PROCESSING POST-DISASTER SOCIAL MEDIA DATA

(AFET SONRASI SOSYAL MEDYA VERISINI IŞLEMEK)

by

ALI BURAK CAN, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Supervisor: Assist. Prof. Dr. Ismail Burak PARLAK

JUNE 2022

ACKNOWLEDGMENTS

I would like to thank my supervisor Assist. Prof. Dr. Ismail Burak Parlak for his thoroughly support during my dissertation process. I also want to express my special thanks to Prof. Dr. Tankut Acarman who has guided me with his experience.

June 2022

Ali Burak Can

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Social media has become a compelling concept in recent years. Contrary to old media, social media ensures bidirectional communication and sharing between sources and consumers. The most instant information shared by all users can be retrieved with a simple search. Although these platforms, where each person has a microphone, seem to be the right place to get instant information when important events occur, it creates chaos when everyone speaks simultaneously. Thus, it is essential to distinguish the desired information from this wild environment, especially in case of any disaster. Within the scope of this research, studies have been carried out for a smart disaster management system with the help of texts and locations shared on microblogging platforms after a disaster. For an application to conduct experimental research, tweets shared on Twitter during and after the Izmir earthquake happened on 30 October, 2020 were collected. The tweets were annotated as fully- and semi-supervised ways to create big and labeled datasets. Datasets are used for both content-based analyzes and classifications. Since the analysis and classification methods results are promising, it has been proven that the study can be useful in post-disaster scenarios.

Keywords : Natural Language Processing, Smart Disaster Management, Social Media

RÉSUMÉ

Les médias sociaux sont devenus un concept très puissant ces dernières années. Lorsqu'un événement se produit, les informations les plus instantanées peuvent être partagées par tous les utilisateurs, tandis que ces informations instantanées peuvent être présentées aux utilisateurs avec une simple recherche par tous les utilisateurs. Bien que ces plateformes, où chaque personne dispose d'un micro, semblent être le bon endroit pour obtenir des informations instantanées lorsque des événements importants se produisent, cela crée le chaos lorsque tout le monde parle en même temps. Ainsi, il est essentiel de bien distinguer les informations recherchées de cet environnement sauvage, notamment en cas de sinistre. Dans le cadre de cette thèse, des études ont été menées pour un système intelligent de gestion des catastrophes à l'aide de textes et de lieux partagés sur des plateformes de microblogging après une catastrophe. Pour l'étude de cas, les tweets partagés sur Twitter pendant et après le tremblement de terre d'Izmir du 30 octobre 2020 ont été collectés. Les tweets ont été annotés comme entièrement et semi-supervisés et sont devenus des ensembles de données étiquetés. Les ensembles de données sont utilisés à la fois pour les analyses et les classifications basées sur le contenu. Étant donné que les méthodes d'analyse et de classification ont été couronnées de succès, il a été prouvé que l'étude peut être utile dans les scénarios post-catastrophe.

Mots Clés : Gestion intelligente des catastrophes, médias sociaux, traitement du langage naturel

ÖZET

Sosyal medya son yıllarda çok güçlü bir kavram hâline geldi. Herhangi bir olay yaşandığında en anlık bilgiler tüm kullanıcılar tarafından verilebiliyor bununla birlikte bu anlık bilgiler tüm kullanıcılar tarafından basit bir aramayla kullanıcılara gösterilebiliyor. Her bir kişinin mikrofonu olduğu bu platformlar önemli olaylar yaşandığında anlık bilgileri almak için en doğru yer gibi gözüküyor olmasına rağmen, tabiri caizse herkesin aynı anda konuşması bir kaos yaratıyor. İstenilen bilginin bu vahşi ortamdan ayırt edilmesi özellikle herhangi bir afetin yaşandığı durumlarda zaruridir. Bu çalışmada afet sonrası mikro blog sitelerinde paylaşılan metinler ve konumlar yardımıyla akıllı bir felaket yönetim sistemi için çalışmalar yapılmıştır. Vaka çalışması için 2020 yılında Izmir depremi sırasında ve sonrasında Twitter üzerinde paylaşılan tweetler toplanmıştır. Tweetler denetimli ve yarı denetimli olarak etiketlenerek veri setlerine dönüştürülmüştür. Veri setleri üzerinde hem metin ve konum çıkarımları içeren içerik tabanlı analizler yürütülmüştür, hem de etiketli veri setleri üzerinde sınıflandırma çalışmaları yapılmıştır. Analiz ve sınıflandırma sonuçları başarı gösterirken yapılan çalışmanın afet sonrası senaryolarda işe yarar olabileceği kanıtlanmıştır.

Anahtar Kelimeler : Akıllı Afet Yönetimi, Sosyal Medya, Doğal Dil İşleme

1 INTRODUCTION

The usage rates of online communication environments, which have taken place in our lives as social media, are gradually reaching high levels. With the developments and prevalent use of information technologies, social media environments that meet the communication needs of people on the internet have a direct or indirect effect on almost everyone's life. Different views have been put forward in the literature about the definitions and functions of social media. Social media is used as an umbrella term and it is a platform of photos, videos, texts and other materials generated to communicate, share information and create content online (Aichner et al., 2021). Offering a wide range of sharing opportunities, social media largely meets the expectation arising from the disadvantages of unilateral old media (Edosomwan et al., 2011). Even though it is still necessary and usable, the old media, in which the information produced from a single center was delivered to the public by mass media, has left itself to the social media, where everyone can produce and disseminate information. Mayfield has summarized the definition of social media under five concepts (Mayfield, 2008). These concepts are briefly as follows ; Participation, everyone contributes. Openness, everyone can benefit from the content. Conversation, anyone can communicate two-way. Community, users with common interests can gather and establish a community. Connectedness, other online resources are also available. Another definition of social media (Liu et al., 2016) describes the social media as good, bad and ugly, shown in Figure 1.1. Social media is "good" because the data is big and connected. It is "bad" because the data is noisy and sparse. And it is "ugly" because the data is heterogeneous, multi-source, partial, and asymmetrical.

The instantaneous reactions of social media users to any incident can create a significant insight into situations that require urgency. Any natural or man-made disaster is a situation that requires swift action. Organizations that need to take quick action should use the power of social media to improve their decision-making processes. When a disaster occurs, the most reliable and immediate data is imperative in order to act in the most accurate and fastest way. Apart from social media, the methods to carry out analysis take a long time to collect people's reactions to disasters, such as making phone calls, observations in close quarters, and meeting one-on-one (Simon et al., 2015). Even

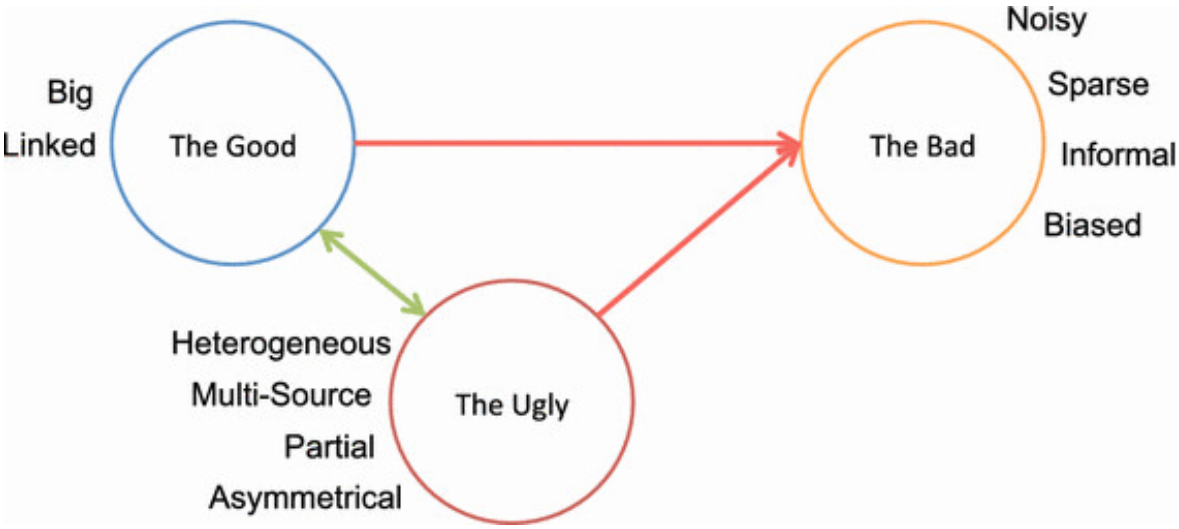

FIGURE 1.1 – The Good, the Bad, and the Ugly of Social Media Data (Liu et al., 2016)

though they can be used to take precautions or prepare for potential future disasters by generating projections of the disasters (Botzen et al., 2019), they are disadvantageous for generating big and enough data and taking action immediately after the disaster.

Social media environments such as Twitter are actively used by both legal and non-institutional users. Online communication between users continues at all stages of a disaster situation. In particular, getting information from places close to the scene is necessary to guide communities better and provides resiliency (Aldrich, 2010). The fact that victims, organizations and people closely related to the disaster communicate on the same platform greatly contributes to the disaster management processes (White et al., 2009). Studies show that assessments of city officials' ability to contain a disaster and the strength of people's responses are positively correlated with social media coverage (Graham et al., 2015).

Emergencies are the events caused by natural or human factors, causing human losses, environmental damage, material losses and psychological effects, threatening and disrupting people's lives and livelihoods. In disaster management, social microblogging becomes a knowledge host where people search for information resources in their community. However, the complex behaviour of the risk is not trivial to handle in massive disasters. It requires the hierarchical organization of locations and the description of the emergency through the feedbacks from the inhabitants. The feedback generated in Twitter about disaster situations contains excessive information for the emergency calls. Nevertheless, unexplained information is considered as noise and true emergency calls might be ignored if the microblogs are not properly classified. Thus, true predic-

tions and classifications apparently accelerates humanitarian and development aid. In text-based disaster analysis, an accurate model is required to tackle with live information in a stack of flooding data for emergency and non-emergency cases. Twitter's role here is to provide all kinds of information about disaster incidents such as cautions, pictures and videos, locations and warnings. Therefore, microblogs can be used to conduct smart disaster management, especially regarding the knowledge it contains. In this case, natural language processing becomes an extension of data extraction for textual data, where all activities are analyzed with earthquake related terms. The lexical analysis arises the semantic gap between emergency/rescue and non-emergency/non-rescue classes. The significance of the microblogs might be ranked through their distance to epicenter of the disaster. The priority of the disaster knowledge is associated to the relevance in social media and mobile applications. In a similar way, collaborative mapping and crowd sensing tools would be applied in data analysis during action planning and emergencies. Our study focuses on this problem by considering the classification routines as a potential premise in disaster analysis given by the community.

The phases of disaster management have been generally defined as four major stages. As shown in Figure 1.2, these stages consist of prevention (mitigation), preparedness, response and recovery. The prevention and preparedness phases basically tell us that we must be ready for any kind of disasters, while the response and recovery stages encompass the actions to be taken for post-disaster scenarios. This research proposes analysis and methods that can be feasible for each phase of disaster management. Processing one of the most alive platforms where people share their quick and powerful responses after a disaster can accelerate the rescue teams, *Response Stage*. Opinions and needs are shared on social media when they demand a remedy for affected people and regions. Acquiring these remedies with the help of processing social media can be useful on *Recovery Stage*. The knowledge extracted from all disasters is viable to project the next potential disasters which means that it can be used for *Prevention Stage* and *Preparedness Stage*.

Within the scope of this study, content-based analysis and disaster-related classification have been carried out to make disaster management smarter. 30 October 2020 Izmir earthquake was chosen as a case study for applying of the analysis. The tweets posted during this earthquake are collected. These tweets are examined spatio-temporally and topic wise. When they are examined spatio-temporally, the tweets containing location info are extracted and their distribution in the rescue and non-rescue classes are obser-

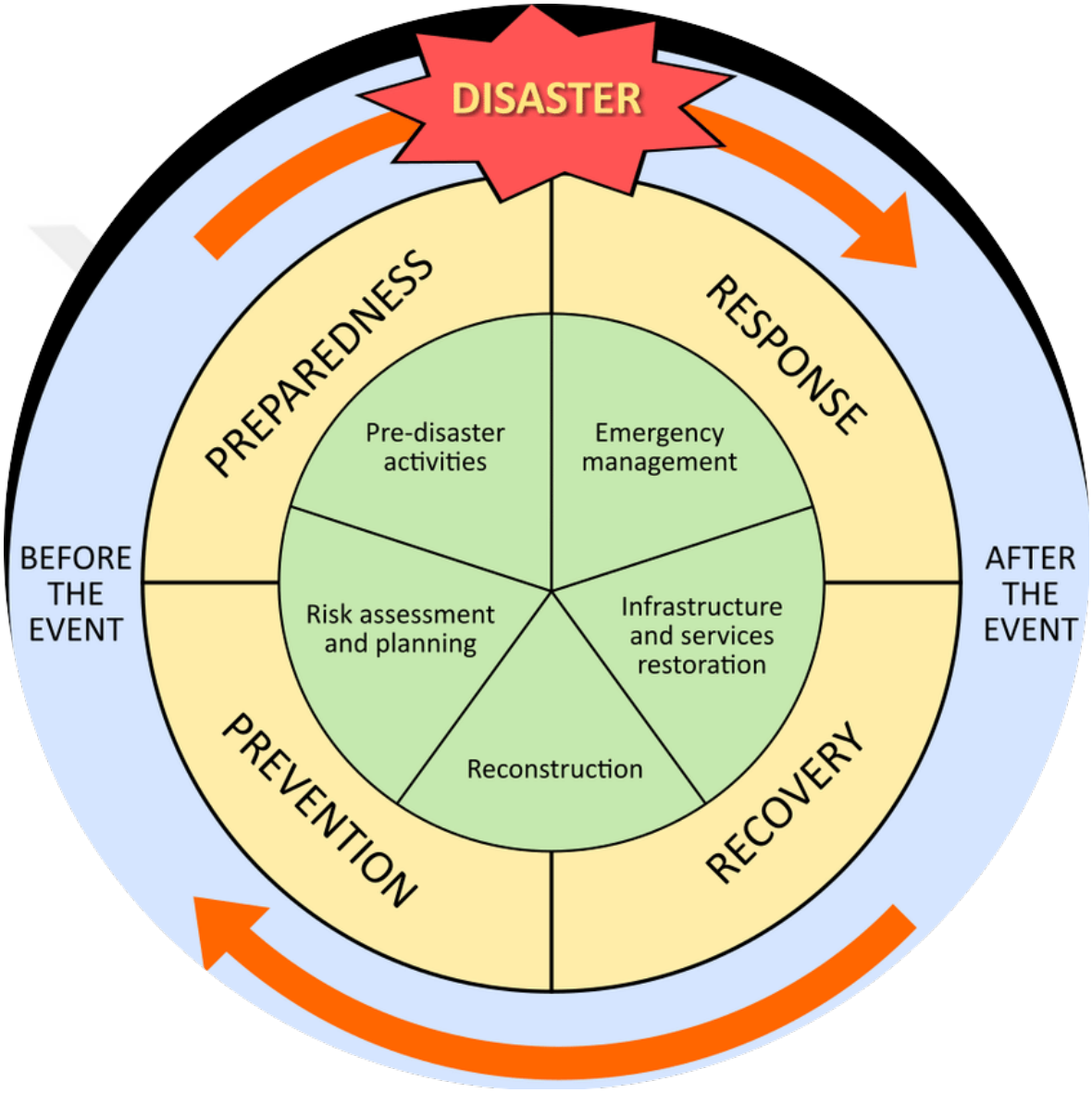

FIGURE 1.2 – Stages of Disaster Management (Erdelj et al., 2017)

ved in both Izmir and Turkey focused. While modeling the topics, the Latent Dirichlet Allocation (LDA) (Blei et al., 2003) algorithm is used, and topic-specific words are extracted specifically for the rescue and non-rescue classes.

This study also proposes a new approach to investigate the social media for disaster management. Twitter usage during an earthquake becomes a multimodal backbone in order to share the knowledge through the different aspects of the disaster. Planning the emergencies is the bottleneck of the rescue organizations in time-limited rescue intervention. Exploring the general population in the epicenter of earthquake would provide vital knowledge in rescue planning. Our method analysis revealed the most important disaster topics that can be derived so that rescue organizations can successfully utilize such data. Our results provide insights into the spatio-temporal distribution of earthquake rescue/non rescue terms to identify Twitter-based discussions related to the 2020 Izmir earthquake.

Our study considers a dataset for the 2020 Izmir earthquake in the post-disaster phase for the processing of responses. In order to analyze the responses two classes, rescue and non-rescue, have been created. Tweets become meaningful to identify the spatio-temporal severity of earthquake fluctuations and validate the rotation of emergency and rescue organizations in disaster zones. Our main contribution is the spatio-temporal analysis of rescue/non-rescue classes by creating manually annotated corpus and classifying rescue tweets. Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018) transformer model has been trained to generate automatic labels. The performance of smart disaster classification has been measured through several methods. These methods consists of a broad range of algorithms from classical machine learning approaches to the state-of-the-art transformer-based language models.

Rest of this dissertation is organized as it follows. Section 2 displays the relevant studies which are conducted before. The datasets used in this study and how they are gathered are explained in section 3. While content-based analysis and results have been thoroughly carried out in section 4, Section 5 encompasses the methodology and the main steps of classifiers and the evaluation process. Moreover, the results are showed and reviewed in their own sections. Consequently, we have concluded our study by highlighting the key points and our future steps in Section 6.

2 LITERATURE REVIEW

Most of the researches related to disaster management are mainly based on traditional operations research and management science practices in the past. However, the use of big data information technology and social media has recently become an integral part of disaster management. Relevant information from social media and social networks increases the situational awareness of decision makers. Disaster management decisions directly or indirectly impact significantly the safety of the victims, the environment, the monetary issues, organizations etc. Reliable, timely, stable, sufficient and qualified information plays a critical role in the stages of disaster management. In this section, a literature study is shown by considering big data and social media in the field of disasters.

Various scholars have investigated the potential of generated big data using different techniques. Table 2.1 and 2.2 summarize the relevant studies that are conducted before making big data usable for smart disaster management. While Table 2.1 mentions the researches on content-based analysis, Table 2.2 points out more on classifications and descriptive analysis. It describes their motivations in the Scope & Aim column. Proposed techniques to deal with the challenges of their goals are given in the Methods column. Dataset tells where the researchers collect their data from. In the disaster column, the occurred catastrophes chosen to be examined by the researchers are listed.

