DECEMBER 2019 M.Sc. – Electronics and Computer Engineering MERVE IŞIK

T.C.

HASAN KALYONCU UNIVERSITY

GRADUATE SCHOOL OF

NATURAL & APPLIED SCIENCES

SOCIAL MEDIA TEXT CLASSIFICATION FOR CRISIS MANAGEMENT

ELECTRONICS AND COMPUTER ENGINEERING

M. Sc. THESIS

MERVE IŞIK

DECEMBER 2019

SOCIAL MEDIA TEXT CLASSIFICATION FOR CRISIS MANAGEMENT

Hasan Kalyoncu University

Electronics and Computer Engineering

M. Sc. Thesis

Supervisor

Assist. Prof. Dr. Saed ALQARALEH

Merve IŞIK

December 2019

GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES INSTITUTE
MSc ACCEPTANCE AND APPROVAL FORM

Electronics-Computer Engineering M.Sc. (Master Of Science) programme student **Merve IŞIK** prepared and submitted the thesis titled "**Social Media Text Classification For Crisis Management**" defended successfully on the date of 26/12/2019 and accepted by the jury as an M.Sc. thesis.

Position	**Title, Name and Surename** **Department/University**	**Signature:**
M.Sc. Supervisor **Jury Head**	Assist. Prof. Dr. Saed ALQARALEH Computer Engineering Department Hasan Kalyoncu University	
Jury Member	Assoc. Prof. Dr. Ahmet Mete VURAL Electrical and Electronics Engineering Department Gaziantep University	
Jury Member	Assoc. Prof. Dr. M. Fatih HASOĞLU Computer Engineering Department Hasan Kalyoncu University	

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

Prof. Dr. Mehmet KARPUZCU
Director

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.

Merve IŞIK

ABSTRACT

SOCIAL MEDIA TEXT CLASSIFICATION FOR CRISIS MANAGEMENT

IŞIK, Merve

M.Sc. in Electronics and Computer Engineering

Supervisor: Assist. Prof. Dr. Saed ALQARALEH

Dec 2019

84 pages

In recent years, impressive attention has been given for mining the publically available huge amount of data to gain situational awareness, which may help in preventing or decrease the effect of some disaster by taking the correct responses. In this study, an effective Convolutional Neural Networks (CNN) tweet classification system that fully supports the Turkish language has been developed.

In addition, the first-ever Turkish tweet dataset for crisis response is created. This dataset has been carefully preprocessed, annotated, well organized and suitable to be used by all the well-known natural language processing tools. Furthermore, the performance of some well-known machine learning algorithms, i.e., K-Nearest Neighbor (KNN), Naive Bayes (NB), and Support Vector Machine(SVM) was investigated. Then, the performances of the ensemble systems Random Forest (RF), AdaBoost Classifier (AdaBoost), GradientBoosting Classifier (GBC), when used for text (tweets) classification, has been also observed.

A wide range of experiments was performed to investigate the performance of the developed system. As a result, the developed approach has achieved very good performance, robustness, and stability when processing both Turkish and English languages.

Key Words: Crises Management Systems; Tweet Classification; Turkish language; Convolutional Neural Networks; Natural Language Processing.

ÖZET

KRİZ YÖNETİMİ İÇİN SOSYAL MEDYA METİN VERİLERİNİN SINIFLANDIRILMASI

IŞIK, Merve

Yüksek Lisans Tezi, Elektronik Bilgisayar Mühendisliği

Tez Yöneticisi: Dr. Öğretim Üyesi Saed ALQARALEH

Aralık 2019

84 sayfa

Son yıllarda bazı felaketlerin etkilerini önlemeye veya azaltmaya yardımcı olmak için durumsal farkındalık sağlamak amacıyla, herkesin erişimine açık olarak bulunan büyük miktarda veri üzerinde veri madenciliğine büyük önem verildi. Bu çalışmada, Türk dilini tam olarak destekleyen etkin bir Evrişimsel Sinir Ağları (CNN) tweet sınıflandırma sistemi geliştirilmiştir.

Ayrıca, kriz yanıtına yönelik ilk Türk tweet veri seti oluşturulmuştur. Bu veri seti dikkatlice önceden işlenmiş, açıklamalı, iyi organize edilmiş ve iyi bilinen tüm Doğal Dil İşleme araçları tarafından kullanılmaya uygundur. Ayrıca, bazı iyi bilinen makine öğrenme algoritmalarının, örneğin K-En Yakın Komşu (KNN), Naif Bayes (NB), Rastgele Orman (RF), AdaBoost Sınıflandırıcı (AdaBoost) ve GradientBoosting Sınıflandırıcı (GBC) algoritmalarının metin (tweet) sınıflandırması konusundaki performansını araştırmak için deneyler yapılmıştır. Ardından, Rastgele Orman (RF), AdaBoost Sınıflandırıcı (AdaBoost) ve GradientBoosting Sınıflandırıcı (GBC) topluluk (ensemble) sistemlerinin metin sınıflandırması konusundaki performansları da gözlenmiştir.

Geliştirilen sistemin performansını ve seçilen makine öğrenme algoritmalarını araştırmak için geniş bir deney yelpazesi yapıldı. Sonuç olarak, geliştirilen yaklaşım hem Türkçe hem de İngilizce dillerini işlerken çok iyi performans, sağlamlık ve istikrar elde etti.

Anahtar Kelimeler: Kriz Yönetim Sistemleri; Tweet Sınıflandırması; Türk Dili; Evrişimsel Sinir Ağları; Doğal Dil İşleme.

To My Family…..

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude and gratitude to my supervisor Assist. Prof. Dr. Saed ALQARALEH, who did not hesitate to share all of his knowledge with me during this study, who supported me on all kinds of issues and who contributed a lot to my thesis and who also contributed a lot to me.

I would like to express my thanks to the members of the ELECTRONICS & COMPUTER ENGINEERING department for continuous encouragement, invaluable suggestions and support at all the stages of my thesis work.

I would also like to thanks to my jury members Assoc. Prof. Dr. Muhammet Fatih HASOĞLU and Assoc. Prof. Dr. Ahmet Mete VURAL for their valuable participance and support. Finally, grateful thanks to my family.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AdaBoost	AdaBoost Classifier
ANN	Artificial Neural Networks
CBOW	Continuous Bag of Words
CMS	Crisis Management System
CNN	Convolutional Neural Networks
GBC	Gradient Boosting Classifier
KNN	K-Nearest Neighbor
ML	Machine Learning
NB	Naïve Bayes
RELU	Rectified Logical Unit
RF	Random Forest
SVM	Support Vector Machine
TF	Term Frequency
TF-IDF	Term Frequency-Inverse Document Frequency

CHAPTER 1

INTRODUCTION

Crisis situations are situations that people may encounter throughout their lives. As in the past, we still often face crisis situations such as disasters and accidents. The damage caused by these situations varies according to the severity of the disaster or accident. Occasions are sometimes effective in large areas, sometimes in small areas. Accidents and disasters, which can occur in many types and forms, can sometimes cause considerable loss of life and material losses. It is very important to take measures to reduce these losses as much as possible and to provide first aid at the time of the disaster or event.

Social media generally refers to online communication between users through online sharing platforms such as social networking sites. It plays an important role on human communication.

In recent years, social media channels have come to the point of being everywhere in our daily lives. Platforms such as Twitter, Facebook, İnstagram are playing a crucial role in sharing information quickly. It will not be a surprise that the social media's data repository is popular in terms of information mining, as it contains valuable information for many fields such as crisis monitoring applications and even marketing may arise to track the perception of a new product. Moreover, it is possible to use social media data to immediately detect a situation that requires an action from a recovery team. For example, during natural disasters, social media can have a crucial role on emergency responses. However, in order to collect the necessary information related to the disaster, it is necessary to analyze the data on social media effectively.

Crisis Management System is a system which aims to detect crises early and respond to events erarly and effectively and provide a safer and more comfortable environment for people affected by the crisis. Our aim in this study is to identify the situations that require help or take precautions on social media and to perform or facilitate the necessary activities. In other words, to create a system that will enable Crisis Management through social media.

1.1 Problem Statement and Motivation

Crisis situations have existed since ancient times and cause negative effects on people's lives. The existence of crisis situations brought the requirement for managing crises in order to combat the crises and minimize the damages. From past to present, various crisis management policies have been pursued and different crisis management systems were created, and these systems have had various successes and failures. Mainly, current crisis management systems can be divided into three categories.

The first one, like the 911, is a very efficient and long-lasting calling system that has saved many people's lives. However, its efficiency is significantly decreased during natural or human-made disasters, as a very large number of calls can be received during such cases, and in order to process these calls in time, we need a large number of employees. Hence, sending these employees to the location is more important and has the priority.

The second category is the manual annotation systems that use some mobile applications, websites and social media to provide information. This system have been used as a result of the availability, easy access and a large number of users. Until this moment, such a system is not frequently used as compared to the first type. This due to the fact that it is totally depending on the human interpretation (as we may have some relevant (disastrous) as well as irrelevant (non-disastrous) information where the non-disastrous could be gossip, rumor, joke or a movie review or something else. In addition, this process can take a long duration, which delays the response time and increase the effect and the damage of the incident. The manual annotation system exists in many studies such as (Nguyen et al., 2016) and (Imran et al., 2016).

The third one is the Automatic Annotation System. With the impressive improvement in the classification and clustering techniques, where some techniques such as CNN was able to achieve a performance similar to the human level, some automated CMSs have been developed. Hence, impressive attention has been given for mining the publically available huge amount of data to gain situational awareness, which may help in preventing or decreasing the effect of some disaster by taking the correct responses. Briefly, this type was developed to overcome the weakness of the second type and work on dividing a large amount of available information into relevant

and irrelevant. In other words, try to detect only the information, say the tweets that related to a specific disaster. This process decreases the human offer required to process the date and inform the people in charge about only the most relevant information. The main problem of this category is that the majority of such systems can work with the English language and very few have been built for other languages such as chines.

Up to our knowledge, there is no such system that can support the Turkish language. Nevertheless, the fact that the structure of the Turkish language is different from English and other languages. In addition, there is no existed any data set for the crisis management system in the Turkish language. Hence, it is difficult to carry out such studies in the Turkish language.

It is worth mentioning that some of the existing approaches can use sensors to identify disaster, and such a system can be considered as a fourth category of CMSs. For instance, sensors such as seismic sensors are widely used in early detection of the earthquake, tsunami, etc. These sensors provide information about the estimated risks and generate early awareness.

Given the disasters and its consequences, is it possible to minimize the damages caused by these disasters and accidents using social media data as much as possible?

The use of social media has almost become a part of our lives. When people are happy, sad, or in other emotions, they share this with their followers on various social media platforms. In addition, they can easily express their ideas about an event through these platforms. Many artists, institutions and organizations are actively using social media. News sites often share their social media accounts as well as their own websites. Given these situations, social media becomes an important repository of data, and studies can be used to identify situations that may or may not require assistance. Thus, by detecting these situations early, various institutions or teams can take action and respond to events more quickly.

The main aim of this study is to improve the Turkish text classification and to build an efficient classification system that can establish a situational awareness by detecting

the text related to the disasters and informing the necessary institutions and organizations ASAP.

1.2 Scope of the Research

This study aims to determine the crisis situations by collecting social media data on Twitter in case of crisis such as disaster and accident. In doing so, we are aiming at achieving optimal results by testing the effectiveness of different machine learning algorithms and deep learning methods for data classification.

1.3 Objectives of the Research

This study primarily aims to collect data on social media and identify situations such as disasters and accidents. For this purpose, the following studies have been carried out:

- Determination and comparison of classification performances on text data of various state of the art machine learning algorithms such as KNN, Random Forest, Ada Boost, etc.
- Classification performance of ensemble systems such as Random Forest, Ada Boost, etc., as compared to ML algorithms' performance.
- Measurement of classification performance on the text data of the system created using CNN, which is a deep learning technique.

The performance of the above-mentioned machine learning algorithms and using deep learning technique are compared using various criteria. These criteria are:

- Investigating the effect of preprocessing vs using the row data on the overall classification performance.
- Investigating the effect of changing the number of samples used for training and testing the system.
- Investigating the performance of state of the art method for extracting features from text.

- Investigating the effect of changing CNN parameters such as the window size and the number and size of the used filters, etc. on the overall classification performance.

In the course of these studies, text data in English and Turkish related to disaster and accident situations collected from Twitter were used.

1.4 Research Contributions and Importance of the Study

Disasters always have an impact on human beings and the destructive effects of disasters have often been serious. Just in 2009, a total of 335 natural disasters occurred worldwide, resulting in material, moral and economic losses. As a result of the publicly documented disasters occurred in 2009, 10.655 people were killed, more than 119 million people were affected by the disasters and economic losses of more than US $ 41.3 billion occurred (Vos et al., 2010).

It is obvious that social media platforms have become very popular especially in recent times and consequently it has become a very important data source. On these platforms, users share a lot of information, especially at the moment that events occur or sometimes just before an event occurs. As soon as many events take place, they are included in the trend topic list in a short time. This has enabled social media to have a very important place in the detection and management of disasters. Considering the statistics mentioned above, we aimed to reduce the damages that may arise by identifying and managing disasters through social media. The proposed classification crisis management system aims to support both English and Turkish languages.

The findings of this study provide a suitable resource especially for researchers who are interested in implementing, discussing and investigating the performance of machine learning and deep learning methods for Turkish.

1.5 Limitations

Tweets related to disasters in English and Turkish language were used in this study. While there is a ready dataset in English, there is no dataset available in Turkish. For this reason, tweets in Turkish language about disaster types were collected by using Twitter api and data set was created for the study. These data were then interpreted by three voluntary decision makers. It took quite a long time. In addition,

since there were not many studies on this subject in Turkish, we did not have much chance to get an idea or to make comparison with previous studies.

1.6 Organization of the Study

The remainder of this study is structured as follows: The second section examines crisis situations and past crisis management systems and covers general information on crisis management through social media. The third section covers the literature review of the studies on crisis management and various data classification methods on social media. The fourth chapter explains briefly the machine learning techniques. In chapter five, according to the system proposed, the crisis management system is implemented. The results of the experiments are presented in the sixth chapter. Finally, the conclusions and future works are presented in the seventh chapter.

CHAPTER 2

DISASTER MANAGEMENT WITH RESPECT TO SOCIAL MEDIA

2.1 Conceptual Disaster Management

The word “disaster” refers to situations that occur suddenly in human life and/or in nature, which is in most cases have a major destructive effect and adversely affect the normal flow of life. Its origin comes from Arabic and it is defined in dictionaries of various languages.

Many definitions of the “disaster” phenomenon have been made in the literature. According to Ergünay, disasters are events that arise from natural and technological reasons, dividing the natural flow of human life and causing physical, economic and social losses. (Ergünay, 2002). Develioğlu defines disaster as a great problem and destruction (Develioğlu, 1978).

The concept of crisis management consists of coordinated rescue, preparedness, mitigation and flexibility efforts planned by governments, voluntary organizations or various local departments (Petak, 1985). The reason why a crisis is important is that it occurs suddenly and in a way that disrupts normal flow and is often uncontrollable. Operational time in crisis management includes the efforts made before, during and after the crisis. A crisis situation attracts the attention of the public and the media and threatens the safety of the public. This sometimes causes changes in public policies. A crisis can trigger rapid changes in public policy as it attracts the attention of the public and the media and threatens the public's trust (Alexander, 2005).

2.2. Disaster Management and Stages

A disaster is, in short, a loss caused by an event or danger. These negative consequences may be physical, social or economic. Disaster management aims preventing disasters or at least reducing post-disaster damages. Therefore, studies should be carried out in order to know the causes of the disaster and how to reduce the damages that may occur after the disaster.

