MACHINE LEARNING OF SOCIAL MEDIA DATA ON A SPATIO - TEMPORAL BASIS

(SOSYAL MEDYA VERİLERİNİN ZAMAN-MEKANSAL TEMELLERE GÖRE MAKİNE ÖĞRENİMİ)

by

B Ü Ş R A Y E Ş İ L B A Ş , 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: Prof. Dr. TANKUT ACARMAN

June 2024

TABLE OF CONTENTS

LIST OF ABBREVIATIONS ..... **iv**
LIST OF FIGURES ..... **v**
LIST OF TABLES ..... **vii**
ABSTRACT ..... **viii**
ÖZET ..... **x**
1. INTRODUCTION **1**
2. LITERATURE REVIEW **7**
2.1 Recent Studies in the Field of Text Classification 8
2.2 Advancements in Machine Learning for Disaster Response and Recovery 9
2.3 Deep Learning Approaches in Disaster and Response and Recovery 15
3. PROCESSING OF THE DATASET **19**
3.1 Data Collection and Description 19
3.2 Preprocessing 21
3.3 Developing a Manually Annotated Dataset 24
3.4 Developing an Automatic Annotated Dataset 24
4. PRESENTED DISASTER CLASSIFICATION **27**
4.1 Machine Learning Methods 28
4.1.1 BERT Models 28
4.1.2 Supervised Learning Algorithms 32
4.1.2.1 Logistic Regression 32
4.1.2.2 Support Vector Machines 33
4.1.2.3 Decision Tree 34
4.1.2.4 Multinomial Naïve Bayes 35
4.1.2.5 Stochastic Gradient Descent 36
4.1.2.6 XGBoost 37
4.2 Implementation of Classification Algorithms 37
4.2.1 K-fold Cross-Validation 39

5. DEEP LEARNING 40
5.1 Deep Learning Algorithms 40
5.1.1 Convolutional Neural Network 40
5.1.2 Deep Neural Network 41
5.1.3 Long Short-Term Memory 42
5.2 Implementations of Deep Learning Algorithms 43
6. NATURAL LANGUAGE PROCESSING 45
6.1 Extracting Keywords 45
6.1.1 Topic Modeling 45
6.1.2 FuzzyTM 47
7. EXPERIMENTAL STUDY 48
8. CONCLUSIONS AND FUTURE WORK 66
REFERENCES 68
APPENDICES 73
Appendix A 73

LIST OF ABBREVIATIONS

AI	**:** Artificial Intelligence
ANN	**:** Artificial Neural Network
ASSLSM	**:** Adjusted Semi-Supervised Learning for Social Media
BERT	**:** Bidirectional Encoder Representations from Transformers
CNN	**:** Convolutional Neural Network
DL	**:** Deep Learning
DNN	**:** Deep Neural Network
ELECTRA	**:** Efficiently Learning an Encoder that Classifies Token Replacements Accurately
LSTM	**:** Long Short-Term Memory
MAE	**:** Mean Absolute Error
ML	**:** Machine Learning
MNB	**:** Multinomial Naïve Bayes
NLP	**:** Natural Language Processing
RF	**:** Random Forest
RNN	**:** Recurrent Neural Network
SGD	**:** Stochastic Gradient Descent
SVC	**:** Support Vector Classifier
XGBoost	**:** eXtreme Gradient Boosting

LIST OF FIGURES

Figure 1.1 Schematic of Social Media Data in Disaster Management Cycle (Joseph et al., 2018) 2
Figure 2.1 Overview of Text Classification Pipeline 8
Figure 2.2 The Architecture of NLP-based VictimFinder Model 11
Figure 2.3 The Proposed Source-Location Prediction Framework Architecture 13
Figure 2.4 The Overall Architecture of CNN Model 16
Figure 2.5 The Multimodal Architecture for the Classification using Text and Image as Input in the System 17
Figure 2.6 CNN Architecture in Transformer-based Bidirectional Attention Model 18
Figure 3.1 An Example of the Data Frame Showing Sample from the Dataset 25
Figure 3.2 Flow Diagram of Manual and Automatic Labeling, Classification of Rescue and Non-rescue Calls, and Natural Language Processing 26
Figure 4.1 The BERT Model Architecture in Sentence Pair Classification 30
Figure 4.2 Support Vector Representation (Kanade et al., 2022) 33
Figure 4.3 Diagram of a Decision Tree Structure 34
Figure 4.4 Gradient Descent Compared with Stochastic Gradient Descent 36
Figure 5.1 The structure of a CNN, consisting of convolutional, pooling, and fully-connected layers (Albelwi et al., 2017) 41
Figure 5.2 General Architecture for a Deep Neural Network with Two Hidden Layers (Shehzad et al., 2021) 42
Figure 5.3 General Architecture of LSTM 43
Figure 5.4 The flowchart of the Deep Learning Implementations 43
Figure 6.1 Visualization of How Topic Modeling Works (Pykes et al., 2023) 46
Figure 7.1 Heat Map of the Clusters and Keywords 58
Figure 7.2 Heat Map of the Clusters-Times and Keywords 59

Figure 7.3 Heat Map of Cities and Keywords 61
Figure 7.4 Number of Coordinates in the Map of Turkiye 62
Figure 7.5 Example Map from the Clusters in Kahramanmaraş City..... 63
Figure 7.6 Example Map from the Cluster in Hatay City..... 64
Figure 7.7 Representation of Words on the Map of Turkiye 65

LIST OF TABLES

Table 2.1 Comparing the Performance of Language Model with Existing Models in the Literature 14
Table 7.1 Performance Metrics of Logistic Regression in Models for Rescue Class 49
Table 7.2 Performance Metrics of Logistic Regression in Models for Non-Rescue Class 49
Table 7.3 Performance Metrics of SVC in Models for Rescue Class 50
Table 7.4 Performance Metrics of SVC in Models for Non-Rescue Class 50
Table 7.5 Performance Metrics of Decision Tree in Models for Rescue Class 51
Table 7.6 Performance Metrics of Decision Tree in Models for Non-Rescue Class 51
Table 7.7 Performance Metrics of MNB in Models for Rescue Class 52
Table 7.8 Performance Metrics of MNB in Models for Non-Rescue Class 52
Table 7.9 Performance Metrics of XGBoost in Models for Rescue Class 53
Table 7.10 Performance Metrics of XGBoost in Models for Non-Rescue Class 53
Table 7.11 Performance Metrics of SGD in Models for Rescue Class 54
Table 7.12 Performance Metrics of SGD in Models for Non-Rescue Class 54
Table 7.13 Performance Metrics of Deep Learning Algorithms for Rescue class 56
Table 7.14 Performance Metrics of Deep Learning Algorithms for Non-Rescue class…56
Table 7.15 Keywords in Texts into the Distinct Clusters 57

ABSTRACT

In recent years, machine learning algorithms have greatly revolutionized the methods used to analyze vast quantities of social media data and extract valuable insights from it. Researchers can leverage algorithms to enhance their understanding of human behaviors, reactions, and emotions, particularly in the context of natural disasters such as earthquakes. Analyzing social media data in terms of both space and time is extremely important. Utilizing machine learning on social media data related to earthquakes in a spatio-temporal manner allows for prompt interventions and allocation of resources, enabling early identification and rapid reaction by evaluating geo-tagged postings and real-time information sharing. It improves communication among individuals or groups involved and supports the ability to endure and recover from natural disasters by offering valuable and significant data on analyzing emotions, long-term strategies for recovery, and geographical patterns of damage. This study exploits crowdsourcing via social media data to extract information about emergency situations and needs after the earthquake.

Using the semi-supervised method, the data has been labeled as either rescue or non-rescue to reach a high level of accuracy in detection. After the earthquake, rescue situations are detected on a spatial and temporal basis, along with location and time information provided by tweets. Two destructive earthquakes of magnitudes of Mw 7.7 and Mw 7.6 occurred on February 6, 2023, in the southeast of Turkiye. 53.537 people died, 107.213 people were injured, and several buildings were damaged. A total of 2.5 million tweets related to these earthquakes were collected from February 6 to February 28, 2023, through the X platform. For labeling purposes, nine BERT language models that are based on attention and transformers were used. Supervised learning methods, including logistic regression, support vector machines, decision trees, multinomial Naïve Bayes, and XGBoost, were applied to assess the precision of the labels and perform classification. Furthermore, the data set was processed with deep learning methods:

convolutional neural networks, deep neural networks, and long short-term memory. A timely and proper response to delivering efficient solutions is possible only when the requests for assistance, rescue, and emergency are promptly and accurately understood. In this thesis, we determined the key terms of each data set through an extensive study of its spatio-temporal dynamics, allowing us to identify urgent supplies and use protection at the appropriate time quickly and clearly. The accuracy of data toward the detection of rescue and non-rescue situations is compared, and keywords on a spatio-temporal basis are extracted to determine hazard situations and emergency needs for coordination purposes. Deep learning and BERT models for detection of rescue and non-rescue classes reach a level of 0.8912 and 0.9792 in recall, respectively.

This study highlights the vital importance of machine learning and deep learning in extracting valuable and applicable insights from social media data, especially in urgent scenarios like natural disasters. It achieves this by providing a thorough comprehension of the changing patterns that occur after the earthquake in Turkiye, incorporating both spatial and temporal factors into the analysis. The results demonstrate the effectiveness of the models in classifying microblogs connected to disasters.

Keywords : Text Classification, Social Media Data, Disaster Management, Machine Learning, Deep Learning, Natural Language Processing.

ÖZET

Makine öğrenimi algoritmaları, son yıllarda çok miktarda sosyal medya verisini analiz etmek ve bu verilerden değerli bilgiler çıkarmak için kullanılan yöntemlerde büyük ölçüde devrim yarattı. Araştırmacılar, özellikle deprem gibi doğal afetlerde insan davranışlarını, tepkilerini ve duygularını daha iyi anlamak için bu algoritmalardan yararlanabilmektedirler. Sosyal medya verilerinin hem mekan hem de zaman açısından analiz edilmesi son derece önemlidir. Depremlerle ilgili sosyal medya verilerinde makine öğreniminin uzay-zamansal bir şekilde kullanılması, coğrafi etiketli gönderileri ve gerçek zamanlı tartışmaları değerlendirerek erken tespit ve hızlı müdahaleyi sağlayarak kaynakların tahsis edilmesine ve hızlı müdahale edilmesine olanak tanır. Ayrıca, bireyler veya gruplar arasındaki iletişimi geliştirir ve duyguların analizi, uzun vadeli iyileşme stratejileri ve coğrafi hasar kalıpları hakkında değerli ve önemli veriler sunarak doğal afetlere dayanmayı ve bu afetlerden kurtulmayı destekler. Bu çalışma, deprem sonrası acil durumlar ve ihtiyaçlar hakkında bilgi elde etmek için sosyal medya verileri aracılığıyla kitle kaynak kullanımından yararlanmaktadır.

Yarı denetimli yöntem kullanılarak veriler, tespitte yüksek doğruluk seviyesine ulaşmak için kurtarma veya kurtarma dışı veriler olarak etiketlenmiştir. Deprem sonrası kurtarma durumları, gönderiler tarafından sağlanan konum ve zaman bilgileri ile mekansal ve zamansal olarak tespit edilmektedir. Türkiye'nin Güneydoğusu'nda 6 Şubat 2023 tarihinde 7,7 Mw ve 7,6 Mw büyüklüğünde iki yıkıcı deprem meydana gelmiş, 53.537 kişi ölmüş, 107.213 kişi yaralanmış, çok sayıda bina hasar görmüştü. 6 Şubat'tan 28 Şubat 2023'e kadar bu depremlerle ilgili toplam 2,5 milyon gönderi X platformu üzerinden toplanmıştır. Sınıflandırma amacıyla toplam 9 tane BERT dil modeli kullanılmıştır. Sınıflandırma işlemlerini gerçekleştirmek için Lojistik Regresyon, Destek Vektör Makineleri, Karar Ağacı, Multinomial Naïve Bayes ve XGBoost gibi denetimli öğrenme yöntemleri uygulanmıştır. Ayrıca veriseti Evrişimli Sinir Ağları, Derin Sinir Ağları ve

Uzun Kısa-Süreli Bellek gibi derin öğrenme yöntemleriyle de işlenmiştir. Etkin çözümlerin zamanında ve doğru şekilde karşılanması ancak yardım, kurtarma ve acil durum taleplerinin zamanında ve doğru anlaşılması ile mümkündür. Bu nedenle, her bir veri kümesinin temel kelimeleri, acil malzemeleri zamanında ve açık bir şekilde belirlememize ve korumayı doğru zamanda kullanmamıza olanak sağlayacak şekilde, uzay-zamansal dinamiklerin kapsamlı bir şekilde incelenmesiyle belirlenmiştir. Kurtarma ve kurtarma dışı durumların tespitine yönelik verilerin doğruluğu karşılaştırılmış ve koordinasyon amaçlarına yönelik tehlike durumlarını ve acil durum ihtiyaçlarını belirlemek için afetin yarattığı yıkım ve tahribat ile ilgili anahtar kelimeler çıkarılmıştır. Kurtarma ve kurtarma dışı sınıfların tespitine yönelik derin öğrenme ve BERT modelleri, duyarlılıkta sırasıyla 0,8912 ve 0,9792 seviyelerine ulaşmıştır.

Bu çalışma, özellikle doğal afetler gibi acil senaryolarda sosyal medya verilerinden değerli ve uygulanabilir bilgiler elde etmede makine öğrenimi, derin öğrenme ve doğal dil işleme konularının önemini vurgulamaktadır. Bunu, Türkiye'de deprem sonrasında meydana gelen değişen modellerin kapsamlı bir şekilde anlaşılmasını sağlayarak, hem mekansal hem de zamansal faktörleri analize dahil ederek sağlamaktadır. Sonuçlar, afetlerle bağlantılı verilerin sınıflandırılmasında modellerin etkinliğini göstermektedir.

Anahtar Kelimeler : Metin Sınıflandırma, Sosyal Medya Verisi, Afet Yönetimi, Makine Öğrenmesi, Derin Öğrenme, Doğal Dil İşleme.