TABLE 2.1 – Related Works of Content-based Analysis

	Disaster	**Dataset**	**Scope & Aim**	**Methods**
(Pourebrahim et al., 2019)	Hurricane Sandy-USA-2012	Twitter	Choice of communication dynamics through the keywords and sentiments	Telephone-based survey Web-based analysis Spatio-temporal analysis Text analysis Sentiment analysis
(Fayjaloun et al., 2021)	Barcelonette earthquake-France-2014 Le Teil earthquake-France-2019	Physical Sensors Twitter	Integration of shakemaps and social sensors	Shakemap analysis Named-entity recognition Text analysis
(Bashir et al., 2021)	Khan Shaykhun chemical attack-Syria-2017	Twitter	Social awareness for civic disaster response	Sentiment analysis Trend analysis Geo-mapping the tweets Word visualization
(Beedasy et al., 2020)	Deepwater Horizon oil spill-Gulf of Mexico-2010	Twitter	Community discourse during social interactions	Social network analysis Trend analysis Cluster analysis
(Behl et al., 2021)	Nepal earthquake-2015 Italy eartquake-2016 COVID-19	Twitter	Emergency reponse for disaster management	Multilayer perceptron based classification Word visualization
(Neppalli et al., 2017)	Hurricane Sandy-USA-2012	Twitter	Emergency response	Sentiment analysis Geo-mapping the sentiments
(Mora et al., 2015)	Canterbury earthquakes-NZ-2011	Twitter	Public perceptions following the earthquakes	Text analysis Focus groups
(Bhavaraju et al., 2019)	Tornado Forest Fire Floods-USA Winter storm	Twitter	Social media sensitivity in natural disasters	Breakout detection Spatial analysis Sentiment analysis Social vulnerability

TABLE 2.2 – Related Works of Predictive Analysis and Classification

	Disaster	Dataset	Scope & Aim	Methods
(Zhai and Yuan, 2020)	Hurricane Florence-USA-2018	Twitter	Measuring the effects of neighborhood equity on disaster awareness	Geo-mapping the tweets Sentiment analysis Latent semantic analysis Spatio-temporal categorization
(Hasfi et al., 2021)	Indonesia 2019 fire and haze	Twitter	Measuring the civic engagement	Text analysis Social network analysis Geo-mapping
(Mangalathu et al., 2019)	South Napa, California earthquake-2014	Twitter	Classification of earthquake-impacted buildings	Text analysis LSTM based classification Word cloud Spatial analysis
(Kankanamge et al., 2019)	South East Queensland Flood-2010	Twitter	Describing the severity of the disaster through local communities	Descriptive analysis Content analysis Spatial analysis
(Kankanamge et al., 2020)	Storm-Fire-Cyclone-Australia-	Twitter Facebook	Engaging disaster management through social communities	Community analysis Cluster analysis
(Eligüzel et al., 2021)	Elazig earthquake-2020	Twitter	Representing the content of earthquake meaning through hashtags	Latent semantic allocation Social spider optimization Cluster analysis
(Cerna et al., 2022)	Firefighter-2012-2020 Emergency Heating Storm-Flood	Interventions of firefighter department	Predicting the number of interventions for capacity planning	CNN LSTM Transformers-BERT models
(Zhu et al., 2022)	Flood	GIS Landsat images	Disaster forecasting through vulnerability zones	Random Forest
(Zhang and Wang, 2022)	Flash flood disaster-2000-2019	Mountain flood IFENG	Developing flood loss assessment model for predicting suitable loss rate	Fuzzy-Rough set RBF neural network

3 CORPUS

The datasets used in this paper are explained in this chapter. The chapter demonstrates the data process from beginning to end. The very first step is obtaining raw data. The analysis, modeling and classification studies we are eager to dig into at first sight need a labeled dataset. Therefore, we have created both manually and automatically annotated corpuses. The preprocessing step is crucial for both topic modeling and classical machine learning approaches. The preprocessing step also gives us promising insights at the end of the study.

3.1 Data Collection

During the data collection process, two different property sets are considered to find earthquake related tweets. Whilst the first property set consists of time and word, the second property set additionally includes popularity and excludes word.

Based on the Izmir earthquake that happened on 30 November 2020, approximately 1 million Turkish tweets containing the word 'deprem'(earthquake) posted on Twitter within a total of 1 year from 30 October 2020 to 30 October 2021 are collected. In order to obtain these data, time- and word-based searches were performed with the help of the Twitter API. The exact search term to be written on Twitter explore page is "deprem since :2020-10-30 until :2021-11-30".

In addition, approximately 250,000 Turkish tweets are acquired by popularity- and time-based search. These tweets have at least 10 retweets that portray popularity and they are posted within 2 weeks, starting from 30 November 2020, which is the occurring time of the earthquake. It can be searched by writing "min_retweets :10 since :2020-10-30 until :2021-11-15".

The two collected datasets containing different earthquake related properties are finally merged with the exception of duplicates. The final dataset possess more than 1,2 million tweets with their corresponding date-time and coordinate(longitude and latitude) information.

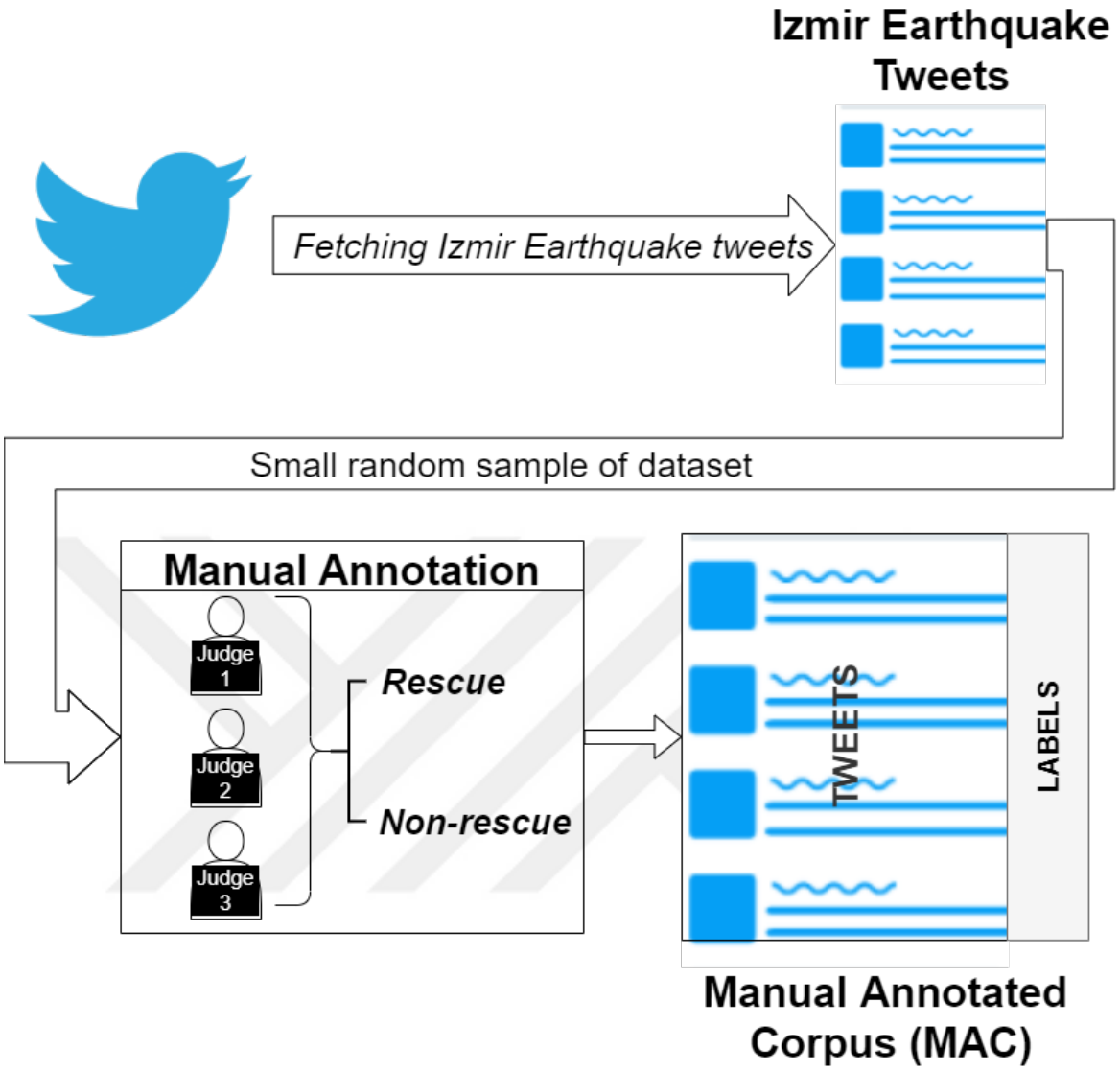

FIGURE 3.1 – Generation of Manual Annotated Corpus (MAC)

Some statistical explorations of the dataset are as follows :

— Even though the tweets have no retweets, 23% of the tweets are comprised of mentions. Mention means that another Twitter account is referred to in a tweet. Similarly, replies are 72% of the mentions. Replies, namely, are the tweets that respond to another one.

— The total number of characters of tweets on average is 146 and the average number of words per tweet is 18.

— Since the anonymity on Twitter, every user does not share their information. Therefore, location included tweets are only 4% of the final dataset. These coordinates come from 97 different cities, majorly Turkish.

— Regarding the total number of tweets per day, the first days after the earthquake

are much higher than the others. Expectedly, 4% of the entire tweets were posted in the first following three days after the earthquake.

3.2 Creating Manual Annotated Corpus

Figure 3.1 describes how the Manual Annotated Corpus(MAC) is created. Among 1,2 million tweets, 3 different judges labeled approximately 10000 tweets as rescue or non-rescue. The tweets that the majority of judges do not agree on their class are filtered out. The remaining 6000 tweets create Manual Annotated Corpus. Class distribution in MAC is equal for classes which are 50% rescue and 50% non-rescue.

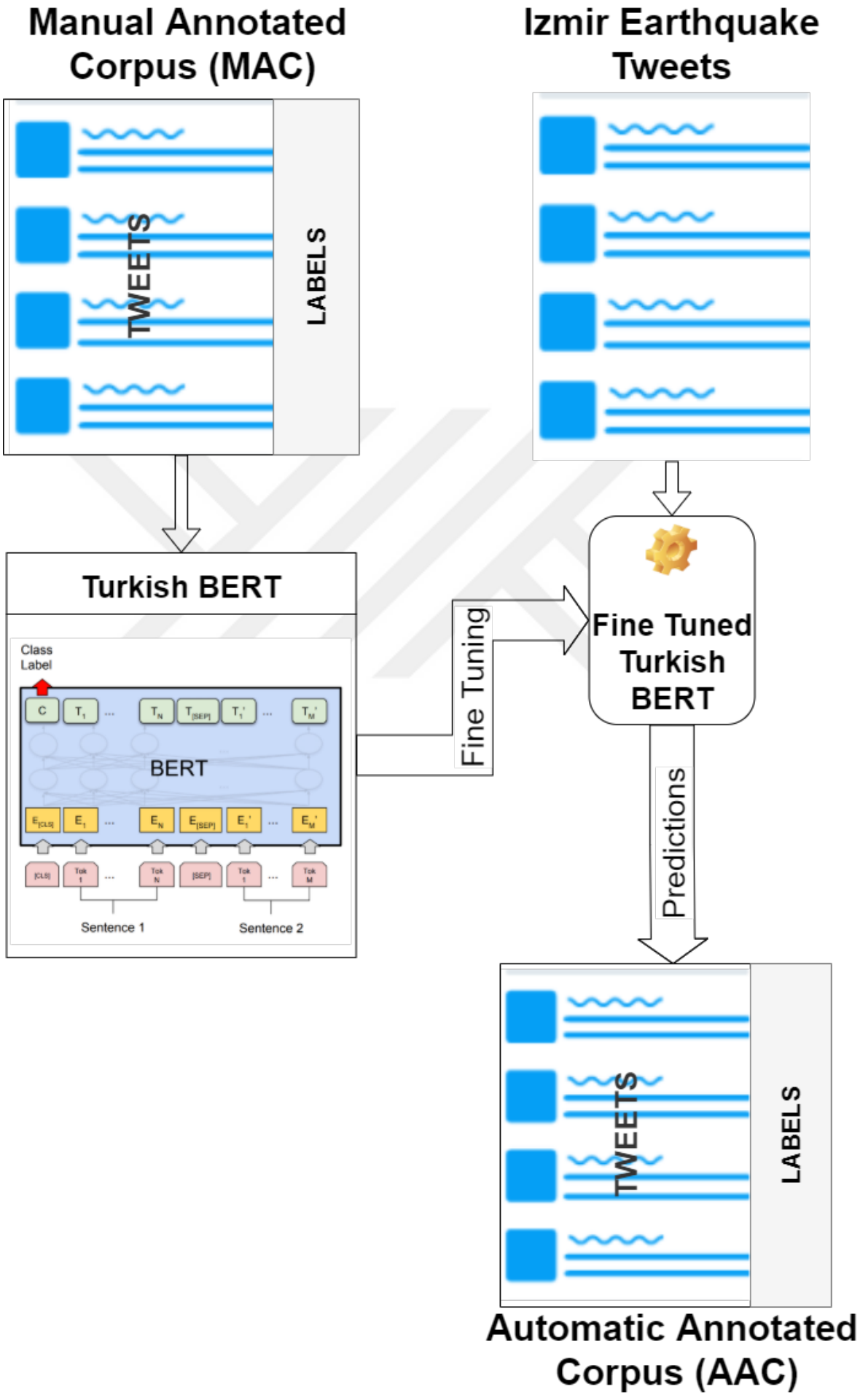

FIGURE 3.2 – Generation of Automatic Annotated Corpus (AAC)

3.3 Creating Automatic Annotated Corpus

How Automatic Annotated Corpus (AAC) is created is shown in Figure 3.2. First, a pre-trained BERT model named Turkish BERT (Schweter, 2020) which is trained on large Turkish corpus is obtained. Then, this pre-trained model fine-tuned by using the MAC dataset containing 6000 annotated tweets. Fine-tuned model predicted all tweets as rescue or non-rescue in collected data from Twitter. In the final step, 1,2 million tweets with corresponding predicted labels, Automatic Annotated Corpus is created.

3.3.1 Preprocessing

1. Extracting contact numbers : By using regular expressions, all texts which match with the turkish phone number format are converted to a single word 'contact-number'.
2. Extracting image links : When collecting tweet, if a tweet contains image, the image link is converted to a single word 'imagelink'. According to the primitive observations, people support their announcements with an image on Twitter. Even though containing 'imagelink' is not distinguishing for tweet classification, the images attached to rescue tweets can be noteworthy. Hence, image links are encapsulated in one word.
3. Extracting announce emojis : According to our observations, people use announce emojis when they tweet about something important. These emojis are represented by a single word 'announcemoji'. This emojis are warning, pushpin, megaphone, loudspeaker and exclamation mark characters which are used for dissemination of warning and informing.
4. Remove mentions : Mentions which start with are removed.
5. Remove links : Except image links, all hyperlinks are deleted.
6. Remove Emojis : Except announce emojis, all emojis are removed.
7. Normalize tweets : Spellings are corrected by using a Turkish Natural Language Processing library named Zemberek (Akın and Akın, 2019).
8. Remove hashtags : Hashtags are the words which starts with # in a tweet. All of them are removed.
9. Remove punctuations : Punctuations in tweets are deleted.

10. Strip numbers : Numbers excluding contact numbers are removed.
11. Strip whitespaces : \n \r \t etc. whitespace characters are removed.
12. Correct accented turkish letters : Old Turkish letters like â, î are converted to the newer ones like a, i.
13. Remove all non-turkish characters : All non-turkish characters are excluded to make sure that unwanted characters don't remain.
14. Remove stop words : Frequent and comparatively meaningless words are removed.

4 CONTENT-BASED ANALYSIS

4.1 Spatio-temporal Analysis

When an earthquake or any disaster occurs, the time and location information is imperative. These two different types of information can be evaluated together and used to guide the rescue teams in the most accurate way. In this study, although the number of tweets containing location information corresponds to a small part of the total dataset, it contains an adequate amount of location information that can be analyzed spatially and temporally. With the help of tweets containing location information in the dataset, we can observe the location of the user who posts tweet. Besides, with the time information in the data set, there is a chance to examine on a time basis on the map. While these figures 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 show that the heatmaps of tweets as 3-day rescue and non-rescue in the focus of Turkey, these figures 4.7, 4.8, 4.9, 4.10, 4.11 and 4.12 focus on Izmir. The heat points shown in red, yellow, green and blue tones on the heatmap represent the regions where the most amount of tweets posted through least, respectively.

On the turkey maps, it can be observed how the rescue and non-rescue tweets changed in the first three days. As can be seen, all tweets about rescue and non-rescue are mostly sent from Izmir, which shows the Izmir region more densely on the heat map. While most of the rescue tweets are seen at high density on İzmir, as can be assumed, the density on the maps of the non-rescue tweets is distributed throughout Turkey along with İzmir and high populated areas such as Istanbul.

The maps focusing on Izmir and its neighboring districts, which is the epicenter of the earthquake, shows how the heat map changes in the first three days. For the maps showing both rescue and non-rescue tweets, it seems that the densest day is the first day, so the density decreases towards the second and third day. While the majority of tweets in the Rescue heat map are concentrated in the center of Izmir for three days, the intensity spreads to the surrounding provinces in the non-rescue heat map.

