.85	8980
Macro avg	0.86	0.86	0.85	8980
weighted avg	0.86	0.85	0.85	8980

Table 5.5: Accuracy for K-Neighbors Classifier-Liar dataset

The KNN algorithm was applied and then the SGD algorithm was applied. Again 70 % of the data set is reserved for training and 30% for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of SGD 90.8 %. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for for SGDClassifier: 0.91

	Percision	recall	f1-score	support
Fake	0.88	0.95	0.92	4763
Real	0.94	0.86	0.90	4217
Accuracy			0.91	8980
Macro avg	0.91	0.91	0.91	8980
weighted avg	0.91	0.91	0.91	8980

Table 5.6: Accuracy for SGDClassifier-Liar dataset

After the SGD algorithm was applied and the Logistic regression algorithm was applied. Again 70% of the data set is reserved for training and 30% for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of Logistic regression 98.90%. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. This one is the closest one to LSTM results.

Accuracy for Logistic Regression: 0.99

	Precision	recall	f1-score	support
Fake	0.99	0.99	0.99	4763
Real	0.99	0.98	0.99	4217
Accuracy			0.99	8980
Macro avg	0.99	0.99	0.99	8980
weighted avg	0.99	0.99	0.99	8980

Table 5.7: Accuracy for Logistic Regression -Liar dataset

The accuracy, precisions, and f1 score can be computed with the help of confusion matrices. A single confusion matrix was created for each model. The values shown are the averaged values over successive trials. Confusion matrix of classification result:

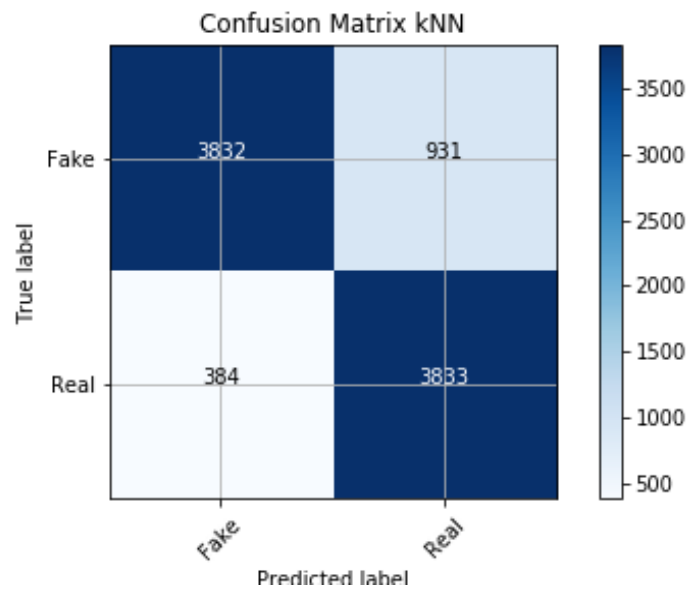


Figure 5.5: Confusion Matrix KNN-Liar dataset

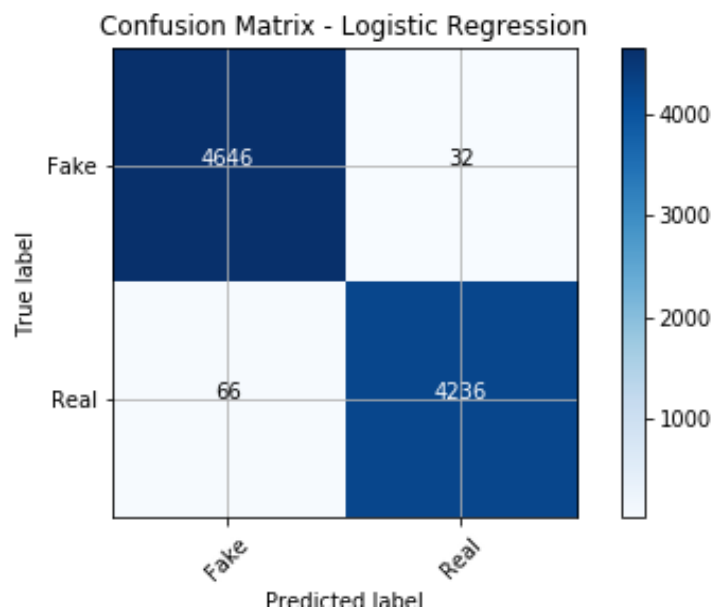


Figure 5.6: Confusion Matrix LR -Liar dataset

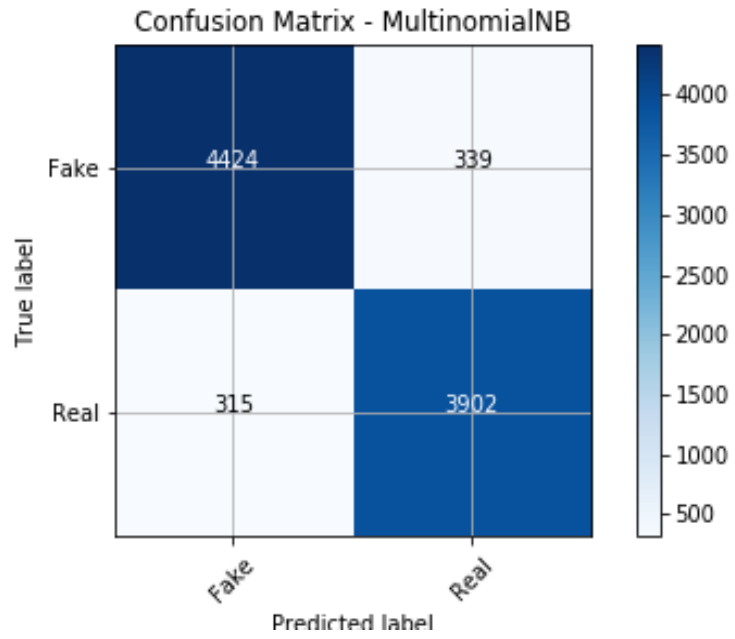


Figure 5.7 Confusion Matrix MNB -Liar dataset

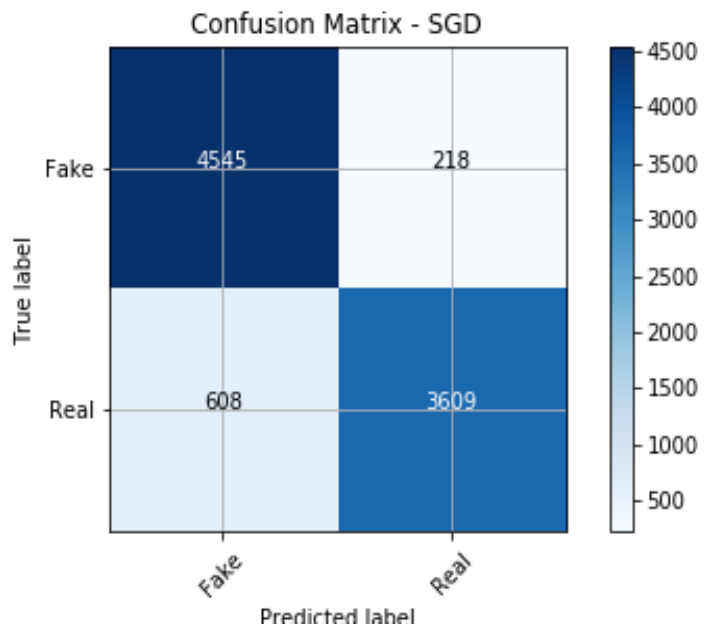


Figure 5.8 Confusion Matrix SGD -Liar dataset

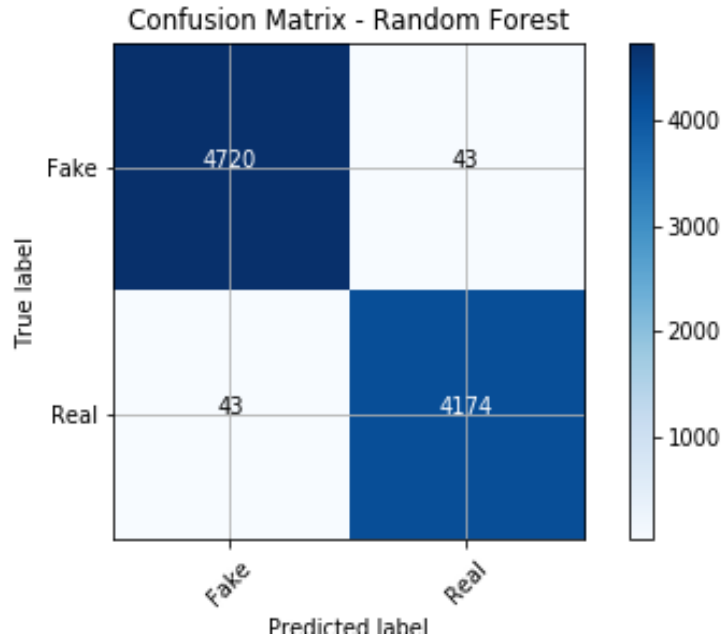


Figure 5.9 Confusion Matrix RF-Liar dataset

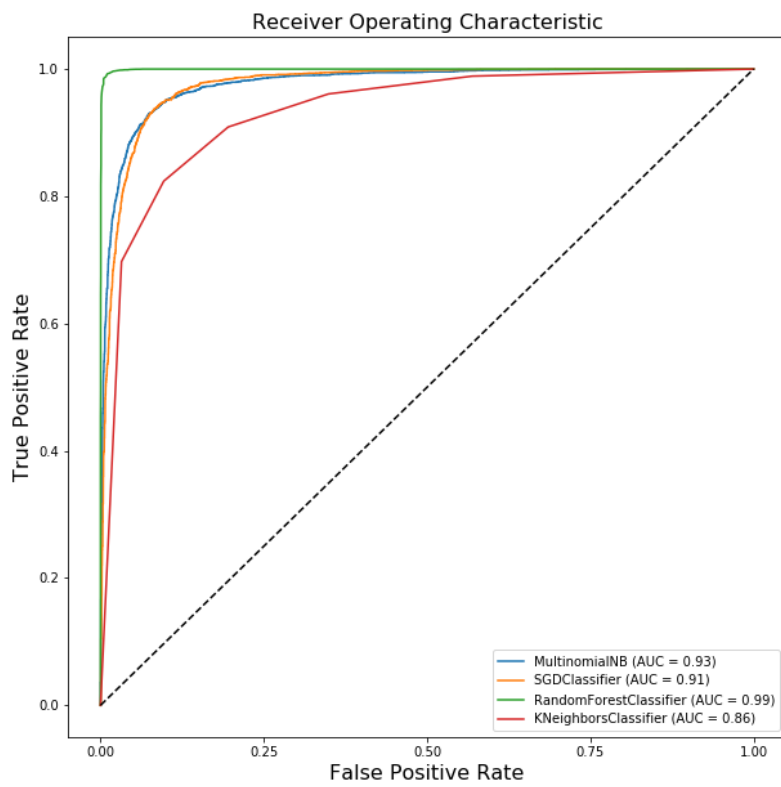


Figure 5.10. ROC -Liar dataset

Based on the results, constructed by taking different algorithms on X-axis and accuracy on the Y-axis. It is inferred that LSTM provides us with the highest accuracy followed by the Keras Neural Network Sequential model. LSTM provides such high results because the text is inherently a serialized object. RNN LSTM algorithm was applied. Again 70 % of the data set is reserved for training and 30 % for the test. Relu and Sigmoid activation functions are used together. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of LSTM 99.8 %. At the end of the classification, a confusion matrix was created.

Layer(Type)	Output Shape	Param#
Embedding	(None, None,128)	13914112
Bidirectional	(None,256)	263168
dense	(None,128)	32896