1. INTRODUCTION

Natural disasters are occurrences that result from specific natural phenomena on Earth and have the potential to cause disastrous effects. The categories of natural disasters are likely to be widely divergent in terms of causality and nature. Many natural disasters include earthquakes, floods, hurricanes, tsunamis, landslides, avalanches, volcanic eruptions, droughts, and heavy rainfall. Some disasters explode quickly, while others build slowly over time due to unfavorable conditions, but most will probably have a devastating impact on society and the environment given the severity of their impacts. Natural disasters disproportionately impact vulnerable populations such as the poor, elderly, children, and marginalized communities due to their vulnerability to factors such as inadequate infrastructure, lack of access to resources, and limited capacity to cope with a disaster. Natural disasters can cause emergency situations because they are sudden and have devastating consequences for humans and the environment. Emergencies have the potential to disrupt and threaten the lives and livelihoods of individuals, resulting in both material and psychological losses. Additionally, they can cause environmental harm. Therefore, disaster preparedness, mitigation, and response approaches play a critical role in minimizing the impact of natural disasters and enhancing resilience in vulnerable communities.

Social media has a substantial impact on natural disaster events because of its ability to quickly distribute information, facilitate coordination and collaboration, enable crowd-sourced reporting, and enhance alerts about needs and situations. By utilizing social media, emergency responders can optimize their understanding of the situation, improve response coordination, and ultimately preserve lives and alleviate the consequences of disasters on susceptible communities. In conclusion, social media data is important in increasing situational awareness, emergency response, community engagement, and disaster operations before, during, and after a natural disaster.

Effective use of social media can help save lives, cut down on damages, and increase the resilience of disaster stricken areas. Currently, the applications found on social media platforms have became the central focus of user interaction and engagement. One of these programs, X, formerly known as Twitter, has a significant role in providing comprehensive information on natural disasters, including alerts, destruction, and casualty figures, accompanied by visual media such as images, text, and videos. The system offers up-to-the-minute data, opportunities for community involvement, extensive reach, and the ability to analyze data.

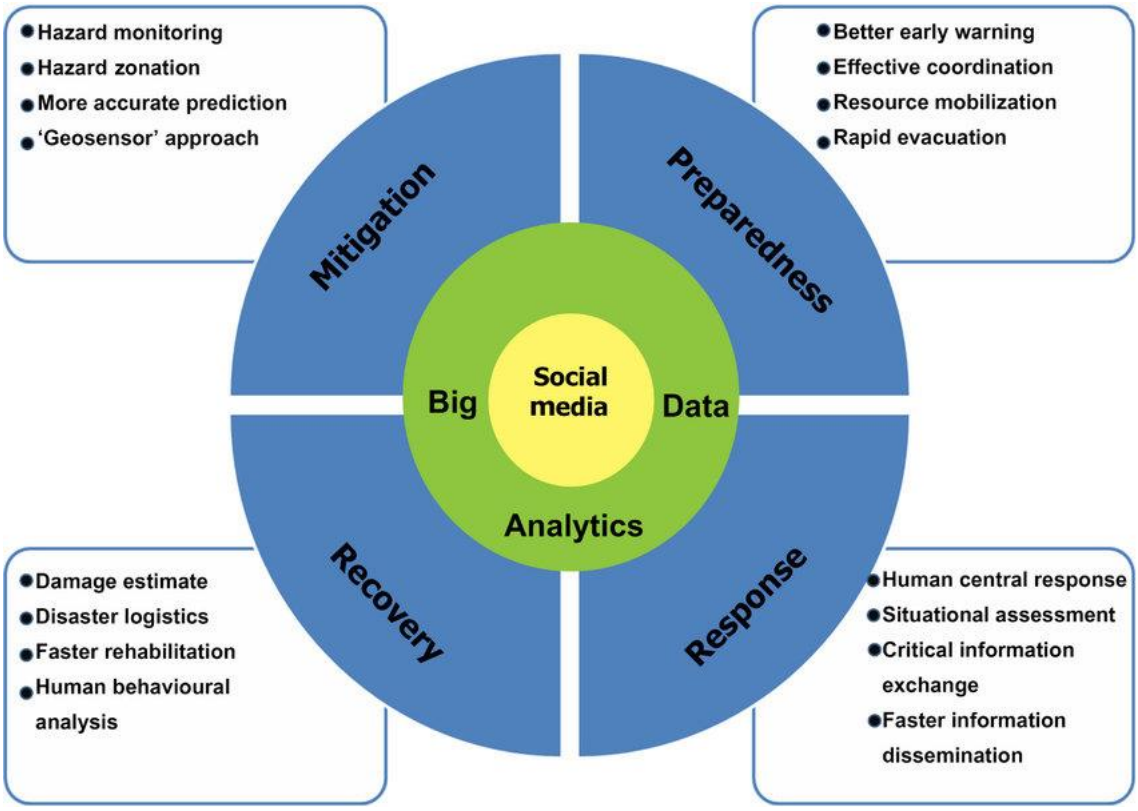

Figure 1.1: Schematic of Social Media Data in Disaster Management Cycle (Joseph et al., 2018)

Disaster management can be defined as the preparation and response to natural or human-made disasters, as well as the subsequent efforts to recover and reduce the impact. In that regard, disaster management is a collaborative process where the government, emergency

services, non-governmental organizations, local communities, and individuals in general work together. In the disaster management cycle, four fundamental stages of big data analytics and social media are generally defined. Figure 1.1 shows that these stages consist of mitigation, preparedness, response, and recovery. The first stage of *Mitigation or Prevention* encompasses activities designed to decrease or remove the risks and vulnerabilities that cause disasters or reduce their effects even if they occur. The typical strategies and actions at this stage are as follows: risk assessment, public awareness and education, improving infrastructure and building standards, environmental protection, investment in technology, and innovation. In general, this phase consists of a set of proactive steps to lower the frequency and outcome of burdens, which subsequently reduce the communities' susceptibility to the exposure and improve their ability to recover from triggers. The second step *Preparedness* highlights planning and preparation activities for a possible disaster or emergency before it happens. National emergencies, telecommunication and response plans, training and exercises, early warning, and resource mobilization are the major activities whose implementation is conducted during the preparedness phase. *Response* is the stage in the disaster cycle where instant measures are taken to mitigate the disaster's impact. Some of the activities to be involved during the response phase include search and rescue operations, communication and coordination with the public and other response agencies, medical aid, sheltering, food and water, restoration of critical services, and managing the immediate consequences of the disaster. The main aim of that phase is to save lives and stabilize the situation with a minimalistic level of suffering. The *Recovery* phase is the time after the immediate response to a disaster. This phase is characterized by a change in focus from saving and shoring up destruction to restoring complexes of communities, infrastructure, and life-supporting services to an original or near-original state. As a rule, recovery requires short- and long-term methodologies that can rebuild, repair, and renew places devastated by human activities. Damage estimate, infrastructure restoration, social and psychological support, and economic recovery are the main activities implemented during the recovery stage.

This research suggests an analysis and methods that are possible to implement in each phase of disaster. The reason of processing a post on one of the most alive platforms is when people eagerly share their quick and powerful responses after a disaster occurs.

When people share their opinions and needs on social media in response to a disaster, it can lead to the development of remedies for affected individuals and regions. Acquiring these remedies with the help of processing social media on the recovery stage can be very helpful. Finally, information from all disasters can be used to predict the next potential disasters, which means that it can be used for the prevention stage and the preparedness stage. When assessing catastrophes using textual data, it is necessary to develop a reliable model that can draw precise conclusions by examining the data in terms of both space and time. The model should be able to determine whether the data suggests rescue or non-rescue situation. Having an effective model to analyze natural disaster data, which focuses on both space and time, is critical for improving predictive ability, risk assessment and mitigation, resource use optimization, communication with the public, and adaptation to new developments. It will enable decision-makers to make assisted decisions, possibly lives, and cultivate resilience in the face of natural disasters

This study aims to use classification, algorithms for machine learning, deep learning, and natural language processing methodologies to evaluate social media data related to disasters. The focus is on understanding the spatial and temporal components of data retrieval and analysis. Through utilizing the enormous quantity of content created by users on platforms like X, previously called Twitter, the objective is to enrich decision-making by leveraging crowd sourcing, enhance awareness of the situation, and optimize the effectiveness of responses to disasters. In this study, we aimed to develop a thorough approach for gathering, preparing, labeling, and analyzing social media data related to emergencies, with a specific focus on the case of the devastating earthquake that occurred in Turkiye on February 6, 2023. The earthquake, with a magnitude of Mw 7.7 and centered in Kahramanmaraş, occurred in Turkiye on February 6, 2023, at 04:17, and affected the cities of Kahramanmaraş, Adyaman, Hatay, Malatya, Gaziantep, Adana, Kilis, Osmaniye, Diyarbakır, Şanlıurfa, and Elazığ. In addition, only nine hours after this earthquake, another earthquake with a magnitude of Mw 7.6 occurred, also centered in Kahramanmaraş. Official statistics in Turkiye indicate that the earthquakes resulted in a minimum of 53,537 deaths and over 122,000 injuries.

Two significant earthquakes resulted in destruction throughout an area of approximately 350,000 and impacted a total of 14 million individuals. On the initial day, around 39,000

buildings collapsed, resulting in the destruction or severe damage to 518,000 residences throughout 11 provinces in Turkey. This included numerous historic landmarks. In addition, at least 518,009 homes and over 345,000 apartments were destroyed. After the earthquake, more than 2 million individuals encountered housing difficulties, while a minimum of 5 million people relocated to other places (Wikipedia, 2023).

Our study began by gathering an extensive data set of tweets that covered the critical time frame from February 6 to February 28, 2023, following the occurrence of the Turkiye earthquakes. More than 2 million pieces of data have been gathered concerning these earthquakes. Our goal was to produce a high-quality data set that concentrated on situations that necessitated immediate assistance, such as identifying trapped individuals or details requiring rescue. The data has been investigated on a spatio-temporal basis by constructing a manually labeled corpus and classifying it. Moreover, specific BERT models have been developed to automatically categorize vast amounts of social media data by utilizing this manually labeled corpus. The performance of these classifications has been tested with various methods, including Logistic Regression, Support Vector Machines, Decision Tree, Multinomial Naïve Bayes, Stochastic Gradient Descent, and XGBoost. Training and analyzing a variety of Bidirectional Encoder Representations from Transformers (BERT) models on a tiny quantity of manually labeled tweets have allowed us to conclude on the most effective strategy for detecting or categorizing social media posts requiring more urgent help. Deep learning approaches have been also applied to better analyze the data set. In addition, the data has been viewed on a spatio-temporal basis and significant word in the era of rescue for each cluster were found using the clustering method.

This study is particularly focused on detecting rescue messages and extracting needs and situations from social media data and its tweets after an earthquake disaster. The extracted information from tweets provides keywords related to rescue situations with their global coordinate position and time, which can be used to improve the coordination of limited rescue resources. This study also suggests a fresh method to determine social media for disaster management. When an earthquake occurred, the use of X, formerly Twitter, is a multi-disciplinary analysis to spread awareness around different parts of the disaster. It is important for rescue organizations to quickly rebuild rescue efforts. Undertaking a

sociological examination of the general public at the epicenter will contribute to more successful relief efforts. The results of this study provide information about the spatio-temporal patterns of rescue and non-rescue labels in the 2023 Turkiye earthquake Twitter conversations.

In conclusion, this study provides a multi-disciplinary methodology for social media data application for both disaster management and disaster response, concentrating on spatio-temporal analysis. We strive to harness machine learning, deep learning, and natural language processing to create more effective, dependable, and adaptable emergency services. As a result, the use of such improved techniques will enable more effective methods for diminishing the repercussions of natural calamities in overly vulnerable areas.

The rest of this dissertation is organized as follows : Section 2 reviews the literature about text classification and learning algorithms applied to disaster management. Section 3 analyzes the creation of the data set used throughout the study, explaining its features, preprocessing details, manual labeling, and automatic labeling processes. BERT models and classification algorithms that are used in the detection of rescue and non-rescue classes are elaborated in Section 4. Deep learning algorithms and their implementations in the study are elaborated in Section 5. In Section 6, the steps taken in natural language processing are explained. Section 7 presents and compares the classification and evaluation results. The success rates of algorithms, reviews, and visual representations obtained throughout the study are presented in this section. Finally, the study is concluded by providing an explanation of the overall key points gained from the study in addition to suggestions for future research in Section 8.

2. LITERATURE REVIEW

The integration of machine learning and deep learning methodologies with the examination of social media data has a critical area of investigation, providing distinctive insights on societal occurrences such as disasters. This review analyzes previous research on the utilization of machine learning and deep learning algorithms to analyze social media data related to spatio-temporal issues, with a specific focus on seismic occurrences.

Research on disaster management has traditionally focused on conventional approaches based on operations research and management science principles. Yet, the advent of social media platforms completely changed the landscape by offering immediate, user-created data that has the potential to provide extremely significant insights on community behaviors and dynamics during crises. Nowadays, social media platforms are vital real-time data acquisition sources accessed during a disaster. This is because they provide insights into the spatial and temporal behavior of emergencies. The increasing amount and variety of data generated from social media during a disaster has sparked a growing interest in machine learning and deep learnin algorithms capable of studying the spatio-temporal behavior of the data. Understanding the dynamics of online community activities is crucial, as accurate and timely information can significantly impact decision-making processes related to victim safety, financial matters, environmental conservation, and organizational actions. Hence, through a review of available literature, this thesis seeks to investigate the existing body of knowledge on the use of machine learning and deep learning algorithms on social media data for disaster management.

2.1 Recent Studies in the Field of Text Classification

(Kowsari et al., 2019) offered a short overview of text classification algorithms, containing information about feature extraction, dimensionality reduction, known algorithms, and methods of evaluation while highlighting their limitations and fields of application in real life. Figure 2.1 illustrates their text classification pipeline's structure and implementations. Even though the study does not provide any experimental results, it offers a general background of available techniques and the outcome of applying such. In their study, (Autelitano et al., 2019) proposed a method for real-time extraction of relevant keywords from social media for the duration of the emergency. They enhanced spatio-temporal features to improve the retrieval of posts and used cross-platform scanning to optimize information extraction.