Disaster management approaches have different perspectives, including pre-disaster, during disaster and post-disaster. As much as post-disaster interventions, the importance of pre-disaster preparations in reducing disaster damage has been understood and this concept has been developed (Akdag, 2002). In addition, disaster events can be grouped under 5 main headings. These; damage reduction, preparedness, recovery and first aid, recovery, rebuild construction stages. Two phases of disaster management system (damage reduction and advance preparation) before disasters, three of them (rescue and first aid, recovery), reconstruction) when and after disasters (Şemgün, 2007; Kurita, 2004).

2.3 Worldwide Disaster Management Development Process

2.3.1 Development of Disaster Management Process in Turkey

The information about the main development in the Crisis management policies implemented in Turkey are summarized as follows:

In early 1943, a new law that is named as "The reduction law (No. 4373)" was adopted for continuous flooding (Çorbacıoğlu et al., 2005), (Ganapati, 2008). During the 1944-1958 period, earthquakes with destructive effects took place, so the government had to adopt laws on disaster management and develop new mitigation policies. During this period, some proactive disaster laws were enacted (Ganapati, 2008). The first proactive law that takes effect before and after earthquakes and called "Measures to be put into effect prior and after earthquakes" (Law no. 4623). In this law, earthquake risks are defined and the necessity of developing intervention and aid programs and the necessity of geophysical investigation for new settlements are included. In the same period, the Ministry of Public Works prepared the first seismic risk map and identified the compulsory earthquake zones (Çorbacıoğlu et al., 2005). The law required the renewal of the seismic data of the Kandilli Observatory.

In 1958-1999, a general law was enacted covering all measures and interventions for all types of disasters. The previous law with law number 4623 covered only interventions and measures for earthquakes, did not include other natural disasters such as floods, landslides and fires. In 1959, a new law called "Measures and assistance to be implemented in relation to natural disasters affecting the lives of the

general public" was enacted and a legal arrangement had provided for several disaster types (Unlu et al., 2010).

Looking at the policies of 1999 and beyond, it is clear that the 1999 earthquake was very effective in these policies. 1999 earthquake, was one of the most violent and destructive in the history of earthquakes in Turkey. Therefore Grand National Assembly of Turkey, gave an exclusive competence of the government to increase measures. Provincial and regional managers have the authority to prepare their own emergency plans (Çorbacıoğlu et al., 2006). In addition, a set of emergency calling numbers have been publicly available and can be called whenever it is needed. Some of them are 112 (Ambulance), 110 (Fire Department) and 155 (Police).

2.3.2 Disaster Management Development Process in Other Countries

1) US: In the United States, the coordinating agency for emergency and disaster management is the United States Federal Emergency Management Agency (FEMA) (Erkal et al., 2009). In addition, other institutions and organizations such as the Federal Government, International resources, Volunteer organizations, Local Government, the United States Government, and various organizations from the private sector are part of disaster management. The rules for the interoperability of institutions are set out in accordance with the Federal Response Plan, which includes 12 emergency functions. These functions include fire fighting, medical services, debris removal, food aid providing.

There is also a "National Earthquake Hazard Reduction Program (NEHRP)' organized under the earthquake hazard mitigation law in the USA. NEHRP is in close contact with four national institutions. These are the Federal Emergency Management Authority (FEMA), the US Geological Survey (USGS), the National Science Foundation (NSF), the National Institute of Standards and Technology (NIST). This program is designed to 1) determine the extent to which settlement and investment areas are vulnerable to earthquake threat, 2) determine the seismic design and building standards, 3) improve earthquake prediction capacity, and 4) educate state governments, business and public on these issues.

2) Japan: Japan is a country with a high probability of tsunami and major earthquakes. Therefore, many studies have been carried out to establish a system called Ocean

Bottom Seismic Sensor System (OBS). With this system, it is aimed to detect seismic fluctuations and tsunami waves that may occur and take precautions. In general, Japanese National Emergency Management Model consists of:

a) National Government level: the Central Disaster Prevention Council convened under the chairmanship of the Prime Minister is obliged to prepare disaster prevention plans and to make the general planning regarding the preparations.
b) Regional Government level: which is responsible for the coordination, implementation and expansion of operations. It chairs the Regional Disaster Prevention Council. The Regional Disaster Prevention Council conducts planning activities in accordance with the general decisions and framework at national level.
c) Municipal Level: it is responsible for preparing and carrying out all kinds of preparations in case of a disaster. The municipality applies to all units with organizational structure. The Municipal Disaster Prevention Council, under the chairmanship of the mayor, is responsible for disaster preparedness and taking precautions in it.
d) People - Individual level aims to make the community resilient and prepared for disasters, to be coordinated with neighborhood organizations and other voluntary organizations.

3) Germany: According to the constitution of Emergency Management in Germany, the duty of the state government is to provide assistance in peace-time disasters, and the duty of the Federal Government is the protection of civil society in case of war. The organizational structure is the inter-ministerial co-ordination group within the Ministry of the Interior, state government, regional or city administration managers, emergency personnel, rescue services, Federal Institute of Technical Assistance (THW), specialized special rescue services, fire brigade and other organizations.

Fire services are responsible for fire fighting and rescue, as well as medical assistance and food distribution to the community. Volunteers are an integral part of civil protection and work at every stage.

The Federal Government assists states with fire protection, health and economic assistance, protection against NBC threats, purchasing special equipment,

providing equipment to volunteers, and training. The Federal Institute for Technical Assistance (THW) was established for rescue services. The Federal Border Police (BGS) provides helicopter support for relief efforts and forms part of Germany's air rescue system. Local authorities can also call the Federal Army for help if necessary. The Federal Government assists states with fire protection, health and economic assistance, protection against NBC threats, purchasing special equipment, providing equipment to volunteers, and training. The Federal Institute for Technical Assistance (THW) was established for rescue services.

4) France: The Directorate for Public Safety is responsible for life and property safety, environmental protection, risk reduction in all kinds of accidents, disasters and disasters. The Directorate manages the national emergency service in public safety management and coordinates local rescue services in relief operations. Operation center CODISC is on duty 24 hours a day to carry out large-scale rescue operations at national level. Responsible to the Ministry of Interior and other state institutions in case of accidents and disasters. Inter-regional centers were established for coordination during the operation. In each region, coordination is provided under the authority of the region responsible.

The task of ensuring public security was shared by local governments and the state in a complementary manner. The Mayor and the District Governorate are responsible for the organization of the mitigation, relief and rescue. The governorship applies the regional plan or other special plans, if any. Governorships ensure interregional harmony for economic assistance, civil defense or protection. Daily public safety is provided by professional and voluntary firefighters.

2.4 Social Media And Crisis Management Process

2.4.1 Crisis Concept

The word crisis is a concept that passes from Greek to medical literature and then to social sciences. It comes from the concept of krisis, which means decision in Greek; In the medical literature, an acute and decisive period in a disease process is a period in which the symptoms are intensified (Hatemi, 1999). In other words, “concept”, which is based on the Greek word “Yunanca Krisis” means decision, which defined as any event that has the potential to affect the integrity and integrity of an

organization. In this respect, it is clear that there are different definitions of the concept of crisis.

2.4.2 Crisis management and Crisis Management Process

Effective crisis management planning is necessary for the successful recovery of crisis situations, crisis management planning, and the establishment of responsible areas and necessary principles in the organization. Crisis management planning consists of a list of potential crisis situations, establishing crisis prevention policies, developing strategies, and tactics to deal with the potential crisis to determine the impact of the crisis. In order to minimize damage channels, it is necessary to communicate effectively with those affected by the crisis (Regester et al., 1997), (Mitroff et al., 1987).

In this context, the crisis management process can be described as "the process in which the organization has to manage and coordinate this situation when it is faced with large incidents that threaten to harm itself and its stakeholders". What is important in the crisis management process is to identify the risks and the types of crises that these risks can create, because different strategies are needed to solve the types of crisis that can create different risks. Based on (Coombs et al., 2006), the first steps that come to mind is the moment of crisis and post-crisis phases. However, an important step to be involved in the crisis management process is the preparation for the crisis, i.e. the preparation for what to do before the crisis.

Augustine (Augustine, 1995) describes the crisis management process in six phases: avoiding the crisis; preparing to manage the crisis; fixing the crisis; freezing the crisis; resolving the crisis and benefiting from the crisis. It is possible to summarize these steps as follows:

A) Avoiding crisis: The simplest and least efficient way to control a potential crisis is taking action. Managers can see crises as an inevitable situation every day, but it is possible to avoid some crisis situations by taking precautions. A situation that has been overlooked or ignored can suddenly become a difficult situation for the organization. It can be said that this stage is a stage that requires attention form organizations.

B) Preparing to manage the crisis: Crisis preparedness activities should be initiated when a crisis prevention activity does not work. The preparatory activities are of great importance. Every person who is in a position to be in a ready-made position during the crisis management process should see and plan for the inevitability of a crisis in the same way as the inevitability of a death or tax.

C) Fixing crisis: Organizations can sometimes focus on technical aspects and ignore perception. However, the situation that causes the crisis is mostly perception of society and this perception can become reality after a while. Organizations that accept the existence of the crisis can carry out the next stages in a more effective and rapid form.

D) Freezing the crisis: When faced with a crisis situation, difficult decisions should be taken and this should be done immediately; the term "bleeding" specified in medical science language should be stopped immediately.

E) Solve the crisis: The crisis occurs suddenly, it won't wait, so it is vital to be quick. Action must be taken immediately to resolve the crisis situation.

F) Benefit from the crisis: All organizations must investigate and study disasters after going back to normal situations. This can increase the experiences and the knowledge about how to deal with such case in the future.

2.4.3 Social Media and Crisis Management

Social media is a form of communication that allows any person to communicate instantly with others anywhere in the world without time and space limitation. According to Kaplan and Haenlein's definition, social media is the name given to the whole of the internet-based applications that allow ideological and technological content to be produced and developed in a user-centric manner on the web 2.0 (Kaplan et al., 2010). According to Safko, the practice of sharing information and ideas through interactive media, which can be described as web-based applications that enable the creation and sharing of words, images, videos and sounds among online groups, brings social media to mind (Safko, 2010).

2.4.4 Social Media and Internet Platforms

2.4.4.1 Blogs

Blogs, a well-known social media platform, are websites that allow content sharing via online media and follow the sharing of verilen bloggers (Mayfield, 2008). Blogs and social networking sites can provide insight into any community with a significant online community, especially in communities with a young population (Mayfield, 2008).

A blog is a personal magazine where people produce and publish their content on any topic they want. In blogs, readers can comment on shared content, so that they can establish an inter-user dialogue.(Özkaşıkçı, 2012). Each user can easily access the blogs and the user's text, images, audio files, and so on. Hence, it does not require a technical knowledge to be create. Blogs have emerged as part of a virtual gathering of people from all over the world who share a common interest in a particular topic.

2.4.4.2 Forums

The discussion site Usenet, founded in 1979 by Tom Truscott and Jim Ellis at Duke University, is considered as the first electronic forum in the history. Forums can be defined as modern versions of bulletin boards. Nowadays, thousands of forum sites that focusing on many issues are available (Başer, 2014). Forum sites usually work with the membership system and by interest they're grouped (Nusair et al., 2012).

2.4.4.3 Content Communities (Flickr, YouTube etc.)

The main purpose of social content communities is to share content among users. Content can be: collections text (e.g. BookCrossing), picture (e.g. Flickr), video (e.g. YouTube)), PowerPoint presentations (e.g. slideshows), etc. (Kaplan et al., 2009).

Content communities that allow sharing and storage of multimedia is one of the fastest growing areas in social media. People could easily access cheap content collections without requiring technical knowledge. Media content can be shared with users in different social networking profiles, blogs and websites. Communities also have the risk of sharing copyrighted content. Although the applications has technical rules that prevent illegal contents, the series and films cannot be prevented from

publishing in these media short time before traditional media channels (Dağıtmaç, 2015).

A) Youtube: YouTube is a platform for users to watch and share videos. It has been founded on February 15, 2005 by a former PayPal employee, then purchased by Google on October 9, 2006. On this platform shared content amateur clips, video clips, movie, TV program snippets and even music files can be shared. Users can vote and comment on the videos. Videos that do not meet the terms of use may be complained and it can be examined and deleted by YouTube officials. It has the world's largest video sharing, where millions of videos are uploaded every day (Nusair et al., 2012).

B) Instagram: Instagram was established as a photo-sharing site in October 2010. It allows people to use various professional filters and then share the images with followers. With the purchase of Facebook on 21 April 2012, Instagram's popularity has increased. Users can make their accounts available to anyone or only to their followers. Posts made publicly available are open to searches under different tags (hashtag) (Benli, 2014).

C) Flickr: Flickr was founded in 2005 and then it was purchased by Yahoo. It is used for storing, editing and sharing photos and videos. Flickr, enables you to see and store uploaded content in large sizes. Popular brands share their contents under Flickr groups of brands, thus, brands achieve content for their own uses while increasing the awareness about that brand. Flickr has integration with many other social media platforms such as Facebook, Twitter, Tumblr and Pinterest and content on Flickr is allowed to share on these social networks (Özkaşıkçı, 2012).

D) Slideshare: Presentation sharing sites are sites where users upload PowerPoint presentations or written documents prepared for different purposes on the internet and share it with other users. one of the most frequently used sites in this area is Slideshare (Daldal, 2013).

E) Pinterest: Pinterest is a visual sharing platform that was launched in 2010. It provides users with the ability to mark, comment, score, and track different users (Başer, 2014). Pinterest, which is derived from the words "pin" and "interest", allows the sharing of the pictures and videos that are liked with other users (Güçdemir, 2012).

Pinterest is a social media network where pictures are pinned to a clipboard based on the interests. Enabling a strong classification and categorization based on the photos' interests, makes it easier for users to organize specific topics based on images (Vardarlıer, 2014).

2.4.4.4 Microblogs (Twitter etc.)

A)Twitter: Twitter was founded in March 2006 by Jack Dorsey, Evan Williams and Biz Stone. Twitter has made the concept of 'instant messaging' popular among millions of people. 'Tweet' which is a text messages can be up to 140 characters, and provides to be followed by other users 'Retweet' to share someone's tweeting with their followers. Users can also Quote content from other users (Özkaşıkçı, 2012).

On Twitter the system which lists the number of messages according to the frequency of usage is called 'Trend topics' is also provided. This system lists the frequently updated most hot topics on Twitter continuously (Güçdemir, 2012). When "hashtag " is used in front of an expression that is meant to be highlighted in a tweet as a # symbol, other users can search for and track other tweets associated with that tag. Twitter allows users who have been friends to communicate through regular tweets and sharing.

Twitter users ' popularity is increasing as the number of followers increases. It is easy for corporate companies and brands to reach and communicate to potential audiences through Twitter. In this context Twitter is considered an effective medium for achieving commercial purposes (Benli, 2014).

B) Tumblr: Tumblr is a service provider established by David Karp in 2007 that allows users to open a personal blog. It was purchased by Yahoo on May 20, 2013 (Tüfekçi, 2015). Tumblr allows you to share text, photos, links, music and video files on your smartphone, computer, email and anywhere else. Users can select themes for Tumblr accounts and customize their accounts as they wish (Bostanci, 2010).

2.4.4.5 Social Networks (Myspace, Facebook etc.)

Social networks are defined as platforms where users on the Internet can communicate with other users around the world and express their feelings and thoughts

by expressing themselves with their profile pages. The number of social networks is quite high in parallel with the density of users using the Internet (Tektaş, 2014).

A) Facebook: Facebook was founded in February 2004 by Mark Zuckerberg. Facebook is currently used all over the world. Facebook enabled the use and production of different applications that can be integrated easily. For example, Facebook users can play chess, poker, etc. with other users. You can play and send virtual gifts. Thus, it can socialize with many tools (Mayfield, 2008).

The uses of Facebook vary from country to country. For example, in France, users use the site to communicate with friends and find old friends, while in Mexico they use to make new friends and find lovers. In our country, the Facebook users' aims are to communicate with friends, find old friends, share videos / photos and get to know their friends better (Şener, 2009).