dense(1)	(None,1)	129

Total params: 14,210,305

Trainable params: 14,210,305

Non-trainable params: 0

Table 5.8. Sequential model -Liar dataset

Confusion matrix, without normalization
[[4757 6]
[5 4212]]

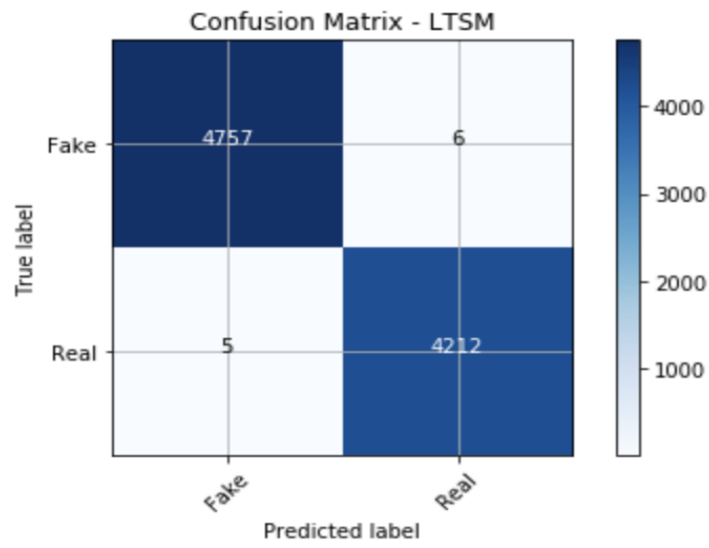


Figure 5.11. Confusion Matrix LSTM -Liar dataset

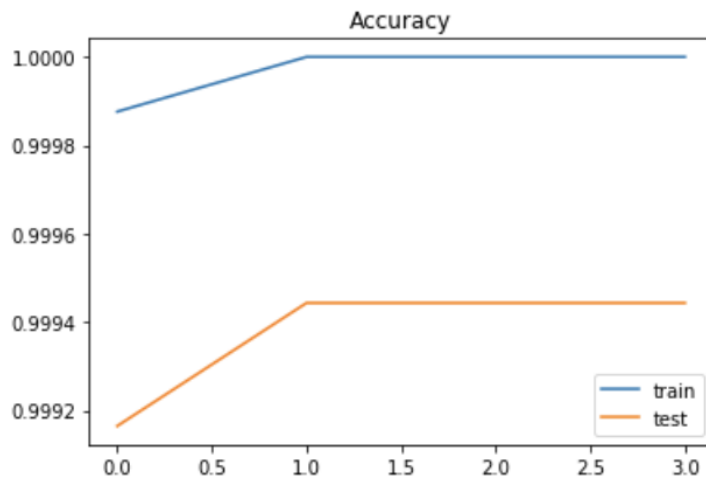


Figure 5.12. Accuracy of LSTM model

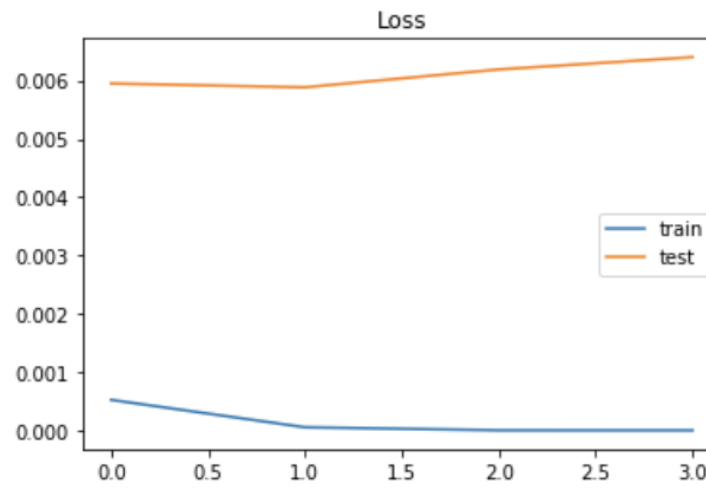


Figure 5.13. Model Loss/Cost-Entropy Loss of LSTM model

The classifications method was applied to the Reuters dataset. Dataset was split into a training set and a test set. %70 of the dataset used for training and %30 for testing. K-Fold cross-validation accepted as 5 for increasing the dataset variance. After preprocessing of datasets, classification algorithms were applied on a liar and Reuters datasets. Firstly, classical algorithms were applied, and then the results were obtained with deep learning methods. Classical methods are K-Nearest Neighbors, Random Forest, Naive Bayes, Stochastic gradient descent, Logistic Regression five different methods in Machine Learning and RNN, LSTM were used Artificial Neural Networks. Afterward, methods were applied, all these results were compared in the tables.