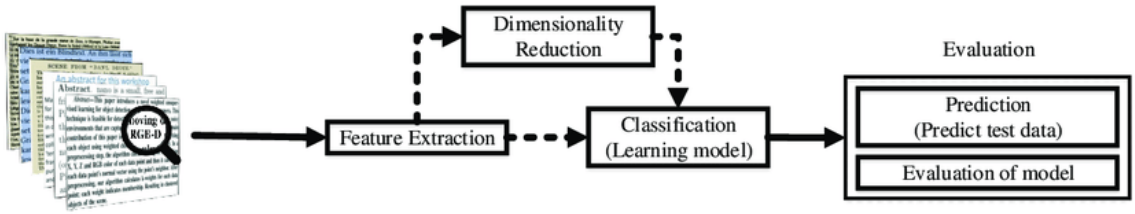

Figure 2.1: Overview of Text Classification Pipeline (Kowsari et al., 2019)

(Scheele et al., 2021) presented a state-of-the-art text-mining method involving spatial data to classify and predict disaster-related social media posts. Furthermore, they concluded that spatial characteristics could be used in conjunction with text data and called for additional investigation of machine learning algorithms and ensemble methods to improve decision-making and information retrieval. (Kumar et al., 2022) introduced a multi - modal approach using a combination of text and image analysis to detect informative content about disasters from Twitter feeds. Researchers created a disaster - oriented system based on information identified from Twitter posts that contain text and images. Using information from the seven data sets, the multi-modality approach hit an F1-score between 0.74 and 0.93, compared to a text-based only F1-score between 0.61 and 0.92. Therefore, using both text and images to qualify information is essential for

identifying informative texts during a crisis. (Contreras et al., 2022) conducted the research looking at the feasibility and accuracy of the sentiment analysis of MonkeyLearn pre-trained model with regard to 2019 Albanian earthquakes on Twitter. In this study, they examined the accuracy of an unsupervised classification algorithm in classifying the polarities of tweeted text. This unsupervised classification was compared with a supervised classification. They established that the pre-trained sentiment analysis model from MonkeyLearn had an accuracy of 0.63, implying that it is effective during the preliminary stage after an earthquake. On the other hand, the researchers raised concerns about the model's tendency to overclassify neutral tweets. Due to the complexity and laboriousness of the process, analyzing and categorizing Turkish texts by their own specific properties is challenging. However, (Savci et al., 2023) concluded that the use of pre-trained language models can be effective in classifying Turkish text, and BERTurk displayed higher accuracy. Turkish NLP text categorization models were evaluated based on training time, accuracy, and CO2. The best model was BERTurk (uncased, 128k) with the highest accuracy and the lowest emissions. The performance was significantly affected by hyperparameters. Additionally, BERTurk (cased) showed better performance than BERTurk (uncased) when the case feature was not even needed. Even though AutoML models learned faster, how they were trained was not explainable.

2.2 Advancements in Machine Learning for Disaster Response and Recovery

(Xiao et al., 2015) examined tweet generation during Hurricane Sandy in New York City. They suggested the MMAM model, which included mass, material, access, and motivation in tweet numbers. Their findings showed correlations with tweet numbers and the size of the population, and landmark sites contributed despite having fewer residents. The study highlighted the importance of social group dynamics for effective social media use in disasters and proposed ground-truth social media-derived situational awareness in future research. A model to assess the impacts of natural catastrophes by using social media articles was demonstrated by (Resch et al., 2018). They then applied machine learning procedures to preprocess the available information and not only extract semantic information patterns but also utilize spatial-temporal research to uncover areas and the timing of the highest frequency of posts.

(Li et al., 2018) investigated a quantification of the patterns and dynamics in social networks generated by retweeting activities at Weibo.com following the Yiliang earthquake. A classification of micro-blog postings was computed using the Multinomial Naïve Bayes Classifier approach to show the strong influence of information type on the diffusion patterns. (Dagaeva et al., 2019) presented a new unique framework and algorithm for analyzing large-scale spatio-temporal datasets. The researchers improved natural language processing combined with weather data aggregation for emergency classification improvement, including the extraction of the most favorable association rules. The authors proposed several algorithms and a framework for big spatio-temporal data mining in the Russian EMIS GLONASS + 112 and provided their evaluation. They presented an improved spatio-temporal co-location pattern mining technique based on DBSCAN, FPGrowth, NLP, spatio-temporal outliers, and spatial autocorrelation detection, as well as algorithm scalability and time performance evaluation. They also presented the use of NLP and emergency data set aggregation with weather to increase the number of emergency categories and improve the quality of rules. As a result, they found significant association rules related to traffic accidents in a dedicated area of Kazan city.

In the study by Eliguzel et al. (2020), practical application of geo-tagged tweets and multiple machine learning approaches, including sentiment analysis and topic modeling, allowed for improving the efficiency of disaster response by reducing time and inaccuracies. Mangalathu et al. (2020) used data from the 2014 South Napa earthquake to analyze the efficacy of various machine learning procedures in predicting the rate of building damage shortly after the calamity. These included k-nearest neighbors, discriminant analysis, random forests, and decision trees.

(Dwarakanath et al., 2021)'s classification of academic research articles on disaster was mapped onto the three distinct stages of disaster. It will help other researchers understand areas that need to be fullfilled by automated information categorization and disaster management. The article justifies the necessity of sifting out useful posts about a disaster from ones of a variety of other categories and to overview the evolution of current machine learning model approaches to provide crisis response based on social media data. Researchers have been examined the role of social media data in systems

prediction and warning at the preparation and early warning stages. One of the studies of this type was performed by (Algiriyage et al., 2021). The research aimed to find the best machine learning or deep learning classifiers of disaster tweets. The scholars investigated three possible situations during the disaster, aftermath of the disaster, and across multiple disasters. They reviewed and compared ML and DL models for disaster tweet classification in various situations. The analysis showed that DL models had higher performance than traditional ML models. In fact, the Bi-LSTM model with Word2Vec outperformed in all situations, including in-disaster, out-disaster, and cross-disaster. The study's results, which used approximately 0.2 million labeled tweets, showed that classifiers could perfectly recognize disaster tweets for all three categories. Nevertheless, it suffers from the limitations of the default parameter setting in ML algorithms and is limited to only English tweets.

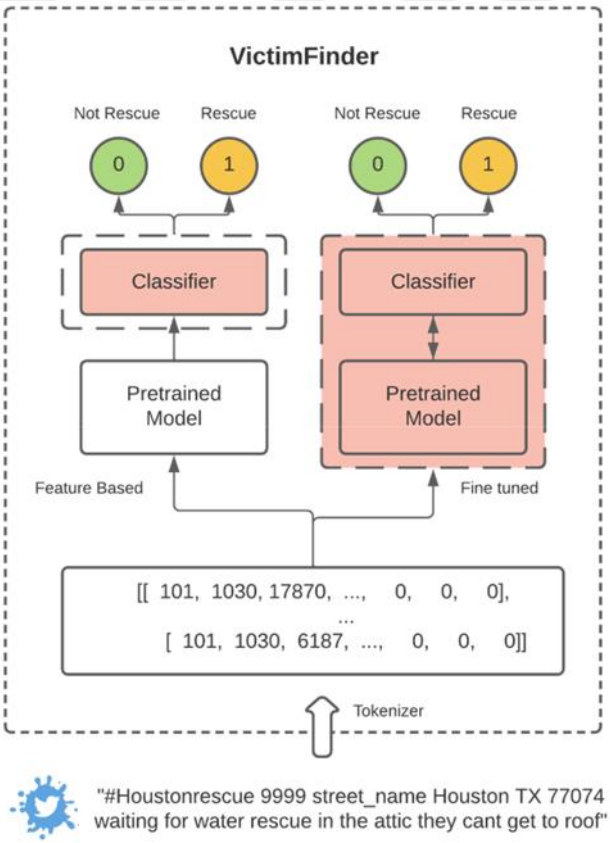

Figure 2.2: The Architecture of NLP-based VictimFinder Model (Zhou et al., 2022)

In the study by (Zhou et al., 2022), they developed and compared ten VictimFinder models for identifying rescue request tweets; three are based on milestone NLP algorithms, and seven are BERT-based. A total of 3191 manually labeled disaster-related tweets posted during 2017 Hurricane Harvey were used as the training and testing data sets. Each model's performance was then evaluated by classification accuracy, computation cost, and model stability. The experimental results clearly showed that all the BERT-based models had significantly improved the accuracy of categorizing rescue-related tweets. The F1-score for the optimal model, a customized BERT-based model with a Convolutional Neural Network, was 0.919, which was 10.6% more than the baseline model. These models can help to facilitate the use of social media in future disaster event rescue operations. Their architecture of the VictimFinder model sis presented in Figure 2.2.

(Linardos et al., 2022) also presented different avenues and case studies of how machine learning could be more successful and used to address problems with disaster response, recovery, and mitigation. The authors of the paper conducted a review of studies published since 2017 on the use of machine learning and deep learning algorithms and methods in the context of disaster management. They divided these techniques according to different phases and subphases of the disaster process and further analyzed the approaches of their application for disaster prediction, risk and vulnerability assessment, disaster detection, early warning, disaster monitoring, damage assessment, and disaster response. The paper summarized the identified problems and prospective research areas and pointed out that the utilization of machine learning and deep learning techniques should contribute to the improvement of disaster recovery operations through better mitigation, vulnerability reduction, and resilience assessment of critical infrastructure. (Saad et al., 2022) proposed another technique; a real-time earthquake localization technique that the authors introduced employs the Random Forest regression and has already demonstrated excellent results in the complicated United seismic region of Japan. Figure 2.3 shows the architecture of their proposed framework. To summarize, the RF model showed the best results when trained and tested on an earthquake data set in Japan. MAE prediction was only 2.88 km, but it remained below 5 km with only 10% of the data and three recording stations. They claimed that it was enough to ensure that the algorithm works effectively. Moreover, the algorithm was also incredibly efficient, adaptable, and

fast, proving to be a useful tool for earthquake location prediction in early warning systems.

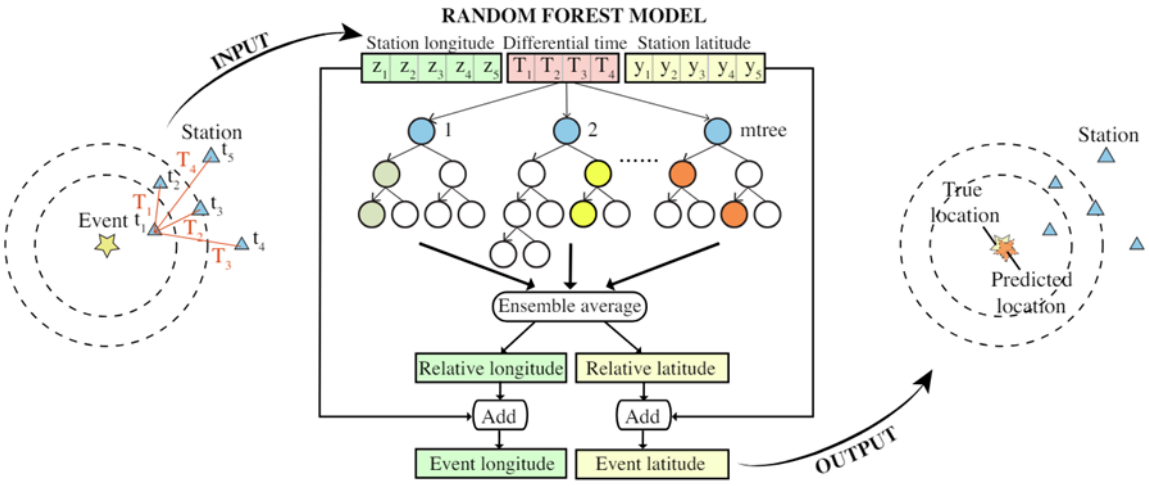

Figure 2.3: The Proposed Source-Location Prediction Framework Architecture (Saad et al., 2022)

(Bensi et al., 2023) explored the possibility of combining social media data with natural language processing techniques and machine learning classifiers to assess earthquake damage. In this study, authors investigated the feasibility of social media data as a secondary data source for damage assessment following an earthquake. They used an intense data set on six earthquake sequences and characteristics of the assessment of damage based on Twitter. In particular, new models using NLP techniques and ML classifiers were developed to classify damage levels from Twitter postings.

Using Twitter data for disaster management, (Manimegalai et al., 2023) applied the BERT model to analyze the information received. The main goal of their work was disaster data analysis, which can be identified as reliable and unreliable, and their further

use in creation schemes with the help of text and natural language processing. The model was trained on 10,000 tweets in this research work. The model with effective performance factors were selected from a set of NLP-based classification approaches. (Elroy et al., 2023) introduced the Adjusted Semi-Supervised Learning for Social Media approach for classifying false earthquake prediction tweets. Specifically, the ASSLSM adjusted the pseudo-labeling constraints based on the metadata and user assumptions to provide the models with more informative guidelines. Their results demonstrated that the ASSLSM methodology further improves the neural network model performance and showed as robust and effective. Moreover, they claimed that using an ASSLSM classifier, we can assign a class to the entire data set and analyze it qualitatively.

Table 2.1: Comparing the Performance of Language Model with Existing Models in the Literature (Powers et al., 2023)

Model	**F1 Score**
BERT - R	0.78
XLNet - R	0.77
BERT - U	0.67
XLNet - U	0.68
CNN + LSTM - KUM	0.77
CNN - BUR	0.84
SVM - BUR	0.83
SVM - YAN	0.69
Logistic Regression - ASH	0.65

(Powers et al., 2023) used machine learning and artificial intelligence to automatically classify Hurricane Harvey tweets relevant to first responders, and they demonstrated the power of language models such as BERT and XLNet in classification tasks and their potential to improve the response phase of disaster response. The results of comparison with their language model performance and existing models in the literature are presented in Table 2.1

Research studies have also focused on the response phase, which involves rapid activities that occurred after the disaster. One of the recent studies that investigated this phase and utilized social media data was conducted by (Cvetković et al., 2023). They outlined several advantages and disadvantages of social media, in particular information sharing and decision-making, but also the risks of spreading misinformation.