B) LinkedIn: LinkedIn was founded in May 2003 by Reid Hoffman to build a professional network in the business community. LinkedIn allows users in a variety of professional fields to find each other, keep track of each other, and allow them to publish a variety of posts to discuss some topics such as changing careers, current professional issues. In addition, it allows users to apply for published job advertisements (Özkaşıkçı, 2012).

LinkedIn enables users to coordinate their professional identities online and aims to make their career processes more successful. That's why LinkedIn is expanding its social network together with users who want to use it more effectively in business (Daldal, 2013). As of March 2017, LinkedIn has 467 million users (106 million active users) in 200 countries. Members in the United States make up almost half of LinkedIn members. India, China, Brazil and the United Kingdom follow this country. In 2018, 5,280,645.00 LinkedIn users are available in Turkey (Thinknum, 2018).

2.4.5 Using Social Media in Crisis Management Practices

If the previously mentioned crisis management process is reduced to three phases (crisis preparedness, crisis situation improvement, crisis response), social media can be used for information dissemination, disaster planning and training, joint problem solving and decision making, etc.

2.4.5.1 Information Dissemination

Using social media is an effective way for organizations to spread the information they have and can be used to quickly providing reliable information. Thus, organizations can use such system to prepare for any crisis situation and intervene quickly (Chan, 2010). Making an immediate statement to the relevant stakeholders mentioned above is an application area of the information dissemination function.

2.4.5.2. Disaster Planning and Training

Disaster training is as important as planning disaster scenarios. Seminars, disaster exercises, informative programs on TV or social media channels provide training in order to raise awareness on disaster. In recent years, technology tools have become popular in disaster planning and training using multiple technologies, such as game practices, also known as (gamification), has been used to provide the necessary training in the event of a disaster or crisis. For example, “Event Commander”, an example of disaster education system, was developed in 2007 by Break Away for the United States Department of Justice. The game is a computer-based simulation designed to prepare for emergencies and crisis situations and allows up to 16 persons to receive game-based training at the same time. In this game, multiple scenarios such as natural disasters, kidnappings, terrorism, building collapse, and many emergency and crisis operation teams are available. Each team leader should immediately recognize the area to be intervened and undertake the process of sharing appropriate resources and allocating tasks to team members. Team members should also follow the team leader's instructions to address the emergency situation (Hanlon, 2006). Game simulations, such as event commander, are both useful and cost-effective for organizations in terms of disaster planning and training (Hanlon, 2006).

2.4.5.3 Common Problem Solving and Decision Making

As stated by Jeff Howe, problem-solving and decision-making can be achieved with organizations that use “crowded resource” (a combination of crowded and outsourcing). With the use of social media features (using the crowded resource), the information needed by organizations for crisis management can be collected using

web-based and mobile technologies. Information exchange and cooperation can help solve problems and make decisions (Chan, 2010).

2.4.5.4 Data Collection

Gathering and disseminating information is crucial for the management of crisis situations. As part of monitoring the crisis situation, social media channels should be monitored to gather news about the crisis and information related to crisis situations. The social media monitoring process can be used to control crisis preparedness and post-crisis periods as well as crisis management efforts (Velev et al., 2012). For example; social geolocation tools, such as Foursquare, can be used to identify and intervene in disaster situations (post-crisis activities).

Many international organizations and public agencies have used social media platforms and technologies to improve their crisis management capabilities. One of the most successful examples of this is Ushahidi, a downloadable software that provides access to reports produced by people who witness an incident during a disaster. The Ushahidi Haiti platform has been created by placing people affected by the earthquake on the map during the Haiti earthquake (2010). This platform can be accessed and used by various organizations.

In addition, social media platforms provide an area where situational awareness can be provided to deal with the situation by solving common problems and making decisions. Examples of platforms that could be most effective and contain intensive information in crisis management are Facebook, Twitter and Instagram. The Twitter Streaming API (Kalucki, 2010) provides real-time access to all public tweets in sampled and filtered form. In Twitter terminology, individual messages define the status of a user. Thanks to the Streaming API, subsets of general status descriptions can be accessed almost in real time, including users' public responses (Bifet et al., 2010).

2.4.6 Social Media during Disasters

Social media leads many users to share information during disasters as this information can be spread rapidly. There are many reasons why social media can be used in disaster situations. Some of them are described in the following.

A) Mitigation of disaster risks: The aim here is to reduce the risks that a disaster may create. In this context, it may be possible to raise public awareness of disaster risks, coordinate and organize discussion platforms to minimize risks and plan events, etc.

B) Emergency Management: Using social media tools, it may be possible to establish an emergency information system through mass sourced resources, to help people to prepare for disaster, make emergency warnings, and coordinate community gathering before and after disaster.

Overall, social media provides the following benefits about disasters: (Washington, 2016)

• Provides important information for disaster victims before and after disaster (via Internet or SMS updates)

• Ensures the presence of volunteers and / or donors and raises awareness for those outside the affected areas

• Establishes links between displaced families and friends

• Provides assistance to affected persons / organizations, centers and other resources

CHAPTER 3

RELATED WORKS

In recent years, social media analytics has become important research fields, which lead to improving many approaches such as CMSs. In addition, impressive attention has been given for mining the publically available huge amount of data to gain situational awareness, which may help in preventing or decrease the effect of some disaster by taking the correct responses. In the following, the recent developments and studies related to the CMS and processing the Turkish language have been summarized.

In (Rogstadius et al., 2013), Incoming tweets are compared to previously collected tweets. Comparison is performed using a Bag of Words approach. Clustering is performed using an algorithm that is based on Locality-sensitive hashing (Petrović et al., 2010). Locality-sensitive hashing algorithm is a probabilistic technique based on hash functions that quickly detect close copies in a property vector stream. A time interval is required to determine whether the incoming message is a copy and it depends on the rate at which similar messages are received (Petrović et al., 2010). General word statistics can be obtained using idf in offline environment, but this is difficult in online environment. Because the frequency of words changes over time. To solve this problem, the algorithm was expanded in two important ways. Firstly, word statistics are generated using both filtered resources and 1% of Twitter is used as base flow. Second, the word distributions are then approached with a simple IIR (infinite impulse response) filter. Words whose spherical frequency is greater than 90% of the maximum frequency are labeled as stop-words unless they follow the keywords. The task of the second extension is to change the oldest hash function per hour with a new function created by the current dictionary. The new hash table is then populated with elements of the removed table. To upgrade the recall, in addition to the normal thread-based clustering algorithm, all new clusters are compared to existing clusters to control the confliction of clusters. This set of clusters is called a story. Then, the stories are sorted according to their dimensions and the reports describing the same

event are grouped together. The story size allows CrisisTracker to predict how important the message is to its source, i.e., the community sharing the message.

In (Girtelschmid et al., 2016), a system that is designed to detecting new emergency situations from Twitter data streams has been implemented. In general, it has four components.

In the first part, FSD is performed using the Locality Sensitive Hashing algorithm which is described in detail in (Petrović et al., 2010). The number of documents to be compared is greatly reduced with this algorithm. Similar tweets are grouped by using the hash tables, so that each tweet that has the same hash has to be compared to each other. The distance between the nearest neighbor and the tweet is calculated using Cosine Similarity. If the distance calculated for the nearest tweet is below a predefined threshold, this algorithm also uses an additional step to compare the distance to a fixed number of recent tweets.

Secondly, the same tweets are then collected in the same bucket. However, before chasing, duplicated and near duplicated tweets are filtered. If (1 - cosineDistance) <threshold , tweet is included in the other processing steps. Then, the fastest growing bucket will be detected in a while. Before the burst content is passed down the processing line, the Burst Event Detector controls the time interval during which the tweets are sent to determine whether this large amount of tweet can be considered an explosion. The explosion content is then consumed by the third component, i.e., the Disaster Detector.

Thirdly, The duty of the Disaster Detector is to find out if the incident is an emergency and where it is located. First of all, all keywords are extracted and topic is identified based on the disaster type dictionary. Then, if the issue matches a disaster situation, it is localized. Finally, once a disaster event has been identified and its location, type, and keywords are identified, the information is transmitted to the Alert component. Keywords are used to launch new data collection sessions from Twitter, both from a current history and a real-time stream.

In (Terpstra et al.,2012), the proposed system that is named as Twitcident, which is a Web-based system that automatically filters and analyzes tweets about

events. The system consists of three parts. The first part keeps a list of events in the Netherlands in real-time. It does this by parsing messages that are publicly available in real-time and informs the emergency services by some information such as disaster type, location, start time. The second part, which is based on the collected information, is to create a Twitter search query that includes the city name and the most commonly used words of people (for example, "earthquake" when an earthquake occurs). The system then uses the Twitter Search API to get historical tweets and Twitter Streaming API to get tweets in real-time. The third component includes analyzing and visualizing the features. Hence, it works on finding the types of tweets (ie retweet, mentioning, answering, singleton (Kwak et al., 2010), the reliable news media accounts, the date-time intervals. Finally, the tool maps the filtered tweets to the map, extracts statistics, and creates a gallery of pictures and videos.

In (Nguyen et al.,2017), a system that analyses social media images about major natural diseasters to detect severity of damage after disasters. This study focuses on Twitter images. Two different settings are used. Firstly AIDR platform (Imran et al., 2014) is used to collect images. In this case, volunteers are employed for labelling images. Secondly, Crowdflower which is a waged crowdsource image annotation platform is used. For damage severity three levels exist: Severe damage, Mild damage and Little-to-no damage. The images that labeled as none are considered as irrelevant. Then the images are processed on a VGG-16 (Simonyan et al., 2014) which is one of the state-of-the-art deep learning object classification models that had the best performance for identifying 1000 object categories in ILSVRC 2014. Then, specific features are generated from each image and a hash value is calculated for each image based on these properties. To determine the level of similarity between images, the resulting hash values are compared. Hamming distance is used to compare the two hash values. If an image with a small distance value is found in the list, the new image is considered a copy.

A new system that uses gazetteer, street map and volunteered geographic information (VGI) data instead of geo-tagged twitter data was introduced in (Middleton et al., 2013). This due to the fact that only 1% of Twitter data has geo-tag information. In this system, some statistical analysis techniques are used to determine “baseline noise”. The system has offline and real-time services. The purpose of offline

services is to prepare a geographic database and to calculate basic statistics in a historical process when there are no disasters. Real-time services take tweets from Twitter simultaneously and identify the tweets and show it on the crisis map. In more detail, the goal of the offline part is to extract geographic data. For this, applications such as OpenStreetMap and GooglePlaces API have been used. Geographic information is stored in a MySQL database with OpenGIS shape data and then visualized on the map. GoogleGeocoding API is used for geocoding the geospatial data. Geocoding is done to fill in missing address fields or to make corrections if there are errors. In the real-time system, a Twitter crawler is used. Keywords are assigned to the Twitter crawler, where multiple words are used for each type of disasters. For example, 'flood', 'tsunami', 'alluvione', 'inondation' are used for the flood disaster. Then, the TwitterStreaming API is used to get the related tweets, which will be saved into a Mysql database. A series of operations are performed when new tweets arrived (collected). Firstly, they are cleaned and tokenized and divided into tokens . Location tokens are determined and matched. All match statistics are saved to the database in OpenGIS format. Finally, the visualization process on the map is performed using Geoserver.

In (Petrović et al., 2010), an FSD system that processes each new incoming document in a fixed time and uses fixed storage is introduced. This fixed processing time is accomplished by using LSH. LSH method divides each query point into parts called buckets. A collision probability is much higher with nearby points. When a new query arrives, it is placed in a bucket and compared to all points in the bucket where the query point is located. As a result, the point closest to the query point is returned. However, based on (Petrović et al., 2010), it has been proven that applying LSH to find the nearest neighbor in an FSD role gives bad results. Because LSH only returns the real close neighbor. İn such case, a point must be close enough to the query point. If the query point is too far from all the other points, the LSH cannot find the real close neighbor. To solve this problem, when a new document is received in this application, it is searched with an inverted index. But the query is only compared to a certain number of sections of the most recent documents. Locality Sensitive Hashing is used to limit space and time. In addition, LSH is working based on keeping the number of stories constant. But only a finite number of buckets are found. Therefore, the number

of documents in a bucket increases over time. Thus, the number of comparisons that need to be made also increases depending on time and this increases the processing time. To solve this problem, the number of documents in each bucket is limited in this study. The oldest document is deleted when the bucket is full.

In (Blum et al.,2014), a system named RtER has the ability to analyze videos, as well as text and photos, was developed. RtER enables people to 1) manage incoming multimedia in parallel, 2) label the incoming multimedia, 3) group the files according to their areas in order to ensure fast decision making in case of emergency. In addition, communication with the people providing video content is available. In this system, the participants have different roles for emergency response task. For example, decision-makers work in command or dispatch centers, the public information officers situate in the emergency operations center, volunteers work on analyzing data from internet-connected media sources, others are categorized as emergency first aid teams, general public, and fans, etc. In addition, multiple streams can be held at the same time on one server and multiple users can view simultaneously.

In (Zielinski et al., 2013), A study was conducted to facilitate the analysis of Twitter data. In general, the system components are: 1)Focused Crawling (FC) that is used to get the relevant information at the highest level. Performed by adding an keyword that is not used in the data collection level. 2)Trustworthiness Analysis (TA) is used to reduce the number of users by classifying users by calculating the reliability and the degree of impact of the tweets. For example, the information shared by a news agency may be more reliable, but not sufficiently effective, while a user's tweet in the event might be more effective. 3)Multilingual Tweet Classification (MTC), is an algorithm that uses Cross-lingual Text Classification (CLTC) algorithms to obtain as much information as possible from monolingual data and to transfer this information to other languages.

In (Rashdi et al., 2019), A system is introduced to integrate the different deep learning architectures (CNN and Bi-LSTM) of different word embeddings (crisis embedding and Glove). In addition, to perform some performance comparisons, the (Nguyen et al., 2016) Crisis Embedding model was re-implemented in this study. The CrisisNLP dataset (Imran et al., 2016), which contains small sub-datasets in which

each one contains annotated tweets about a crisis event is used to perform the evaluation. F1 score was used to evaluate and compare the models. As a result, Bi-LSTM with GloVe embedding achieved the highest score among the four performed experiments. CNN with Crisis embedding model got score higher than Bi-LSTM with same embedding type, and Bi-LSTM with GloVe embedding got higher score than CNN with GloVe .

In (Alam et al., 2018), a CNN model is combined with a graph-based network and a pre-trained word embedding is used. In addition, for the large crisis dataset, a continuous bag-of-words(CBOW) word2vec (Mikolov et al. 2013) model was developed. In this study, Two real-world Twitter datasets were collected using the Twitter streaming API during the 2015 Nepal earthquake and 2013 Queensland. To obtain the labeled dataset, 2 paid annotators were employed to label data into categories relevant or irrelevant. In the graph construction, k-nearest neighbor-based approach has used. In this approach, if a graph consists of n vertices, each vertex has n edges. The edge between the two nodes is defining the distance between the two nodes. The value of the distance shows the similarity between these nodes. Then, the K-d tree data structure is used to describe the nearest instances. Euclidean Distance is used to compute distances between nodes. For graph construction, each tweet has defined by word embedding vectors. The models are trained using the adadelta (Zeiler 2012) algorithm. As a result, the system has reached 93.54 F-measure value when all labeled data and 21K unlabelled data were used.

In (Kumar et al., 2019), A convolutional neural network-based system is introduced to predict locations in the tweets by looking at the tweet text. The presented system consists of 3 parts: 1) word embedding which is representing words in a vector space. GloVe embedding which is a pre-trained embedding model was used in this section. 2) Convolutional model designed to learn the necessary features in tweets, 3) fully connected layer that provides an estimation of results by evaluating the extracted features. In (Kumar et al., 2019), some English tweets that have various location references such as street names, building names, cities, region and even country names have been used. In addition, the other tweet's metadata has been removed.

Unfortunately, Up to the time of writing this thesis, we could not find any study related to crisis management systems related to the Turkish language. However, this study can be considered as an implementation of a Turkish text classification system for crisis response. In the following, we have summarized some recent studies about Turkish text classification.