Model	Score
LSTM	86.533666
Random Forest	95.635910
Naive Bayes	92.144638
SGD	93.516209
KNN	62.219451
Logistic Regression	95.635910

Table 5.9. : Model and Scores

First, the Naive Bayes classification algorithm was applied to the Reuters dataset. 70 % of the dataset is reserved for training and 30% for the test. After training, the algorithm was tested for 30% test data and classification was completed with an accuracy of Naive Bayes 92.01 %. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for MultinomialNB: 0.92

	Precision	recall	f1-score	support
Fake	0.93	0.92	0.93	440
Real	0.90	0.92	0.91	362
Accuracy			0.92	802
Macro avg	0.92	0.92	0.92	802
weighted avg	0.92	0.92	0.92	802

Table 5.10: Accuracy for MultinomialNB - Reuters dataset

After the Naive Bayes, the Random Forest algorithm was applied. Again 70 % of the data set is reserved for training and 30 % for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of Random Forest 95.38 %. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for Random Forest: 0.95

	Percision	recall	f1-score	support
Fake	0.97	0.94	0.96	440
Real	0.93	0.97	0.95	362
Accuracy			0.95	802
Macro avg	0.95	0.96	0.95	802
weighted avg	0.95	0.95	0.95	802

Table 5.11: Accuracy for Random Forest- Reuters dataset

After the Random Forest, the KNN algorithm was applied. Again 70 % of the data set is reserved for training and 30 % for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of KNN 64.21 %. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for KNN: 0.64

	Precision	recall	f1-score	support
Fake	0.61	0.96	0.75	440
Real	0.85	0.25	0.39	362
Accuracy			0.64	802
Macro avg	0.73	0.61	0.57	802
weighted avg	0.72	0.64	0.59	802

Table 5.12: Accuracy for KNN - Reuters dataset

The KNN algorithm was applied and then the SGD algorithm was applied. Again 70 % of the data set is reserved for training and 30% for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of SGD 93.01%. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for SGD: 0.93

	Precision	recall	f1-score	support
Fake	0.95	0.92	0.94	440
Real	0.90	0.94	0.92	362
Accuracy			0.93	802
Macro avg	0.93	0.93	0.93	802
weighted avg	0.93	0.93	0.93	802

Table 5.13: Accuracy for SGD Classifier - Reuters dataset

After the SGD algorithm was applied and the Logistic regression algorithm was applied. Again 70% of the data set is reserved for training and 30% for the test. Initially, the number of trees was randomly assigned to 100. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of Logistic regression 96.0 %. At the end of the classification, a confusion matrix was created and precision, recall, f-measure, and support values were calculated. Additionally, micro, macro, and weighted average and K-Fold validation, ROC Curve also calculated.

Accuracy for Logistic regression: 0.96

	Precision	recall	f1-score	support
Fake	0.95	0.97	0.96	411
Real	0.97	0.94	0.95	391
Accuracy			0.96	802
Macro avg	0.96	0.96	0.96	802
weighted avg	0.96	0.96	0.96	802

Table 5.14: Accuracy for Logistic regression- Reuters dataset

After the accuracy, precisions, and f1 score were computed with the help of confusion matrices I created a single confusion matrix for each model. The values shown are the averaged values over successive trials. Reuter dataset confusion matrix of classification result:

```
Confusion matrix, without normalization
[[398  13]
 [ 22 369]]
```

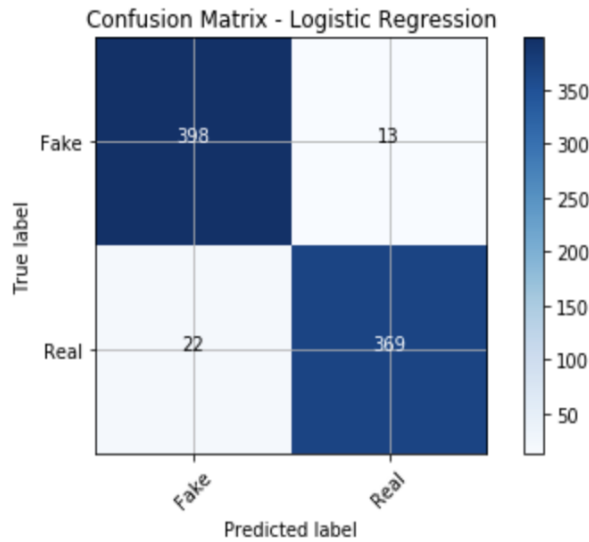


Figure 5.14: Confusion Matrix Logistic regression- Reuters dataset

```
Confusion matrix, without normalization
[[405  35]
 [ 29 333]]
```