2.3 Deep Learning Approaches in Disaster and Response and Recovery

Deep learning algorithms, on the other hand, are a vast avenue of research in natural disaster management and response. (Nyugen et al., 2016) created an online learning model named CNN to classify tweets in a disaster response scenario. They suggested an original online learning algorithm to train CNN in online fashion. They found that this learning technique was the best match for disaster response. They assumed that this base model trained on past crisis labeled data and a small batch of event-specific labeled data, which is used for online learning. They claimed that neural network models had the advantage of learning features automatically, which simplifies the training process when compared to offline learning models such as SVM and logistic regression. They also provided source codes for the online learning of CNN models for further studies. A CNN architecture is introduced by (Huang et al., 2019) to automatically retrieve and label relevant social media posts posted during disasters by integrating visual and textual information. The results indicated that both CNNs performed well in learning visual and textual characteristics. The fused feature showed that incorporating an additional visual feature enhances resilience compared to solely using the textual feature. The on-topic posts were automatically labeled based on their texts and pictures and depict disaster documentation, which happened in real-time during an event. These social media posts, when geotagged with rich spatial context, can significantly assist in multiple disaster mitigation measures.

In another research, (Yu et al., 2020) applied an effective CNN model to Twitter topic classification after various scenarios of hurricanes.

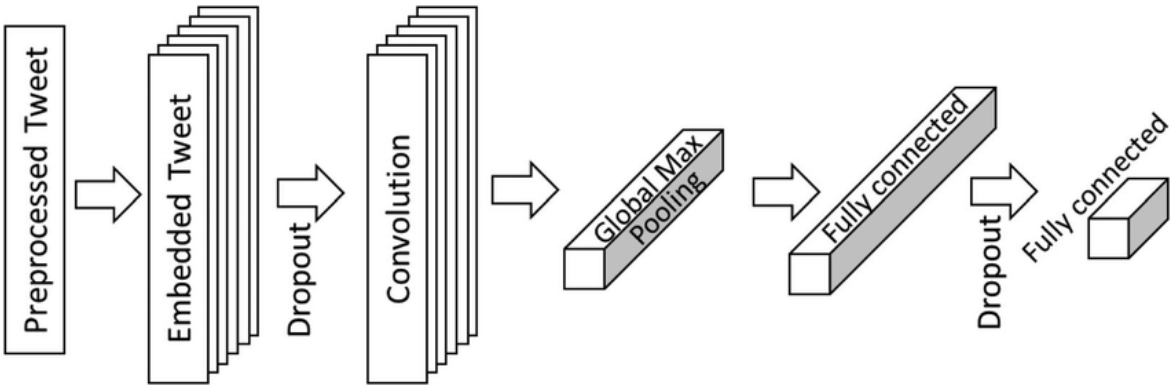

Figure 2.4 The Overall Architecture of CNN Model (Yu et al., 2020)

CNN architecture, which was the definition of the model and included the CNN configuration used to classify tweet themes, is plotted in Figure 2.4. The results indicated that the CNN model outperformed traditional machine learning methods, which indicates its potential to improve situational awareness.

Wang et al. (2020) suggested utilizing spatio-temporal data mining and LSTM networks to predict earthquakes. The possibility that LSTM networks can learn complex spatio-temporal patterns has promising implications for earthquake prediction. According to their research, LSTM networks have the ability to understand complex structures in space and time, leading to significant findings in earthquake prediction. A hybrid approach was also developed by (Ofli et al., 2020) aimed to integrate text and visual components of social media data through a multimodal deep learning framework. They developed a multimodal deep learning architecture along with its modality-agnostic shared representation based on CNNs. Their research demonstrated through extensive experiments with real-world disaster datasets that the proposed multimodal architecture outperforms these models trained on only one modality. The architecture of the multimodal deep neural network that they utilized in their experiment is shown in Figure 2.5. In the image modality, the conventional VGG16 network was used, while a CNN-based architecture was used in the text modality.

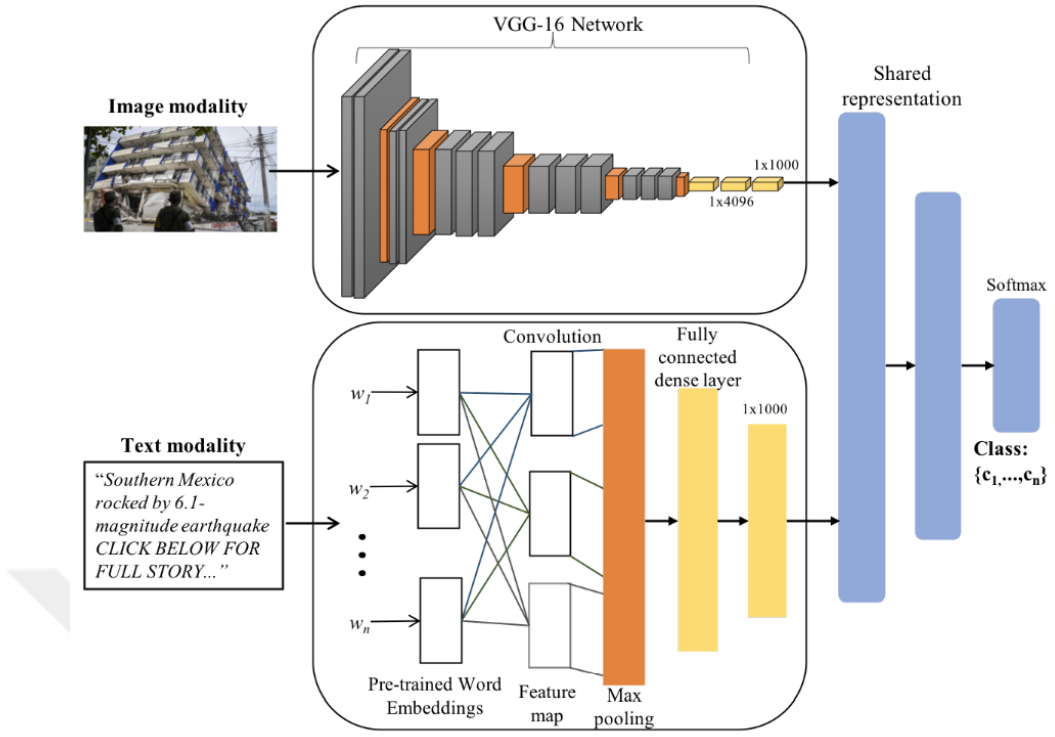

Figure 2.5: The Multimodal Architecture for the Classification using Text and Image as Input in the System (Ofli et al., 2020)

(Jia et al., 2023) conducted a review of the literature from all kinds of aspects, including different assessment objects, data modalities, and deep learning models on the application of deep learning technology in earthquake disaster assessment and protection. They particularly updated the information on the use of CNNs for image classification in structural damage detection. A deep learning framework by (Koshy et al., 2023) classified informative tweets during emergencies using textual and visual aspects and exponential fusion to increase emergency responders' situational awareness. Their system architecture is plotted in Figure 2.6.

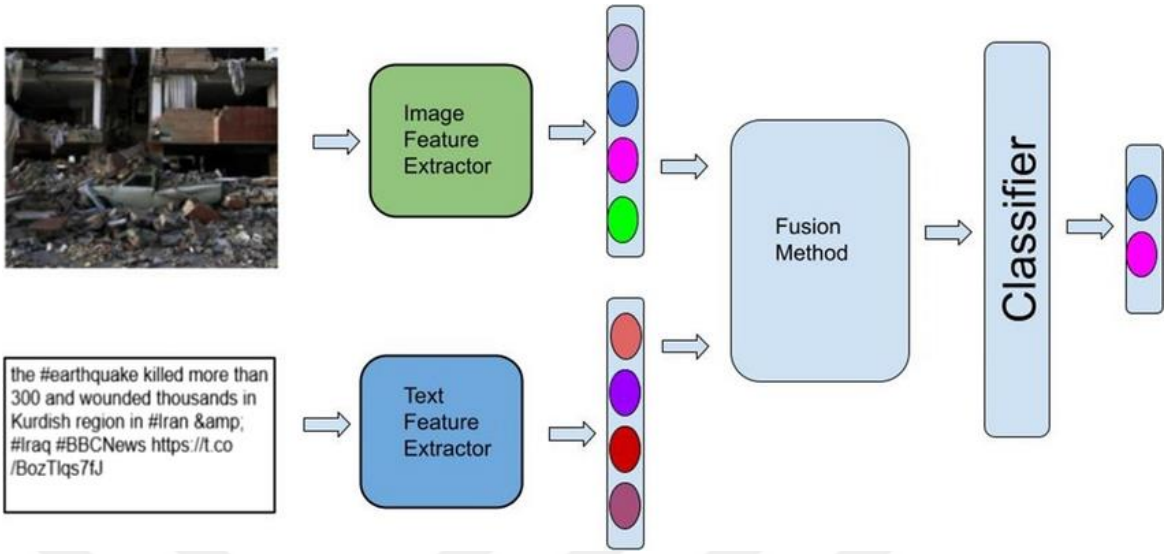

Figure 2.6: CNN Architecture in Transformer-based Bidirectional Attention Model (Koshy et al., 2023)

They specifically employed neural models, specifically a fine-tuned RoBERTa model for texts and a Vision Transformer model for images, along with other BiLSTM and attention components. The accuracy of their approach ranged from 94% to 98% on top of seven data sets of actual disasters, including wildfire, hurricane, earthquake, and flood, and outperformed numerous state-of-the-art methods. Combining text and image information significantly improved their classifiers' performance.

3. PROCESSING OF THE DATASET

Social media data denotes the huge volume of texts, images, videos, links, and all other material that is being generated and spread through social networks by their users. This data is an invaluable source of information about people's moods, trends, and behaviors and it is used in the following ways: marketing, research, and every other feasible goal. Researchers collect and format social media data into distinct collections for study or research purposes, known as social media data sets. This may include items like posts, comments, likes, shares, profiles, or other user-expressed data from a single or multiple platforms. A social media data set is commonly used for research purposes into user interactions, feelings driving factors, social media analysis, and the way in which data is disseminated or misinterpreted.

This chapter presents the data sets used in this thesis. The data processing is demonstrated, starting from the first point and ending with the final one. The first issue concerns receiving raw data. Analysis, modeling, and classification research that we want to conduct from the very beginning require data sets with labels. That is why we have both manually and automatically labeled corpuses. Preprocessing is the key for topic modeling and classical machine learning and deep learning techniques.

3.1 Data Collection and Description

Data collection entails sourcing relevant information from various sources in a systematic manner, while data description entails providing comprehensive summaries and characteristics of the collected data to help in the analysis and interpretation of the results. This approach is effective in ensuring the quality and suitability of the data for further analysis.

Firstly, we outline the research questions and the targeted outcomes. The objective is to obtain the most accurate and truthful information or observations, while discarding unnecessary data. Our objective is to create a high-quality data set focusing on instances in which assistance was urgently required, whether citing that people were trapped or details calling for rescue. For collecting data, X platform, which has a significant role in providing comprehensive information on natural disasters, including alerts, destruction, and casualty figures, accompanied by visual media such as images, text, and videos, is chosen. The data set comprised Turkish posts posted by people about the major earthquake that happened in Turkiye on February 6, 2023, after obtaining the data publicly shared on X. We extracted tweets containing specific terms related to the earthquakes using the twscrape and snscrape Python libraries. Such terms included ‘deprem’,’sarsıntı’, ‘enkaz’, ‘arama’, ‘kurtarma’, ‘yardım’, ‘acil’, ‘ambulans’, ‘iftaiye’, ‘ahbap’, ‘afad’, and ‘bina’. We conducted data collection between February 6, 2023 and February 28, 2023.

The command example with the snscrape library used in collecting tweets in the command line is as follows:

snscrape --jsonl --max-results 200000 twitter-search "deprem OR yardım OR acil OR afad OR ahbap OR bina OR ambulans OR kurtarma OR maraş OR hatay OR adana OR malatya OR antep OR elazığ OR osmaniye since:2023-02-06 until:2023-02-07" > tweets_2023-02-06.json

After finishing the process of collecting data, an overall total of 2.5 million tweets are obtained with their corresponding date-time, text, coordinates (longitude and latitude), and location information.

Some statistics about the data set are as follows:

- The first step involves checking for duplicate data. We found approximately 9% of all collected data to be repetitive. Retweet is one of the features on X that allows users to share someone else’s tweet directly with their followers. Retweets can also be referred to as duplicate data.

- Some tweets were irrelevant with Turkish language. Approximately 8% of the collected data was non-Turkish.

- The first few days immediately after the earthquake are more important in terms of the number of tweets per day. Naturally, 18% of all the tweets were shared in the first three days immediately after the earthquake.

- The tweets had an average total character of 138 and an average tweet word of approximately 20 words.

- On X, not all users post their information. Therefore, location-included tweets make up only 4% of the final dataset. These coordinates are from 97 different locations, mostly in Turkiye.

The data source offers a rich and timely source of data, capturing the real-time reactions and discussions between a broad range of users within the affected areas. To analyze the available data, we applied extensive filtering to confirm the quality and relevance of the data. As a result, we eliminated all spam, irrelevant posts, posts in non-Turkish languages, and all duplicate records. Our analysis methods included topic modeling and sentiment analysis. Our dataset's strengths include the real-time aspect, which allowed it to capture people's reactions to the earthquake crisis in a timely manner. However, due to the nature of active social media users, there are also some weaknesses, such as the possibility of sampling bias and the inability to verify shared information. Nevertheless, these did not diminish the value of enriching the study's results with the resulting dataset.