In (Ayata et al., 2017), A study was conducted to perform sentiment analysis of Turkish tweets using Machine Learning and Word Embedding. Word2Vec (Mikolov et al., 2013) model was used for the conversion of words into a vector form. The vector representation of the tweets was created by using the Total Based Representation Model and the Product-based Representation Model separately. Then, the performances were compared using Support Vector Machine (SVM) and Random Forest (RF) classifiers. The results of (Ayata et al., 2017) showed that classifier performance varies according to the used data set.

In (Kılınç, 2016), A study was conducted to measure the effect of Ensemble Learning on Turkish text classification. Ensemble Learning aims to achieve more accurate results by using the results obtained by using different classifier types together. In more details, the Bagging (Breiman, 1996), Boosting (Freund et al., 1996), Rotation Forest (Rodriguez et al., 2006) ensemble learning methods were used with Naïve Bayes, J48 - Decision Tree, K-Nearest Neighbor (K-NN), and Support Vector Machine (SVM) classification algorithms. In (Kılınç et al., 2017), a new Turkish dataset named TTC-3600 was introduced. this dataset consisted of categorized data of several newspapers. The results of this study showed that the Rotation Forest algorithm can improve the performance of the KNN algorithm, while other ensemble models didn't have a significant impact on KNN. Related to the SVM, ensemble systems didn’t have a positive effect on it, and both Bagging and Boosting ensemble systems have decreased the accuracy of SVM. On the other hand, all the studied ensemble systems were able to improve the performance of both the J48 and Naïve Bayes models.

In (Baygın, 2018), the classification of Turkish documents using the Naïve Bayes algorithm was performed. A total of 1150 documents in 5 different categories was used. Firstly, the documents were preprocessed and then feature extraction was performed using the n-gram technique. Different models were created by 2-gram, 3-

gram, and 4-gram respectively. As a result, the system reached a total of 92% success, but the model with 3-gram showed the best performance in terms of time and classification.

CHAPTER 4

MACHINE LEARNING APPROACHES

Machine learning technology has been incorporated into our lives in many areas in the modern world: it is actively involved in many applications such as online searches, shopping sites, cameras, smartphones, and systems. Machine learning systems enable image processing, face recognition systems, creating and sending personalized data by recognizing user interests.

4.1 Machine Learning for Text Classification

Machine Learning has become popular in classification of textual data as well as other areas mentioned above. This process consists of the following stages:

4.1.1 Data aggregation

Data aggregation is the compiling of information from multiple sources such as databases with the intent to prepare combined datasets. In some cases, ready datasets are available, however, for some other topics, it is not. The reasons for this may be the language for the study or the study subject. If datasets are not existed, the required amount of data should be collected. Data collecting can be done by classical methods such as observation, survey, interview, scanning written resource or by using various software or platforms. In this study, data was captured from Social media using Twitter Search API. This API allows collecting public tweets. The tweets contain user information, tweet date, location, favorite count, etc. The API also allows filtering tweets by keywords.

4.1.2 Data pre-processing

The “Pre-processing” stage regulates the data to ensure an efficient configuration of the classification system. In general, it has been proved that the preprocessing of the text data is an important and essential step and if we skip such a step, the chance of the system to be badly affected by the noisy and inconsistent data is increased. The objective of this step is to clean data by eliminating the noise and irrelevant part of tweets such as punctuation, special characters, numbers, and terms

which don’t carry much weight in context to the text. The main steps of Data pre-processing are:

A) Tokenization: Tokenization is the step of splitting longer text strings into smaller pieces or tokens. Large text sets can be divided into sentences, words, and characters. Tokenization is also called text segmentation or lexical analysis.

B) Stop-word Elimination: Stop words are functional words specific to language that commonly used (pronouns, prepositions, conjunctions) such as 'the', 'of', 'and', 'to', etc. (Aït-Sahalia et al., 2019). In general, stop words are removed in Natural Language Processing tasks. Hence, it has been proven that this step is very important for Information Retrieval (IR) and text mining (Rogstadius et al., 2013, Sharma et al., 2015) .

C) Normalization: Before further processing the text data, it must be normalized. Normalization generally consists of a series of related tasks to put all text on a flat playing field, and it is consists of: converting all text to the same letter size (upper or lower), removing punctuation, converting numbers to word equivalents, etc. In addition, stemming, and lemmatization, spelling and grammar correction can also be performed in this step.

4.1.3 Feature Extraction

Words in a text are usually discrete, and categorical features. Therefore it is necessary to convert the text to the Vector Space Model (VSM) (Li et al, 2018 ; Christian et al., 2016). The vector space model is an algebraic model that converts text into a vector of words and then converts words into a vector format. This process consists of several steps: First: create a dictionary with terms in the texts (in our case, the data set of tweets). In other words, each of the terms in the data set is defined in the vector space, each having a different identity. The second step is to obtain a vector representation of each term and add it to the vector space. It is worth mentioning that this is done after all tweets are preprocessed. Therefore, all stop words were previously deleted. Secondly, obtain the representative of each term and add it to the vector space. The followings are some well-known methods that can be used for representing each term in our vector space:

A) Term Frequency (TF) :

Term Frequency , which represents the term by its frequency in the document, in our scenario, each term is represented by the frequency of the set of vocabularies in the previously created vector of words) (Christian et al., 2016). This value can be calculated using Equation 4.1.

$$TF\ (word_i) = \frac{F(wordi)}{N_W} \quad (4.1)$$

Where F (wordi) refers to the number of occurrences of the i^{th} word, and N_W is the total number of words in our VSM.

B) Term Frequency - Inverse document frequency (TF-IDF) [Christian et al., 2016]:

In general, the Term Frequency work on assigning the highest score to the most frequent word, i.e., if a word occurs frequently in the VSM, it will have a high TF. In contrast, the IDF measures how rare (important) the term is. Hence, the IDF assigns the highest score to the rare words, and the score decrease whenever the frequency of words increases. For instance, the terms Deprem (“earthquake”), Yangın(“fire”), and Sel (“Flood”) are technical terms, which its exists can be very useful, will have a high IDF, while, terms such as "Yapacak"(doing something), "Ederim" (doing, practicing something), "Hatta"(even), “olduklarını” (being something), which appears in most texts many times, and have little importance will have a low IDF. IDF can be represented and found using Equation 4.2.

$$IDF\ (word_i) = \log \frac{N}{N_T} \quad (4.2)$$

Where N_T is the number of tweets contains the ith word, and N is the total number of tweets. Finally, the TF-IDF is calculated by multiplying the TF and IDF values:

$$TF\text{-}IDF\ (word_i) = TF\ (word_i) * IDF\ (word_i) \quad (4.3)$$

C) Word Embedding [Li et al., 2018 – Naili et al., 2017]:

It can be considered as the state of the art method that can represent words in low-dimensional vector space (Karlik et al., 2011), while effectively preserving contextual similarity. One of its important advantage, its ability to obtain almost the same representation for similar meaning words. But embedding requires a huge amount of text data (millions) to be useful. The most popular embedding approaches are Word2Vec (Mikolov et al., 2013, 2017), GloVe (Pennington et al., 2014), and FastText (Joulin et al., 2016).

1) Word2Vec: Word2Vec(Mikolov et al., 2013, 2017) is one of the popular state-of-the-art word embedding methods. Word2vec is a combination of models used to represent distributed representations of words in a corpus.

Word2vec can use one of two model architectures to produce a scattered representation of words: continuous bag-of-words (CBOW) or a continuous skip-gram (Mikolov et al., 2013). In the continuous bag-of-words architecture, the model estimates the current word from a window of its surrounding words. In this model, the order of context words does not affect the prediction result. In Continuous Skip-Gram architecture, the model summarizes the words of each sentence and uses the existing word to estimate the window (neighbors) around the context words (Mikolov et al., 2013). Continuous Skip-gram assumes nearby context words more important than further context words, so the nearby words are more weighted. The general architectures of the Word2Vec models are shown in Figure 4.1.

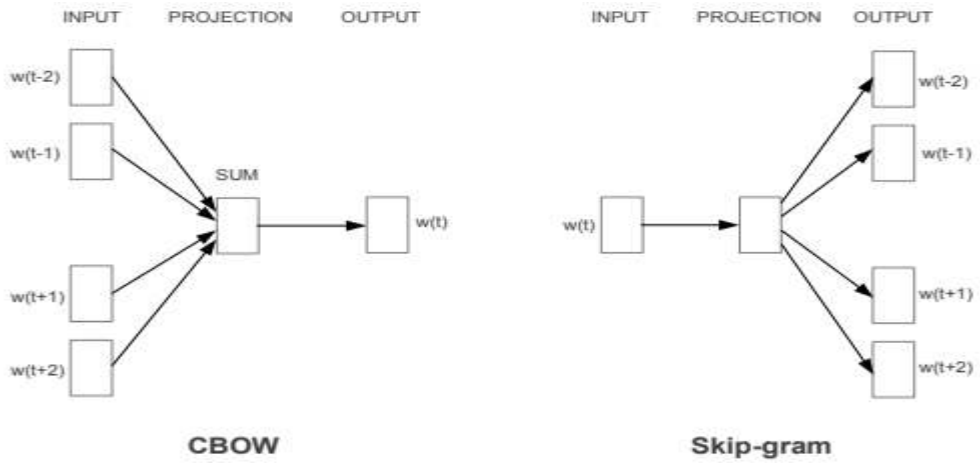


Figure 4.1. The architecture of the CBOW and Skip-gram models (Mikolov et al., 2013).

2) GloVe: In general, the GloVe algorithm consists of the following steps (Pennington et al., 2014):

1. The word co-occurence statistics is collected as co-occurence matrix and it is defined as *X*. Each element in *X* is represented as X_{ij} that represents how often word i appeared in context of j. Usually, corpus are scanned in the following manner: for each term we look for context terms within some area defined by a *window_size* before the term and a *window_size* after the term. The distance between the term and neighbor is defined as $offset$. Less weight is assigned to more distant words($decay$) as shown in Equation 4.4:

$$decay = 1/offset \quad (4.4)$$

2. Define soft constraints for each word pair:

$$w_i^T w_j + b_i + b_j = \log(X_{ij}) \quad (4.5)$$

Here w_i - vector for the main word, w_j- vector for the context word, b_i, b_j are scalar biases for the main and context words.

3. Define a cost function (J) using:

$$J = \sum_{i=1}^{V} \sum_{j=1}^{V} f(X_{ij}) \left(w_i^T w_j + b_i + b_j - logX_{ij}\right)^2 \quad (4.6)$$

Then, the GloVe function is represented as:

$$f(X_{ij}) = \begin{cases} \left(\frac{X_{ij}}{x_{max}}\right)^{\alpha} & if\ X_{ij} < x_{max} \\ 1 & otherwise \end{cases} \quad (4.7)$$

3) FastText: FastText (Joulin et al., 2016) is an extension of Word2Vec recommended by Facebook in 2016. The feature of this model is that it separates sentences into words and then separates words into n-grams. It then sends these sub words to the artificial neural network to be defined in the vector space. For example, the 3-gram of the word "phone" is "pho", "hon", and "one". The embedding vector of "phone" will be the sum of all these n-grams. After the Artificial Neural Network training, the training data set will have word placements for all the given n-grams (Joulin et al., 2016).

4.1.4 Classification

In this study, the performance of some well-known machine learning approaches, i.e., K-Nearest Neighbors (KNN), Naïve Bayes(NB), Support vector

machine (SVM), Random Forest, AdaBoost Classifier, and GradientBoosting Classifier. In addition to the convolutional neural networks (CNN), that can be used in classifying the information available before or even during any crisis have been investigated. In the following, information related to these approaches is summarized.

4.2 Machine Learning Approaches for Text Classification

In this thesis the following Machine Learning algorithms for text classification are used and its performance was investegated.

A) K-Nearest Neighbor (KNN)

K-nearest neighbors (Nikhath et al.,2016), is a simple algorithm that stores all existing neighbors and classifies new data according to the used similarity criteria (ex, distance functions as Manhattan, Euclidean distance) (Nikhath et al.,2016). KNN was used as a non-parametric technique for statistical estimation and pattern recognition tasks in the early 1970s.

K-Nearest Neighbor (KNN) algorithm is a frequently used classification algorithm in data mining and machine learning (Nikhath et al.,2016). This algorithm is frequently preferred because of its flexibility in data types (Cunningham et al., 2007). However, there are several factors that influence the superior classification performance of k-NN. Such as , the distance matrix used to calculate the distance between two data points (Nikhath et al.,2016). A second factor is the selection of the appropriate k value. A small selection of k means that there are few neighbors. This increases both 1) the effect of noise on the result, 2) the cost of calculation. In addition, the small number of neighbors provides low deviation and high flexibility. On the other hand, larger k leads to diversity decreases and deviation increases (DataCamp, 2018). It is worth nothing mentioning that in our case each tweet is assigned to the most common class among its' nearest neighbors (Cunningham et al., 2007). Psedocode of the KNN Algorithm is in the following:

BEGIN

Input: D = {(x1,c1), . . . , (xN, cN)} classified data

x = (x1, . . . ,xn) new instance to be classified

FOR each labeled instance (xi, ci) where i = (1, . . . ,N)

Calculate distance(xi, x)

Order d(xi, x) from lowest to highest

Select the K nearest instances to x:

Assign x to the most frequent class in K

END

In other words, KNN gets the labelled data instances {(x1,c1), . . . , (xN, cN)} and x = (x1, . . . ,xn) which is the data that will be classified, i.e., input. After that the algorithm calculates distances of each data in x with each labelled data, and selects K samples that have most lowest distance to x. As aresult, data is assigned to most frequent class in the K samples.

B) Naïve Bayes (NB)

Naïve Bayes is a probabilistic classifier that based on Bayes theorem with naive independence assumption. Naïve Bayesian model is easy to build, with no complicated iterative parameter estimation, which make it particularly useful for very large datasets. Naïve Bayes operates by assuming independence, i.e., the presence of some feature will not affect the other features (Frank et al., 2006). In other words, Naïve Bayes classifier relies on the probability model, and in many practical applications, the method of maximum likelihood is used in parameter estimation for "Naïve Bayes" models (Dietterich et al., 2000). Applications of Naïve Bayes includes real-time prediction, multi-class prediction, text classification, spam filtering, sensitivity analysis and suggestion systems.

The basis of the Naïve Bayes classifier is based on the Bayes theorem. It is a lazy learning algorithm that can also work with unstable data sets. The operation of the algorithm calculates the probability of each case for an element and classifies it according to the highest probability value (Chen et al., 2009). With very little training data, it can do very successful work. If a value in the test set has an unobservable value in the training set, it returns 0 as the probability value, i.e., this value cannot be predicted. This is commonly known as Zero Frequency. Correction techniques can be used to solve this situation. One of the simplest correction techniques is known as Laplace estimation (Chen et al., 2009). The goal of Naïve Bayes Classifier is to calculate conditional probability:

$$p(C_k|x_1, x_2, \dots, x_n) \tag{4.8}$$

for each of *k* possible outcomes or classes *Ck*.