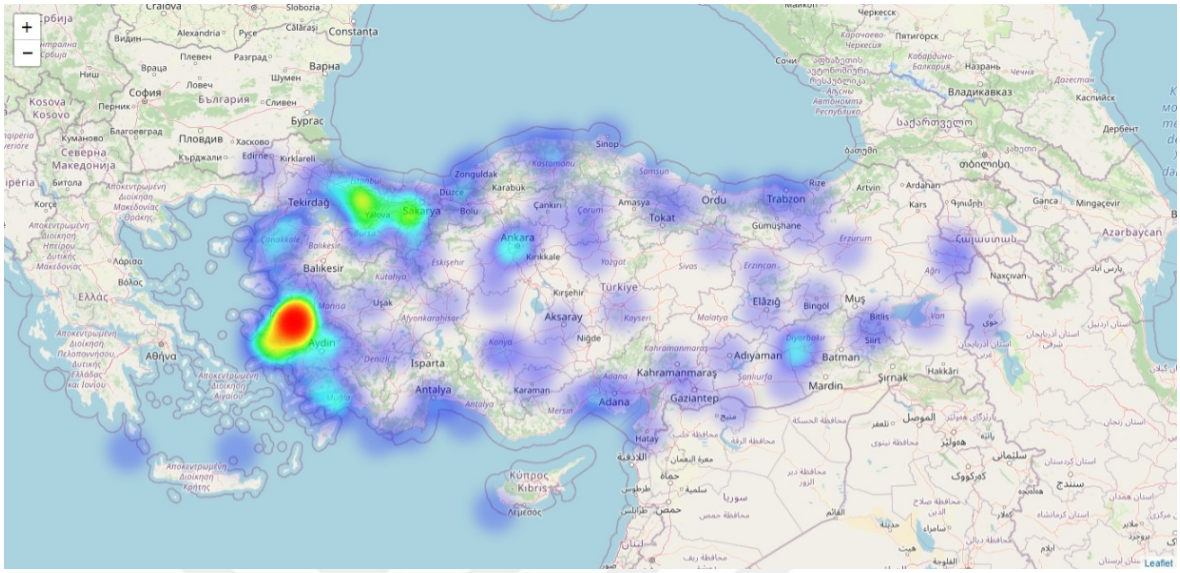

FIGURE 4.1 – Map of Turkey for First Day Rescue Class

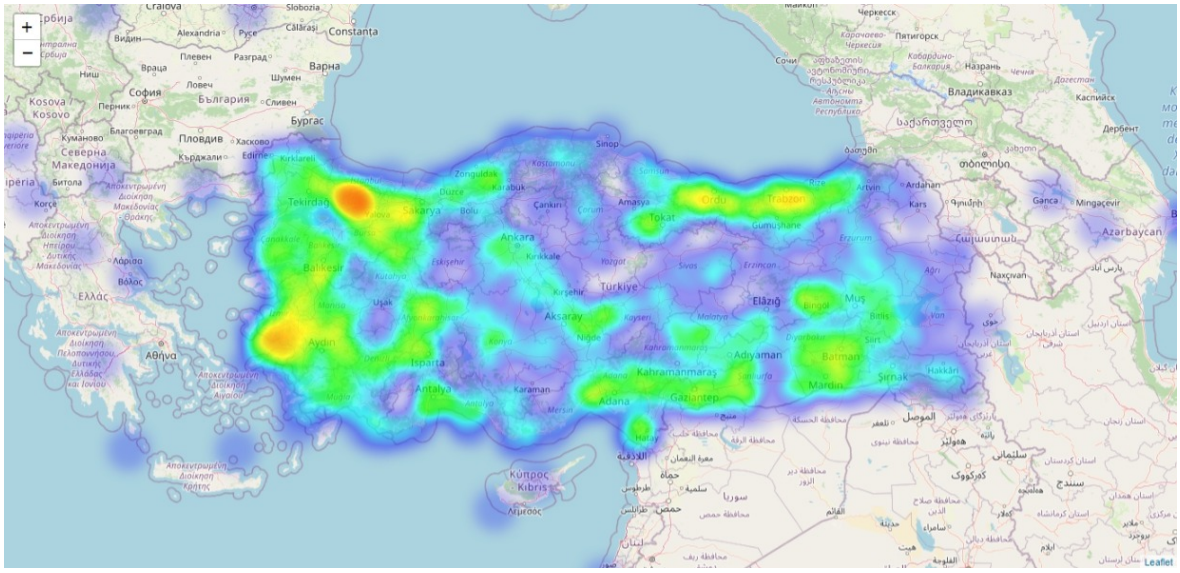

FIGURE 4.2 – Map of Turkey for First Day Non-rescue Class

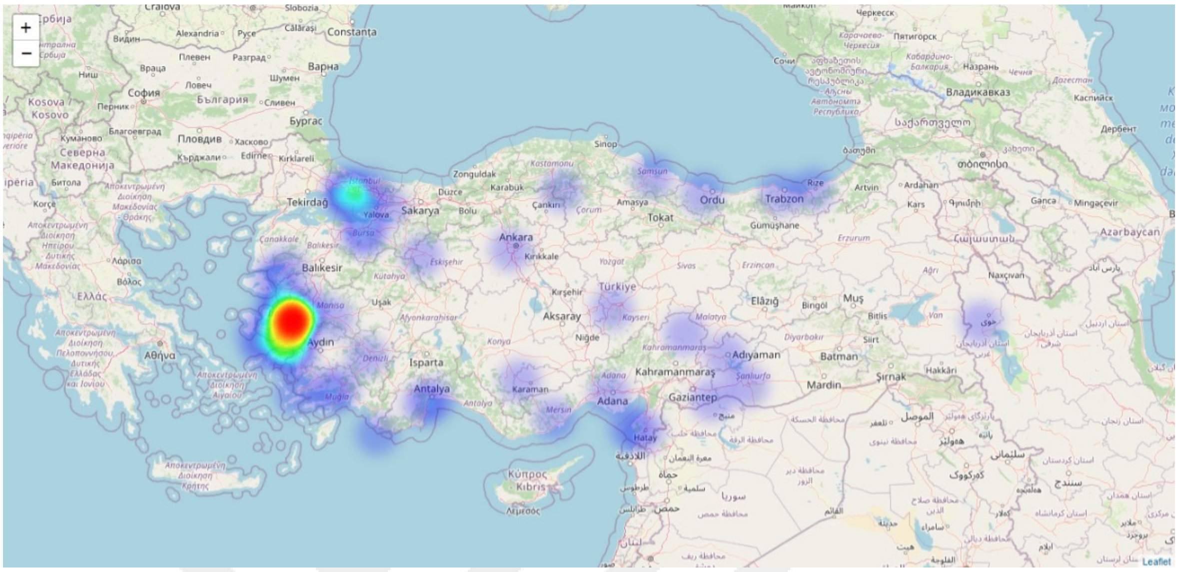

FIGURE 4.3 – Map of Turkey for Second Day Rescue Class

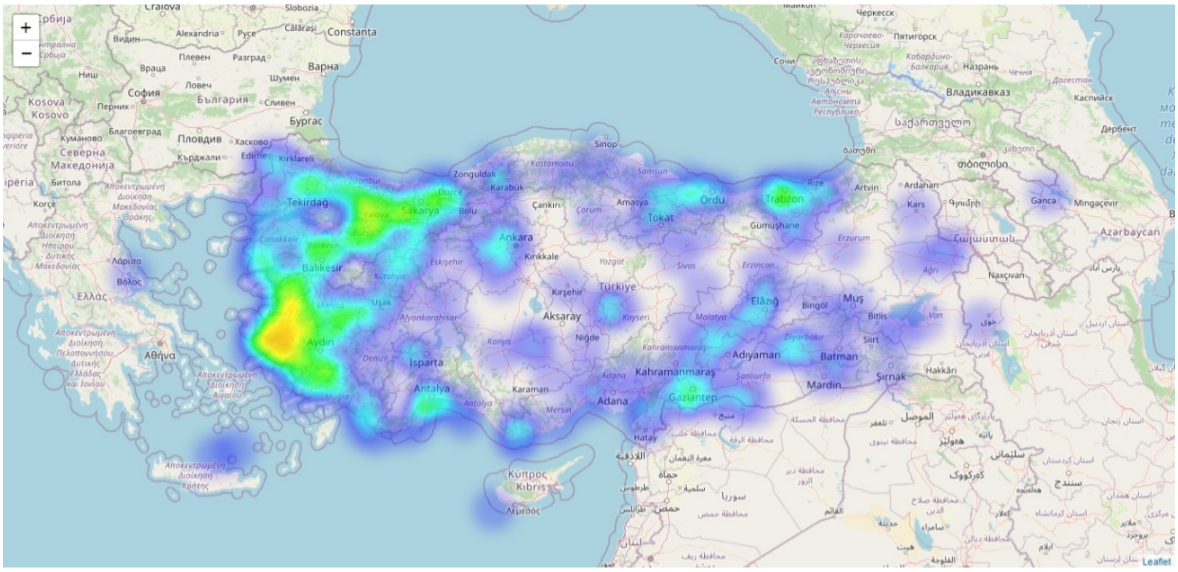

FIGURE 4.4 – Map of Turkey for Second Day Non-rescue Class

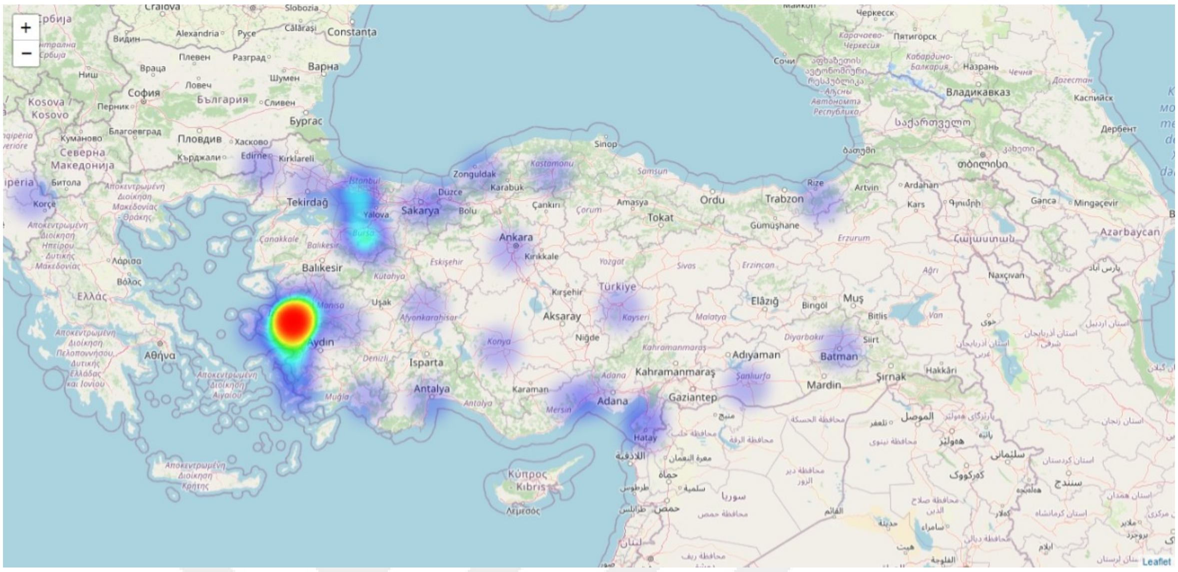

FIGURE 4.5 – Map of Turkey for Third Day Rescue Class

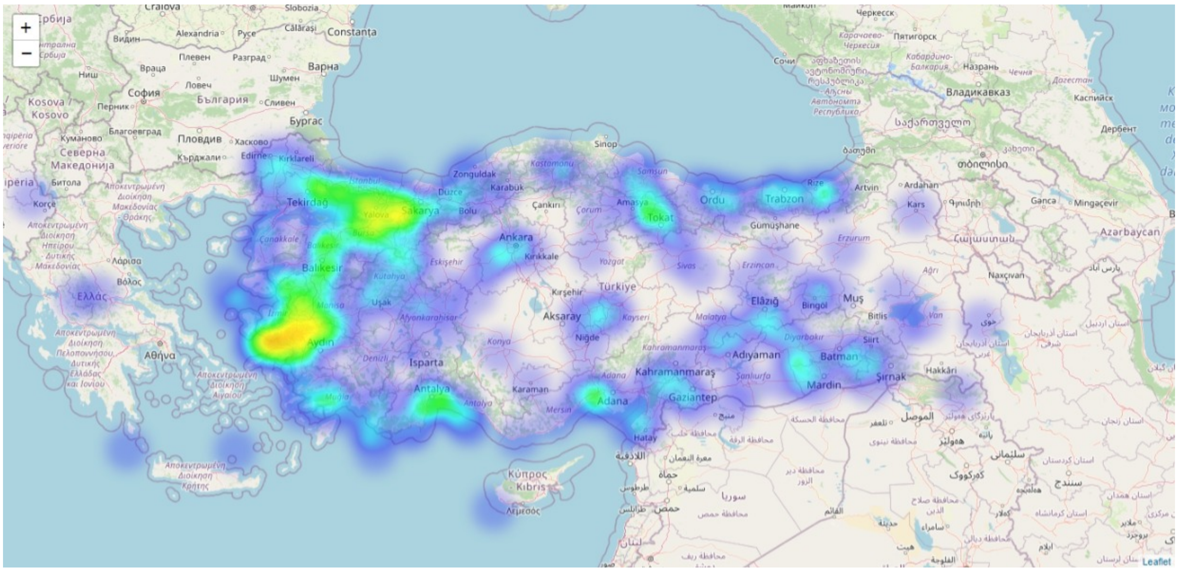

FIGURE 4.6 – Map of Turkey for Third Day Non-rescue Class

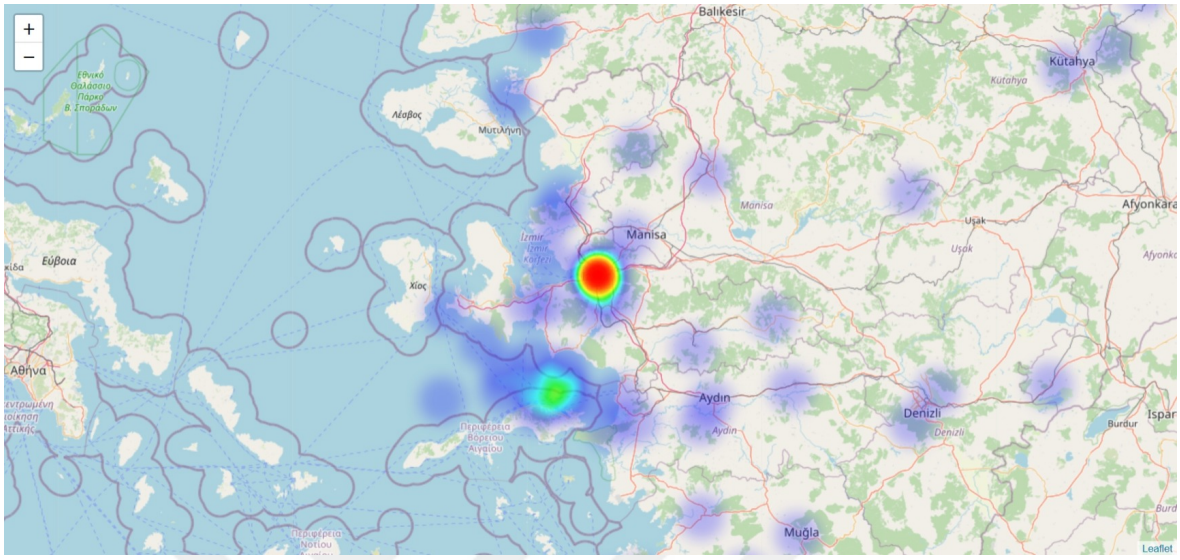

FIGURE 4.7 – Map of Izmir for First Day Rescue Class

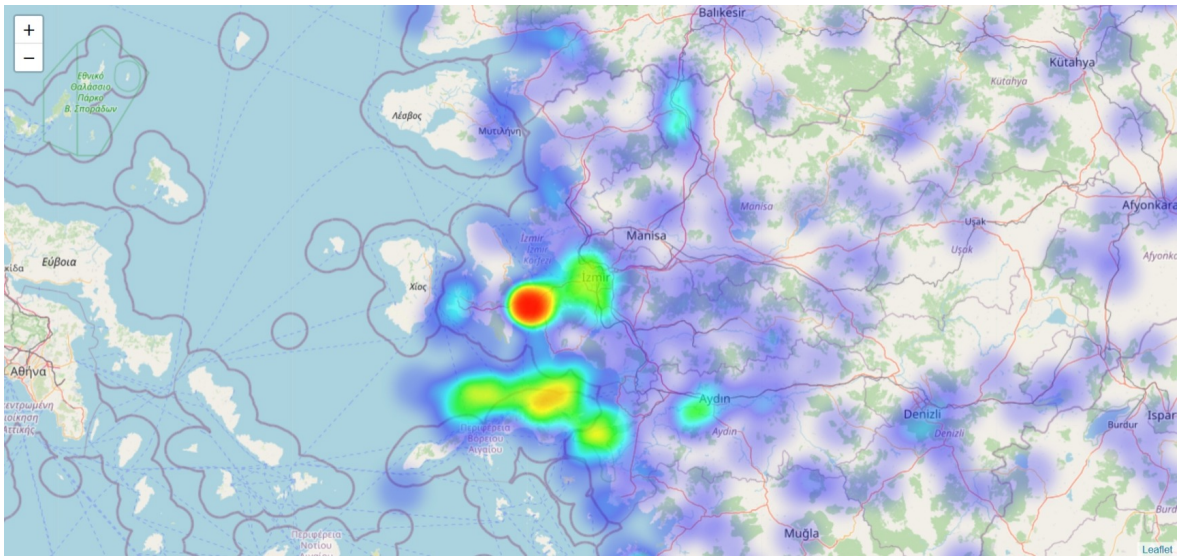

FIGURE 4.8 – Map of Izmir for First Day Non-rescue Class

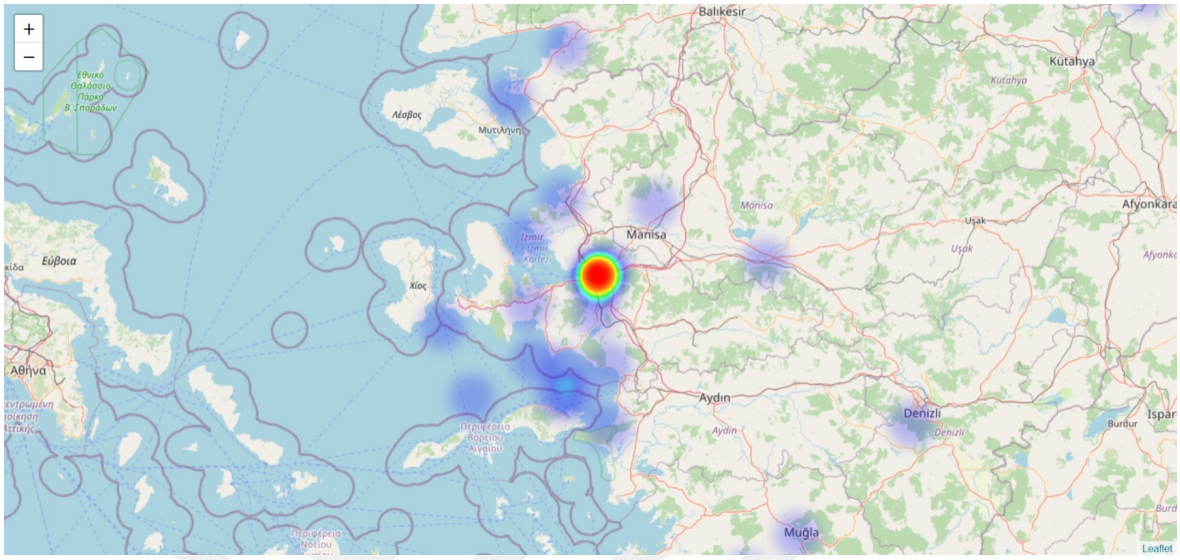

FIGURE 4.9 – Map of Izmir for Second Day Rescue Class

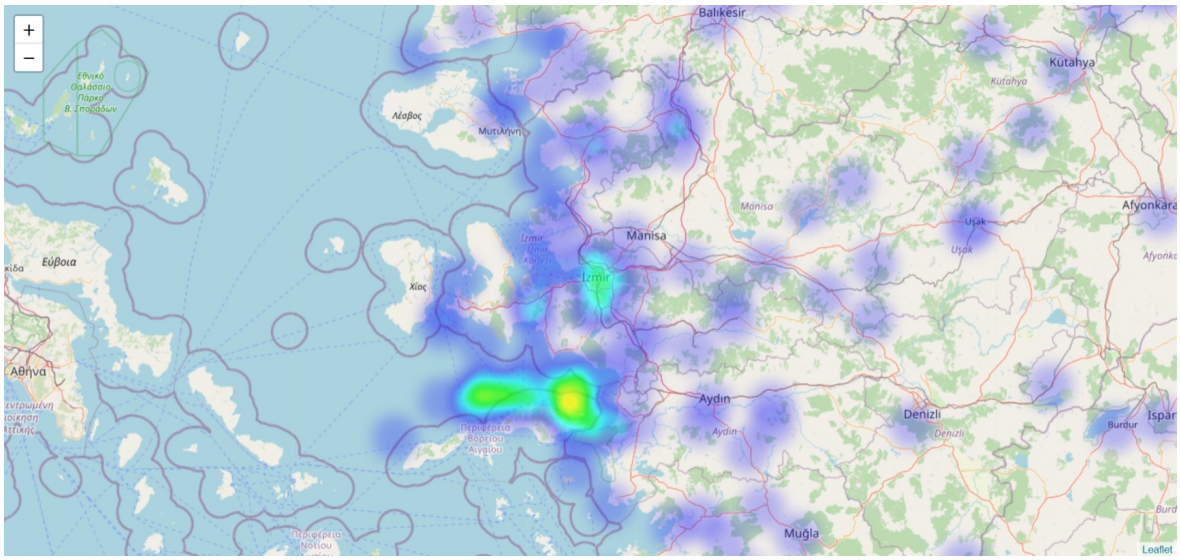

FIGURE 4.10 – Map of Izmir for Second Day Non-rescue Class

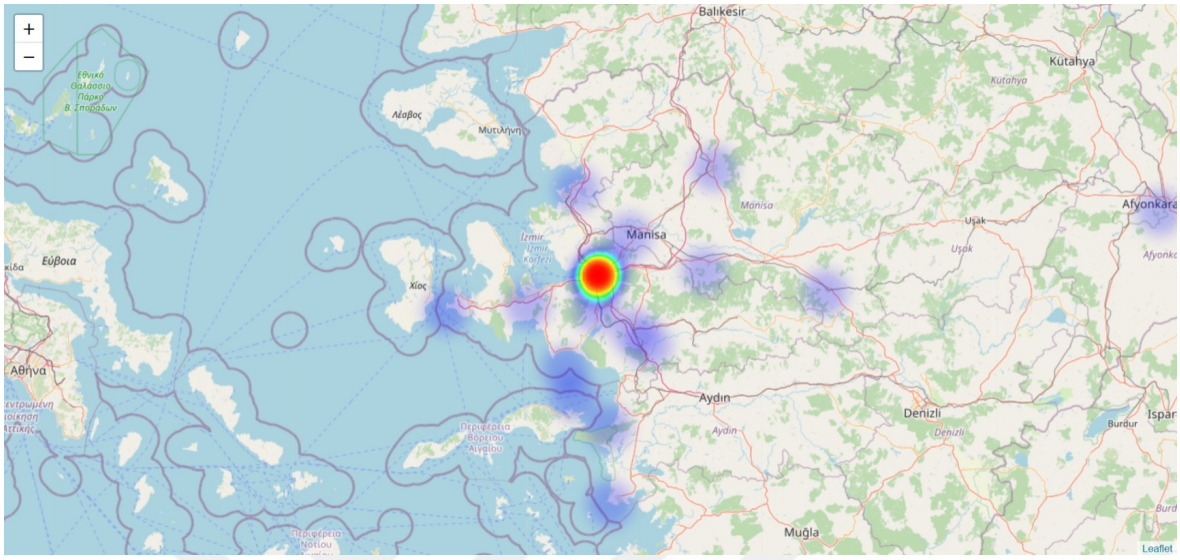

FIGURE 4.11 – Map of Izmir for Third Day Rescue Class

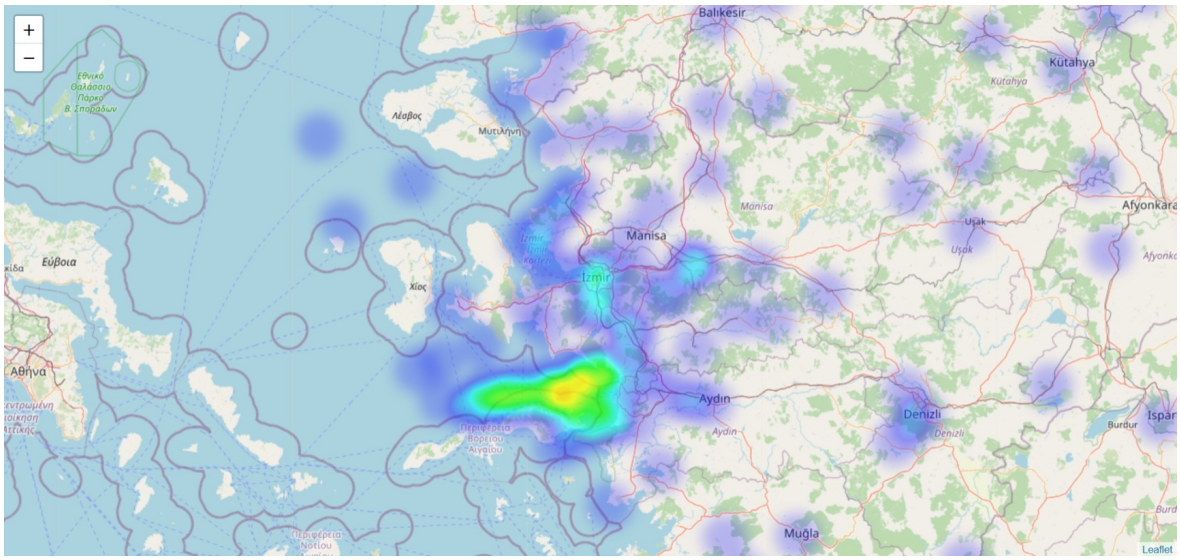

FIGURE 4.12 – Map of Izmir for Third Day Non-rescue Class

4.2 Word Clouds

Word clouds of each day after earthquake are shown in figures 4.13, 4.14, 4.15, 4.16, 4.17 and 4.18. While non-rescue tweets usually wish to get well soon for victims and contain more general things, rescue tweets contain address information, asking for help and expressing urgency. When we examine it on a day-to-day basis, another point for rescue tweets is that while there are more words about those who are under the collapsed buildings on the first day, as the days progress, there are more words related to post-earthquake problems such as the need for blood, the need for blankets, the need for donations.

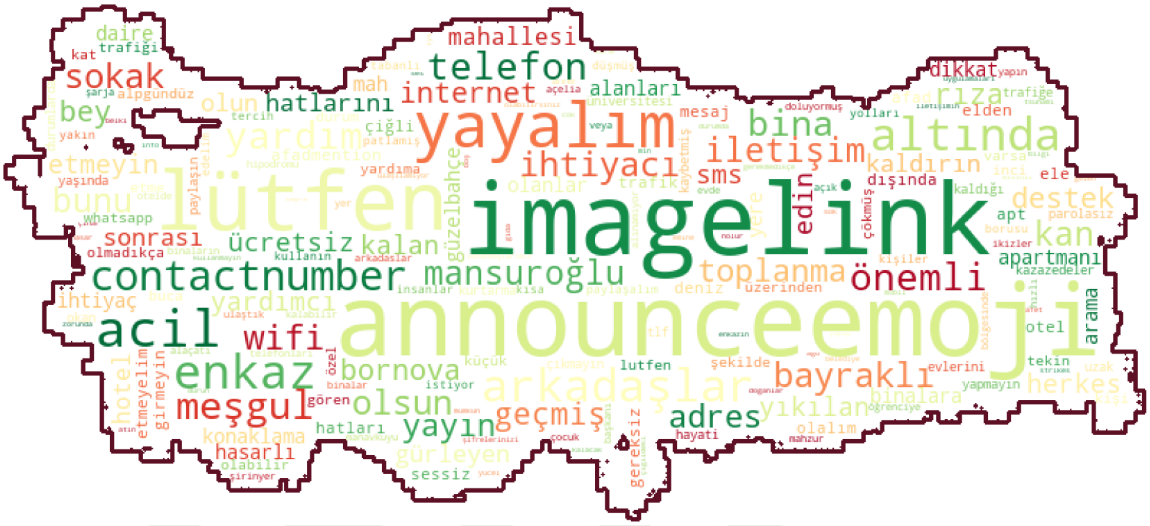

Figure 4.13 – Word Cloud for First Day Rescue Class

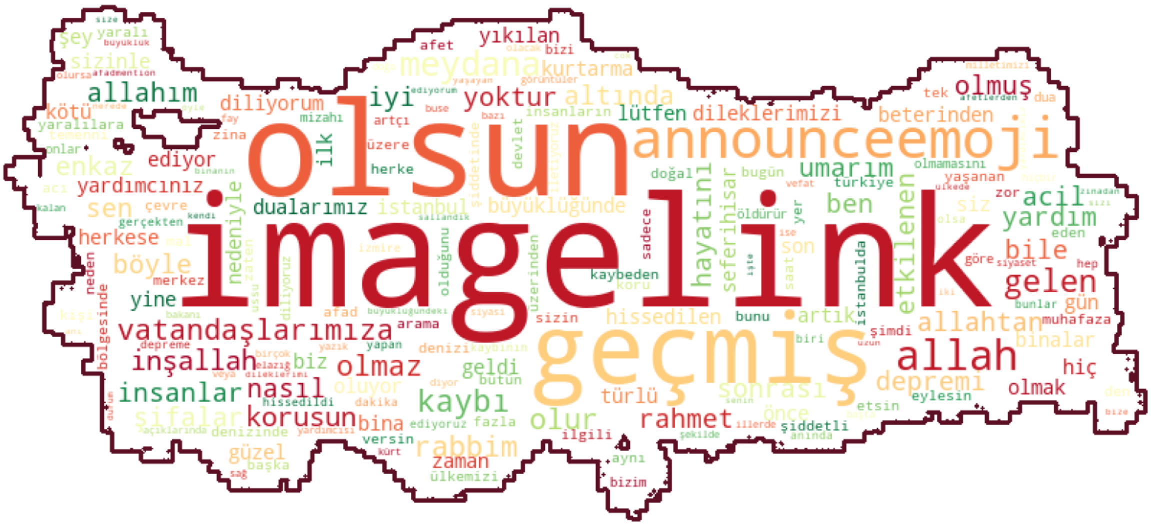

Figure 4.14 – Word Cloud for First Day Non-rescue Class