3.2 Pre-processing

Data preprocessing is an important phase in the data mining and machine learning processes. Data preprocessing includes the tasks of cleaning, converting, and structuring unprocessed data into a format that is appropriate for analysis and modeling. Data preprocessing is intended to improve the quality of the data and prepare it for further

analysis or modeling. This study included carrying out the following steps to preprocess the data:

1. **Removing duplicated data :** The duplicated data, which had the exact same wording in the data set, is eliminated. Retweet is one of the features on X that allows users to share someone else's tweet directly with their followers. Retweets can also be referred to as duplicate data. Furthermore, this process is repeated after each preprocessing step. The objective is to generate a dataset that has unique data inputs.

2. **Extracting contact numbers :** The various formats of Turkish phone numbers in the data set are replaced with the string "contact-number" by using regular expressions. We normalized the data by converting all phone numbers to this standard format, making it easier to process and analyze. This also ensures consistency in the data set. Furthermore, substituting phone numbers with a string enables us to focus on the pertinent details without becoming overwhelmed by details.

3. **Extracting image links :** Some users also post images in their posts on social media platforms. Using regular expressions, image URLs are detected, which are links that terminate with file extensions such as .png, .jpg, .jpeg, .gif, or .bmp. Afterwards, the image links found in the textual content of every tweet within the DataFrame are replaced with the string "imagelink".

4. **Extracting announce emojis :** Users commonly include announcement emojis in their postings to indicate the occurrence of natural disasters, such as earthquakes. The emojis in question are caution sign (⚠️), pushpin (📌), loudspeaker (📢), speaker with two sound waves (🔊), red exclamation mark (❗), police car light (🚨), SOS button (🆘), and collision symbol (💥). These kinf of emojis are replaced with the string "announcemoji".

5. **Removing mentions :** A mention is an abbreviation starting with an "@" and ending with the individual or entity's given username. All mentions in texts are removed.

6. **Removing emojis :** Emojis are the term used to designate small digital icons that express an idea, feeling, concept, or gesture in digital communications. All emojis, except announce ones, are removed.

7. **Removing links :** The links may point to a posted URL. All hyperlinks, with the exception of the image links, are removed.

8. **Removing hashtags :** A hashtag is a word or phrase that is preceded by the hash sign "#" and is used on social media and similar sites to identify digital content about a particular topic. Such symbols enable us to find tweets about a specific topic, so they are more accessible to like-minded users. All of hashtags are removed.

9. **Removing punctuations :** Punctuations are symbols used in written language to mark intermission, prioritization, or other linguistic characteristics. These letters include commas, punctuation marks, question and exclamation digits, quotation symbols, parentheses, and others. All punctuations in the texts are removed.

10. **Removing non-Turkish characters :** All non-Turkish characters are eliminated to ensure there are no unused characters left.

11. **Removing stop words :** Stop words refer to normal words or frequent words that any known fact filter is left out of the analysis of texts because they minimize the meaning of a sentence. These words are removed.

12. **Stripping numbers :** All numbers except contact numbers are eliminated.

13. **Stripping whitespaces :** Whitespaces are any kind of space character within the text of a tweet. They are used to separate words, phrases, or any other type of element within the information of a tweet. All whitespace characters like \n , \r , \t etc. are removed.

14. **Normalization :** Data normalization is the systematic procedure of arranging and managing data in a database to reduce redundancy and reliance, therefore enhancing data integrity and effectiveness. Processing library called as Zemberek is used to normalize Turkish tweets (Akın and Akın, 2019).

15. **Correcting accented Turkish letters :** Several distinct Turkish characters have been replaced with their contemporary equivalents, such as “Â” with “A”, “Î” with “I”, “î”with “I”, “â” with “a”, “û” with “u”, and “Û” with “U”.

16. **Lemmatization :** Data lemmatization is the procedure of transforming words into their fundamental form in order to simplify analysis by considering different versions of a word as equivalent. The words in Turkish tweets have been lemmatized using a pre-built vocabulary provided by Zemberek.

3.3 Developing a Manually Annotated Dataset

Two categories are utilized for labeling the data: rescue and non-rescue labels. Rescue data commonly refers to tweets that contain information pertaining to rescue activities, such as calls for aid, accounts of those trapped or injured, updates on rescue operations, or offers of support. Non-rescue data would include tweets that do not pertain to rescue actions. This includes broad talks about earthquakes, news updates, unrelated content, or personal opinions and tales. In order to generate a labeled data set, an early step involved extracting a small sample from the gathered data. The contents of 4000 data points are read carefully and logically; 2156 of them are labeled by hand, with an equal distribution of half labeled as “rescue” and the other half as “non-rescue” (1078 each). The remaining data out of 2156 are excluded from the analysis.

3.4 Developing an Automatic Annotated Dataset

The 2156 tweets that are manually annotated are subsequently employed alongside the BERT models for annotating all the data in a semi-supervised way. Semi-supervised learning refers to a machine learning approach that involves training a model using a data set containing both labeled and unlabeled data. The objective of semi-supervised learning is to learn a function that can accurately predict the output variable based on the input variables. In this way, it is like supervised learning. However, the algorithm is trained on a data set that contains both labeled and unlabeled data unlike supervised learning. While

learning in a semi-supervised manner, the model leverages both the labeled and unlabeled data at the same time. The labeled data facilitates the learning process and provides the model with clear examples from which to learn. The unlabeled data, in turn, allows the model to draw inferences and recognize better patterns and relationships in the data distribution (GeeksforGeeks, 2023).

First of all, pre-trained BERT models are fine-tuned and trained by using the manually annotated corpus with 2156 labeled tweets. Then, these models are trained on the large Turkish corpus. Such a fine-tuned model labeled each tweet from the collected data from X as rescue or non-rescue. Therefore, in the final step, approximately 1,6 million tweets with predicted labels are obtained, and an automatic annotated corpus is created.

Keeping a close watch on efficiency and accuracy, nine Bert models are gently trained, each containing the kernel of our dataset explicitly built by hand. These models can annotate the core dataset on their own, as they could identify tweets that expressed urgency or encompass rescue-related tweets. Moreover, in the next phase, the annotated dataset converged supervised learning algorithms, classifying the annotated datasets of the various Berts systematically. The chosen annotated dataset belonged to the Bert that classified the dataset the most successfully after thorough examination. Classification steps are explained in more detail in the next step.

As a result of all these processes, 5 data examples showing the final version of our dataset are shown in Figure 3.1.

	date	rawContent	coordinates	place	location	languages	processed1	processed2	processed3	processed4	...	processed6	processed7	processed8	processed9	processed10	processed11	processed12	processed13	text	label
0	2023-02-10 23:27:52+00:00	Deprem bölgesinde Suriyeliler gasp yapıyor, i...	{'_type': 'snscrape.modules.twitter.Coordinate...	{'_type': 'snscrape.modules.twitter.Place', 'i...	Yozgat	tr	Deprem bölgesinde Suriyeliler gasp yapıyor, i...	Deprem bölgesinde Suriyeliler gasp yapıyor, i...	Deprem bölgesinde Suriyeliler gasp yapıyor, i...	deprem bölgesinde suriyeliler gasp yapıyor ...	...	deprem bölgesinde suriyeliler gasp yapıyor ...	deprem bölgesinde suriyeliler gasp yapıyor ...	deprem bölgesinde suriyeliler gasp yapıyor ilg...	deprem bölgesinde suriyeliler gasp yapıyor ilg...	deprem bölgesinde suriyeliler gasp yapıyor ilg...	deprem bölgesinde suriyeliler gasp yapıyor ilg...	deprem bölgesinde suriyeliler gasp yapıyor ilg...	deprem bölgesinde suriyeliler gasp ilişkilerini...	['deprem', 'bölge', 'suriyeli', 'gasp', 'ilş',...	0
1	2023-02-10 23:28:28+00:00	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	{'_type': 'snscrape.modules.twitter.Coordinate...	{'_type': 'snscrape.modules.twitter.Place', 'i...	Mardin	tr	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	iyilik hiçbir zaman boşa gitmiyen bir yatırımd...	...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	İyilik hiçbir zaman boşa gitmiyen bir yatırımd...	iyilik zaman boşa gitmiyen yatırımdır geçmişte...	['iyi', 'zaman', 'boşa', 'git', 'yatırım', 'ge...	0
2	2023-02-10 22:56:40+00:00	announcemoji Hayırsever vatandaşlarımızın ya...	{'_type': 'snscrape.modules.twitter.Coordinate...	{'_type': 'snscrape.modules.twitter.Place', 'i...	NaN	tr	announcemoji Hayırsever vatandaşlarımızın ya...	announcemoji Hayırsever vatandaşlarımızın ya...	announcemoji Hayırsever vatandaşlarımızın ya...	announcemoji hayırsever vatandaşlarımızın yard...	...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	announcemoji hayırsever vatandaşlarımızın yard...	['an', 'hayırsever', 'vatandaş', 'yardım', 'ka...	1
3	2023-02-10 23:31:32+00:00	@ahtap AFAD DA BİZİM AHBAP DA!! \n\nSalih arti...	{'_type': 'snscrape.modules.twitter.Coordinate...	{'_type': 'snscrape.modules.twitter.Place', 'i...	Penarth, Wales	tr	AFAD DA BİZİM AHBAP DA!! \n\nSalih artık su ...	AFAD DA BİZİM AHBAP DA!! \n\nSalih artık su ...	AFAD DA BİZİM AHBAP DA!! \n\nSalih artık su ...	afad da bizim ahtap da ! ! salih artık şu güze...	...	afad da bizim ahtap da salih artık şu güze...	afad da bizim ahtap da salih artık şu güze...	afad da bizim ahtap da salih artık şu güzelim ...	afad da bizim ahtap da salih artık şu güzelim ...	afad da bizim ahtap da salih artık şu güzelim ...	afad da bizim ahtap da salih artık şu güzelim ...	afad bizim ahtap salih artık güzelim insanları ...	afad ahtap salih güzelim insanları sadece yard...	['af', 'ahtap', 'sali', 'güzelim', 'insan', 's...	0
4	2023-02-10 22:26:31+00:00	İçim yanıyor içim ben burada duramıyorum enkaz...	{'_type': 'snscrape.modules.twitter.Coordinate...	{'_type': 'snscrape.modules.twitter.Place', 'i...	Germany	tr	İçim yanıyor içim ben burada duramıyorum enkaz...	İçim yanıyor içim ben burada duramıyorum enkaz...	İçim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim ben burada duramıyorum enkaz...	içim yanıyor içim duramıyorum enkaz altında al...	['içim', 'yan', 'içim', 'dura', 'enkaz', 'altı...	1

Figure 3.1: An Example of the Data Frame Showing Sample from the Dataset

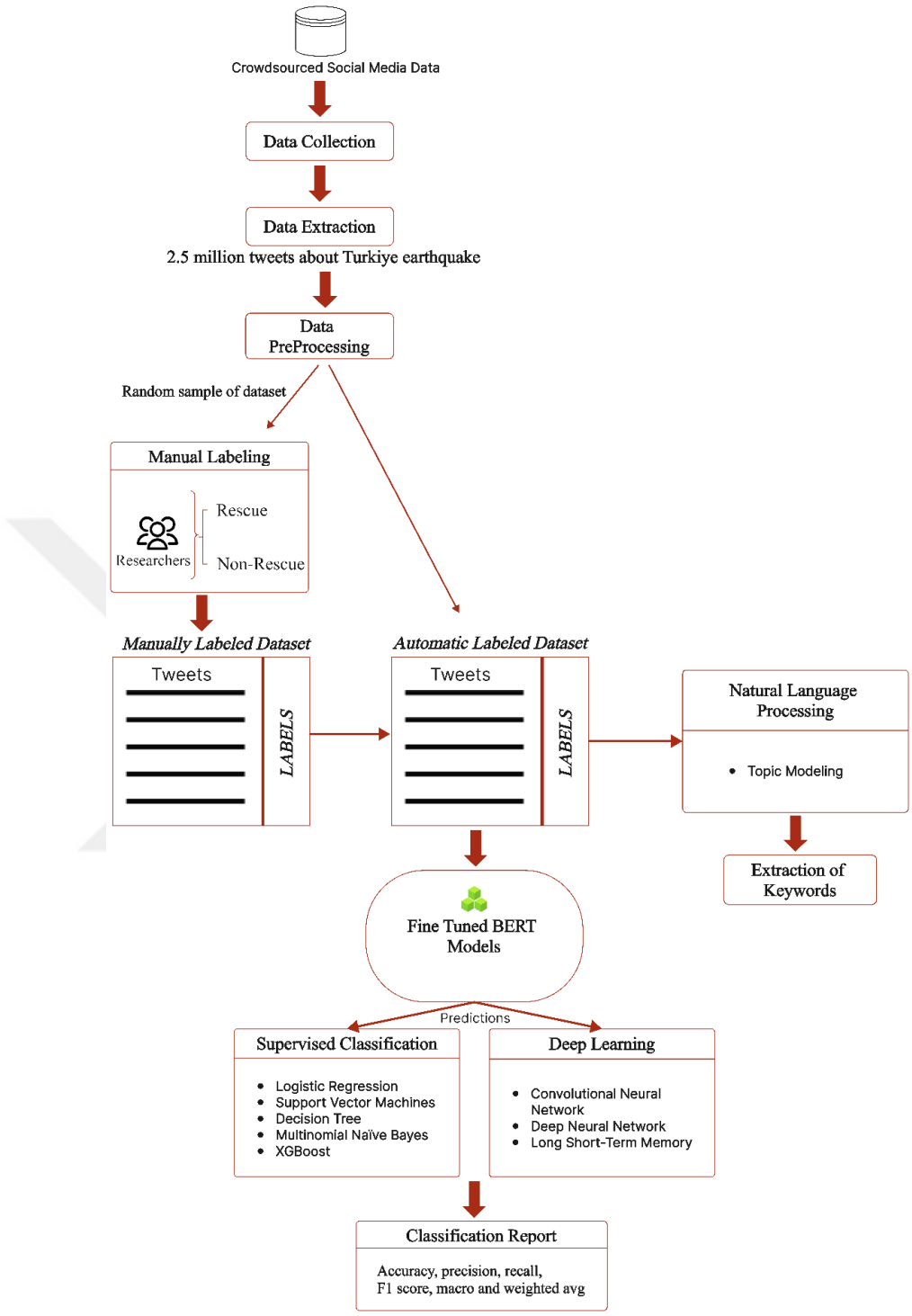

Figure 3.2: Flow Diagram of Manual and Automatic Labeling, Classification of Rescue and Non-rescue Calls, and Natural Language Processing

Figure 3.2 presents the flow diagram of our research study. After the data collection and processing steps, classification algorithms are applied, and performance metrics are compared. Important words are also extracted with NLP.