Let x=($x_1, x_2, \dots, x_n$). Using Bayesian theorem, we can get:

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)} \propto p(C_k)p(x|C_k) = p(C_k, x_1, x_2, \dots, x_n) \tag{4.9}$$

And, the joint probability can be written as:

$$\begin{aligned} & p(C_k, x_1, x_2, \dots, x_n) \\ &= p(x_1|x_2, \dots, x_n, C_k) \cdot p(x_2, \dots, x_n, C_k) \\ &= p(x_1|x_2, \dots, x_n, C_k) \cdot p(x_2|x_3, \dots, x_n, C_k) \cdot p(x_3, \dots, x_n, C_k) \\ &= p(x_1|x_2, \dots, x_n, C_k) \cdot p(x_2|x_3, \dots, x_n, C_k) \dots \cdot p(x_n|C_k) \cdot C_k \end{aligned} \tag{4.10}$$

Assume that all features x are mutually independent, we can get:

$$p(x_1|x_2, \dots, x_n, C_k) = p(x_1|C_k) \tag{4.11}$$

And Naïve Bayes formula can be written as:

$$\begin{aligned} & p(C_k|x_1, x_2, \dots, x_n) \\ & \propto p(C_k, x_1, x_2, \dots, x_n) \\ &= p(x_1|C_k) \cdot p(x_2|C_k) \dots \cdot p(x_n|C_k) \cdot p(C_k) \\ &= p(C_k) \prod_{i=1}^{n} p(x_i|C_k) \end{aligned} \tag{4.12}$$

C) Random Forest (RF)

The random forest is an ensemble classifier based on a series of decision tree models (Onan et al., 2016). The random forest consists of a large number of individual decision trees working as an ensemble. Each tree in the random forest predicts a class, and the class with the most votes becomes the prediction of the model. These trees that the Random Forest model consists of are relatively uncorrelated. In addition, the Random forest applies the technique of bagging (bootstrap aggregating) to decision tree learners. Hence, Bootstrapping enables the random forest to work well on relatively small datasets. However, it is getting more complicated and its execution time increases when it is required to deal with a large number of samples. The equation of Random Forest is as follows:

$$\{h(x, \theta_k), k = 1,2 \dots i \dots\} \tag{4.13}$$

Where h represents the Random Forest classifier, x specifies the input variable, and $\{\theta_k\}$ is independently distributed random prediction variables used to generate each tree (Breiman, 2001). Compared to other machine learning methods, such as the support vector machine and the artificial neural network, Random Forest is simpler in computation and insensitive to multivariable linear variables and outliers (Rodriguez-Galiano et al., 2012)

D) AdaBoost Classifier (AdaBoost)

The AdaBoost or Adaptive Boost classifier is an iterative ensemble method that boosts the performance of a weak classifier by using it within an ensemble structure. The classifiers in the ensemble are added one at a time so that each subsequent classifier is trained on data which have been "hard" for the previous ensemble members. In other words, AdaBoost trains the machine learning model by selecting the training set based on the estimation of the last training (Rodriguez et al. 2006).

E) GradientBoosting Classifier (GBC)

Gradient boosting is an ensemble technique for regression and classification problems. This classifier produces a prediction model by sequentially fitting the base

learner to current “pseudo”-residuals by least squares at each iteration. Based on (Friedman et al., 2002), the pseudo-residuals are the gradient of the loss functional being minimized, with respect to the model values at each training data point evaluated at the current step.

Determining the weaknesses of weak learnings can be considered as the major difference between the AdaBoost and the Gradient Boost Algorithm. The AdaBoost model identifies weaknesses using high-weight data points, while gradient boosting carries this by using gradients in the loss function (Friedman et al., 2002).

F) Convolutional Neural Networks

Recently, in addition to traditional approaches for text classification, there are methods developed using deep learning. Some of these approaches were described in the following.

Deep convolutional neural networks (CNN) is a specialized Artificial Neural Network (ANN) type that uses convolution in at least one of its multiple layers rather than the general matrix multiplication (Namatēvs et al., 2017). Simple neural networks consist of one or several layers, which are usually hidden layers. CNNs are composed of several layers. With this feature, CNNs can represent a variety of high-grade nonlinear functions. These are convolution layer, pooling layer, flatten layer etc. Deep learning is used to learn complex features that can represent high-level abstractions (Namatēvs et al., 2017).

More details about the main components (layers) of the CNN has been summarized in the following.

A) Convolutional layer: Convolution is generally a combination of two mathematical functions to create a new function. The resulting new function shows the mapping ratio of the used two functions. The convolution between two functions in a dimension can be expressed as:

$$g(x) = f(x) \odot h(x) = \int_{-\infty}^{\infty} f(s)h(x-s)ds \qquad (4.14)$$

where f(x) and g(x) are the two used functions and s is a dummy variable of integration (takes values 0 or 1) (Namatēvs et al., 2017). The convolutional layer(s) can be

considered as the main building block of the CNN (Namatēvs et al., 2017), and mainly aims to detect local conjunctions of features and mapping their appearance to a feature map. In addition, the convolution layer is a layer that consisted of filters and learnable kernels that extract local properties (Krig, 2016). The first convolutional layer provides the most basic features such as edge, corner, while the following convolutional layers reveal features in more detail. Thus, the most advanced properties appear in the final convolutional layer. The kernel size parameter on this layer specifies the size of the filter used. The filter is shifted around the feature map in the input and the features are exposed.

B) Embedding layer: Another type of layer that mainly developed for natural language processing is the embedding layer. This layer very popular as it can represent the words in low dimensional vector space (Karlik et al., 2011), while efficiently preserving the contextual similarity (ALRashdi et al, 2019). Its main advantage is that it allows words with similar meaning to have a similar representation.

C) Pooling layer: Pooling layers that responsible for reducing the size of the activation maps by down sampling the feature maps through summarizing the presence of features in the feature map patches and reducing the dimensionality of the feature maps that will be used by the following layers. Average, max and hybrid are very frequent used pooling methods (Namatēvs et al., 2017).

D) Flatten layer: The flatten layer converts the output from the previous layer into a 1-dimensional array. In most cases, this type of layer is set as the layer before the last one, i.e., the classification layer, where the output of the previous layer is corrected to form a single long feature vector and finally connected to some fully connected layers that perform the classification task (Jin et al, 2014).

F) Dropout layer : The Dropout layer performs the function of neglecting some randomly selected neurons. In this case, the neglected neurons need to be predicted, which provides multiple representations for the neural network and reduces the sensitivity to weights. Hence, this layer can efficiently solve and decrease system overfitting (Brownlee, 2016).

G) Non-linearity layer: In general, artificial Neural Networks (ANN) are designed as universal functions that has the ability to calculate and learn any function. Non-linear activation functions enable networks to learn more efficiently. The most popular activation functions used in neural networks are:

1) Sigmoid Function: It is a nonlinear activation function. This function has two different types, i.e., the Uni-Polar Sigmoid Function and the Bi-Polar Sigmoid Function. The Uni-Polar Sigmoid Function is particularly useful in networks that are trained by back propagation algorithms. This due to the fact that it has an easy to distinguish formula and provides ease of calculation during the training of neural networks (Karlik et al., 2011). This function produces an output value in the range [0, 1]. On the other hand, the Bi-Polar Sigmoid Function found to be more flexible and returns output in the range [-1, 1] (Karlik et al., 2011).

2) Rectified Logical Unit (RELU) Function: which can be considered as special implementation aims to combine non-linearity and rectification layers. The output of the ReLU function takes values in the range $[0, +\infty)$.

It is worth mentioning that, Sigmoid function can outperform RELU when integrated into a normal size neural network, but it is not preferred to be used with a large neural network with too many neurons, as this function can cause almost all neurons to be activated, which requires complicated processing. However, for such a scenario, RELU is less computational complexity (Ramachandran et al., 2017) as it inactivated some neurons, i.e., the RELU function assign the value zero for all negative inputs.

3)Softmax Function: The Softmax function takes a vector of real numbers as input and normalizes these inputs proportionally to their exponential functions (Liu et al., 2016). It has been proven that Softmax performs very well when used as a classifier. In addition, this method performs a probabilistic interpretation to find the membership (probability) of the input to each class.

CHAPTER 5

PROPOSED CRISIS MANAGEMENT SYSTEM

This chapter aims to provide information about the design of the proposed system and the methods used for crisis management through social media. As mentioned before, the crisis management system aims to minimize material and moral damages that may occur by detecting possible crisis situations early or collecting the necessary information in case of a crisis situation and making due diligence. In general, this work aims at classifying and identifying social media data related to the crisis efficiently, and inform the authority asap (almost a real-time response) to gain situational awareness, which may help in preventing or decrease the effect of some disaster by taking the correct responses. For this purpose, since our proposed system is machine learning based, it consists of two separate parts: the training part where the system is trained and the real-time part that makes the detection of crisis situations.

5.1 The Proposed System

The proposed system that aims to detect crisis situations is implemented by using Convolutional Neural Networks (CNN). The system consists of two main parts.

5.1.1 Mechanism of System Training

Since the developed system is machine learning based, the system must be trained with the training data first. Some of the collected data related to various disaster types was used to train the system. Figure 5.1 shows the structure of training the proposed approach, and its main components are:

5.1.1.1 Data Collection

In order to test and implement the proposed system, first of all it is necessary to have relevant data and train it. At this stage, for the English part of this study, we used a ready dataset consisting of pre-collected tweets related to disaster types. Related to the Turkish, some Turkish tweets were collected with Twitter Search API by using keywords related to disasters. In this thesis, three sub-datasets were constructed from our newly developed Turkish dataset. These sub-datasets were balanced, i.e., number of disaster and non-disaster tweets is equal, where, the first one is related to the

Deprem, i.e., “earthquake”, the second one is related to the Yangın, i.e., “fire”, and the third one is related to the trafik kazası “traffic accident” that contains 4300, 2000, and 5700 tweets respectively.

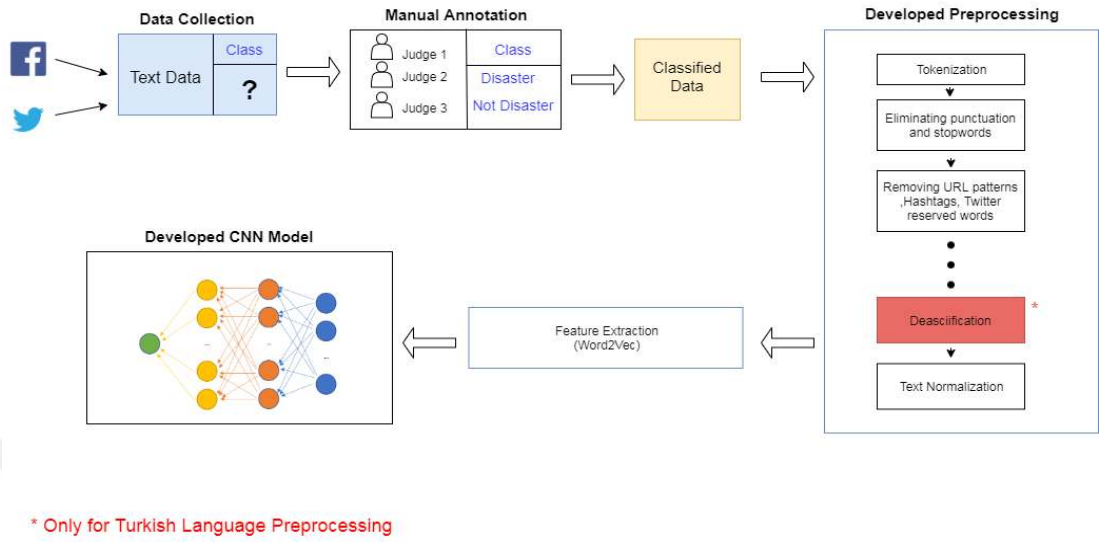


Figure 5.1. Structure of the proposed CNN-based approach (training).

5.1.1.2 Manual Annotation

Since the purpose of our machine learning model is tweet classification, the data used in training should be labeled. Tweets should, therefore, be classified after collection. For the classification task, three people were assigned as judges (annotators), and each tweet was classified separately by these judges through reading the tweets and vote whither each tweet is relevant or not, and then majority voting was applied for the final decision.

5.1.1.3 Preprocessing

Preprocessing of the Turkish language is different than English because Turkish is agglutinative and has a different structure. The main problems that make Turkish hard to process are:

i) Turkish is an agglutinative language

Turkish is an agglutinative language, i.e., adding suffixes to a root word can generate new and arbitrarily long words. Hence, some suffixes may change the Part-of-Speech (POS) tagging and the semantic orientation of the word. Overall, it is a hard task and practically limited to build a lexicon that contains all variants of Turkish words.

ii) Negations

In general, a small modification in a review can change the whole meaning. In addition, negation also expresses by sarcasm and implicit sentences that don't contain any negative words. Furthermore, in Turkish, words can be negated using many ways, such as the affixes me/ma or siz/sız or using a separate word such as "değil" or "yok". Whereby, each way can change the word meaning.

iii) Turkish Alphabet

"ğ", "ç", "ı", "ö", "ş", and "ü" are characters that do not exist in the English alphabet. In the informal domain, people tend to substitute these Turkish letters by the closest ASCII characters, for instance, "ş "is written as "s" and "ö" is written as "o". Therefore, the semantic analysis of Turkish is more risky to be defeated by the erroneous writings.

In this work, the pre-processing consists of the following steps. However, it is important to note that the first two steps are the same for almost all languages, and the remaining ones are specific for Turkish.

A) Tokenization: Tokenization is the step of splitting longer text strings into smaller pieces or tokens. Large text sets can be divided into sentences, words, and characters. Tokenization is also called text segmentation or lexical analysis. In our study, each tweet is tokenized into words. The purpose of this part is to break down tweets to make our machine learning model more accurate.

B) Eliminate and delete the useless information: In classification tasks, it is very important to clean up noisy data and to extract and use the useful information. In general, some part of the tweet is useless (does not give a specific meaning and may degrade the performance of the system) in our dataset. This step contains the following substeps:

1) **Eliminating the Stop words and Punctuation:** To eliminate stop words, the list of Turkish and English stop words which already defined in Nltk corpus (Perkins, 2014) is used. In addition, punctuations are also removed from the text content.

2) **Removing URL patterns and Hashtags:** hashtag are used to specify important topics in a tweet. The hashtag is a '#' (eg #earthquake), which makes it easier for users to notice a tweet on Twitter. The fact that a topic is a trend topic also depends on the large number of uses of a hashtag. But sometimes users send tweets that are irrelevant to the subject of the hashtag they use, so hashtag can be misleading. Hence, using the tweet text instead of totally depends on the hashtag allows us to find the desired information. In addition, the URLs contained in tweets do not provide us with useful information. In this study, The URLs and hashtags existed in the tweet are removed.
3) **Removing Twitter Reserved Word:** Twitter has some abbreviations, i.e., special words such as "RT, FAV, VIA". Removing such text or tag helps in getting the actual informative content.

In addition to the above sub steps: extra space, single-character word, and emoji are eliminated.

F) Deasciification: As mentioned before, there are six Turkish language-specific characters which are 'ğ', 'ç','ş','ö','ü','ı'. People tend to substitute these Turkish letters by the closest ASCII characters, and even some devices allow the user to use only standard ASCII characters. In this case, the data should be converted into its correct form. This operation is called deasciification. Hence, deasciification is applied to the Turkish language only in this study.

G) Text Normalization: This step consisted of multiple sub steps such as converting all text to the same letter size (upper or lower), converting numbers to word equivalents, handling the misspellings (spatially the ones have done intentionally such as "geliyoruuumm" should be normalized as "geliyorum"), etc.

5.1.1.4 Feature Extraction

Based on our preliminary investigation, the Word2Vec with a pre-trained word vector outperforms others when processing the Turkish language. In general, Word2vec is a two-layer neural network. In our case, the collection of input tweets and the output is a feature vector for each word in the collection as shown in Equation 5.1.

$$\text{Word Embedding } (\text{word}_i) = [f_1, f_2, f_3, \ldots\ldots\ldots f_k] \quad (5.1)$$

Where k is set based on the used approached, however, based on the experimental investigation k=300 was able to achieve the highest performance in most cases, and f is a float number. Hence, k=300 means that each word is represented by a vector of 300 float numbers.

5.1.1.5 Classification

In the proposed system, Convolutional Neural Networks (CNN), which is a deep learning technique is used. The tweets whose features have been extracted and vectorized with Word2Vec are sent to the CNN based system and the system is trained in this section. Seperate CNN models are developed for English and Turkish.

5.1.2 System Testing and Real-time Tweet Classification

Figure 5.2 shows the mechanism of testing and/or using the approach in real-time. In general, after the training process of the system is completed, it will be ready to be used in real time.