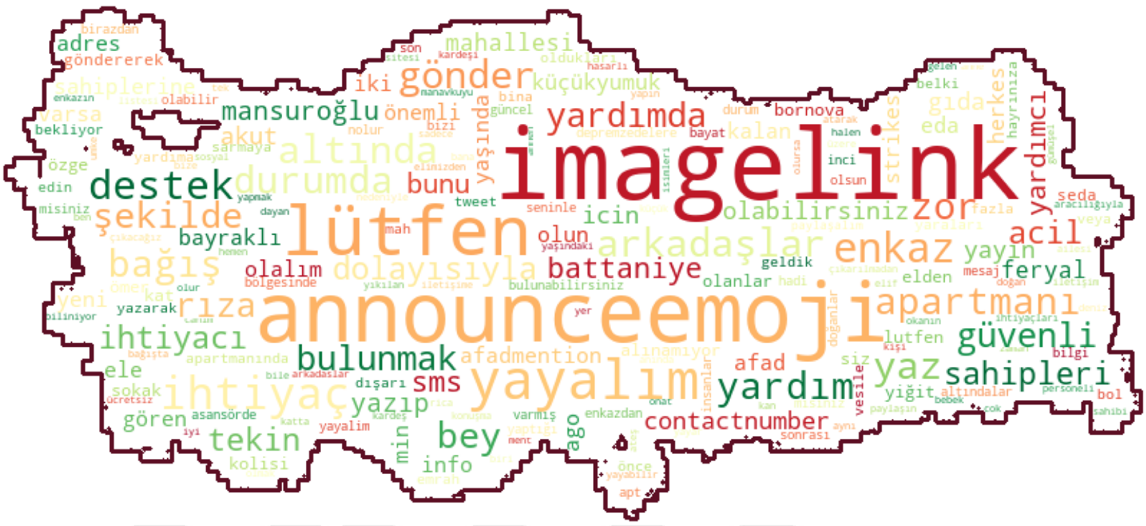

FIGURE 4.15 – Word Cloud for Second Day Rescue Class

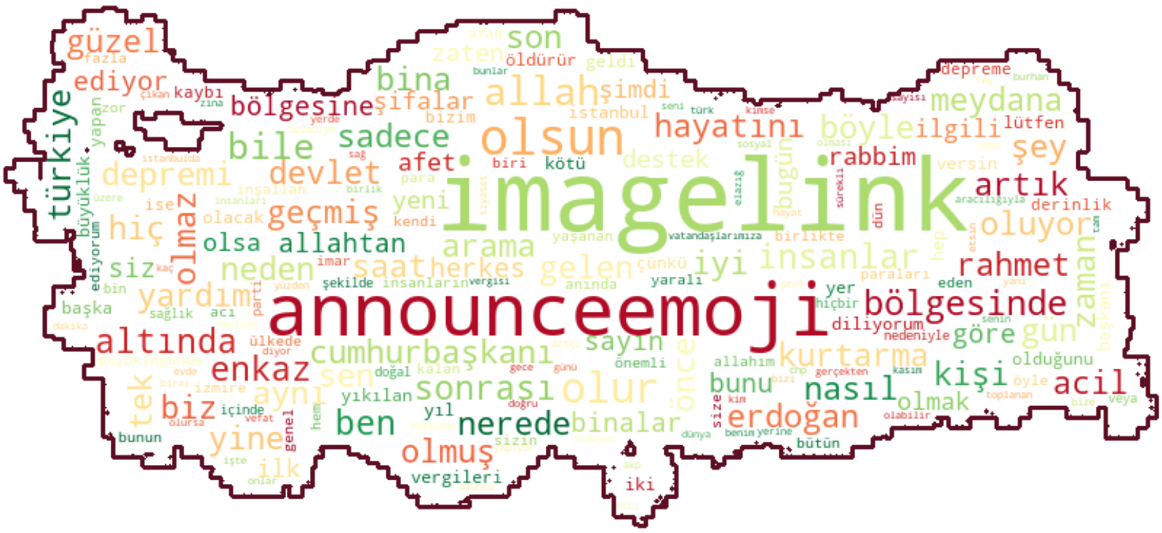

FIGURE 4.16 – Word Cloud for Second Day Non-rescue Class

FIGURE 4.17 – Word Cloud for Third Day Rescue Class

FIGURE 4.18 – Word Cloud for Third Day Non-rescue Class

4.3 Visualising Co-occurrences

When generating the co-occurrence network graph shown in the figures 4.19, 4.20, 4.21 and 4.22, both nodes and edges were weighted for the 40 most frequent words. The radius of a node determines how much it occurs in the data set. In addition, the edges are weighted according to the similarity of the co-occurrence of the words in the nodes. While calculating the widths of the edges, first the co-occurrence matrices of the words are created, and then the relation between the words are calculated by measuring the cosine similarities of the vectors of words. According to this calculation, the widths of the edges are weighted.

The difference in the co-occurrence network diagram for both rescue and non-rescue in AAC and MAC, while the graphs in MAC seem more compact, they seem more distributed for AAC. This is because there are far fewer tweets in MAC than in AAC. Both AAC and MAC show parallelism in rescue and non-rescue tweets. For the Rescue class, the correlations between words related to topics of urgent importance such as addresses and important warnings are wider, while in the non-rescue class, the links between the words of general information sentences that are related to earthquakes but do not contain urgency are more frequent. In addition, separate word groups appearing in AAC graphs are seen. Among these, the words ‘bağış’ (donation) and ‘gönder’ (send) for rescue refer to tweets about fundraising. The separate group in the non-rescue class which consists of ‘vergiler’ (taxes) and ‘nerede’ (where ?), on the other hand, refers to tweets that question where the collected earthquake taxes go.