4. PRESENTED DISASTER CLASSIFICATION

Data classification involves arranging and grouping data according to its various characteristics. The process involves the creation of different classes or groups of data to ensure that information is well managed and identified. It also makes data collection, analysis, and retrieval much simpler and more secure.

In this section, the machine learning models are trained using the tweets in the sizable dataset that we labeled by the semi-supervised method. The steps for text classification are as follows: collecting and preprocessing the text data, feature extraction, and classification algorithm selection. The process is initiated by manually categorizing 2156 texts, with an equal distribution between rescue and non-rescue classifications (1078 tweets each). Following that, BERT transformer models are implemented and trained with this manually labeled dataset to automatically assign labels using a semi-supervised approach. The BERT models are trained on manually labeled data and evaluated to obtain accuracy, precision, recall, and F1-score scores. The *ClassificationModel* from the *Simple Transformers* library in Python (Simple Transformers, 2020) is used to implement the BERT models. Classification is carried out using supervised learning algorithms including Logistic Regression, Support Vector Machines, Decision Trees, Multinomial Naïve Bayes, Stochastic Gradient Descent, XGBoost. The *CountVectorizer* class from scikit-learn in Python is used to transform the text data into a sparse matrix. Labels are converted to binary values, and these values are put into the array list. While the models are trained on 75% of the dataset, their performance is assessed using the remaining 25%. The most effective labeling model is then selected among this group of models. We have used a variety of methods, from classical ML approaches to more modern transformer methods.

Transformer-based language models from state-of-the-art language model libraries are used to maximize the scores obtained. These are as follows:

- convbert-base-turkish-cased,
- convbert-base-turkish-mc4-cased,
- electra-base-turkish-cased-discriminator,
- electra-base-turkish-mc4-cased-discriminator,
- bert-base-turkish-cased,
- bert-base-turkish-128k-cased,
- distilbert-base-turkish-cased,
- bert-base-cased,
- bert-base-multilingual-cased.

The best model for labeling is selected after a careful evaluation of performance metrics, computational efficiency, and robustness. The criteria are based on the highest accuracy, efficiency, and management of data distribution differences.

4.1 Machine Learning Methods

Machine learning algorithms used to classify data are described in this section.

4.1.1 BERT Models

BERT : Bidirectional Encoder Representations from Transformers are a pre-trained natural language processing model developed by Google. It belongs to the group of

transformer-based models, meaning that to process a written word's meaning in a specific context, it considers both the direction from left to right and from right to left. BERT is trained on a huge amount of text data using unsupervised learning algorithms, allowing the model to develop both detailed word units and the semantics of entire sentences (Devlin et al., 2018). BERT is a generalized language model with 24 Transformer blocks, 1024 hidden layers, and 349M parameters. As pre-trained with large text data, it becomes transferable to other small language processing activities than small languages as required. For this reason, it has resulted in state-of-the-art achievement for natural language processing activities such as Multi-Genre Natural Language Inference, Quora Question Pairs, Question Natural Language Inference, Stanford Sentiment Treebank, Microsoft Research Paraphrase Corpus, Recognizing Textual Entailment, Winograd Schema Challenge, SQuAD v1.1, and SQuAD v2.0.

BERT has been one of the leading models in a plethora of natural language tasks, including text classification, question answering, named entity recognition, and many others. From all nine different BERT models used for data tagging, bert-base-turkish-cased, bert-base-turkish-128k-cased, bert-base-cased, and bert-base-multilingual-cased are the applied models whose names start with BERT.

The name "base" states that this model is a variation stemming from the original BERT architecture, which is classified according to its layer and parameter numbers. The "cased" feature means that the model perceives the difference between uppercase and lowercase. When the name contains "128k", it suggests that this model is based on a substantially larger corpus of data compared to the amount of typical data used in the creation of BERT models. According to the title bert-base-multilingual-cased, the word "multilingual" means that the target is developed and optimized in response to various languages.

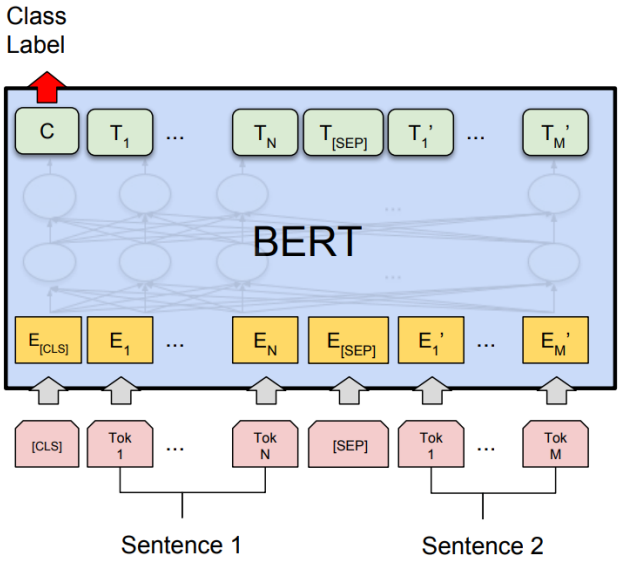

Figure 4.1: The BERT Model Architecture in Sentence Pair Classification (Shenfield et al., 2020)

The BERT model architecture in classification tasks is illustrated in Figure 4.1. One of the key features distinguishing BERT models from all the previous NLP models is that they are built on the transformer architecture from the depths of the model. This robustly bidirectional and unsupervised process allows BERT to take into account both the context to the left and right of each word. BERT's method enables the creation of state-of-the-art models for a wide variety of tasks, including language inference and question answering and, with only one additional layer of output without requiring remarkable task-specific architectural changes.

ConvBERT : ConvBERT is the state-of-the-art pre-trained model that integrates a novel propagation-based dynamic convolution module in place of the common self-attention mechanisms, providing a way to capture local dependencies inherently adapted in BERT architectures (Jiang et al., 2020). In this study, the convbert-base-turkish-cased and convbert-base-turkish-mc4-cased models are deployed. Convbert-base-turkish-cased model maximizes BERT with dynamic convolution at the span level and is pre-trained on the same dataset as bert-base-turkish-cased. Convbert-base-turkish-mc4-cased is pre-trained on a different data set using (mc4 corpus) compared to convbert-base-turkish-cased. The name of the model “convbert-base-turkish-mc4-cased” means that the

ConvBERT model is trained with the MC4 corpus of text data from many languages, including Turkish.

ELECTRA : This language model is created using a new pre-training approach called self-monitoring, consisting of concurrent training of two transformer models, the generator and discriminator. The input tokens are then replaced with tokens generated using a generator model. The primary role of the discriminator model is to decide if tokens in the input sequence are real or have undergone modifications by the generator. Some types of the ELECTRA model trained on Turkish texts for different tasks, including electra-base-turkish-cased-discriminator, electra-base-turkish-mc4-cased-discriminator are applied in this study. Electra-base-turkish-cased is an ELECTRA model variant specific for the Turkish language that differs from the BERT in the task itself. In the basic BERT model, training is asked to predict only masked 15% of the tokens, while in the ELECTRA model, every token is used to classify whether the token remained the same or was replaced. In addition, Electra-base-turkish-mc4-cased is trained on a bigger Turkish MC4 data set.

DistilBERT : DistilBERT is essentially a more compressed, shorter, lighter, and cheaper version of the BERT device. In contrast to the original BERT model, the motivation of this model is to enhance computing speed and the ability to use it when sources are limited. The prior section in comparison to the BERT framework is that this model focuses on the distilbert-base-turkish-cased model, which is a special form of the DistilBERT architecture designed for processing Turkish texts. It provides computational efficiency while still managing case sensitivity in tasks involving the Turkish language. Distilbert-base-turkish-cased is the distilled version of the bert-base-turkish-cased model; the distilled model of general-purpose performance can be followed with high performance using several downstream tasks. For instance, it is observed as possible to run it on mobile devices.

4.1.2 Supervised Learning Algorithms

In this research, some of the supervised learning algorithms are implemented. Supervised learning is a type of machine learning in which the algorithm is developed using data that is "pre-tagged," in other words, each input data point has its own output label. The algorithm is trained with a dataset consisting of input-output pairs, learning a mapping function from the input to the output. In supervised learning, the algorithm modifies its parameters several times to reduce the gap between its predictions and the real labels in the training part. Classification and regression are two terms frequently used in supervised learning. In classification tasks, one wants to predict a categorical label or class for each input data, while in regression tasks, the algorithm tries to predict a continuous numerical value.

4.1.2.1 Logistic Regression

As its name suggests, logistic regression is a statistical method that is widely used to solve classification problems. However, despite the name, logistic regression is not truly a regression model; in fact, it is instead a classifier. The linear combination of inputs is computed with weights; the calculations are then run through a logistic function, also called a sigmoid layer as given by Equation 4.1.

$$S(x) = \frac{1}{1+ e^{-x}} = \frac{e^{x}}{1+ e^{-x}} = 1 - \sigma(-x) \tag{4.1}$$

The goal of the sigmoid function is to map the output to a value between 0 and 1, which indicates the probability that the input belongs to a given category.

4.1.2.2 Support Vector Machines

SVM is a machine learning algorithm that helps us to solve complex classification, regress ion, and outlier's identification challenges using supervised learning models. SVM spatially dictated optimal data transformation decides the margins between data points of various predefined classes and labels or outputs. SVM algorithms are used in various applications including healthcare, natural language processing, signal processing, and speech & image recognition fields. The primary aim of a SVM algorithm is to find a hyperplane separate clearly differentiates the data points within the classes. The proper place and margin so that the largest margin among the classes can be considered. The support vector representation is plotted in Figure 4.2.

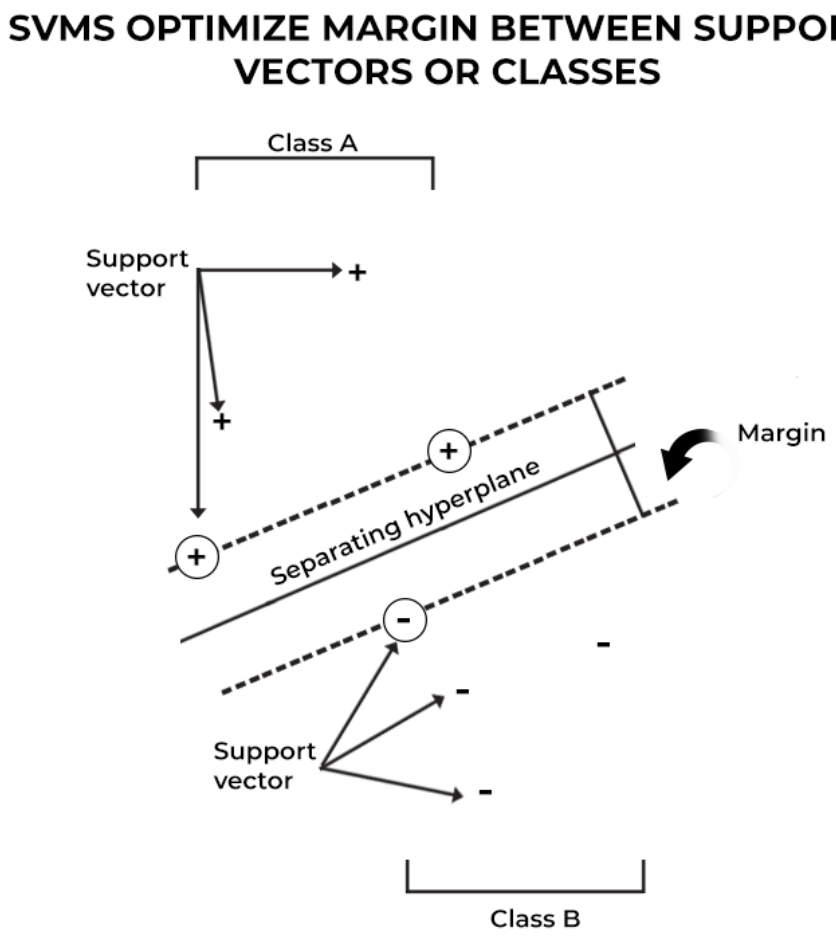

Figure 4.2: Support Vector Representation (Kanade et al., 2022)

4.1.2.3 Decision Tree

A decision tree is a non-parametric supervised learning algorithm for both classification and regression tasks. This is commonly used for the identification of the target variable value by learning the simple rules based on its statistics and features. The tree can be described as an approximately piecewise constant function with a hierarchical, tree-like structure: the root node, branches, internal nodes, and leaf nodes. These are presented in the diagram in Figure 4.3. Decision tree learning uses a divide-and-conquer approach, using a greedy search to find the optimal split points within a tree. This splitting is then recursively performed top-down on the training data until all, or as many records as possible, can be classified under certain class labels. Whether all data points end up being classified into homogeneous sets depends largely on the size of the decision tree.