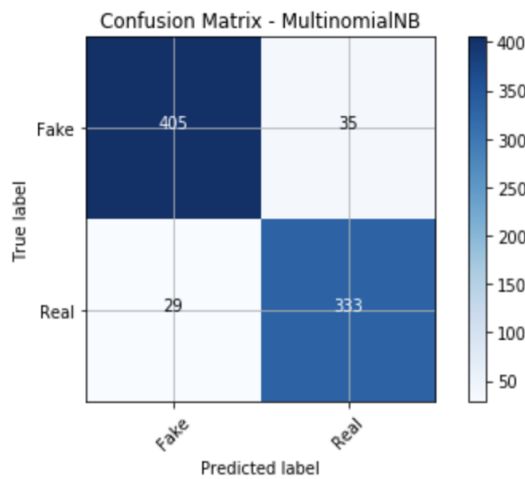


Figure 5.15: Confusion Matrix MNB - Reuters dataset

Confusion matrix, without normalization
[[404 36]
 [20 342]]



Figure 5.16: Confusion Matrix SGD Classifier - Reuters dataset

Confusion matrix, without normalization
[[414 26]
 [11 351]]

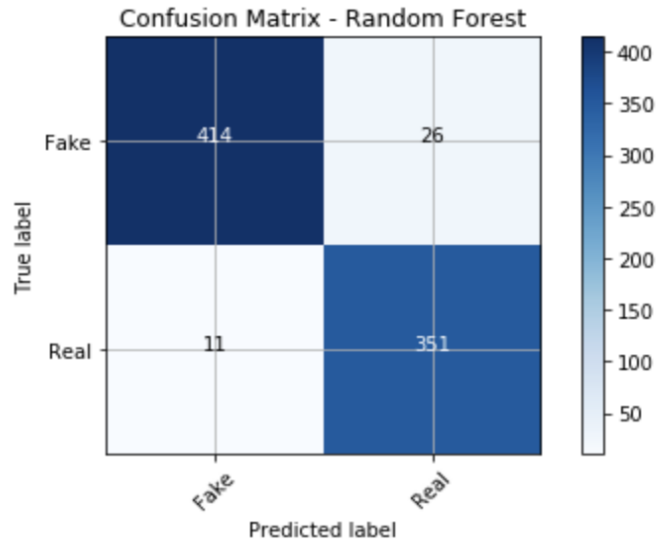


Figure 5.17: Confusion Matrix RF- Reuters dataset

Confusion matrix, without normalization
[[424 16]
[271 91]]

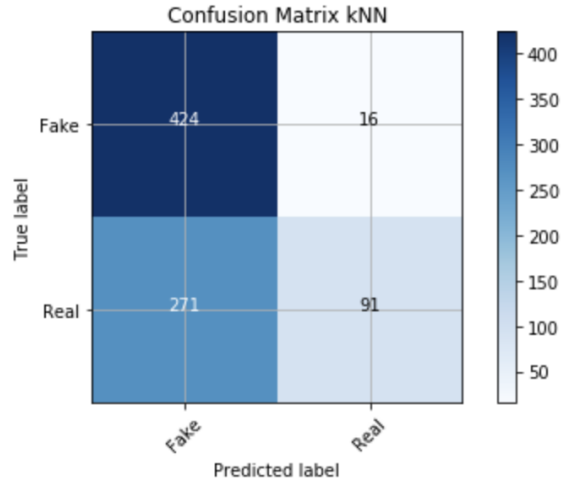


Figure 5.18: Confusion Matrix KNN- Reuters dataset

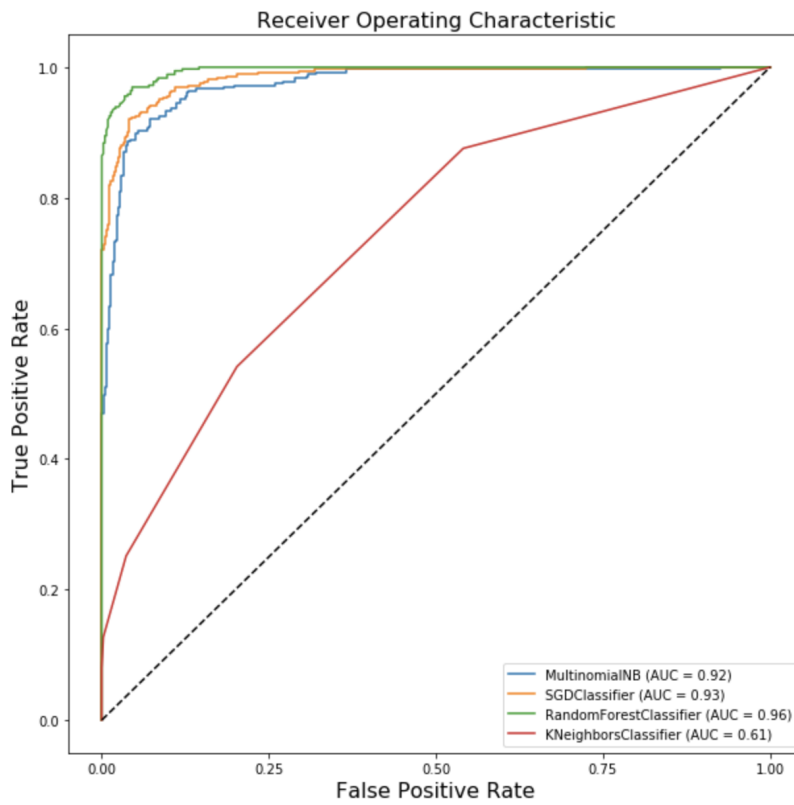


Figure 5.19: ROC Curve -Reuters dataset

Model	Score
Random Forest	0.959156
Naive Bayes	0.937641
SGD	0.926108
KNN	0.640474