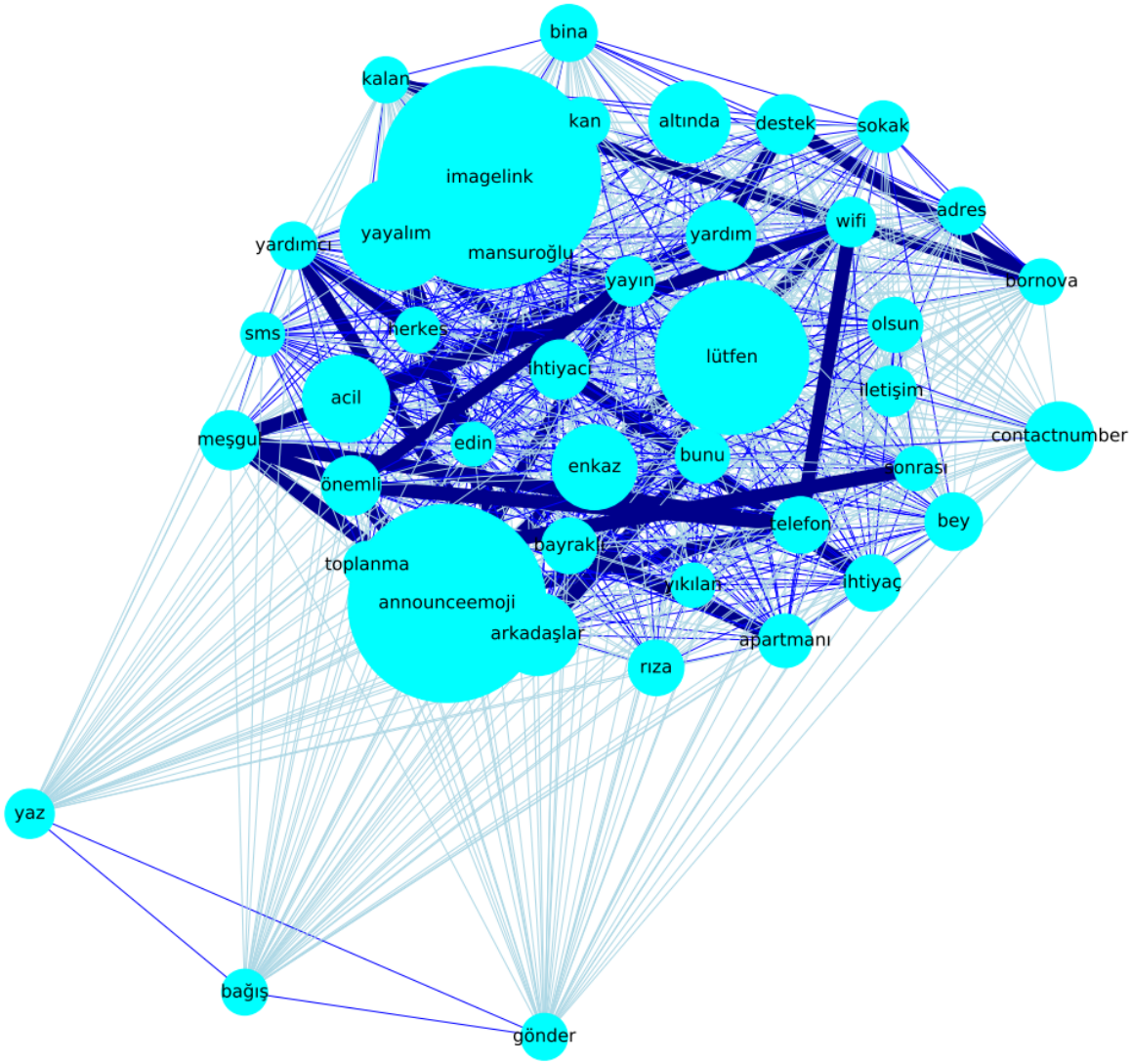

FIGURE 4.19 – Visualization of Words Cooccurrences for Rescue Class in AAC dataset

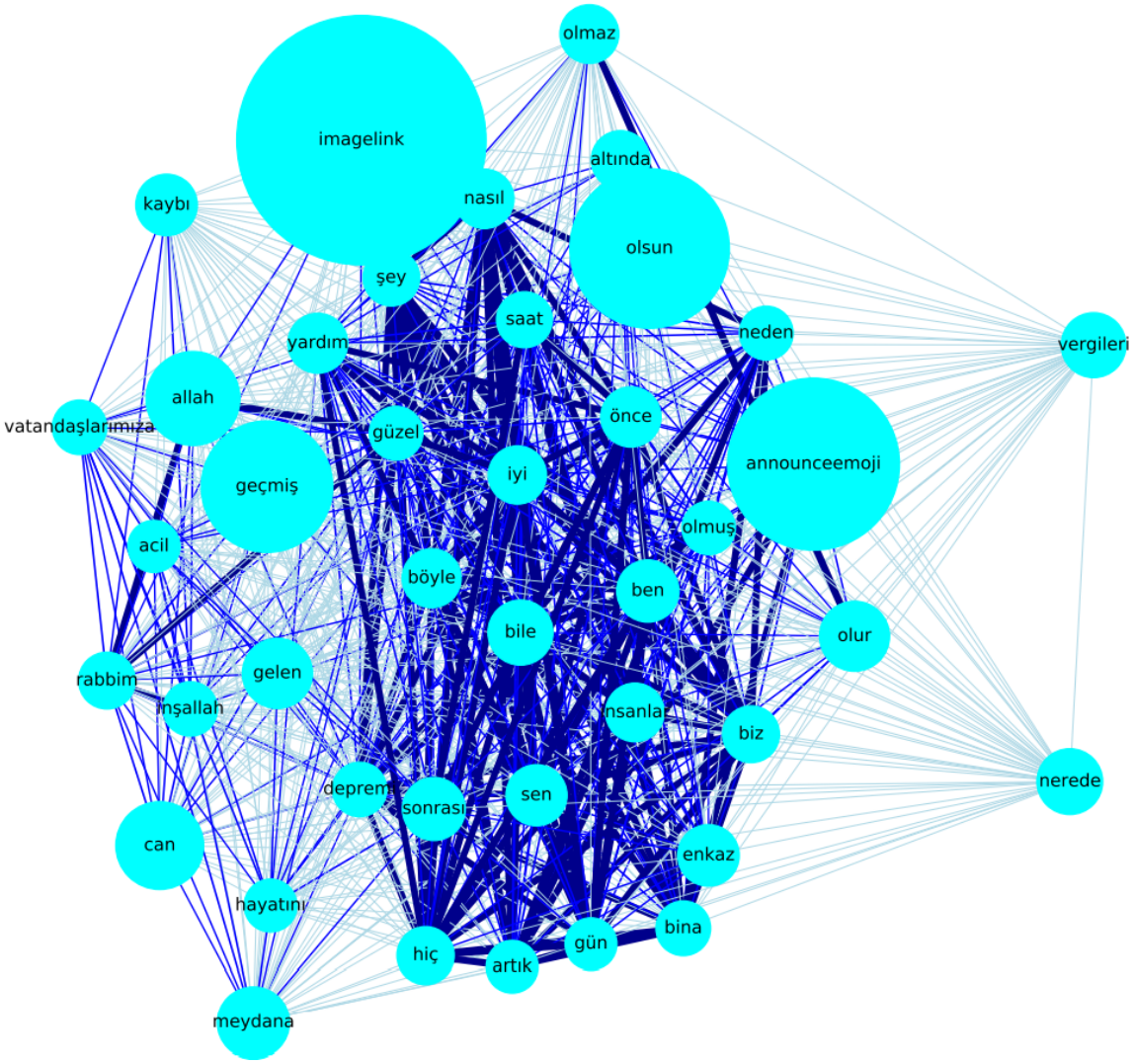

FIGURE 4.20 – Visualization of Words Cooccurrences for Non-rescue Class in AAC dataset

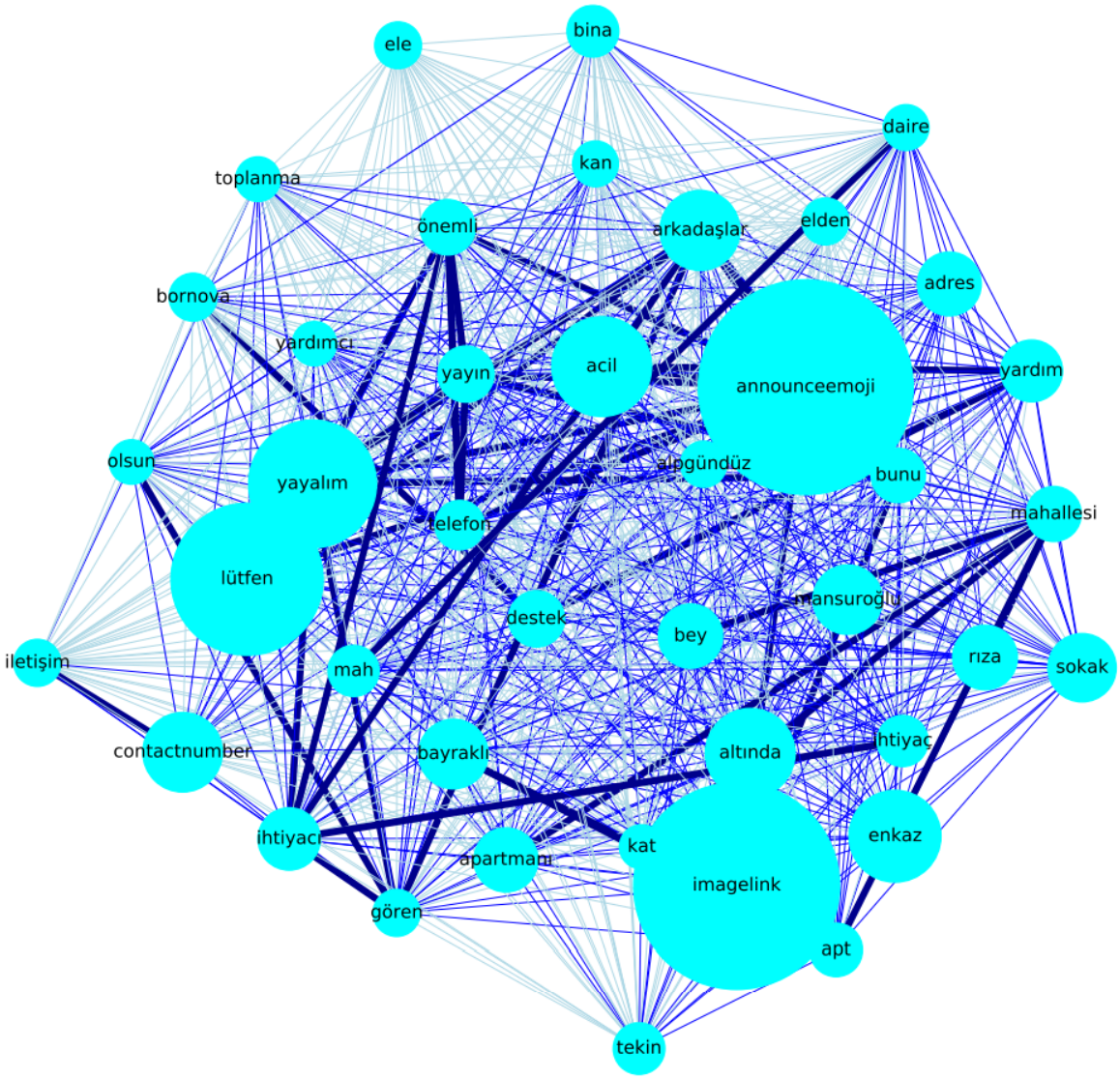

FIGURE 4.21 – Visualization of Words Cooccurrences for Rescue Class in MAC dataset

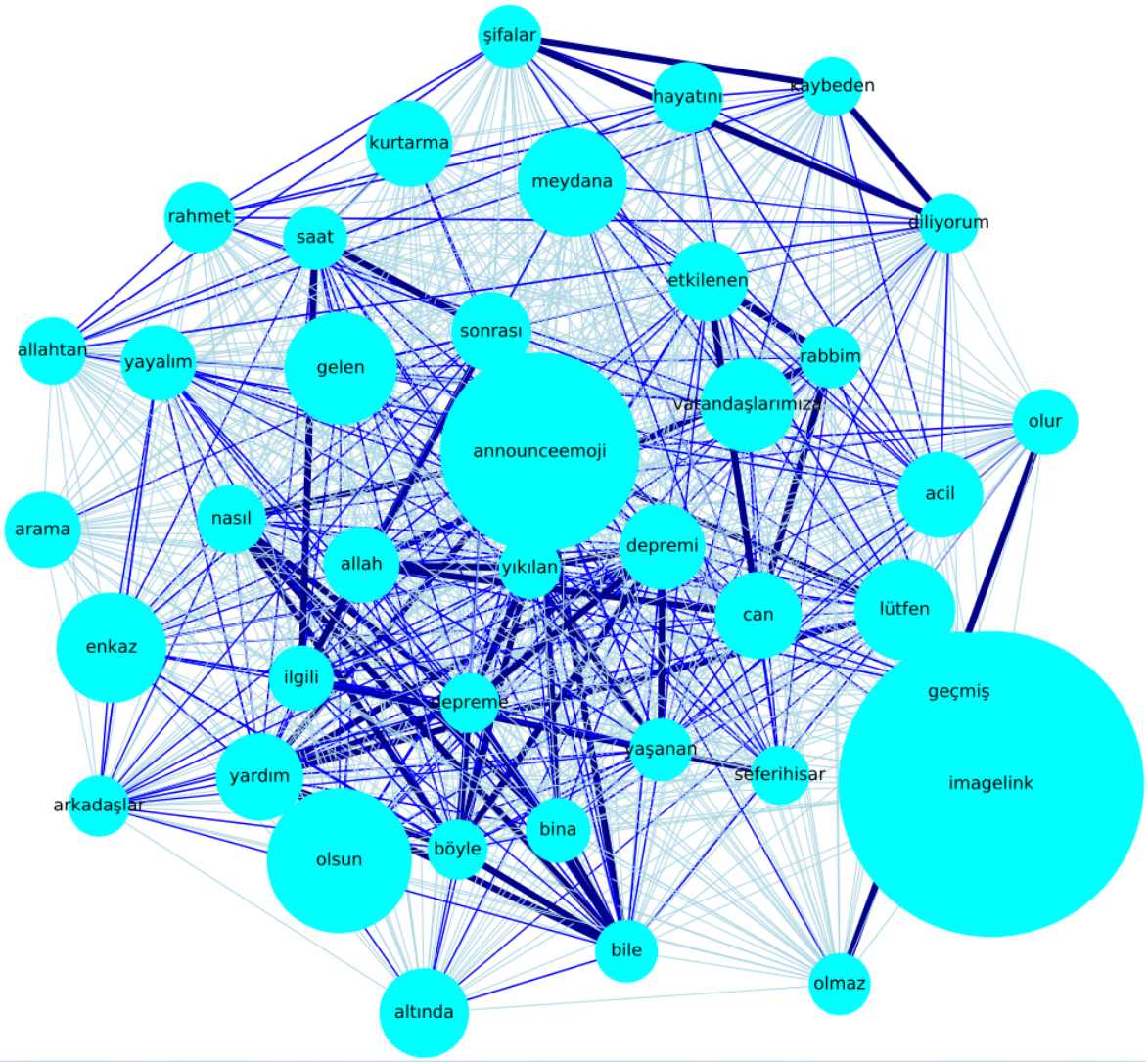

FIGURE 4.22 – Visualization of Words Cooccurrences for Non-rescue Class in MAC dataset

4.4 Visualising Word Embeddings

Word2Vec (Mikolov et al., 2013) is an unsupervised neural network based model to represent words in vector space. It has 2 sub-methods : CBOW (Continuous Bag of Words) and Skip-Gram. CBOW uses a context surrounding words to predict a word, while Skip-gram guesses the context by using a word. As a result of the Word2Vec operation, a dictionary is obtained in which each word has a vector.

By finding general features in high-dimensional data, it reduces the number of dimensions and compresses the data. The purpose of the Principal Component Analysis (PCA) method is to find a new set of dimensions that will better capture the diversity of the data. Since it can be reduced to 2 or 3 dimensions, it makes visualization much easier.

One of the content-based analysis methods in this study is to visualize the word embeddings in AAC dataset. Word2Vec is implemented and trained with rescue and nonrescue tweets to obtain word embeddings. Then, these embedding vectors' dimensions are reduced to three by one of the dimensionality reduction techniques called PCA.

It is an important step to examine the proximity of words to each other in order to be able to do context analysis of words and beyond. In this study, vectors of all words in the corpus were created by training on the AAC dataset with the Word2Vec method. The dimensions of these vectors were visualized by reducing them to 3 dimensions with the help of PCA. In order to increase readability, the 40 most common words for rescue and nonrescue were selected separately for visualization and placed in these 3 dimensions.

When the embeddings of Rescue tweets are examined, as it can be seen in the figure 4.23, words with similar context are very close to each other. For example, "rıza", "bey", "apartmanı" words refer to Rıza Bey Apartment, which collapsed in the earthquake. Other obvious examples are the word-pairs "acil"(urgent) and "yardım"(help), "enkaz"(debris) and "yıkılan"(collapsed), "ihtiyaç"(need) and "kan"(blood), "contact-number"(see Preprocessing) and "iletişim"(contact) close together.

On the other hand, looking at the nonrescue tweets in figure 4.24, it is displayed that the words that are likely to be included in the same sentence are close to each other