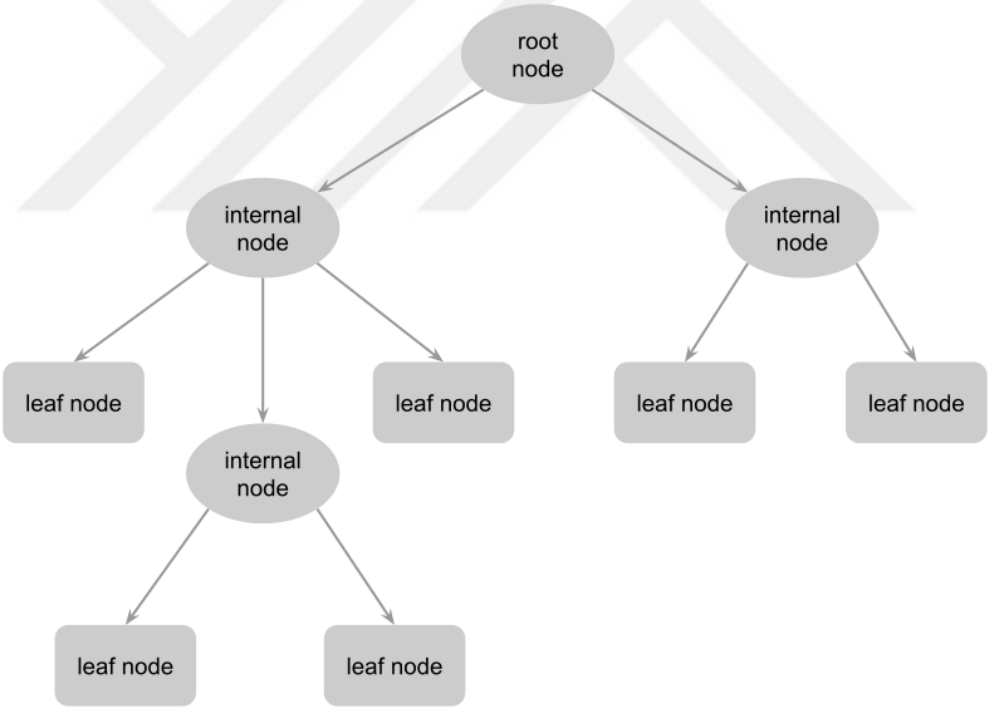

Figure 4.3: Diagram of a Decision Tree Structure (Sanchez et al., 2020)

4.1.2.4 Multinomial Naïve Bayes

Naïve Bayes is a family of probabilistic algorithm based on Bayes Theorem that is "Naive" due to its usual independence sub-predicate, loosely labeled as "no feature would make a difference than another when a class label is known." Bayes Theorem was formulated by Thomas Bayes and used to calculate the probability of an event given the prior knowledge of conditions that might be related to the event. Thus, this theorem can be formulated as Equation 4.2.

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{4.2}$$

- *P(A)* is the prior probability of class A,
- *P(B)* is the prior probability of B,
- *P(B|A)* is the probability of occurrence of B given class A
- *P(A|B)* is the probability of class A when predictor B occurs.

Multinomial Naïve Bayes is a popular and one of the most efficient machine learning algorithms that is fundamentally based on Bayes Theorem. Multinomial Naïve Bayes is a probabilistic classifier using probability distributions to determine the classification of text data, making it suitable for data that characterize discrete event frequencies or count from several tasks in natural language processing. It is well-used in text classification when the data at hand is in the mode of discrete possibilities, such as word appearances within the body of documents.

4.1.2.5 Stochastic Gradient Descent

SGD is utilized for optimizing machine learning models. It is a variant of the Gradient Descent algorithm. SGD addresses the computational inefficiency of traditional Gradient Descent methods when dealing with large datasets in machine learning projects. Instead of using the entire dataset for each iteration, only a single random training example is selected to update the model parameters and calculate the gradient in SGD. The advantage of using SGD is computational efficiency, especially when dealing with large datasets. When using a single sample or a small batch, the computational cost per iteration is significantly reduced compared to traditional Gradient Descent methods that require processing the entire dataset.

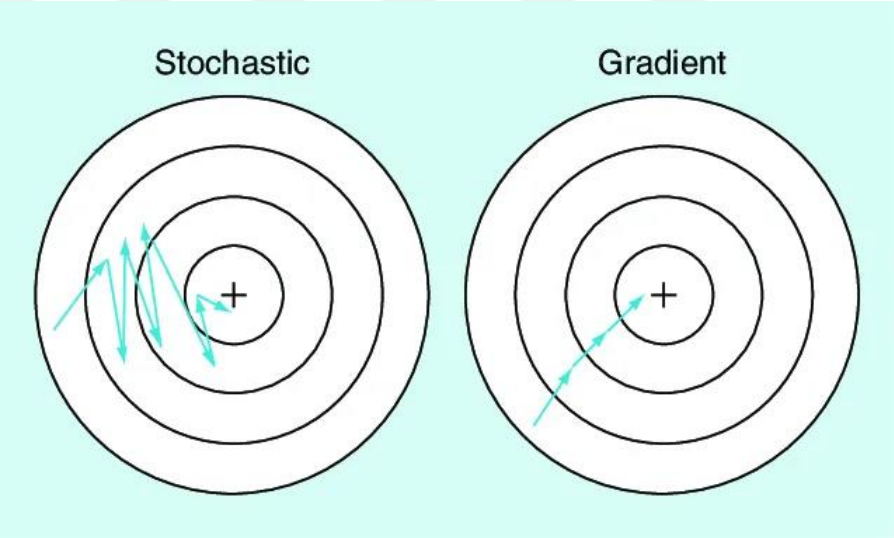

Figure 4.4: Gradient Descent Compared with Stochastic Gradient Descent (Carpenter et al., 2018)

SGD performs an operation on one randomly received data at each step. The points aim to reach the optimum point by constantly changing. It is faster than Batch Gradient Descent. The comparison of Gradient Descent and Stochastic Gradient Descent algorithm is illustrated in Figure 4.4.

4.1.2.6 Extreme Gradient Boosting

XGBoost is one of the best and most popular machine learning programs used for most supervised learning functions, such as regression, classification, and ranking. Notably, it utilizes the gradient-boosting architecture, and it is very popular because of its great accuracy and scalability. XGBoost is a data structure designed to work on large-scale datasets that integrate perfectly with other Python machine learning libraries while also offering various applications. Classification is one of its most frequent applications. It predicts a separate class label for an input characteristic. The specially designed XGBClassifier module is applied to handle classification work. The classification formulation for XGBoost is provided in Equation 4.3.

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{4.3}$$

where:

- $\hat{y}_i$ is the predicted probability that the sample x_i belongs to the positive class or the predicted class label.
- $\phi(x_i)$ is the score function, which predicts the likelihood of the sample's class membership.
- K is the sum of all trees that are present in an ensemble.
- $f_k(x_i)$ stands for the probability of the k - th tree.

4.2 Implementation of Classification Algorithms

In the development of BERT models, similar steps are applied to nine models. We performed text classification process using the pre-trained BERT models and supervised

learning classifiers, and then evaluated the performance of these classifiers utilizing cross-validation. A summary of our improvements is given below:

1. Training the BERT Models:

- Necessary libraries including Pandas for data handling, SimpleTransformers for BERT model training, and scikit-learn for evaluations are imported.

- Training arguments are set for the BERT models, such as the number of training epochs as 5.

- Then, BERT model for text classification using the ClassificationModel class from SimpleTransformers library is initialized.

- Next, the BERT model is trained on manually labeled data set which contains texts and corresponding labels.

2. Batch Processing:

- A function named batch_generator is defined to generate batches of data for processing.

- Using the trained BERT model, the function is iterated over all data set in batches, and the predictions are created. The DataFrame is then updated with the predictions. Thus, automatic labeling is achieved in a semi-supervised way.

3. Vectorization:

- CountVectorizer from scikit-learn is used to convert text data into a numerical representation.

- The number of features (words) is limited to 3000 and transformed the text data into a sparse matrix of word counts.

- The labels data is converted to a binary array.

4. Model Evaluation:

- Supervised classifications are implemented.

- K-fold cross-validation is performed using the classification models and it is evaluated using accuracy, precision, recall, and F1-score metrics.

- Classification reports are generated including precision, recall, and F1-score for each class.

4.2.1 K-fold Cross-Validation

K-fold cross-validation is a technique used for evaluating the effectiveness of machine and deep learning models. This technique partitions the original dataset into K folds, with each fold being about the same size. In K iterations, the model is bench-marked, and in each iteration, K-1 folds are used for training and the remaining fold for validation. In our case, five iterations are performed, and each fold is used as the validation data exactly once. The measures of model performance, including accuracy or loss, are combined over K iterations to obtain a definitive and reliable evaluation of its efficacy (Lemos, 2022). This method is used in the execution of machine learning and deep learning algorithms in this thesis.

5. DEEP LEARNING

Three deep learning algorithms used in this study including Convolutional Neural Network (CNN), Deep Neural Network (DNN), and Long Short-Term Memory (LSTM).

5.1 Deep Learning Algorithms

In this section, the general definition and features of the deep learning algorithms used throughout the study are given.

5.1.1 Convolutional Neural Network

CNN is commonly used as one of the advanced deep learning algorithms in image processing. It is developed to handle grid-like data, such as photos. CNN is a technique that uses different procedures to collect information from photos and then categorizes them. This framework is composed of convolutional layers, pooling layers, and robustly connected layers. CNNs have been used in anomaly detection more broadly since the late 2000s. It is appropriate for managing high-dimensional data, since it can identify features with a robust spatial relationship, and fewer variables are necessary compared to other approaches. Nonetheless, it is not possible to encode the specific object position, and working with long data strings to process is beneficial. General structures of the CNN architecture are displayed in Figure 5.1.

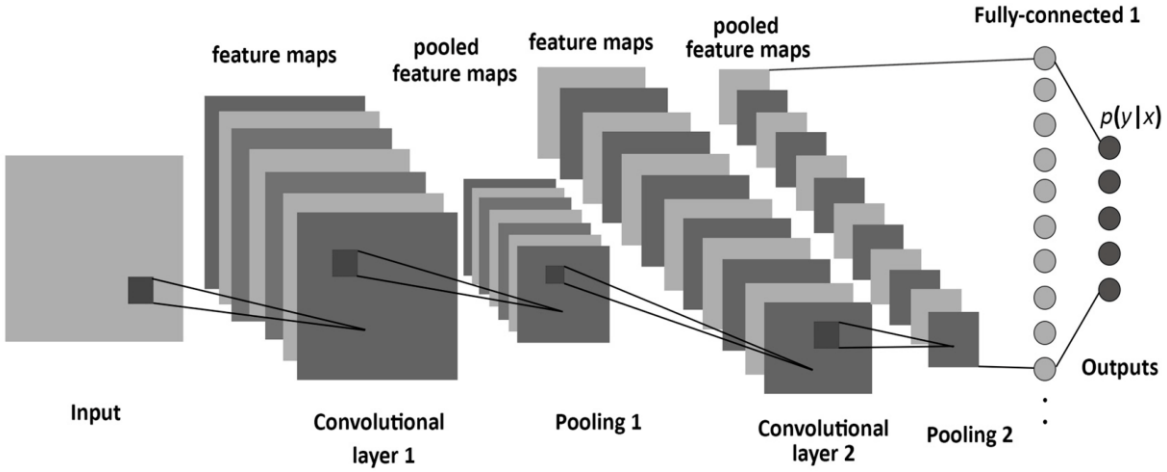

Figure 5.1: The structure of a CNN, consisting of convolutional, pooling, and fully-connected layers (Albelwi et al., 2017)

5.1.2 Deep Neural Network

Deep Neural Networks are a type of artificial neural network where in-depth refers to several layers residing between the input and output layers. Several layers of linked neurons, where each link has a certain weight, compose this neural network. Each neuron with weighted inputs operates a non-linear activation function on the inputs. Meeting the requirements of complex data signals, DNNs can process different types of data signals, including but not limited to time-series data, structured data, and sequential data. This increased ability of DNNs to process complex data signals has provided the framework for their application beyond only image processing. Machine learning has been effectively established using these techniques, including dissimilar types like natural language processing, picture and audio recognition, recommendation engines, and many others. The general structure of a DNN architecture is presented in Figure 5.2.

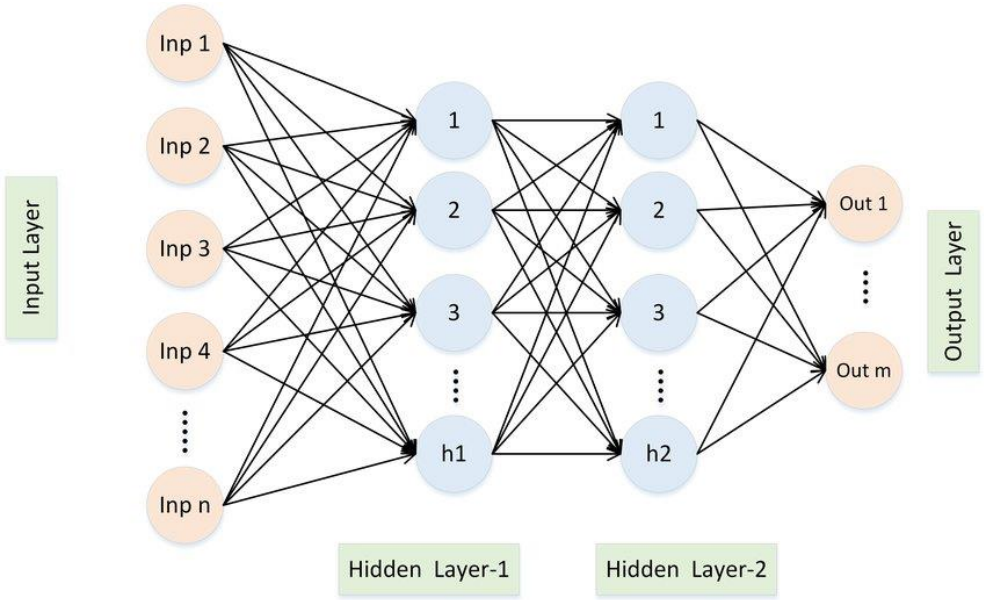

Figure 5.2: General Architecture for a Deep Neural Network with Two Hidden Layers (Shehzad et al., 2021)

5.1.3 LSTM

LSTM is an architectural variant of RNNs that has been tremendously successful in learning long-term dependencies. It is suitable for processing sequential data. In recent years, it has grown rapidly in the last few years, specifically in the field of anomaly detection from sequential data. As LSTMs are a variant of RNN architecture, they incorporate specialized memory cells equipped with gating mechanisms. Diverse sequential data tasks, including speech recognition, natural language processing, handwriting recognition, and time series prediction, extensively employ LSTM networks. Moreover, LSTM, unlike traditional neural networks, has feedback connections, allowing it to process entire sequences of data, not just individual data points. As a result, they can be very good at understanding time series, text, speech, or any other sequence pattern. Figure 5.3 shows the general architecture of LSTM.