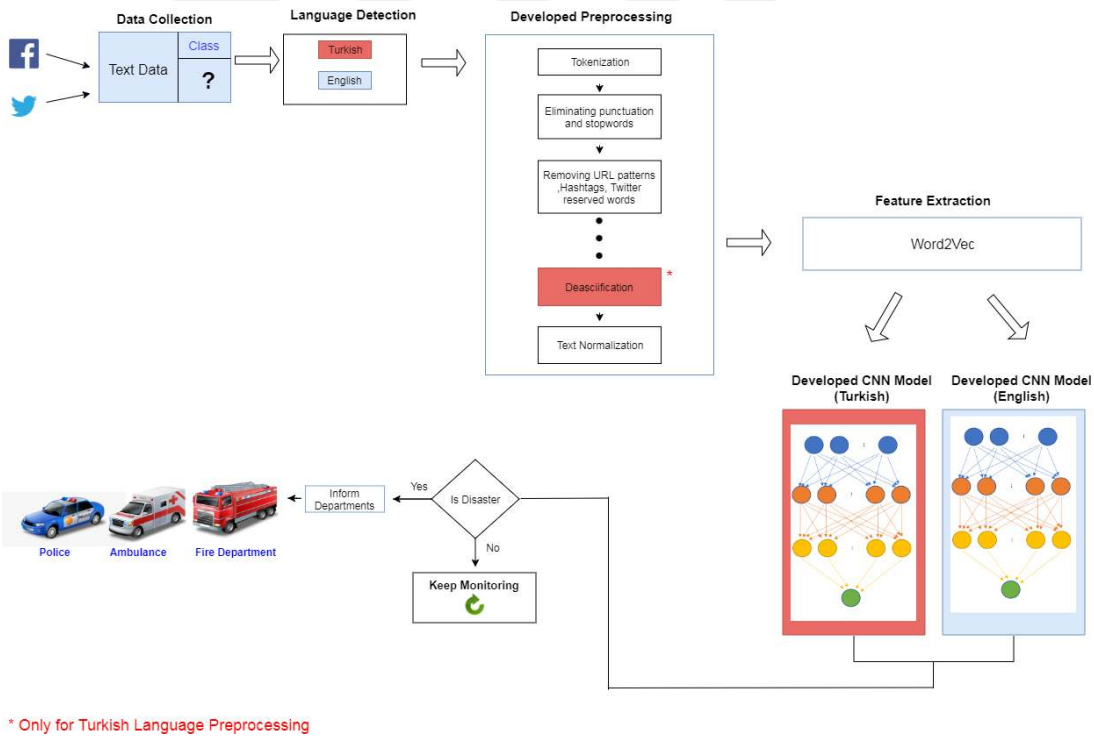


Figure 5.2. Structure of proposed CNN-based approach (testing and real-time)

It is worth mentioning the following: 1) the real-time data can be retrieved by using the Twitter Streaming API. The real-time system aims to detect the classes of tweets in real-time as “disaster” or “not disaster”. 2) The proposed system aims to identify crisis situations by classifying tweets in English and Turkish, and the system performs separate operations for these two languages. Therefore, the language of the data should be determined before proceeding to other stages. Likely, the collected tweeter data has a property that refers to the text language. Hence, this property is used to determine whether the data is in Turkish or English and then the system proceeds to other stages. 3) preprocessing and feature extraction steps are performed same as in the training part.

Finally, in real-time usage of the system, if any crisis situation is detected, the related departments are informed. Otherwise, the system continues to receive data and repeat the same operations until it detects a crisis event.

CHAPTER 6

EXPERIMENTAL STUDY

In this section, the performance of the proposed system for supporting both Turkish and English languages was investigated through multiple experiments. In addition, the performance of the K-Nearest Neighbor (KNN), Naïve Bayes (NB), Support Vector Machine(SVM), and the Random Forest (RF), AdaBoost Classifier (AdaBoost), GradientBoosting Classifier (GBC) ensemble systems have been investigated. It is important to note that related to the "K" value for the KNN, we have tested the values in the range [1, 25] and the Grid Search was used to find the best K value. The same process was done to detect the value of the "number of estimators" for the RF Classifier by testing the values {25, 50, 75,100, 150, and 200}.

To ensure the robustness of the experiments, multiple databases have been used in this study and its details as shown in the following sub-section.

6.1 Datasets

We have reconstructed sub-datasets from the two main datasets to perform the experimental study. In more detail, in the experiments performed to measure the effectiveness of ML methods, four sub-datasets were constructed from the English dataset ("socialmedia-disaster-tweets-relevent") and the number of the samples in the sub-datasets is 2.500, 5.000, 7.500 and 10.860 respectively. In the experiments related to the Turkish language, three sub-datasets were constructed from the newly developed Turkish dataset. These sub-datasets were balanced, i.e., number of disaster and non-disaster tweets is equal, where, the first one is related to the Deprem, i.e., "earthquake", the second one is related to the Yangın, i.e., "fire", and the third one is related to the trafık kazası "traffic accident" that contains 4300, 2000, and 5700 tweets respectively. Related to the English language, another three sub-datasets that contains 3.000, 6.000, and 10.860 tweets respectively were constructed. The details about these two datasets are summarized in the following sections.

6.1.1 English Dataset

The “socialmedia-disaster-tweets-relevent” dataset (Figure Eight, 2015), which includes 10.860 tweets collected using multiple search keywords such as “ablaze”, “quarantine”, and “pandemonium”, then hand classified as relevant or irrelevant, i.e., two classes: disaster and non-disastrous, where the non-disastrous could be a gossip, rumor, joke or a movie review or something else. It is worth mentioning that a very few tweets in this dataset were classified as “Can't Decide” and these ones are ignored. Some samples from this dataset are shown in Table 6.1.

Table 6. 1 Sample of the English dataset's Tweets (Figure Eight, 2015).

Keyword	Tweet	Class
Ablaze	huge fire at Wholesale markets ablaze	Relevant
	#nowplaying Alfons - Ablaze 2015 on Puls Radio #pulsradio	Not Relevant
Quarantine	Officials: Alabama Home Quarantined Over Possible Ebola Case - ABC News"	Relevant
	Are Users of this Sub to be Quarantined?	Not Relevant
Pandemonium	@KhalidKKazi mate they've taken another 2 since I posted this tweet it's pandemonium	Relevant
	Cyclists it is pandemonium on the roads today. Drive carefully!	Not Relevant

6.1.2 Turkish Dataset

Since we could not find any dataset in Turkish, we had to create the dataset ourselves. A few steps have been followed to construct Turkish dataset for training and testing the proposed machine learning-based system:

A) Identification of Disaster Types: In order to collect and create a disaster dataset, the searching keywords were detected related to the following crisis earthquake, fire, traffic accident, flood, etc disaster types.

B) Collection of Tweets: In this step, data is captured from Social media using Twitter Search API. This API allows to collect public tweets. The tweets contain user information, tweet date, location, favorite count, etc. The API also allows filtering tweets by keywords.

C) Manual Classification of Tweets: Three judges (annotators) were responsible for reading the tweets and vote whether each tweet is relevant or not, and then majority voting was applied for the final decision. For example, suppose that a tweet is classified by 3 annotators as 'disaster', 'disaster', 'not disaster' respectively. The class of this tweet is considered 'disaster'. Some samples of the Turkish tweets are shown in Table 6.2.

Table 6. 2 Samples of the collected Turkish tweets.

Topic	Tweet	Meaning	Class
Deprem	İZMİR'DE DEPREM! İzmir merkezde hissedilen şiddetli bir deprem meydana geldi. Ayrıntılar birazdan...	EARTHQUAKE IN IZMIR! A severe earthquake occurred in the center of Izmir. The details are soon ...	disaster
	Galatasaray maçı öncesi yine deprem heyecan dalgası #deprem	The wave of earthquake excitement before Galatasaray match again #earthquake	not disaster
	İzmir sallandı yine #deprem	Izmir rocked again #earthquake	disaster
	Bugün deprem oldu sandım meğersem kalbim atıyormuş	I thought there was an earthquake today, and I realized this happens since my heartbeat.	not disaster
Yangın	Bu aralar içimde bir yangın var. Hem yorgunum. Biraz da suskun	There's a fire inside me lately. And I'm tired. A little reticent	not disaster
	Bu gece dokunmayın bana yüreğim yangın yeri.	Don't touch me tonight, my heart is a fire place.	not disaster
	#Sondakika İstanbul'un Tuzla ilçesindeki bir deri fabrikasında yangın çıktı. Çok sayıda itfaiye ekibi yangına müdahale ediyor	#Newsbreak A fire broke out in a leather factory in Tuzla, Istanbul. A large number of firefighters respond to fire.	disaster
	Andırın'da yıldırım düştü ormanlık alanda yangın çıktı	A fire broke out in the forest because lightning fell in Andirin	disaster
	Hayat zincirleme trafik kazası gibi.. Canım sıkılıyor pazardan dolayı mı acaba ??	Life is like a chained car accident. I'm bored because of the Sunday?	not disaster
	Emniyet şeridi trafik kazası, arıza hâlleri, acil yardım, kurtarma veya kaza incelemesi amacıyla kullanılır. Emniyet şeridini ihlâl etmenin cezası 1.002 TL, vicdani sorumluluğu ise	Safety strip is used for traffic accidents, breakdowns, emergency aid, rescue or accident investigation. The penalty for violating the safety strip is 1.002 TL, and the conscientious responsibility is	not disaster

Trafik Kazası	paha biçilemezdir. Emniyet şeridini gerekmedikçe kullanmayın.	invaluable. Do not use the safety strip unnecessarily.	
	Isparta Lise GFB üyesi Samet Ercan kardeşimiz geçirdiği trafik kazası sonucu yoğun bakıma alınmıştır. Kardeşimize acil şifalar dileriz.	Isparta High School GFB member Samet Ercan was taken to intensive care due to a traffic accident. We wish our brother urgent recovery.	disaster
	Karabük'te zincirleme trafik kazası: 1 ölü, 6 yaralı	A traffic accident in Karabük: 1 dead, 6 injured	disaster
İş Kazası	Aydın'da tren istasyonunda işçi olarak çalışan babası bir kaza sonucu vefat etti. Sonra evleri bir yangında kül oldu. Anne çocuğunu alıp iş bulma ümidiyle İzmir'e taşındı. Ama iş bulamayınca çocuğunu yetimhaneye bırakmak zorunda kaldı.	His father, who worked as a worker at the train station in Aydın, died as a result of an accident. Then the houses became ash in a fire. The mother took her child and moved to Izmir in the hope of finding a job. But when he could not find a job, he had to leave his child in an orphanage.	disaster
	Bununla birlikte iş kazası geçiren çırak yada stajyer kaza sonrası istirahat raporu almışsa; çalışılmadığına dair bildirim, işverenler tarafından değil, primlerinin bildirildiği okul veya işkur tarafından yapılacaktır.	However, if the apprentice or trainee who had an occupational accident received a rest report after the accident; not to be notified by employers, but by the school or employer whose premiums are notified.	not disaster
Sel	Gaziantep'te, kırsal mahallerde yağan sağanak sele dönüştü, hayvanların olduğu ağılı su bastı. Selde maddi hasar meydana gelirken, vatandaşlar taşan dere yatağından geçemedi.	In Gaziantep, the heavy rainfall in rural areas became flood, the barn with animals flooded. While material damage occurred, the citizens did not pass the overflowing stream bed.	disaster
	Haziran ve Ağustos aylarında yaşanan sel felaketinin ardından ilçemiz de ciddi zararlar meydana gelmiştir.	After the flood disaster in June and August, serious damages occurred in our district.	disaster
	Kalbimin yarısı sular altında	Half my heart flooded	not disaster
	Sonra gittin Çocuk oldum bir daha ağladım. Kaç şiir kaç kere sular altında kaldı...	Then you went I became a child I cried again. How many	not disaster

		poems have been flooded, how many times ...	

However, annotators experienced some difficulties in classifying some of the collected tweets. One of these difficulties is tweets sent for mocking or just for criticism purposes, although it is related to the type of disaster. The annotators found it difficult to classify these tweets. Some examples of these tweets are given in Table 6.3.

Table 6. 3 Sample of tweets sent for criticism or joke purposes.

Tweet	Meaning	Disaster
Bu da öyle..Sıradan orman yangını sabotaj olsa bile ABD gibi yangın söndürme filoları olan bir devletin anında söndürmesi gerekir.. Hemde en zenginlerin beldesi. Sözde bazı eyaletler bağımsız olmak istiyormuş. Teksas Kaliforniya ...Onun için vuruyorlardır.	This is the case..Even if ordinary forest fire is sabotaged, a state that has fire fighting fleets like the US should extinguish it immediately. It is the town of the richest. Some states supposedly wanted to be independent. Texas California ... Hitting for this reason.	Yangın
Annemle deprem oldu iddiasını kazanmışımdır. #deprem hissedince evdeki en hisli insan rozetini de kazandım.	I won the claim that there was an earthquake with my mother. When I felt the #earthquake, I won the most sensible human badge in the house.	Deprem
Jungkook trafik kazası geçirdi. armyler hassas olduğu için bir süre kimse trafik, araba, sürücü kelimeleri kullanamaz çünkü dünyada tek araba kullanan kaza yapan jungkook o kadar kaza haberi olan idol gördüm ilk kez oluyo bu	Jungkook was in a car accident. This is the first time I've seen the idol with so much accident news that jungkook is the only car driving accident in the world because nobody can use traffic, car, driver words for a while	Trafik Kazası

6.2 Evaluation

In this thesis, we have used the following standard metrics that can be used for evaluating classification systems:

A) Accuracy: which is the ratio of the total number of correctly classified tweets divide by the total number of tweets and calculated using Equation (6.1)

$$\text{Accuracy} = \frac{TP+TN}{\text{Total}} \quad (6.1),$$

B) Precision: it refers to the ratio between the correct predictions and the total predictions and can be obtained using Equation (6.2).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6.2),$$

C) Recall: Recall, represents the ratio of the correct predictions and the total number of correct tweets in the dataset (Equation (6.3)).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6.3),$$

D) F1 score: F1 score, which is a weighted average of Precision and Recall, consider both false positives and false negatives into account. F1 can be calculated using Equation (6.4).

$$\text{F1 Score} = \frac{2*\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (6.4).$$

6.3 Experiments and Results Analysis

This section provides information about experiments and results analysis of them. Our experimental study consists of two parts. The first part includes experiments for investigating the performance of machine learning algorithms, and the second part of the experiments aims to observe the performance of the proposed CNN classification system.

6.3.1 Investigating the Performance of Machine Learning Algorithms

Experiment 1: The Effects of Pre-processing the Tweets on the Performance of the Studied ML Algorithms

In this experiment, we have investigated the effect of preprocessing the tweets on the performance of each of the studied ML techniques. The main aims of this experiment are to find the degree that performance can be affected by either preprocessing the data or use it directly. In addition, it is expected that some algorithms

can be more sensitive to the noise and unprocessed data than others. In other words, we will find the percentage of improvement, if existed, for each algorithm. It is worth mentioning that the sub dataset, i.e., 10.860 tweets were used in this experiment. The result is shown in Table 6.4.

Table 6. 4 The accuracy of the studied ML algorithms with and without the pre-processing operation.

Algorithm	**Accuracy (%)**		**Improvement (%)**
	Without pre-processing	With pre-processing	
KNN	80.53	84.55	7.9
NB	60.31	60.98	0.5
RF	86.28	91.17	5.9
AdaBoost	80.69	83.96	7.9
GBC	82.50	91.06	8.2

As shown in Table 6.4, it is clear that the preprocessing has improved the performance of almost all the algorithms, and Gradient Boosting Classifier achieved 8.2% as improvement, i.e., it can be considered as the most sensitive. In addition, although, the NB achieved the lowest performance (Accuracy), it has been found that it is the least affected by the noise as its performance with and without the preprocessing is almost same.