Table 5.15. : K-Fold Validation - Model and Score Reuters dataset

Based on the results, constructed by taking different algorithms on X-axis and accuracy on the Y-axis. It is inferred that LSTM provides us with the highest accuracy followed by Keras Neural Network. I used a Sequential model. LSTM provides such high results because the text is inherently a serialized object. Bi-Directional RNN and LSTM algorithms were applied. Again 70% of the data set is reserved for training and 30% for the test. Also, Relu and Sigmoid activation functions are used together inside the Dense Layer. Then the model was applied on the train set and then tested. As a result of the test, the model reached an accuracy rate of LSTM 99.8%. At the end of the classification, a confusion matrix was created.

Layer(Type)	Output Shape	Param#
Embedding	(None, None,128)	5371904
Bidirectional	(None,256)	263168
dense	(None,128)	32896
dense(1)	(None,1)	129

Table 5.16: Sequential model Reuters dataset

Total params: 5,668,097

Trainable params: 5,668,097

Non-trainable params: 0



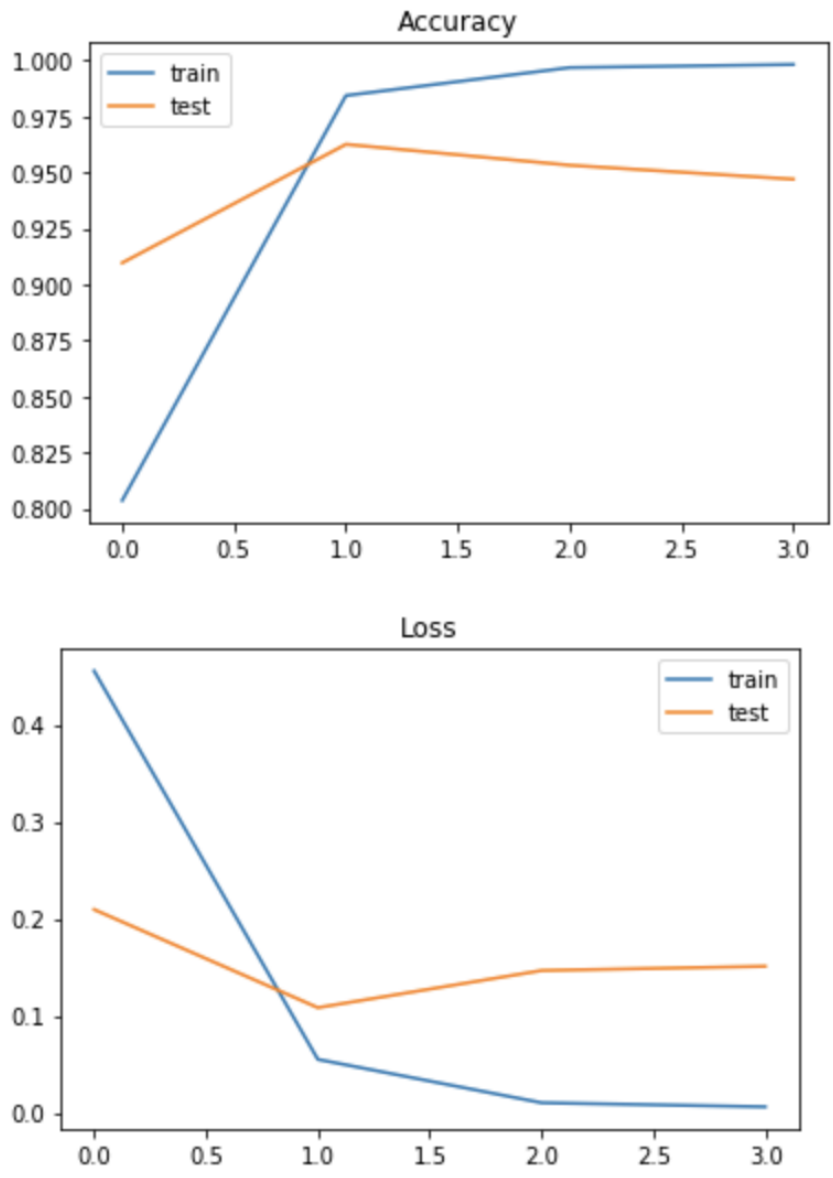


Figure 5.20: Accuracy and Model Loss/Cost-Entropy Loss of LSTM model

Confusion matrix, without normalization
[[430 10]
 [59 303]]