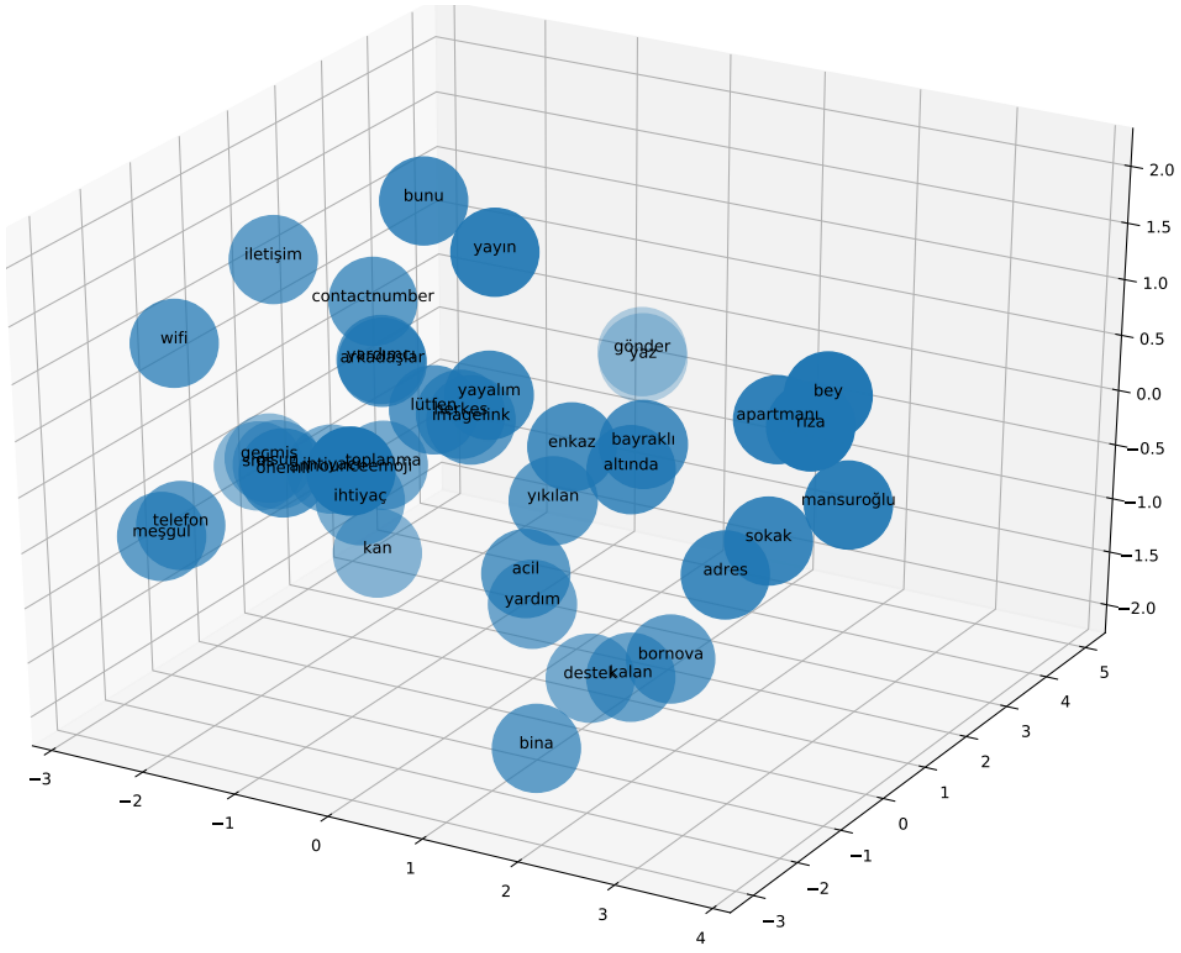

FIGURE 4.23 – Visualization of Rescue Embeddings

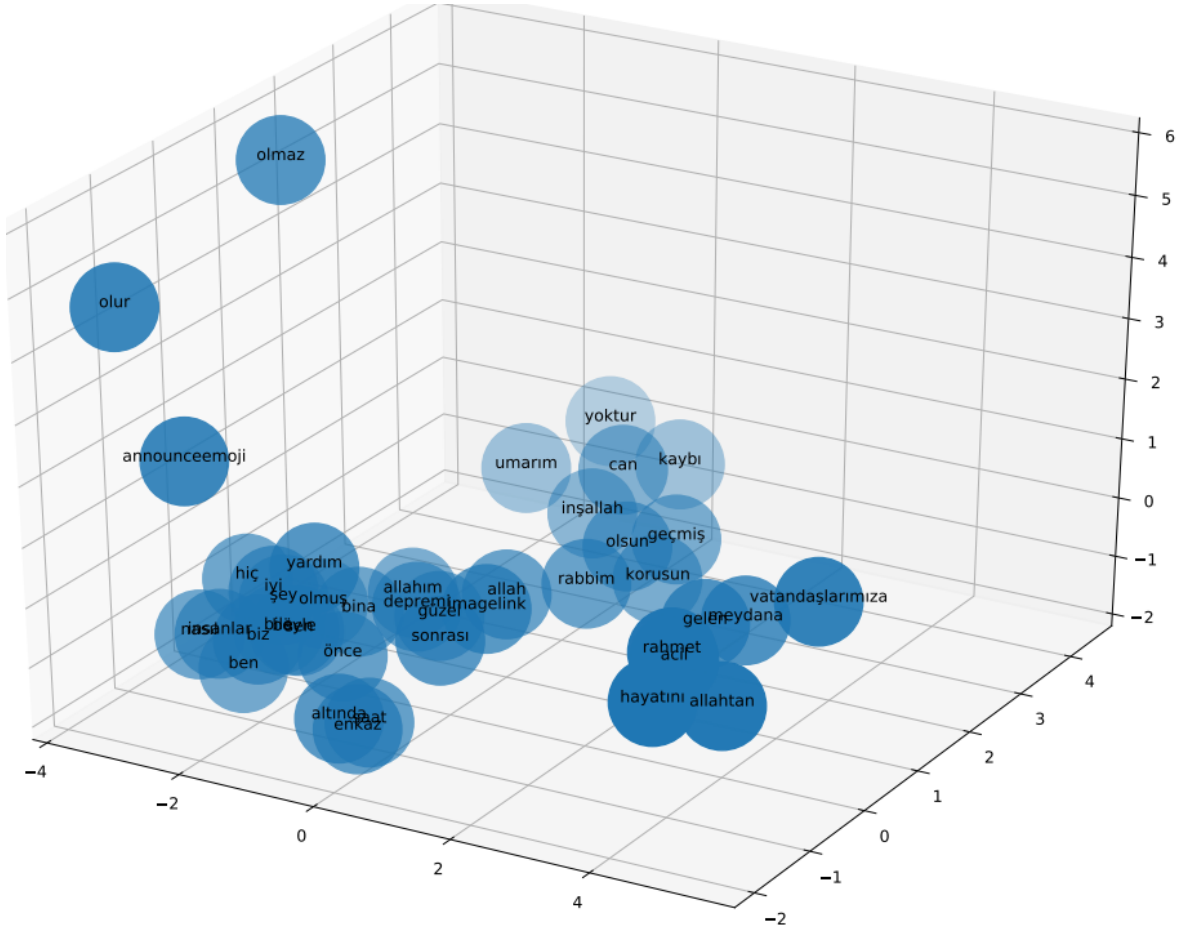

FIGURE 4.24 – Visualization of Non-rescue Embeddings

on the plane. The words "geçmiş" and "olsun" refer to getting well soon. Also, in nonrescue tweets, the words that are likely to appear in condolence tweets, which are frequently seen after the earthquake, are close to each other. For example, " İzmir'de meydana gelen depremde ölen vatandaşlarımıza Allah rahmet eylesin, acil şifalar." (May God have mercy on those who died in the earthquake that took place in Izmir, and get well soon to the injured.), this sentence expresses condolences and as can be seen "meydana", "gelen"(occurred), "vatandaşlarımıza"(our citizens), "acil"(soon) and "rahmet"(mercy) are close together in the plane.

4.5 Topic Modeling

Under this section, we will see topic analyzes for AAC dataset generated within the scope of the study. In order to model rescue and non-rescue topics in created corpuses, Latent Dirichlet Allocation method is implemented. LDA algorithm modeling the topics help to analyze which is employed by Gensim library.

Manuel Annotated Corpus is divided into two classes as rescue and non-rescue. Then, for each rescue and non-rescue topics, most appropriate word sets are extracted. It is also analyzed by examining most appropriate different level of word sets for each topic. Each level has 5, 10, 20, 50 words respectively.

Automatic Annotated Corpus is firstly filtered to eliminate tweets which don't possess location information and to obtain first 72 hours since earthquake has occured. Then, it is categorized as rescue and non-rescue topics. Each topic splitted into 30 minutes intervals. For each topic and for each interval, implemented Latent Dirichlet Allocation extracts most appropriate words for the topics and the intervals.

Latent Dirichlet Allocation (LDA) is a generative probabilistic model used for topic modeling which is defined unsupervisedly extract topics from documents. It is one of the most popular methods among Topic Modeling algorithms, as it provides comparatively accurate results and additionally it can be trained. The main assumption of LDA is that each document consists of topics and each topic has a probability distribution of words. Briefly, it strives to illuminate the semantic relations between words by taking topics into account. This following 4 steps are carried out through LDA algorithm process :

1. Assign a topic to each word in each document randomly

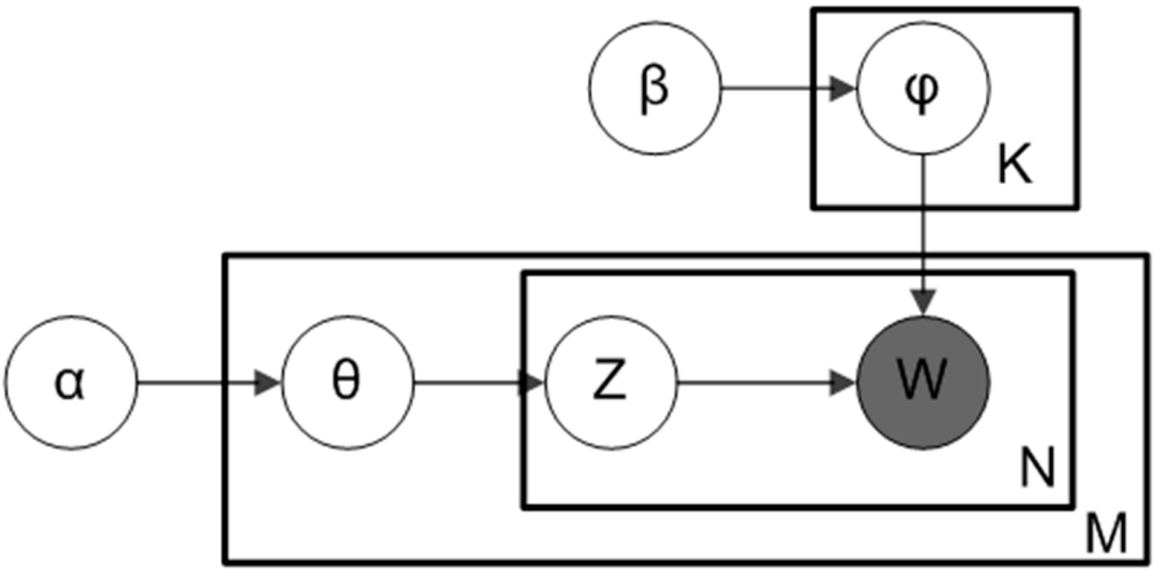

FIGURE 4.25 – Latent Dirichlet Allocation

2. Update the assigned topic for a single word in a single document by using Dirichlet distributions
3. Repeat Step 2 for all words in all documents

With iteration, the words will reach a state of equilibrium.

In the original paper of LDA illustrates the method in a diagram shown in Figure 4.25.

It is the diagram of an LDA model where ;

— α is a parameter representing the word distribution per-document,

— β is word distribution per-topic,

— θ is a vector for the topic distribution on a document M,

— Z is a topic for a chosen word in a document,

— W refers to specific words in N,

— The plate M is the length of the documents,

— ϕ is the word distribution for topic K,

— The plate N is the number of words in the document.

In the tables 4.1, 4.2, we see the output of 25 words for the first 3 days of the rescue and non-rescue topics modeled using LDA. The words are weighted according to their relation to the subject and shown in the tables in order from top to bottom.

TABLE 4.1 – LDA Outputs for Rescue Tweets First Three Days

First Day	**Second Day**	**Third Day**
announceemoji	imagelink	imagelink
imagelink	announceemoji	announceemoji
lütfen	lütfen	lütfen
yayalım	yayalım	elif
enkaz	ihtiyaç	yaz
altında	enkaz	yardım
arkadaşlar	arkadaşlar	gönder
acil	altında	arkadaşlar
contactnumber	yaz	enkaz
apartmanı	apartmanı	yayalım
ihtiyaç	destek	bağış
yayın	gönder	kat
sokak	yardım	altında
bey	bey	apartmanı
rıza	rıza	bey
yardım	bağış	ihtiyaç
destek	zor	destek
bunu	durumda	rıza
ihtiyacı	şekilde	bebek
bayraklı	acil	kedi
mansuroğlu	battaniye	acil
adres	güvenli	apartmanında
mahallesi	yardımda	ihtiyacı
önemli	bulunmak	hayata
tekin	tekin	yaşındaki

TABLE 4.2 – LDA Outputs for Non-rescue Tweets First Three Days

First Day	**Second Day**	**Third Day**
imagelink	imagelink	imagelink
announceemoji	announceemoji	announceemoji
olsun	olsun	olsun
olur	can	saat
allah	allah	elif
geçmiş	bile	vergileri
can	olur	allah
enkaz	sonrası	nerede
olmaz	ben	can
bile	enkaz	sonrası
gelen	bina	enkaz
altında	nasıl	bile
meydana	altında	güzel
sonrası	iyi	sen
insanlar	insanlar	hiç
devlet	hiç	bina
yardım	önce	yardım
saat	böyle	altında
bina	biz	olur
ben	yardım	nasıl
böyle	rahmet	ben
hiç	şey	yaşındaki
önce	son	kurtarma
iyi	hayatını	iyi
nasıl	gün	biz

Let's exemplify the first 3-day LDA outputs of the AAC dataset in order to better explain the two topics. In other words, how these outputs would look if they were one or more tweets, the following results are approximately the tweets of topics :

— First Day - Rescue : "Acil ihtiyaç var arkadaşlar enkaz altında kalan var. Lütfen yayın. Adres : Mansuroğlu Mahallesi Rıza Bey Apartmanı Bayraklı/izmir."
(There is an urgent need, friends. There is someone left under the rubble. Please disseminate. Address : Mansuroğlu Mahallesi Rıza Bey Apartment Bayraklı/İzmir.)

— First Day - Non-Rescue : "İzmir'de meydana gelen depremde ölenlere Allah rahmet eylesin, yaralılara geçmiş olsun."
(May God have mercy on those who died in the earthquake that took place in Izmir, and get well soon to the injured.)

— Second Day - Rescue : "Lütfen yayalım acil battaniye ihtiyacı bulunmaktadır."
(Please disseminate. It's urgent ! Blankets are needed.)
"SMS göndererek bağış yaparak destek olalım."
(Let's support by sending an SMS.)

— Second Day - Non-Rescue : "Herkese geçmiş olsun meydana gelen deprem sonrası enkaz altında aramalar sürüyor. Her bir can çok önemli."
(Get well soon after the earthquake. Searches continue under the collapsed buildings. Every life is very important.)

— Third Day - Rescue : "Elif bebek enkaz altında lütfen yayalım arkadaşlar."
(Elif is under the rubble, please disseminate it)
"İhtiyaçlar için yardım ve bağış gönderelim."
(Let's send aid and donations for the needs.)

— Third Day - Non-Rescue : "İzmir'e geçmiş olsun deprem sonrası saatlerdir kurtarma çalışmaları sürüyor."
(Get well soon after the earthquake in Izmir, rescue efforts have been going on for hours.)
"Toplanılan deprem vergileri nerede ?"
(Where are the earthquake taxes collected ?)

As exemplified after Tables 4.1 and 4.2, the problems occurring right after the earthquake in the first days are naturally discussed on social media, while the following

days evolve into post-earthquake problems. In the first day's search and rescue efforts, people are instructed to prefer the internet or text messaging, and in the following days, the topics related to the blanket and shelter needs of the victims after the earthquake are discussed. Non-rescue tweets generally contain statements from politicians or officials who try to inform. However, in rescue tweets, whatever is needed at that moment is popularized by spreading over social media from all over the world.

5 SMART DISASTER CLASSIFICATION

In this section, we will train machine learning models with the tweets in the large AAC dataset that we have labeled by a semi-supervised method. These models will be trained on 75% of the dataset, and the remaining 25% will be split for evaluation. AAC dataset includes approximately 1 million rescue or non-rescue tweets. A broad range of methods from classical machine learning methods to state-of-the-art transformer-based methods have been implemented and used.

We have decided to use classical machine learning algorithms such as Logistic Regression, Linear SVC, Decision Tree Classifier, Random Forest Classifier, Multinomial NB as baselines for comparative analysis with other methods. In order to achieve the highest success, the state-of-the-art transformer-based language models have been implemented. These models are :

— Bert base turkish 128k cased

— Distilbert base turkish cased (Sanh et al., 2019)

— Bert base cased

— Bert base multilingual cased (Devlin et al., 2018)

— Bert base turkish cased

— Convbert base turkish cased (Jiang et al., 2020)

— Convbert base turkish mc4 cased

— Electra base turkish cased discriminator (Clark et al., 2020)

— Electra base turkish mc4 cased discriminator

A language model is trained unsupervisedly. It stores the most fundamental representations of words and sentences depending on the context. In order to use it in our problem, we need to proceed with the approach called transfer learning. In other words, we will fine-tune our AAC dataset by supervised retraining using pre-trained BERT models.

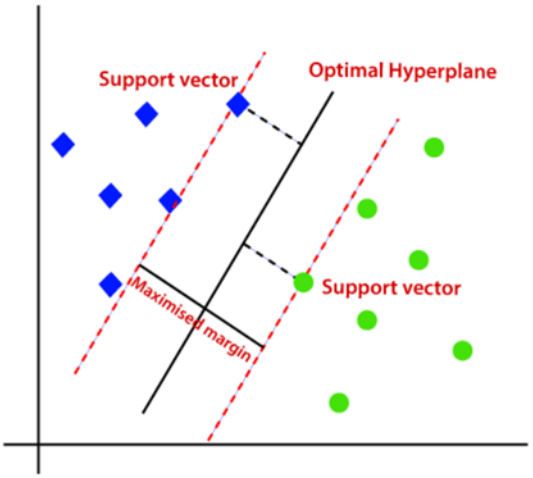

FIGURE 5.1 – Support Vector Machine (SVM)

5.1 Methods

5.1.1 Support Vector Machine

Support Vector Machine is a supervised algorithm that aims to create a machine learning model for prediction or classification. In the classification process, considering K-dimensional data, SVM finds K-1 dimensional hyper-plane. If there is a linearly separable data set in the binary classification problem, there are infinitely many hyperplanes that can separate this dataset. However, there is only one hyperplane with a maximum boundary between these planes. While this hyperplane is called the optimal hyperplane, the support vectors are those data points that lie on the borders of the margins. Using these support vectors, the margin of the classifier is maximized. Figure 5.1 displays how SVM creates support vectors and optimal hyperplane.

5.1.2 Logistic Regression

$$S(x) = \frac{1}{1 + e^{-}x} \tag{5.1}$$

Logistic regression is a machine learning method used for classification tasks. It got this name because it uses a logistic function, and this function is also called the sig-

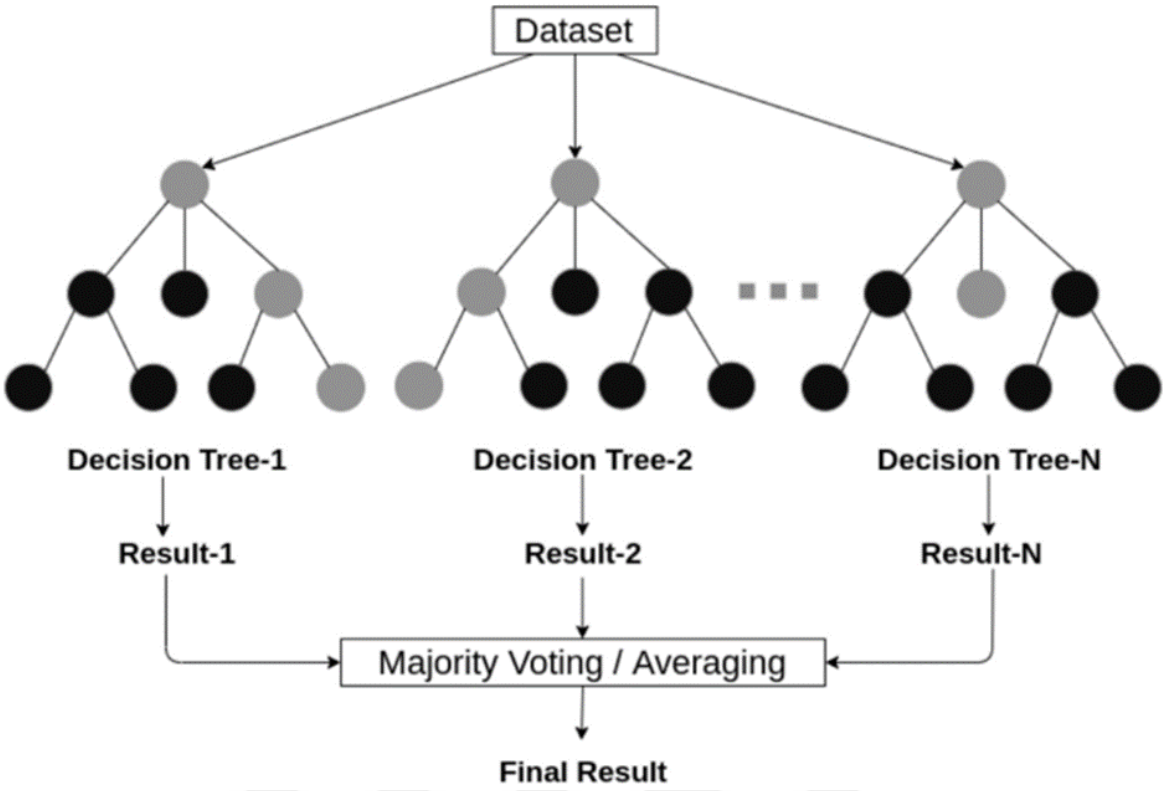

FIGURE 5.2 – Random Forest

moid function and is shown in Formula 5.1. A Logistic Regression model can predict the dependent variable based on multiple independent variables. Because logistic regression turns complex calculations into a simple arithmetic problem, it significantly simplifies analyzing the impact of multiple variables and helps minimize the impact of confounding factors.