Experiment 2: Robustness and Scalability of the KNN, NB, RF, AdaBoost and GBC for English Language

In this experiment, the robustness and scalability of the mentioned techniques were investigated using multiple different size of datasets. Also, in addition to the accuracy, the F1 score, which as mentioned before considers both false positives and false negatives into account has been calculated. Figure 6.1 shows the results of this experiment using the first and second datasets, and Figure 6.2 shows the results using third and fourth datasets, and the results can be summarized as follows.

1) KNN was able to achieve above 80% accuracy for all the datasets, however, it has been defeated by the RF, AdaBoost, and GBC algorithms. On the other hand, the NB, it has achieved the worst performance as compared to the other algorithms using all the four datasets.

2) Related to the Robustness and Scalability while increasing the number of processed tweets, all these classifiers showed that it has stability and can handle all the datasets efficiently.

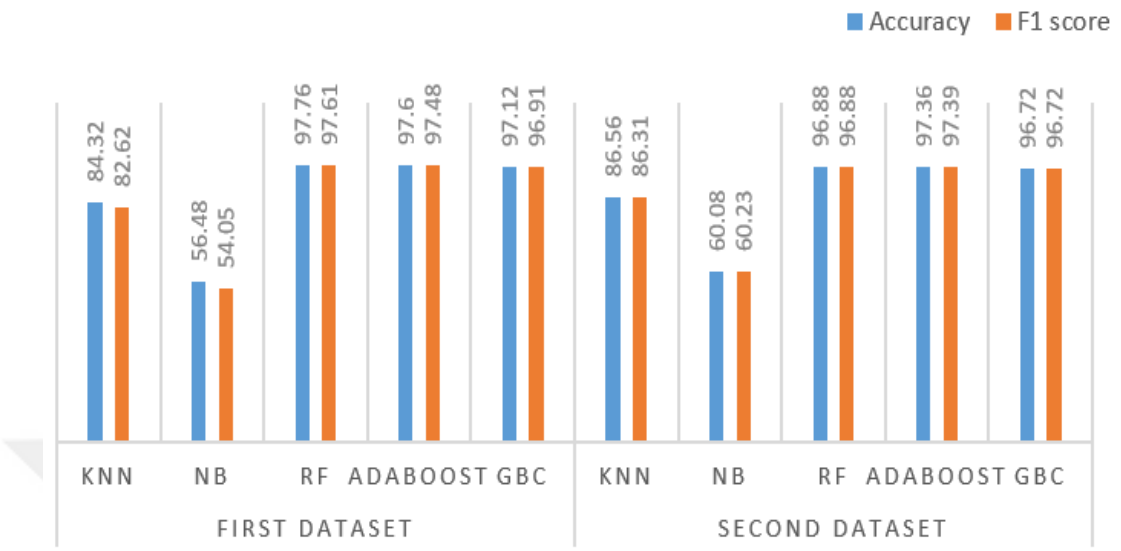


Figure 6.1. The accuracy and F1 score of the studied algorithms using the first and second datasets.

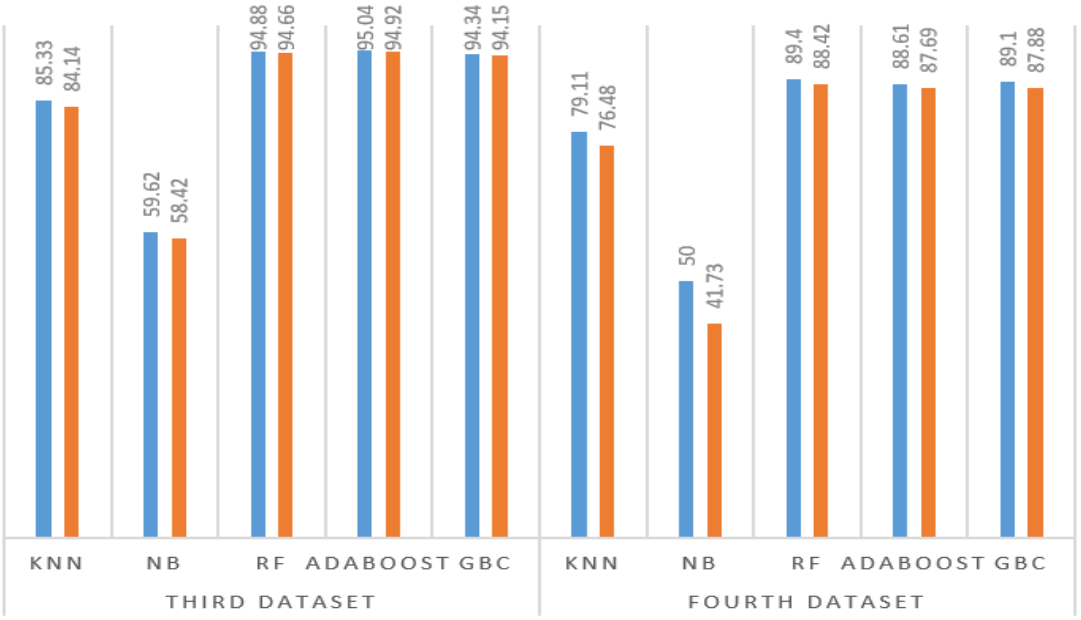


Figure 6.2. The accuracy and F1 score of the studied algorithms using the third and fourth datasets.

Experiment 3: Performance of the Ensemble Systems for English Language

It has been proven that ensemble learning can overcome the weakness of using the classifiers individually and it has the ability to make the system more accurate,

robust and scalable. In this experiment, the performance of the RF, AdaBoost, Bagging Classifier ensemble systems and the system that can be created using the KNN, NB, and SVM have been investigated. The majority voting, i.e., soft, was used to make the final decision on the class that each sample (tweet) belongs to. Based on Tables 6.5 and 6.6, the followings can be derived:

1) All the studied ensemble systems achieved good performance and it is clear that such a system is more stable as compared to any single classifier.

2) Overall, the 4th ensemble system (RF) can be considered as the best one, as it has achieved the highest accuracy for all the datasets.

3) Based on the execution time, shown in Table 6.6, it is clear that RF required the least execution time. Overall, the RF ensemble system has achieved the best performance based on both performance and execution time. It is worth mentioning that the execution time, shown in Table 6.6 includes using the Grid Search algorithm to find the best values for the used classifiers.

Table 6. 5 The accuracy of the studied Ensemble systems using English datasets.

Ensemble Learning #	**Used Algorithms**	**Accuracy (%)**			
		First Dataset	Second Dataset	Third Dataset	Fourth Dataset
1st	{KNN, NB, SVM}	75.84	75.76	74.08	69.50
2nd	AdaBoost	98.08	97.68	95.15	89.78
3rd	Bagging Classifier	98.08	96.96	94.03	90.80
4th	RF	99.20	97.52	95.20	90.84

Table 6. 6 The execution time of each ensemble system for English.

Ensemble Learning #	**Used Algorithms**	**Time (seconds)**			
		First Dataset	Second Dataset	Third Dataset	Fourth Dataset
1st	{KNN, NB, SVM}	20.55	81.45	188.97	378.06
2nd	AdaBoost	7.81	13.20	19.28	28.85

3rd	Bagging Classifier	20.62	39.81	55.09	110.44
4th	RF	3.37	4.99	7.71	13.43

Experiment 4: Robustness and Scalability of the KNN, NB, RF, AdaBoost, Bagging and SVM for Turkish Language

In this experiment, the robustness and scalability of the studied algorithms for Turkish text classification were investigated using multiple different sizes of datasets. Table 6.7 shows the results of this experiment using all the datasets, and Table 6.8 shows the algorithm's execution time. Unlike the previous experiment, in order to get a more clear idea about the execution time for each of the used algorithms, the Grid Search algorithm was not used in this experiment. Hence, the execution time presented in Table 6.8 is the execution time when using the default version of the studied algorithms.

1) Without using the Grid Search, both Random Forest and Bagging have the best performance and showed performance very close to each other when processing the first dataset. On the other hand, for the second and third datasets, AdaBooost has got the highest accuracy among the 6 classifiers.
2) Without using the Grid Search, the performance of Naïve Bayes and SVM classifiers has decreased which leads to obtaining a low accuracy.
3) As shown in Table 6.8, it is clear that the NB algorithm is the fastest, however, when considering both accuracy and execution time, there is not doubt that RF has significantly beaten the others when processing the Turkish language.

Table 6. 7 The accuracy of ML algorithms for Turkish without Grid Search

Used Algorithms	**Accuracy (%)**		
	First Dataset	Second Dataset	Third Dataset
KNN	71.30	86.76	80.56
NB	54.95	72.02	50.10
SVM	50.39	67.89	56.11
AdaBoost	89.05	95.59	94.99
Bagging Classifier	90.60	93.33	93.99
RF	91.79	93.62	93.39

Table 6. 8 Execution time for ML algorithms for Turkish without Grid Search

Used Algorithms	Time (seconds)		
	First Dataset	Second Dataset	Third Dataset
KNN	0.37	0.14	0.05
NB	0.01	0.01	0.01
SVM	2.80	1.50	0.33
AdaBoost	0.44	0.29	0.21
Bagging Classifier	0.42	0.20	0.13
RF	0.08	0.04	0.03

Experiment 5: Performance of Ensemble Systems for Turkish Language

In this experiment, the performance of the possible ensemble systems that can be created using the KNN, NB, SVM, RF, AdaBoost, and BaggingClassifier machine learning algorithms on Turkish have been investigated, and the results are shown in Tables 6.9 and 6.10. It is very clear that the Classifier created from {KNN, NB, SVM} was not able to compete with the RF, AdaBoost, and BaggingClassifier ensemble systems and achieved the worst performance. On the other hand, its execution time was the second-longest time.

Table 6. 9 The accuracy of the studied Ensemble systems using Turkish datasets.

Ensemble Learning #	Used Algorithms	Accuracy (%)		
		First Dataset	Second Dataset	Third Dataset
1st	{KNN, NB, SVM}	68.56	85.16	70.34
2nd	AdaBoost	95.30	95.87	94.99
3rd	Bagging Classifier	92.98	93.71	94.39
4th	RF	94.53	93.90	94.79

Table 6. 10 The execution time of each ensemble system for Turkish datasets.

Ensemble Learning #	Used Algorithms	Time (seconds)		
		First Dataset	Second Dataset	Third Dataset
1st	{KNN, NB, SVM}	106.03	58.16	13.73
2nd	AdaBoost	240.35	856.504	989.11
3rd	Bagging Classifier	75.25	34.21	23.20
4th	RF	9.99	4.50	2.84

6.3.2 Investigating the Performance of Developed CNN Model

In this section, the experiments for investigating the performance of the developed CNN model and its results have been analyzed in detail.

Experiment 1: The Effects of the Developed Pre-Processing Model

The main aims of this experiment are to find the degree that performance can be affected by either preprocessing the data or use it directly. The results are shown in Table 6.11.

Table 6. 11 The accuracy of classification with and without the pre-processing operation.

Dataset	Accuracy (%)		Improvement (%)
	Without pre-processing	With pre-processing	
1st Turkish dataset	86.4	90.5	4.75
2nd Turkish dataset	83.36	87.29	4.71
3rd Turkish dataset	86.13	91.52	6.26
1st English dataset	76	79.67	4.83
2nd English dataset	75.6	78.27	3.53
3rd English dataset	79.56	83.7	5.20

According to the results in Table 6.11, preprocessing improves performance when processing both Turkish and English languages. Another conclusion that can be drawn from this experiment is the improvement of the pre-processing on the

classification performance for the Turkish language reached 6.26%. Therefore, since Turkish is an agglutinative language, preprocessing is a must step to have an efficient system.

Experiment 2: Comparing the Performance of Feature Extraction Techniques for Processing Turkish Language

The aim of this experiment is to observe the performance the Term Frequency (TF), Term Frequency-Inverse document frequency (TF-IDF) and Word2Vec, when used as a feature extraction methods These methods were applied to the system separately and the most suitable method for the Turkish language was determined. The results are shown in Table 6.12. It is clear that the Word2Vec method has signifcantly outperforms the others. Thus, to create an effective system for processing both Turkish and English languages, it is best to use Word2Vec method among these three approaches.

Table 6. 12 Accuracy of TF, TF-IDF, and Word2Vec Approaches when Processing the Turkish and English languages.

Dataset	**Accuracy (%)**		
	TF	TF-IDF	Word2Vec
1st Turkish dataset	57.12	51.53	93.21
2nd Turkish dataset	61.60	55.80	90.00
3rd Turkish dataset	57.32	49.46	88.44
1st English dataset	52.06	48.34	76.03
2nd English dataset	50.30	49.70	77.35
3rd English dataset	59.01	55.47	80.33

Experiment 3: Investigating the Effect of the Input Window Size.

The aim of this experiment is to investigate the effect of changing the number of words in the processed text window for processing both Turkish and English languages. Based on our preliminary experiments, window sizes 1, 3, 5 and 7 were

selected and the performance of the proposed CNN model was measured with the mentioned window sizes. İt can be seen from the results in Table 6.13, that system performance was very close to each other for all the tested values. However, it is important to note that increasing the window size significantly reduces the processing time.

Table 6. 13 The effect of changing the number of words in the processed text window.

Dataset	**Accuracy (%)**			
	window size =1	window size =3	window size =5	window size=7
1st Turkish dataset	92.37	92.09	93.77	92.19
2nd Turkish dataset	90.60	90.60	88.00	86.80
3rd Turkish dataset	88.23	88.18	88.86	87.90
1st English dataset	74.83	76.56	75.50	75.77
2nd English dataset	79.41	77.55	78.28	79.75
3rd English dataset	80.77	80.52	80.37	80.00

Experiment 4: Effect of Changing the Number of CNN's Filters

One of the factors that significantly affects the overall performance of any CNN model is the number of filters used in the model. In this experiment, the effect of changing the number of filters was examined when processing both Turkish and English texts. The number of filters is set to 32, 64, 128 and 256. As in the previous experiment, the presented CNN model was used to investigate the performance of all values tested. As shown in Table 6.14, in most cases, 64 were found to be the best number of filters for the Turkish language. However, in the case of the English language, none of the numbers beat others significantly. On the other hand, related to the processing time 32 filter requires the least amount of time, and 64 requires the second least amount of time.

Table 6. 14 The effect of changing number of used filters.

Dataset	Number of Filters			
	32	64	128	256
	Accuracy (%)			
1st Turkish dataset	92.00	92.65	91.72	92.65
2nd Turkish dataset	90.40	90.00	89.60	89.00
3rd Turkish dataset	88.51	88.23	88.44	87.92
1st English dataset	77.23	75.23	77.23	73.37
2nd English dataset	77.81	77.28	76.82	78.08
3rd English dataset	80.92	79.45	81.29	80.59

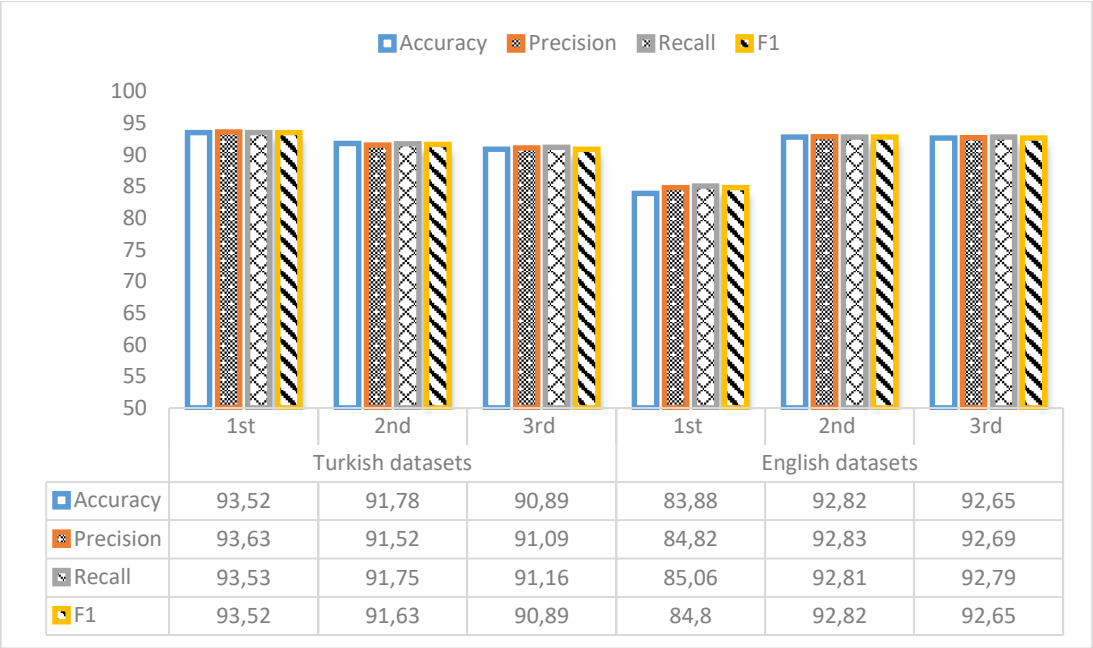
Experiment 5: Robustness and Scalability of the Proposed System

In this experiment, we aimed to observe the robustness and scalability of our developed CNN system. In addition to the accuracy, the system's Precision, Recall, F1 Score values were also calculated. The presented system has been tested on all English and Turkish datasets. Figure 6.3 shows the results and the findings of this experiment are as follows:

1) Over 90% accuracy has been achieved in the system developed for the Turkish. Hence, the proposed system has the ablility to process Turkish datasets effectively.