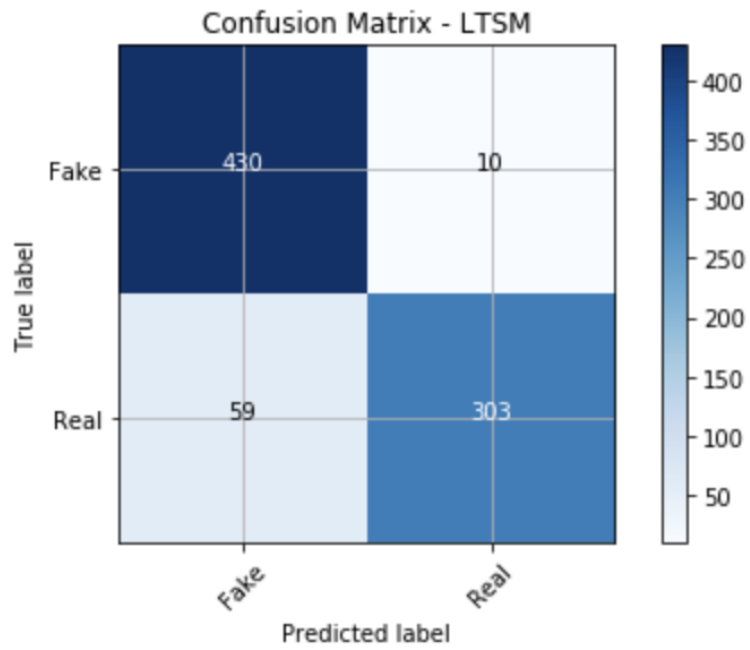


Figure 5.21: LSTM Confusion Matrix

5.2 Future Work

Studies have shown how important this issue is. It has the power and influence to directly concern the politics of the countries in which I studied my thesis, the economy and societies. The futures of countries and societies can be changed using artificial intelligence technologies. He emphasized the importance of the thesis and the importance of future issues in these days when the events of the digital coup are more and more. In the continuation of this thesis, it is planned to take the data from the fixed source only as variable and real-time. The developed model has the power to detect fake news in real-time. Along with great and exciting developments, the study is of advanced difficulty. Determining how fake news is fake is a multi-step process. While the social structure and history of countries can change whether a news is fake or not, it is necessary to improve the artificial intelligence model produced in this thesis and to strengthen its analysis capability. In the future, user profiles-based features can also be included for better prediction of news articles. An approach with multiple parallel channel-based deep convolutional neural networks of different kernel sizes can also be helpful to perform the classification of news articles. Fake news detection is an emerging research area that has a scarce number of datasets. It has been observed that basing fake news detection only on controlled models is not sufficient in all cases. To solve this problem, most research focuses on additional information such as author information. I think the most successful approach would be the automatic fact check model, that is to force the model with some kind of knowledge base, the purpose of the model will then be to extract information for the text and verify the information in the database. The problem with this approach is that the knowledge base has to be constantly and manually updated to stay up to date. There is no data on real-time news or regarding current affairs. The current model is run against the existing dataset, showing that the model performs well against it. In my future work, news article data can be considered related to recent incidents in the corpus of data. The next step then would be to train the model and analyze how the accuracies vary with the new data to further improve it.

CHAPTER 6

6.CONCLUSION

Fake news has multiple parameters that require a large amount of data and a demanding analysis process. It is necessary to pay attention to many factors during analysis. As I have shared before, in the studies conducted in this field, some studies have used data from very limited sources. The diversity of the news source shows us how authentic the data to be processed is. For these reasons, I used datasets that have been proven to be of good quality in international standards. I explained the part about the quality of the datasets in the dataset section. There are many definitions of even fake news. "According to what is fake?" Need to answer the question well. Before processing the data, I explained my research on this part that changes the direction of work in the fake news types section.

In this thesis, I have presented a model for fake news detection through different machine learning and deep learning techniques. Also, the article investigated the five Machine learning methods RNN, LSTM and compared their accuracy. In addition to these, the data preprocessing stages were meticulously done and the best methods were chosen to improve the results. Various procedures were applied to ensure that the model did not have vanishing gradient and overfitting problems. Since the Sigmoid activation function caused the vanishing gradient problem, the Relu activation function was also used and the error was resolved. Two different data sets were used in this study. It was explained how good the data sets used in the previous chapters are in the literature. The model that provides the highest accuracy of the LIAR dataset is LSTM and the highest accuracy score is 99.8%. However, in the Reuters dataset, which is also a quality dataset, we have seen that the Random Forest model is the method with the highest result with 95.38%. Despite the high performance of our classifier, there are areas for improvement. Evaluating our models using binary class datasets to predict the content of any news article as real or fake. My other focal point, echo chamber defined as a group of people who have the same understanding of social issues or direction would be to look for any news from the perspective of counterfeit detection, for example, political echo chamber. The primary motivation for including echo rooms is that any user is not isolated on any social media platform but exists in the form of a community. In terms of the future, focusing in-depth on the news-Send relations is an important area to be discovered, and it monitors

the inclusion of different echo chamber for better classification of news articles.