5.1.3 Random Forest

The Random Forest algorithm is based on combining many decision trees instead of creating a single decision tree. Each of these decision trees are independent of each other and are created from different samples selected from the training set with the bootstrap technique. Random Forest algorithm uses parallel ensembling which executes multiple decision trees at once and majority voting is used for classification. Thus, it is expected that many decision trees predict better than only one decision tree. For different training sets, random feature selection from the same distribution is used. While creating the decision trees, all trees are scanned for relevant attributes. The attributes from other trees are then combined. In this way, the most used attribute is

selected. After the selected attribute is included in the decision tree, the same processes are repeated in the other stages. Figure 5.2 demonstrates the architecture of Random Forest algorithm.

5.1.4 Naive Bayes (NB)

$$\mathrm{P(c \mid x)} = \frac{\mathrm{P(x \mid c)}P(x)}{\mathrm{P(x)}} \quad (5.2)$$

Naive Bayes classification is based on Bayes' theorem and tells what the probability of the target class is. In equation 5.2 ;

- c, the class to be guessed
- $P(c \mid x)$, probability of event c occurring when x event occurs.
- $P(x \mid c)$, probability of occurrence of x when event c occurs
- $P(x)$ is the probability of occurrence of x

The multinomial NB method can be used to construct the polynomial Naive Bayes classifier. It is fast, easy to apply and highly efficient. Multinomial Naive Bayes models the distribution of words in a document as polynomial. The document is treated as a string of words and it is assumed that the position of each word is formed independently of the other.

5.1.5 Decision Tree

The decision tree algorithm classifies the data with a top-down distribution, as shown in Figure 5.3. Nodes are placed hierarchically according to their information gains.

Among the information gains obtained for all features, the feature that provides the highest information gain takes place in the first step of the decision tree. The highest information gain is achieved by using the entropy values at each node. The final shape of the decision tree is obtained by pruning backwards. A decision tree provides powerful and fast ways to make sense of the data structure. It is preferred because it requires less time to prepare data, provides ease of interpretation and is not affected by non-linear relationships. The disadvantages are that the decision tree drawings with too many

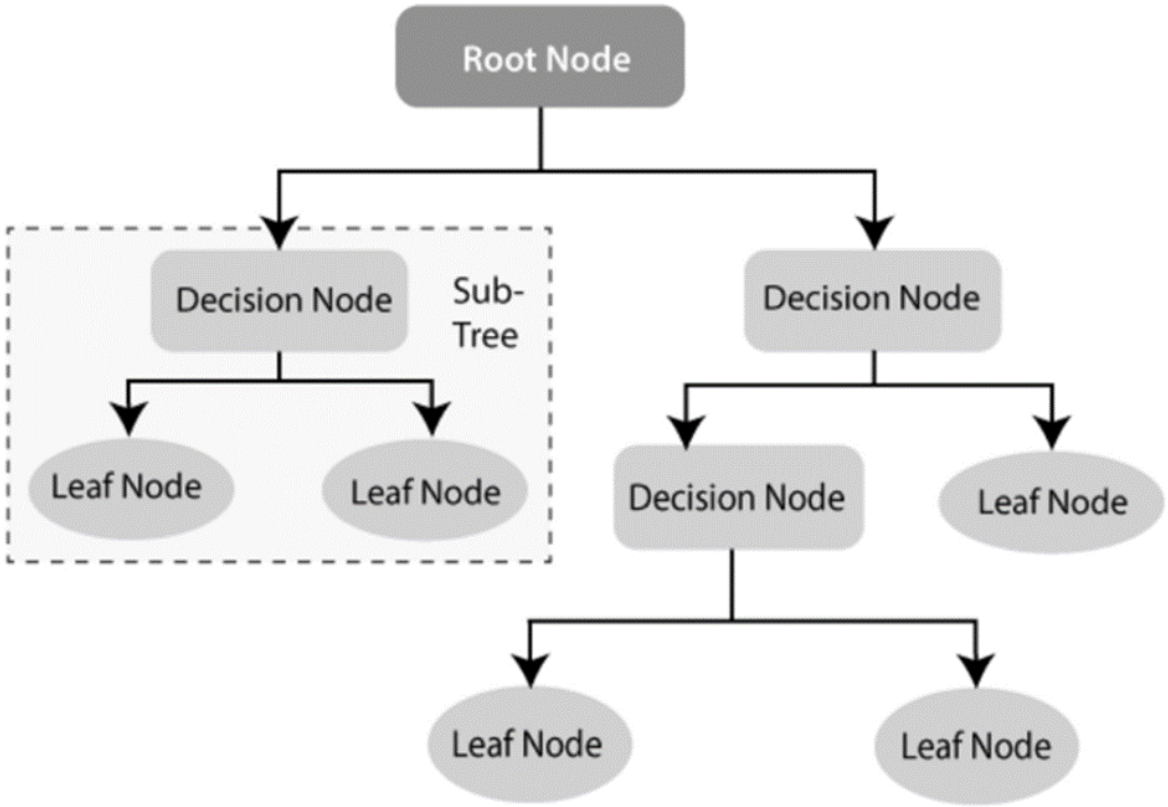

FIGURE 5.3 – Decision Tree

branches present a complex structure and are time consuming. Large trees are difficult to present and understand.

5.1.6 BERT

BERT is a generalized language model with 24 Transformer blocks, 1024 hidden layers, and 349M parameters. Pre-trained with large text data, BERT allows it to be transferred for other small language processing tasks as needed (Fine-tuning). Thus, it has demonstrated state-of-the-art success for 11 natural language processing tasks .

Unlike other previous learning models used in natural language processing, it can process multiple words simultaneously. This both supports parallel work and contributes to the capture of the context. Two different unsupervised learning methods were used for the training of BERT. 15% of the words were masked and these masks were to be predicted bidirectionally. The other method is to try to predict the next sentence, so that it can be learnt the relations between sentences.

— *Bert Base Cased* is the base model of BERT architecture.

— *Bert Base Multilingual Cased* uses basic BERT architecture and it is trained on multilingual data.

— *Bert Base Turkish Cased* is trained on Turkish data and has same basic Bert architecture.

— *Bert Base Turkish 128k Cased* is bigger model compared to Bert Base Turkish Cased, since it is trained on larger Turkish data.

— *DistilBert Base Turkish Cased* is a distilled version of Bert Base Turkish Cased model. The general-purpose of DistilBert can be fine-tuned with good performances on several downstream tasks, keeping the flexibility of larger models. It has been shown that the model is small enough to run on the edge, e.g. on mobile devices.

— *ConvBert Base Turkish Cased* uses a span level dynamic convolution to improve BERT. It is trained on same dataset as used in Bert Base Turkish Cased

— *ConvBert Base Turkish mc4 Cased* is trained on different dataset (mc4 corpus) compared to ConvBert Base Turkish Cased.

— *Electra Base Turkish Cased* is a different Turkish model than basic BERT architecture. While, in the basic BERT method is to predict only masked 15% of the tokens, ELECTRA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately) method uses every token to classify whether the token has been replaced or kept the same.

— *Electra Base Turkish mc4 Cased* is trained on bigger mc4 dataset for Turkish.

5.2 Evaluation

For evaluation purposes, several metrics are assessed. These metrics are Precision, Recall and F1 scores for both rescue and non-rescue class. Besides, accuracies of the models are also calculated. Additionally, Receiver operating characteristic (ROC) curves have been plotted and Area Under Curve (AUC) scores are computed to compare all models according to their true positive and false positive rates. Explanations of metrics are below :

— Precision : It is a measure of how accurately a class is predicted.

— Recall : It is a measure of how correctly the classifier predicted the true positive value.

TABLE 5.1 – Top Five Models for Imbalance Rate : 8

Model	Non-rescue				Rescue				Accuracy
	Precision	Recall	F1	Support	Precision	Recall	F1	Support	
DistilBERT Base Turkish Cased	99	99	99	203891	92	88	90	25609	98
Logistic Regression	97	99	98	203891	91	79	84	25609	97
Random Forest	97	99	98	203891	90	75	82	25609	96
Naive Bayes	96	98	97	203891	87	73	80	25609	95
ConvBERT Base Turkish mc4 Cased	96	98	98	203891	91	72	80	25609	95

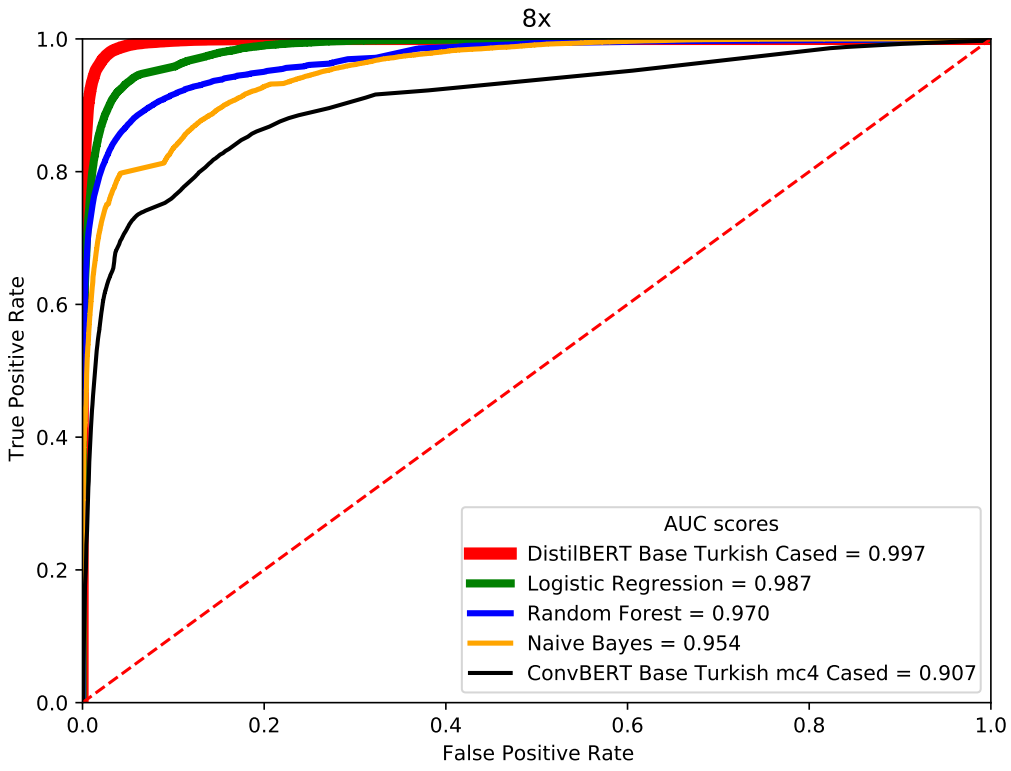

FIGURE 5.4 – Roc Curves of Top Five Models for Imbalance Rate : 8

— F1 score : It is the harmonic mean of Recall and Precision values.

— Support : The number for a class in the dataset split for evaluation.

— Receiver operating characteristic (ROC) Curve : This is a chart used to summarize the classifier's performance over all possible values. It plots True Positive Rate versus False Positive Rate at different classification thresholds.

— The Area Under Curve (AUC) score : It is a measure of how well a parameter can be distinguished between two classes.

One of the ultimate aims of this study is to discern the important tweets for post-earthquakes. The Rescue class is the class we are eager to guess most accurately. Although it can be assessed as one of the indicator of success factors to guess the non-rescue class very accurately, it is the recall value of the rescue class that will show how far we have achieved our ultimate goal.

Table 5.1 presents the results of the best five models trained with a whole AAC dataset, sorted by the recall value of the rescue class. All models performed near perfect for the non-rescue class. The overall accuracy performance of the models is also close to perfect. However, the recall value of the rescue class, which is the most important for us, gave worse results compared to other values. The ROC curves and AUC scores in Figure 5.4 show that they should be improved.

5.2.1 Downsampling

In order to measure and improve the success of the models we use, experiments were carried out from different perspectives. In the AAC dataset, the tweets belonging to Non-rescue class are eight times more than the tweets belonging to rescue class. To handle this type of highly imbalanced data, downsampling method was performed for AAC dataset. The changes of the results were observed, when imbalance rate of dataset has been gradually reduced. While the imbalance rate was eight in the first AAC version, the models were retrained and evaluated in cases where the imbalance rate was four, two and equal by eliminating the non-rescue class.

While the ratio of the number of non-rescue and rescue classes is four in Figure 5.2,

while it is two in Figure 5.3 and one in Figure 5.4, the results of the best five models are shown according to the recall value of the rescue class. By looking at the results