2) The system has showed good performance generally for the English language. The accuracy of the system using the second and third datasets was more than 90%. This indicates that increasing number of tweets has a significant positive impact on the overall performance of the system.

3) With regard to Robustness and Scalability, while increasing the number of tweets processed, our system has shown that it is stable and can handle all datasets efficiently.



	1st	2nd	3rd	1st	2nd	3rd
	Turkish datasets			English datasets		
Accuracy	93,52	91,78	90,89	83,88	92,82	92,65
Precision	93,63	91,52	91,09	84,82	92,83	92,69
Recall	93,53	91,75	91,16	85,06	92,81	92,79
F1	93,52	91,63	90,89	84,8	92,82	92,65

Figure 6.3. The Accuracy, Precision, Recall, and F1 Score for the Developed System.

CHAPTER 7

CONCLUSIONS AND FUTURE WORKS

In recent years, social media data become one of the main resources for researchers in many areas and lead to improving many approaches such as text classification and all its sub-fields like crisis management systems. In addition, impressive attention has been given to create crisis management systems by mining the publically available huge amount of data on social media to gain situational awareness, which may help in preventing or decrease the effect of some disaster by taking the correct responses.

This study presents the new CNN-based effective Turkish tweet classification system for crisis response. This system can classify data efficiently before or during any crisis. Since the Turkish language is an agglutinative language, an effective preprocessing model has also been developed and integrated as part of the developed system.

In this study, the first Turkish language dataset for crisis responses that can be used in future studies has been created. This dataset is carefully preprocessed and adapted to Natural Language Processing methods.

A wide range of experiments was conducted using the developed Turkish language dataset in addition to the “socialmedia-disaster-tweets-relevent” English dataset. In addition, studies were carried out with the most advanced Vector Space Model techniques to find the most suitable technique for the Turkish language.

Testing and comparing the performance of existing word embedding systems can be a topic for our future research. System development using multiple CNN models in parallel may be another research topic. In addition, the use of RNN for the purpose of developing this study is one of the possible future research topics in this area.

REFERENCES

Aït-Sahalia, Y., Xiu, D. (2019). Principal component analysis of high-frequency data. Journal of the American Statistical Association. **114(525)**, 287-303.

Alam, F., Joty, S., Imran, M. (2018). Graph Based Semi-Supervised Learning with Convolution Neural Networks to Classify Crisis Related Tweets. In *Twelfth International AAAI Conference on Web and Social Media*.

Alexander, D. (2005). Towards the development of a standard in emergency planning. *Disaster Prevention and Management: An International Journal*, **14(2)**, 158-175.

ALRashdi, R., O'Keefe, S. (2019). Deep Learning and Word Embeddings for Tweet Classification for Crisis Response. *arXiv preprint arXiv*. **1903.11024**.

Augustine, N. R. (1995). Managing the crisis you tried to prevent. Harvard business review, **73(6)**, 147.

Ayata, D., Saraçlar, M., Özgür, A. (2017). Turkish tweet sentiment analysis with word embedding and machine learning. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*. 1-4.

Başer, A. (2014). Sosyal medya kullanıcılarının kişilik özellikleri, kullanım ve motivasyonlarının sosyal medya reklamlarına yönelik genel tutumları üzerindeki rolü: Facebook üzerine bir uygulama.

Baygın, M. (2018). Classification of Text Documents based on Naive Bayes using N-Gram Features. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*. 1-5. Ieee.

Benli, M. (2014). *Sosyal medyanın insan kaynakları yönetiminde işe alım sürecinde etkisi: Örnek bir çalışma* (Master's thesis, Maltepe Üniversitesi, Sosyal Bilimler Enstitüsü).

Bifet, A., & Frank, E. (2010). Sentiment knowledge discovery in twitter streaming data. In *International conference on discovery science* (pp. 1-15). Springer, Berlin, Heidelberg.

Blum, J. R., Eichhorn, A., Smith, S., Sterle-Contala, M., & Cooperstock, J. R. (2014). Real-time emergency response: improved management of real-time information during crisis situations. *Journal on Multimodal User Interfaces*, **8(2)**, 161-173.

Breiman, L. (1996). Bagging predictors. Machine learning, **24(2)**, 123-140.

Breiman, L. (2001). Random forests. Machine learning, **45(1)**, 5-32.

Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications*, ***36*(3)**, 5432-5435.

Christian, H., Agus, M. P., & Suhartono, D. (2016). Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*. **7(4)**, 285-294.

Cunningham, P., Delany, S. J. (2007). k-Nearest neighbour classifiers. Multiple Classifier Systems. **34(8)**, 1-17.

Çorbacıoğlu, S., & Kapucu, N. (2005). Critical evaluation of Turkish disaster management: historical perspectives and new trends. *The Turkish Public Administration Annual*, **29(31)**, 53-72.

Çorbacıoğlu, S., & Kapucu, N. (2006). Organisational Learning and Selfadaptation in Dynamic DisasterEnvironments. *Disasters*, **30(2)**, 212-233.

Dağıtmaç, M. (2015). Sosyal medya tercihlerinde kullanıcıyı etkileyen faktörler (Yıldız Teknik Üniversitesi, Sosyal Bilimler Enstitüsü Doktora Tezi), İstanbul.

Daldal, K. M. (2013). Sosyal medyada mobil etiketleme farkındalığı: Sosyal medya tüketicileri üzerinde bir araştırma (Master's thesis, Hitit Üniversitesi Sosyal Bilimler Enstitüsü).

DataCamp. 2018. KNN classification using Scikit-learn. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn. 19.11.2019.

Develioğlu F. 1978. Ottoman Turkish Encyclopedic Dictionary. 30th edition. Ankara: Doğuş Ltd. Şti. Press.

Dietterich, T. G. (2000). Ensemble methods in machine learning. In International workshop on multiple classifier systems (pp. 1-15).

Ergünay, O. (2002). Afet Yönetimi ve Yerel Yönetimler, Ankara, Türk Belediyeler Derneği –ERGÜNAY Konrad Adenauer Vakfı Ortak Yayını.

Erkal, T., & Değerliyurt, M. (2009). Türkiye'de afet yönetimi. Doğu Coğrafya Dergisi, **14(22)**,147-164.

Figure Eight. 2015. Disasters on Social Media, https://www.figure-eight.com/data-for-everyone/ . 10.11.2019.

Frank, E., Bouckaert, R. R. (2006). Naive bayes for text classification with unbalanced classes. In European Conference on Principles of Data Mining and Knowledge Discovery. 503-510.

Freund, Y., Schapire, R. E. (1996). Experiments with a new boosting algorithm. In icml. **96**, 148-156.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, **38(4)**, 367-378.

Ganapati, N. E. (2008). Disaster management structure in Turkey: away from a reactive and paternalistic approach?. *Publıc Admınıstratıon And Publıc Polıcy-New York,* **138**, 281.

Girtelschmid, S., Salfinger, A., Pröll, B., Retschitzegger, W., Schwinger, W. (2016). Near real-time detection of crisis situations. In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).* 247-252.

Güçdemir, Y. (2012). Sanal ortamda iletişim bir halkla ilişkiler perspektifi. İstanbul: Derin Yayınları.

Imran, M., Castillo, C., Lucas, J., Meier, P., Vieweg, S. (2014). AIDR: Artificial intelligence for disaster response. In *Proceedings of the 23rd International Conference on World Wide Web* . 159-162.

Imran, M., Mitra, P., Castillo, C. (2016). Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages. *arXiv preprint arXiv*. 1605.05894.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., Mikolov, T. (2016). Fasttext. zip: Compressing text classification models. *arXiv preprint arXiv*. 1612.03651.

Kalucki, J. 2010. Twitter streaming API. http://apiwiki. twitter. com/Streaming-API-Documentation. 14.07.2018.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, **53(1)**, 59-68.

Kaplan, A. M., Haenlein, M. (2009). The fairyland of second life: About virtual social worlds and how to use them. *Business Horizons*, **52(6)**, 563—572.

Karlik, B., Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*. **1(4)**, 111-122.

Kılınç, D. (2016). The Effect of Ensemble Learning Models on Turkish Text Classification. *Celal Bayar Üniversitesi Fen Bilimleri Dergisi*. 12(2).

Kılınç, D., Özçift, A., Bozyigit, F., Yıldırım, P., Yücalar, F., Borandag, E. (2017). TTC-3600: A new benchmark dataset for Turkish text categorization. *Journal of Information Science.* **43(2)**, 174-185.

Krig, S. (2016). Computer Vision Metrics. Survey, Taxonomy and Analysis of Computer Vision. *Visual Neuroscience, and Deep Learning*. 637.

Kumar, A., Singh, J. P. (2019). Location reference identification from tweets during emergencies: A deep learning approach. *International journal of disaster risk reduction*. **33**, 365-375.

Kurita, T. (2004). Total Disaster Risk Management and the Importance of International Cooperation. Asian Disaster Reduction Center, Japan.

Kwak, H., Lee, C., Park, H., Moon, S. (2010). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web . 591-600.

Li, H., Caragea, D., Li, X., Caragea, C. (2018). Comparison of Word Embeddings and Sentence Encodings as Generalized Representations for Crisis Tweet Classification Tasks. en. In: New Zealand.13.

Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016). Large-margin softmax loss for convolutional neural networks. In *ICML*. **Vol. 2, No. 3**, 7.

Mayfield, A. 2008. What is Social Media. https://www.icrossing.com/uk/sites/default/files_uk/insight_pdf_files/What%20is%20Social%20Media_iCrossing_ebook.pdf. 11.03. 2019.

Middleton, S. E., Middleton, L., & Modafferi, S. (2013). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*. **29(2)**, 9-17.

Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv*. 1301.3781.

Mikolov, T., Chen, K., Corrado, G., Dean, J., Sutskever, L., Zweig, G. 2013. word2vec. https://code. google. com/p/word2vec. 28.09.2019

Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. arXiv preprint arXiv. 1712.09405.

Mitroff, I. I., Shrivastava, P., & Udwadia, F. E. (1987). Effective crisis management. Academy of Management Perspectives, 1(4), 283-292.

Naili, M., Chaibi, A. H., & Ghezala, H. H. B. (2017). Comparative study of word embedding methods in topic segmentation. Procedia computer science. 112, 340-349.

Namatēvs, I. (2017). Deep convolutional neural networks: Structure, feature extraction and training. Information Technology and Management Science. 20(1), 40-47.

Nguyen, D. T., Alam, F., Ofli, F., Imran, M. (2017). Automatic image filtering on social networks using deep learning and perceptual hashing during crises. *arXiv preprint arXiv*. 1704.02602.

Nguyen, D. T., Joty, S., Imran, M., Sajjad, H., Mitra, P. (2016). Applications of online deep learning for crisis response using social media information. *arXiv preprint arXiv.* 1610.01030.

Nikhath, A. K., Subrahmanyam, K., Vasavi, R. (2016). Building a K-Nearest Neighbor Classifier for Text Categorization. International Journal of Computer Science and Information Technologies. **7(1)**, 254-256.

Nusair, K., Erdem, M., Okumus, F., & Bilgihan, A. (2012). Users attitudes toward online social networks in travel. Social media in travel, tourism and hospitality: Theory, practice and cases, 207-224.

Onan, A., Korukoğlu, S., Bulut, H. (2016). Ensemble of keyword extraction methods and classifiers in text classification. Expert Systems with Applications. **57**, 232-247.

Özkaşıkçı, I. (2012). Sosyal Medya Pazarla (ma). Yeni Çağda Sosyal Medya Kullanımı ve Performans Ölçümü, Le Color/Levent Print City, İstanbul.

Pennington, J., Socher, R., Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 1532-1543.

Perkins, J. (2014). Python 3 text processing with NLTK 3 cookbook. Packt Publishing Ltd.

Petak, W. J. (1985). Emergency management: A challenge for public administration. *Public Administration Review*. **45**, 3-7.

Petrović, S., Osborne, M., Lavrenko, V. (2010). Streaming first story detection with application to twitter. *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics*. 181-189.

Ramachandran, P., Zoph, B., Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.

Regester, M., & Larkin, J. (1997). Issue and crisis management: fail-safe procedures. Kitchen, PJ, Public relations: Principles and practice, Thomson, London, 212-222.

Rodriguez, J. J., Kuncheva, L. I., Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. IEEE transactions on pattern analysis and machine intelligence. 28(10), 1619-1630.

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, **121**, 93-107.

Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., Laredo, J. A. (2013). CrisisTracker: Crowdsourced social media curation for disaster awareness. *IBM Journal of Research and Development*. **57(5)**, 4-1.

Safko, L. (2010). The social media bible: tactics, tools, and strategies for business success. John Wiley & Sons.

Sharma, D., Jain, S. (2015). Evaluation of stemming and stop word techniques on text classification problem. *International Journal of Scientific Research in Computer Science and Engineering*. 3(2), 1-4.

Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*. 1409.1556.

Şemgün, H. (2007). Afet yönetimi sistemi ve Marmara depremi sonrasında yaşanan sorunlar. Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, Ankara.

Terpstra, T., De Vries, A., Stronkman, R., Paradies, G. L. (2012). Towards a realtime Twitter analysis during crises for operational crisis management ,1-9.

Thinknum. 2018. LinkedIn country-by-country audience visualized. https://media.thinknum.com/articles/tracking-linkedins-growth-from-american-icon-to-global-player/. 12.12.2019

Timothy Coombs, W., Holladay, S. J. (2006). Unpacking the halo effect: Reputation and crisis management. *Journal of Communication Management*, **10(2)**, 123-137.

Tüfekçi, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colo. Tech. LJ*, **13**, 203.

Unlu, A., Kapucu, N., & Sahin, B. (2010). Disaster and crisis management in Turkey: a need for a unified crisis management system. *Disaster Prevention and Management: An International Journal*, **19(2)**, 155-174.

Vardarlıer, P. (2014). İnsan kaynakları yönetiminde sosyal medyanın rolü. Yayınlanmamış Doktora Tezi, Beykent Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul.

Velev, D., & Zlateva, P. (2012). Use of social media in natural disaster management. *International Proceedings of Economic Development and Research*. 39, 41-45.

Vos, F., Rodríguez, J., Below, R., Guha-Sapir, D. (2010). Annual disaster statistical review 2009: the numbers and trends. *Centre for Research on the Epidemiology of Disasters (CRED)*.

Zielinski, A., Middleton, S. E., Tokarchuk, L. N., Wang, X. (2013). Social media text mining and network analysis for decision support in natural crisis management. *ISCRAM*.