REFERENCES

Kai Shu, Amy Sliva, Suhan Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1):22–36, 2017

World Wildlife Fund (WWF) two images provided to show deforestation between two dates 2019 <https://factcheck.afp.com/these-two-pictures-do-not-show-same-rainforest-ten-years-apart>

As Fake News Spreads Lies, More Readers Shrug at the Truth, Sabrina Tavernise 2016 <https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html>

Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate, Craig Silverman et al., 2016

<https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis#.vn9qGBEY7>

Claire Wardle and Hossein Derakhshan. Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe Report, 27, 2017. <https://arxiv.org/pdf/2001.00623.pdf>

Liang Wu, Fred Morstatter, Kathleen M Carley, and Huan Liu. Misinformation in social media: Definition, manipulation, and detection. ACM SIGKDD Explorations Newsletter, 21(2):80–90, 2019. <https://arxiv.org/pdf/2001.00623.pdf>

Xinyi Zhou and Reza Zafarani. Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315, 2018. <https://arxiv.org/pdf/2001.00623.pdf>

7 types of misinformation in the German election Ingrid Brodnig,2017

<https://firstdraftnews.org/latest/7-types-german-election/>

Data Cleaning: Problems and Current Approaches (Erhard Rahm et al.,2020)

<https://www.coursehero.com/file/46312933/DataCleaning-Problem-and-Approachespdf/>

Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says, Gil Press,2016

<https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/?sh=735fbce6f637>)

Classification Model for Prediction of Heart Disease using Correlation Coefficient Technique - Sireesha Moturi 1, Dr. Srikanth Vemuru2 , Dr.S.N.Tirumala Rao 3

GeeksforGeeks, 2018 Data Cleansing | Introduction <https://www.geeksforgeeks.org/data-cleansing-introduction/>

Preprocessing Techniques for Text Mining Dr.S.Kannan 2014

Preprocessing Techniques for Text Mining - An Overview Dr. S. Vijayarani 2015

S.Jusoh and H.M. Alfawareh, Natural language interface for online sales, in Proceedings of the International Conference on Intelligent and Advanced System (ICIAS2007). Malaysia: IEEE, November 2007, pp. 224–228.

Text Preprocessing in Natural Language Processing , Harshith,2019

<https://towardsdatascience.com/text-preprocessing-in-natural-language-processing-using-python-6113ff5decd8>

Dropping common terms: stop words , 2009

<https://nlp.stanford.edu/IR-book/html/htmledition/dropping-common-terms-stop-words-1.html>

Evaluation of Preprocessing Techniques for Text Classification, Ammar Ismael Kadhim, 2018

Stemming and lemmatization, 2009

<https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html>

Neural Network Methods in Natural Language Processing (Synthesis Lectures on Human Language

Technologies) Paperback , Yoav Goldberg Page 92 April 17, 2017

Understanding NLP Word Embeddings — Text Vectorization ,Prabhu, 2019

Perception Management and Manipulation - Mücahit Gültekin/ Algı Yönetimi ve Manipülasyon, 2016

Beyond One-hot Encoding: lower dimensional target embedding Pau Rodríguez et al.,2018

Logistic Regression in Python, Mirko Stojiljković, 2020 <https://realpython.com/logistic-regression-python/>

Machine Learning Classification KNN, Ekrem Hatipoğlu, 2018 <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-k-nn-k-en-yakin-komsu-part-9-6f18cd6185d>

Stochastic Gradient Descent - <https://scikit-learn.org/stable/modules/sgd.html>

L. Deng, A tutorial survey of architectures, algorithms, and applications for deep learning, APSIPA Trans. Signal Inf. Process. 3 (2014)

What Deep Learning?,<https://www.mathworks.com/discovery/deep-learning.html>

Neural Networks, <https://www.ibm.com/cloud/learn/neural-networks>

Yapay Sinir Ağı(Artificial Neural Network) Nedir?, 2020

<https://www.veribilimiokulu.com/blog/yapay-sinir-agiartificial-neural-network-nedir/>

Yapay Sinir Ağları Nilgun Sengoz ,2017) <https://www.derinogrenme.com/2017/03/04/yapay-sinir-aglari/>

Recurrent Neural Network Nedir? Hamza Ergüder,2018

<https://medium.com/@hamzaerguder/recurrent-neural-network-nedir-bdd3d0839120>

Understanding RNN and LSTM Aditi Mittal 2019

<https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e>

The Vanishing Gradient Problem [Chi-Feng Wang](#), 2019

RNN (Recurrent Neural Network) Nedir ve Nasıl Çalışır? , Azer Çelikten,2020
<https://medium.com/@azercelikten/rnn-recurrent-neural-network-nedir-ve-nasil-calisir-d1246b1a61fb>

Comparison of Activation Functions for Deep Neural Networks Ayyüce Kızrak, 2019
<https://towardsdatascience.com/comparison-of-activation-functions-for-deep-neural-networks-706ac4284c8a>

Stehman, Stephen V. (1997). "Selecting and interpreting measures of thematic classification accuracy". *Remote Sensing of Environment*. **62** (1): 77–89. Bibcode:1997RSEnv..62...77S. doi:10.1016/S0034-4257(97)00083-7.

Powers, David M. W. (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". *Journal of Machine Learning Technologies*. **2** (1): 37–63. S2CID 55767944.

What is the best metric precision recall f1 and accuracy to evaluate the machine learning model for imbalanced data , Ibrahim mohamed Gad, 2020

(https://en.wikipedia.org/wiki/Receiver_operating_characteristic)

Cross Validation, 2017

<https://veribilimcisi.com/2017/07/13/capraz-dogrulama-cross-validation-nedir/>

The Reuters Dataset ,Martin THOMA 2017 <https://martin-thoma.com/nlp-reuter>

Experiments in Open Domain Deception Detection, Ott et al., 2011; Pérez-Rosas & Mihalcea, 2015

"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection ACL 2017 • William Yang Wang,