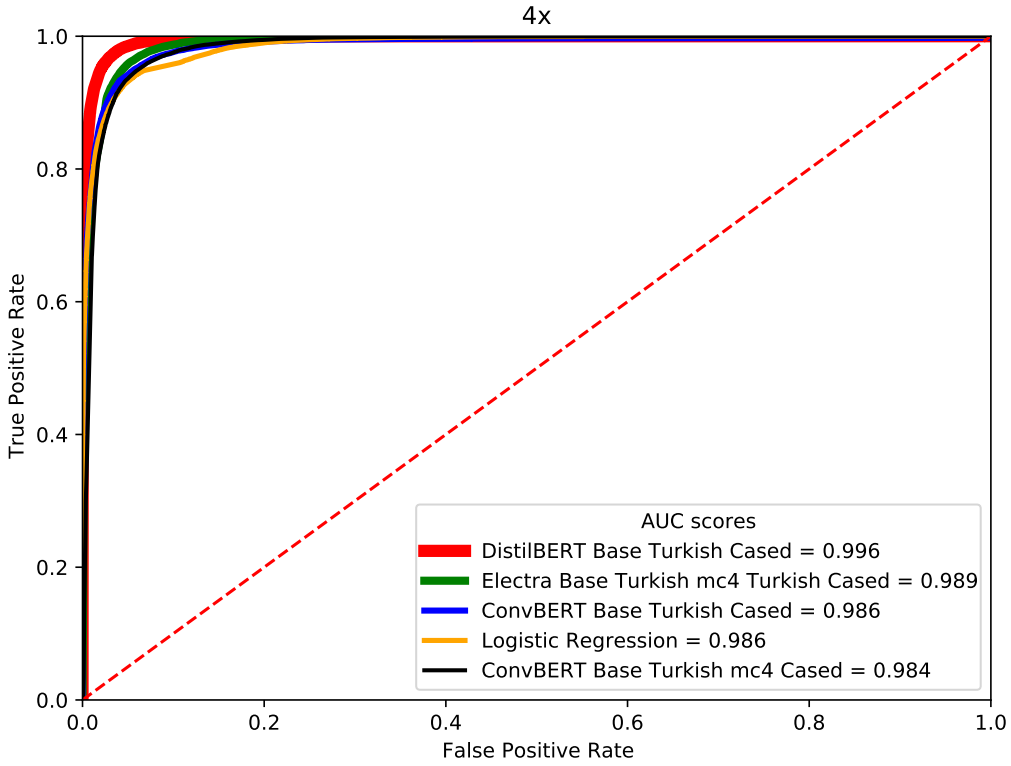

FIGURE 5.5 – Roc Curves of Top Five Models for Imbalance Rate : 4

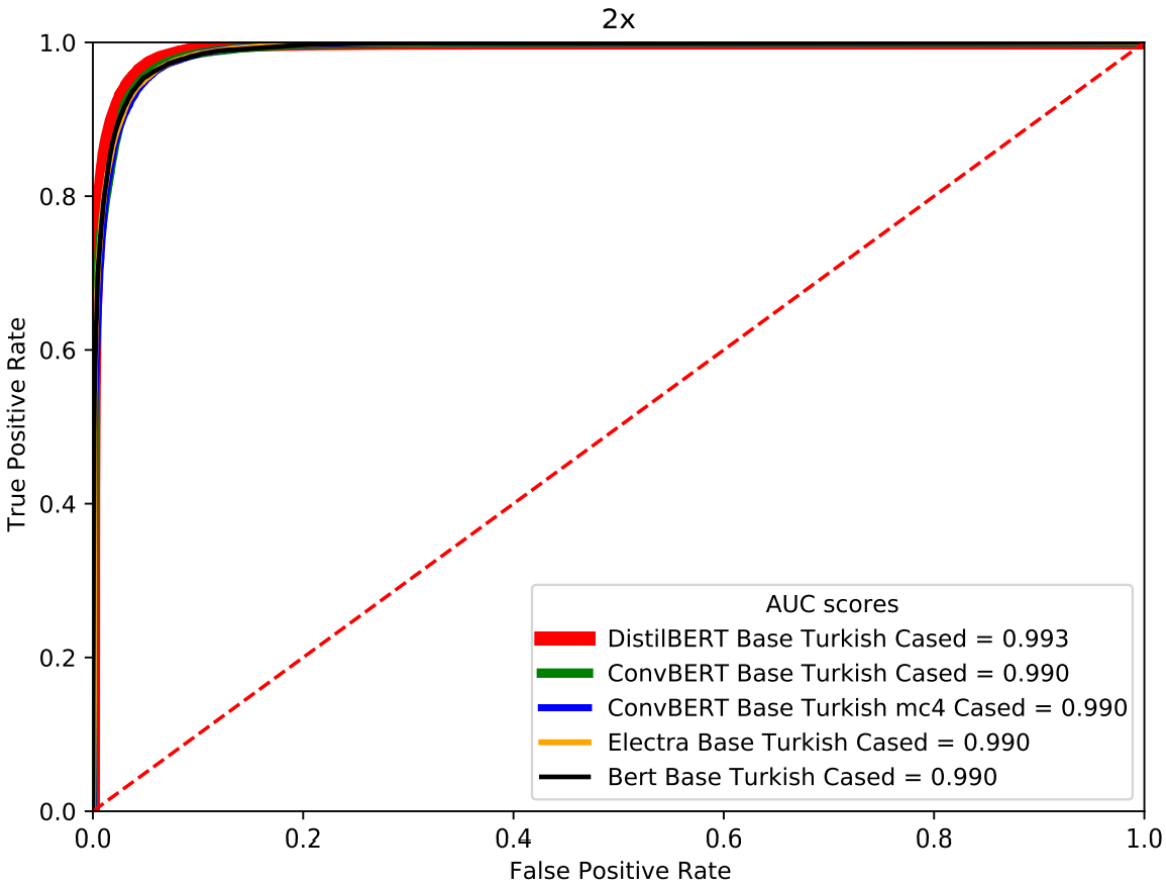

FIGURE 5.6 – Roc Curves of Top Five Models for Imbalance Rate : 2

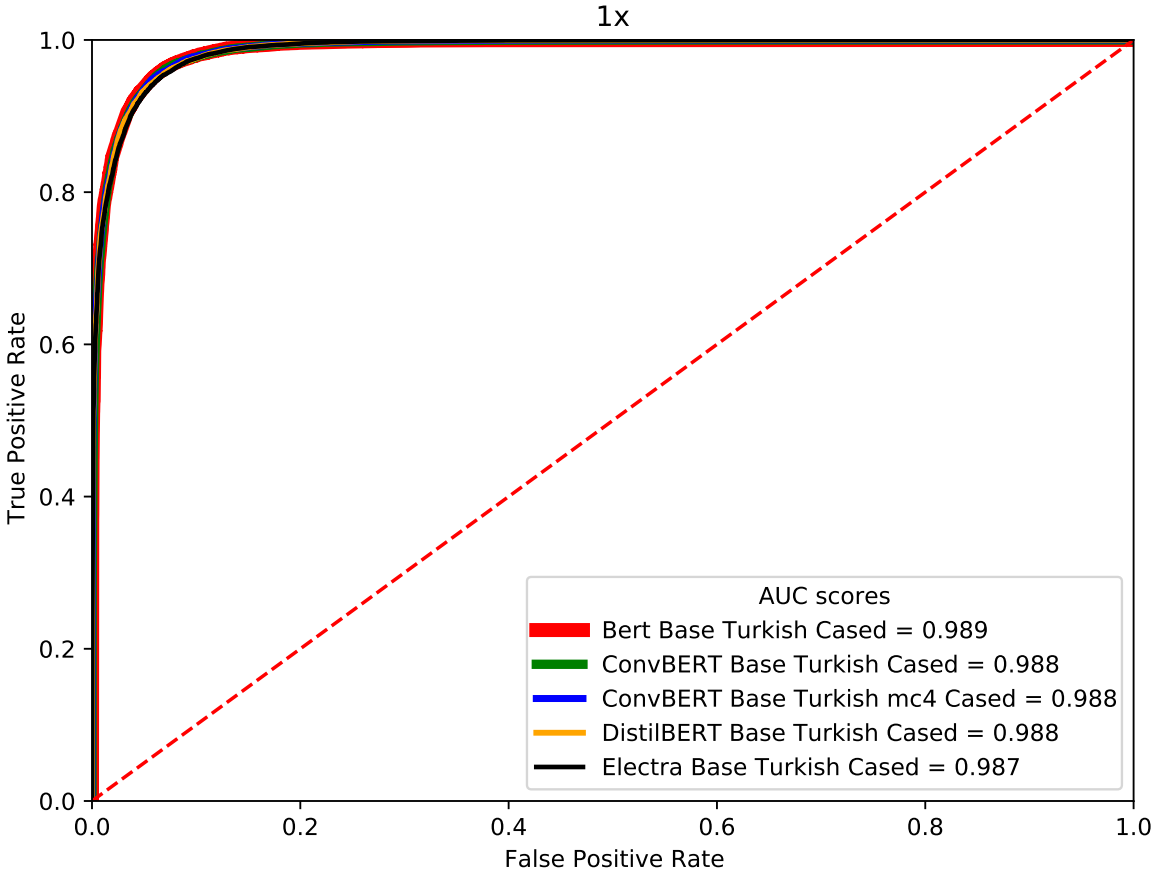

FIGURE 5.7 – Roc Curves of Top Five Models for Imbalance Rate : 1

TABLE 5.2 – Top Five Models for Imbalance Rate : 4

Model	Non-rescue				Rescue				Accuracy
	Precision	Recall	F1	Support	Precision	Recall	F1	Support	
DistilBERT Base Turkish Cased	98	98	98	102069	95	93	94	25431	97
Electra Base Turkish mc4 Turkish Cased	97	98	98	102069	93	91	92	25431	96
ConvBERT Base Turkish Cased	93	98	96	102069	93	90	92	25431	96
Logistic Regression	94	93	94	102069	94	88	91	25431	95
ConvBERT Base Turkish mc4 Cased	94	93	94	102069	93	85	89	25431	94

in the tables, it can be noticed that we are able to increase the metric that is most significant for us with the downsampling method we use to deal with the imbalance dataset. Even though, there are fluctuations and decreases in the non-rescue class and overall accuracy of the models, this is not essential for the study to measure.

The ROC curves plotted as a result of gradual downsampling are displayed in Figure 5.5, 5.6 and 5.7. The class imbalance rates are four, two, and one, respectively. The curves paralleling those in the tables show that the downsampling method also directly affects the AUC scores.

Despite the various differences of results according to the imbalance rates, the DistilBERT model, which is the distilled version of the Turkish BERT model, gives the best results in almost all of them. Since it was created for faster inference, training and prediction, it is expected to give comparatively faster but worse results when compared to other BERT versions in the literature. However, in this study, it has been observed that he can generalize much better in Turkish and learn the characteristics of the rescue class more accurately.

TABLE 5.3 – Top Five Models for Imbalance Rate : 2

Model	Non-rescue				Rescue				Accuracy
	Precision	Recall	F1	Support	Precision	Recall	F1	Support	
DistilBERT Base Turkish Cased	97	97	97	51151	94	96	95	25349	96
ConvBERT Base Turkish Cased	96	94	95	51151	94	96	95	25349	95
ConvBERT Base Turkish mc4 Cased	95	93	94	51151	93	95	94	25349	94
Electra Base Turkish Cased	95	93	94	51151	94	93	94	25349	94
Bert Base Turkish Cased	93	96	95	51151	92	86	89	25349	93

TABLE 5.4 – Top Five Models for Imbalance Rate : 1

Model	Non-rescue				Rescue				Accuracy
	Precision	Recall	F1	Support	Precision	Recall	F1	Support	
Bert Base Turkish Cased	96	95	96	25532	94	97	95	25468	96
ConvBERT Base Turkish Cased	95	95	95	25532	93	97	95	25468	96
ConvBERT Base Turkish mc4 Cased	96	95	96	25532	94	97	95	25468	96
DistilBERT Base Turkish Cased	94	94	94	25532	93	97	95	25468	96
Electra Base Turkish Cased	93	94	94	25532	94	96	95	25468	95

6 CONCLUSION

The capability of disaster management becomes more efficient with the accurate and relevant information resources. The synthesis of disaster information channels is considered as the most challenging issues in smart city analysis. Even if multi-scale and multi-platform resources increase the capacity of information flood, they introduce irrelevant noise that reduces the planning of service quality. In this case, automatic analysis with big data analytics becomes crucial regarding the disaster affected people. Our study focused on the 2020 Izmir earthquake in order to reveal the contents of rescue and non-rescue events.

This study presents two different corpuses for post-disaster tweets which are Manually Annotated Corpus (MAC) and Automatic Annotated Corpus (AAC). Gathering relevant crowd generated data from Twitter is the first step of this study. Then, a small part of the raw and wild data is manually annotated to label the whole dataset semi-supervisedly. A pre-trained language model called BERT is fine-tuned with manually annotated data. The fine-tuned model becomes to able to classify rest of the dataset that has not been labeled yet.

Content-based analysis consists of various methods and techniques such as spatio-temporal analysis and word analysis. In an emergency situation, the most vital informations for rescue teams are the answers to when and where. Locations shared by tweet owner enables us to know where it is posted. The heatmaps of the first three days after earthquake focusing both Turkey and Izmir show that the densest area for rescue tweets is Izmir and neighboring cities which are the nearest districts to the epicenter of earthquake. Word analysis also provides significant insights to separate rescue tweets from non-rescue tweets. The word analysis includes word clouds, co-occurrences of words, word embeddings visualization and topic modeling for both rescue and non-rescue. Word clouds display the most frequent words and phrases for our MAC and AAC datasets. Co-occurrences of words are plotted regarding their co-existing in a tweet. Word embeddings are the vector representations of the words as they are created by the Word2Vec method. These word analysis are applied for both AAC and MAC datasets and they let us to know our semi-supervised automatic annotation method is

adequate for smart disaster planning, since both AAC and MAC outputs after word analysis are pretty similar. Additionally, topic modeling by LDA gives the words for each topic according to their probability. The topics can even be distinguished intuitively by examining the words.

On the other hand, from classical machine learning approaches to currently best language models are implemented to conduct classification process in our semi-supervisedly labeled Automatic Annotated Corpus. To cope with the imbalanced data, downsampling technique have been gradually tested. The rescue and non-rescue classification results have been evaluated through Precision, Recall and F1 metrics, their ROC curves have been plotted and AUC scores have been computed. The most direct-to-the-point metric is decided as recall value of rescue class, since it is more important to predict rescue tweets correctly. Regarding this value, the downsampling method rises the success of the models. Even though the other top five models differ in the changes of imbalance rate of dataset, DistilBert-Turkish provided more accurate classification in terms of rescue class almost every time. We have observed that fine-tuned DistilBert would tackle the generalization of the rescue class.

For the future works, it is planned to generate a larger model to work on all kind of disasters by gathering crowd generated data for different disasters. This larger model enables us to develop a real time dashboards where all the analyses for any disaster are displayed. With the location information, real time maps can show the exact location where needs help. Another goal we want to achieve is extracting locations from the content(image, text, video) to collect more consistent coordinates.

REFERENCES

Aichner, T., Grünfelder, M., Maurer, O. and Jegeni, D. (2021). Twenty-five years of social media : A review of social media applications and definitions from 1994 to 2019, *Cyberpsychology, Behavior, and Social Networking* **24**(4) : 215–222. PMID : 33847527.
URL: *doi.org/10.1089/cyber.2020.0134*

Akın, A. and Akın, M. (2019). Zemberek-nlp.

Aldrich, D. (2010). Fixing recovery : Social capital in post-crisis resilience, *Political Science Faculty Publications* **6**.

Bashir, S., Bano, S., Shueb, S., Gul, S., Mir, A., Ashraf, R., Shakeela and Noor, N. (2021). Twitter chirps for syrian people : Sentiment analysis of tweets related to syria chemical attack, *International Journal of Disaster Risk Reduction* **62** : 102397.

Beedasy, J., Petkova, E., Lackner, S. and Sury, J. (2020). Gulf coast parents speak : children's health in the aftermath of the deepwater horizon oil spill, *Environmental Hazards* **20** : 1–16.

Behl, S., Rao, A., Aggarwal, S., Chadha, S. and Pannu, H. (2021). Twitter for disaster relief through sentiment analysis for covid-19 and natural hazard crises, *International Journal of Disaster Risk Reduction* **55** : 102101.

Bhavaraju, S., Beyney, C. and Nicholson, C. (2019). Quantitative analysis of social media sensitivity to natural disasters, *International Journal of Disaster Risk Reduction* **39** : 101251.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent dirichlet allocation, *J. Mach. Learn. Res.* **3**(null) : 993–1022.

Botzen, W. J. W., Deschenes, O. and Sanders, M. (2019). The economic impacts of natural disasters : A review of models and empirical studies, *Review of Environmental Economics and Policy* **13**(2) : 167–188.
URL: *doi.org/10.1093/reep/rez004*

Cerna, S., Guyeux, C. and Laiymani, D. (2022). The usefulness of nlp techniques for predicting peaks in firefighter interventions due to rare events, *Neural computing amp ; applications* **34**(12) : 10117—10132.
URL: *europepmc.org/articles/PMC8881897*

Clark, K., Luong, M.-T., Le, Q. V. and Manning, C. D. (2020). Electra : Pre-training text encoders as discriminators rather than generators.
URL: *arxiv.org/abs/2003.10555*

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018). Bert : Pre-training of deep bidirectional transformers for language understanding.
URL: *arxiv.org/abs/1810.04805*

Edosomwan, S., Prakasan, S., Kouame, D., Watson, J. and Seymour, T. (2011). The history of social media and its impact on business, *Journal of Applied Management and Entrepreneurship* **16** : 79–91.

Eligüzel, N., Çetinkaya, C. and Dereli, T. (2021). A state-of-art optimization method for analyzing the tweets of earthquake-prone region, *Neural Computing and Applications* **33**.

Erdelj, M., Król, M. and Natalizio, E. (2017). Wireless sensor networks and multi-uav systems for natural disaster management, *Computer Networks* **124**.

Fayjaloun, R., Gehl, P., Auclair, S., Boulahya, F., Guérin-Marthe, S. and Roullé, A. (2021). Integrating strong-motion recordings and twitter data for a rapid shakemap of macroseismic intensity, *International Journal of Disaster Risk Reduction* **52** : 101927.
URL: *www.sciencedirect.com/science/article/pii/S2212420920314291*

Graham, M. W., Avery, E. J. and Park, S. (2015). The role of social media in local government crisis communications, *Public Relations Review* **41**(3) : 386–394.
URL: *www.sciencedirect.com/science/article/pii/S0363811115000077*

Hasfi, N., Fisher, M. and Sahide, M. (2021). Overlooking the victims : Civic engagement on twitter during indonesia's 2019 fire and haze disaster, *International Journal of Disaster Risk Reduction* **60** : 102271.

Jiang, Z., Yu, W., Zhou, D., Chen, Y., Feng, J. and Yan, S. (2020). Convbert : Improving bert with span-based dynamic convolution.
URL: *arxiv.org/abs/2008.02496*

Kankanamge, N., Yigitcanlar, T. and Goonetilleke, A. (2020). How engaging are disaster management related social media channels ? the case of australian state emergency organisations, *International Journal of Disaster Risk Reduction* **48** : 101571.

Kankanamge, N., Yigitcanlar, T., Goonetilleke, A. and Kamruzzaman, M. (2019). Determining disaster severity through social media analysis : Testing the methodology with south east queensland flood tweets, *International Journal of Disaster Risk Reduction* **42** : 101360.

Liu, H., Morstatter, F., Tang, J. and Zafarani, R. (2016). The good, the bad, and the ugly : uncovering novel research opportunities in social media mining, *International Journal of Data Science and Analytics* **1**.

Mangalathu, S., Sun, H., Nweke, C., Yi, Z. and Burton, H. (2019). Classifying earthquake damage to buildings using machine learning, *Earthquake Spectra* **36**.

Mayfield, A. (2008). What is social media ? icrossing.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013). Distributed representations of words and phrases and their compositionality, *Neural and Information Processing System (NIPS)*.
URL: *papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf*

Mora, K., Chang, J., Beatson, A. and Morahan, C. (2015). Public perceptions of building seismic safety following the canterbury earthquakes : A qualitative analysis using twitter and focus groups, *International Journal of Disaster Risk Reduction* **13**.

Neppalli, V., Caragea, C., Squicciarini, A., Tapia, A. and Stehle, S. (2017). Sentiment analysis during hurricane sandy in emergency response, *International Journal of Disaster Risk Reduction* **21** : 213–222.

Pourebrahim, N., Sultana, S., Edwards, J., Gochanour, A. and Mohanty, S. (2019). Understanding communication dynamics on twitter during natural disasters : A case study of hurricane sandy, *International Journal of Disaster Risk Reduction* **37** : 101176.
URL: *www.sciencedirect.com/science/article/pii/S2212420918310434*

Sanh, V., Debut, L., Chaumond, J. and Wolf, T. (2019). Distilbert, a distilled version of bert : smaller, faster, cheaper and lighter.
URL: *arxiv.org/abs/1910.01108*

Schweter, S. (2020). Berturk - bert models for turkish.
URL: *doi.org/10.5281/zenodo.3770924*

Simon, T., Goldberg, A. and Adini, B. (2015). Socializing in emergencies—a review of the use of social media in emergency situations, *International Journal of Information Management* **35**(5) : 609–619.
URL: *www.sciencedirect.com/science/article/pii/S0268401215000638*

White, C., Plotnick, L., Kushma, J., Hiltz, S. and Turoff, M. (2009). An online social network for emergency management, *International Journal of Emergency Management* **6**.

Zhai, W. and Yuan, F. (2020). Examine the effects of neighborhood equity on disaster situational awareness : Harness machine learning and geotagged twitter data, *International Journal of Disaster Risk Reduction* **48** : 101611.

Zhang, M. and Wang, J. (2022). Global flood disaster research graph analysis based on literature mining, *Applied Sciences* **12** : 3066.

Zhu, W., Liu, K., Wang, M., Ward, P. J. and Koks, E. E. (2022). System vulnerability to flood events and risk assessment of railway systems based on national and river basin scales in china, *Natural Hazards and Earth System Sciences* **22**(5) : 1519–1540.
URL: *nhess.copernicus.org/articles/22/1519/2022/*

APPENDIX A TRANSLATION OF SIGNIFICANT WORDS FOR THIS STUDY

Turkish	**English**
acil	Urgent
adres	Address
allah	God
allah rahmet eylesin	Rest In Peace
altında	Under
announceemoji	See preprocessing section
apartman	Apartment
araştırma	To search
arkadaşlar	Friends
battaniye	Blanket
bayraklı	Bayraklı Neighbourhood (Proper Noun)
bağış	Donation
bey	Rıza Bey Apartment (Proper Noun)
birlikte	Together
bornova	Bornova District of Izmir
can	Life
contactnumber	See preprocessing section
cumhurbaşkanı	President
daire	Flat
destek	Support
devlet	State
elden ele	hand to hand
elif	One of the sufferers (Proper Noun)
enkaz	Debris
etkilenen	Affected
geçmiş olsun	Get well soon
gönder	To send
güvenli	Secure
ihtiyaç	Need

Turkish	**English**
iletişim	Contact
imagelink	See preprocessing section
internet	Internet
inşallah	Hopefully
kan	Blood
kayıp	Loss
kişi	Person
kurtarma	Rescue
lütfen	Please
mahalle	Neighborhood
mansuroğlu	Mansuroglu Apartment (Proper Noun)
meydana gelen	Occurring
meşgul	Busy
rabbim	God
rıza	Rıza Bey Apartment (Proper Noun)
sms	SMS
sokak	Street
son	End
sonra	After
tekin	Tekin Apartment (Proper Noun)
telefon	Phone
toplanma	Assembly (area)
vergiler	Taxes
yardım	Help
yayalım	let's spread (Referring a tweet contains rescue info)
yayın	let's spread (Referring a tweet contains rescue info)
yiğit	One of the sufferers (Proper Noun)
yıkılan	Collapsed
çökmüş	Collapsed
önemli	Important
özge	One of the sufferers (Proper Noun)

BIOGRAPHICAL SKETCH

Ali Burak Can has studied Computer Engineering in Istanbul University between 2013 - 2018. In 2019, he enrolled Galatasaray University Computer Engineering Department for master study. He is currently working as a Machine Learning Engineer in a tourism company.