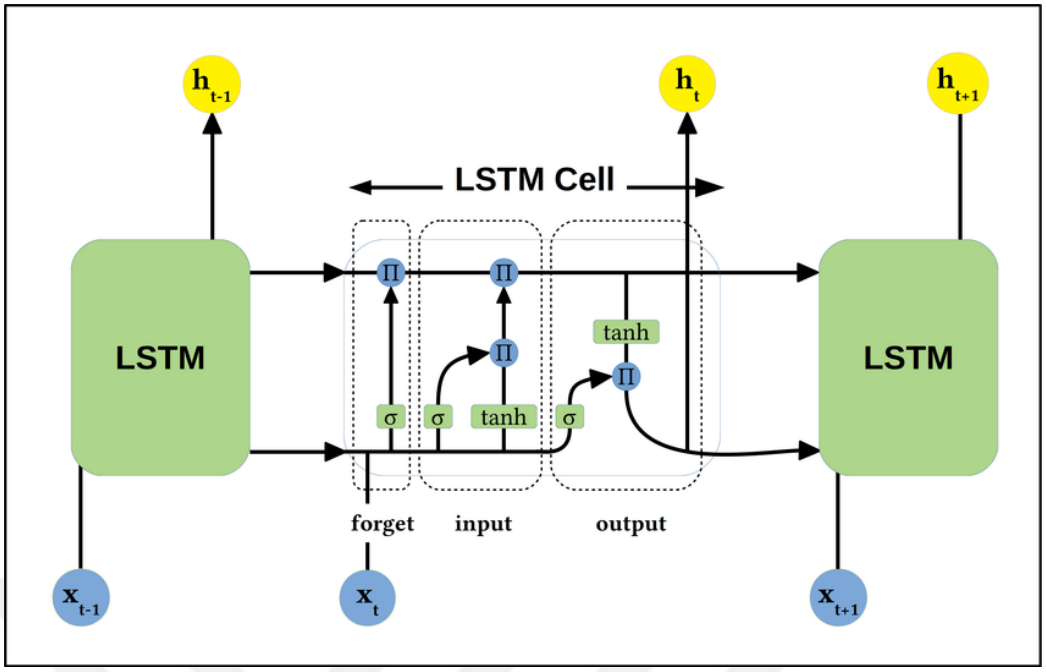

Figure 5.3: General Architecture of LSTM (Shenfield et al., 2020)

5.2 Implementations of Deep Learning Algorithms

The same procedures have been utilized to generate and train these models. Nevertheless, it is obvious that the model structure varies for each model. Figure 5.4 illustrates the flowchart of the implementations.

Figure 5.4: The flowchart of the Deep Learning Implementations

1. CNN Implementation : To begin, the data set is splitted into train and test sets and the data is processed by tokenizing and padding text sequences. Next, the CNN architecture is built, which includes one embedding layer to turn text into dense vectors, two pairs of convolutional and max-pooling layers to pull out features, one flattening layer to turn the

output into a 1D array, and two dense layers for classification. The model is compiled with appropriate loss and optimization functions and trained using K-fold cross-validation. Finally, the model's performance is evaluated on the test set using metrics such as accuracy, precision, recall, and F1-score.

2. DNN Implementation : Firstly, data set is splitted into training-validation and test sets. The text sequences are then tokenized and padded to prepare them for model input. One embedding layer, one flattening layer, four dense layers with ReLU activation functions, and dropout layers for regularization make up our DNN architecture. The model is compiled using the Adam optimizer and binary cross-entropy loss function. K-fold cross-validation training is used to evaluate performance on different subsets of the training data, iterating over epochs. Finally, the trained model is evaluated on the test set and the metrics such as accuracy, precision, recall, and F1-score are calculated to assess its performance.

3. LSTM Implementation : Data set is splitted into training-validation and test sets. Then, the text sequences are tokenized and padded to ensure uniform length inputs for the model. Our LSTM model architecture consists of one embedding layer to transform text into dense vectors, followed by one LSTM layer with 64 units to capture sequential dependencies in the data. One dense layer is also added with a sigmoid activation function for binary classification. The model is built using the Adam optimizer and the binary cross-entropy loss function. The K-fold cross-validation is used to test how well the model worked on different groups of training data, doing this over epochs. Finally, the model is evaluated on the test set and the metrics such as accuracy, precision, recall, and F1-score are computed.

6. NATURAL LANGUAGE PROCESSING

NLP is a field of artificial intelligence that centers on teaching computers to accurately understand and interpret human language and gather the acquired information in a manner that is both valuable and actionable. NLP is a discipline that involves developing algorithms and methods to allow computers to process and interpret substantial volumes of natural language data, including statistics, content, and voice recording.

NLP applications have many processes, including text interpretation, named entity recognition, sentiment analysis, and language generation. In other words, natural language processing strives to bring human interactions and computer comprehension together. NLP systems can develop chatbots, have language translation capabilities, and provide text summarization.

6.1 Extracting Keywords

In this section, how the important words in the data set are extracted and what methods are applied when finding these words are explained.

6.1.1 Topic Modeling

In this study, an unsupervised natural language processing technique through a machine learning algorithm, topic modeling, is used to tag and group text clusters that have the same topic.

In NLP, topic modeling is a technique that automatically identifies the topics or themes within a collection of documents. Topic modeling is particularly suited for analyzing large corpora of text data to extract hidden patterns and structures. Specifically, it is statistical modeling that uses unsupervised machine learning to examine and sort sets of related words in a text.

Topics are views of a corpus of text. Naturally, documents on the same topic contain some words more frequently. For instance, “dog” and “bone” are more frequently appearing words in a document about dogs than in one about cats. The same way, “cat” and “meow” are more likely in a document about cats. As a result, topic modeling's task is to read the documents and receive clusters of similar words. In other words, topic models include guessing at words and grouping them into topics to produce a topic cluster. An example of how topic modeling works is presented in Figure 6.1.

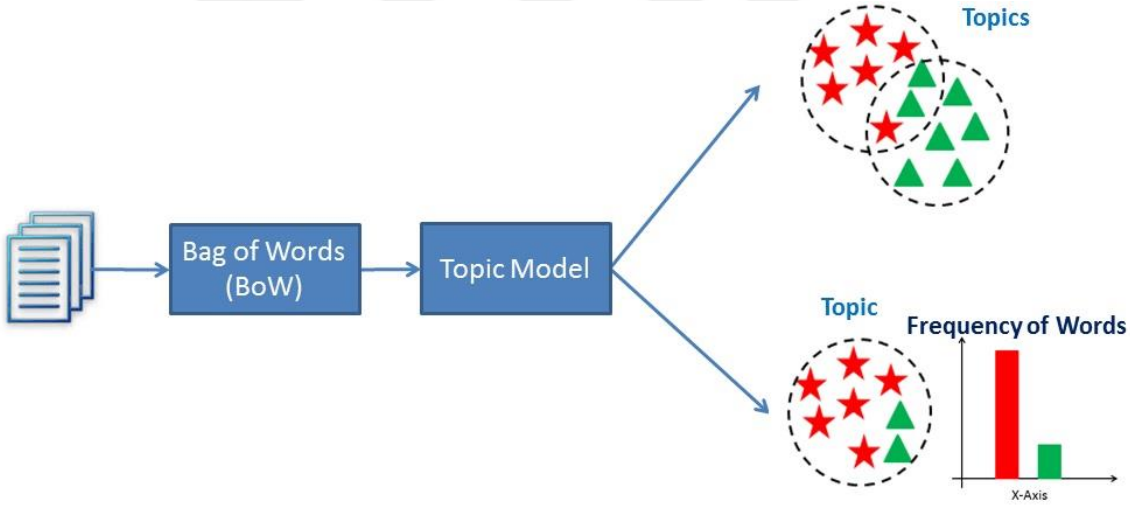

Figure 6.1: Visualization of How Topic Modeling Works (Pykes et al., 2023)

Topic modeling is an essential tool in multiple domains and helps facilitate the meaningful interpretation of extensive volumes of textual data. It helps to reveal the central themes inherent to large volumes of text data and allows researchers, businesses, and organizations to explore various sources quickly. It is possible due to the fact that topic modeling reveals the most relevant and applicable topics, which might be unknown to a reader at the time of analysis. Finally, topic modeling serves as a tool to employ in

the process of topic categorization, trend generation, emotion analysis, and information recommendation.

6.1.2 FuzzyTM

(Rijcken et al., 2020) presented FLSA-W, an innovative modeling technique that greatly outperforms existing models in terms of coherence, interpretability, and diversity. They also released the FuzzyTM module, which includes FLSA-W but also FLSA and FLSA-V, which are modeled on fuzzy logic. The initial FLSA approach attempts to locate clusters in the projected document space. The original FLSA approach tries to find clusters in the projected space of documents. However, they are difficult to cluster because documents might contain multiple topics, often being very sparse. Then, it makes more sense to cluster words than documents. This is what we do with FLSA-W. FLSA-W clusters on a space of words projected from word frequencies, implicitly assuming the projections make sure related words are eventually near each other. The optimization functions determine how we achieve this. With FLSA-V(os), we acquire the output of Vosviewer as input to our model. Vosviewer is an open-source software tool for bibliometric mapping, and the projections have the explicit guarantee that analogous words are put near each other (Rijcken et al., 2022).

In this study, the FLSA_W library is utilized from the FuzzyTM package to identify significant keywords in tweets that convey a sense of urgency. Some of the data obtained via X, formerly known as Twitter, lacks location metadata. Thus, in this section, the data is exclusively analyzed that have known coordinates. We employed temporal and spatial parameters in data analysis to identify the words. Initially, a clustering technique is used to consolidate tweets originating from close coordinates. Key terms are identified from each cluster of tweets, retrieving ten significant terms for each cluster. Furthermore, these clusters are also grouped based on spatial coordinates and time intervals of 30 minutes, 1 hour, and 2 hours. Once again, ten keywords are extracted for the clusters split according to time and coordinates.

7. EXPERIMENTAL STUDY

Originally, we implemented a series of stringent data preprocessing techniques to enhance the quality, processability, and comprehensibility of the dataset. Executing the data preprocessing stages carefully and precisely improved the performance of the models, leading to more dependable outcomes.

The focus of the classification is on recognizing urgent pleas for aid, reports related to debris, and imminent assistance needs. The complexity of textual input necessitated a comprehensive analysis of machine learning and deep learning architectures for classifying. We consider BERT models that capture intricate semantic relationships in textual data and supervised learning classification. Training of BERT models with fine-tuning in a semi-supervised approach on a manually annotated dataset illustrates performance progress that can be refined with training modification.

Nine BERT models are trained using a dataset that is tagged manually in order to automatically classify all texts. Semi-supervised learning method is used during the training process. We were able to decrease our reliance on labor-intensive manual labeling and enhance productivity through the implementation of semi-supervised learning, resulting in significant time and cost savings. We chose the dataset that the BERT model labeled and had the best performance metrics for further research. To accurately assess the efficacy of BERT models, we performed five training iterations.

Table 7.1 throughout Table 7.12 presents the results of the precision, recall, F1 score, and accuracy of all BERT models for rescue and non-rescue classes.

Table 7.1: Performance Metrics of Logistic Regression in Models for Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9170	0.8767	0.8964	343076	0.9555
convbert-base-turkish-mc4-cased	0.9136	0.8692	0.8909	371314	0.9494
electra-base-turkish-cased-discriminator	0.9020	0.8381	0.8689	352320	0.9430
electra-base-turkish-mc4-cased-discriminator	0.9167	0.8689	0.8922	372868	0.9499
bert-base-turkish-cased	0.9081	0.8536	0.8800	346589	0.9484
bert-base-turkish-128k-cased	0.9115	0.8670	0.8887	362559	0.9496
distilbert-base-turkish-cased	0.9201	0.8814	0.9003	313947	0.9608
bert-base-cased	0.9122	0.8498	0.8799	371036	0.9449
bert-base-multilingual-cased	0.9276	0.8912	0.9091	362718	0.9586

Table 7.2: Performance Metrics of Logistic Regression in Models for Non-rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9657	0.9777	0.9717	1220051	0.9555
convbert-base-turkish-mc4-cased	0.9599	0.9744	0.9671	1191813	0.9494
electra-base-turkish-cased-discriminator	0.9539	0.9735	0.9636	1210807	0.9430
electra-base-turkish-mc4-cased-discriminator	0.9596	0.9753	0.9674	1190259	0.9499
bert-base-turkish-cased	0.9590	0.9754	0.9671	1216538	0.9484
bert-base-turkish-128k-cased	0.9604	0.9746	0.9674	1200568	0.9496
distilbert-base-turkish-cased	0.9705	0.9808	0.9756	1249180	0.9608
bert-base-cased	0.9542	0.9745	0.9643	1192091	0.9449
bert-base-multilingual-cased	0.9675	0.9790	0.9732	1200409	0.9586

Table 7.3: Performance Metrics of SVC in Models for Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9174	0.8744	0.8954	343076	0.9551
convbert-base-turkish-mc4-cased	0.9148	0.8650	0.8892	371314	0.9488
electra-base-turkish-cased-discriminator	0.9045	0.8315	0.8665	352320	0.9422
electra-base-turkish-mc4-cased-discriminator	0.9184	0.8648	0.8908	372868	0.9494
bert-base-turkish-cased	0.9101	0.8489	0.8784	346589	0.9479
bert-base-turkish-128k-cased	0.9127	0.8635	0.8874	362559	0.9492
distilbert-base-turkish-cased	0.9210	0.8799	0.9000	313947	0.9607
bert-base-cased	0.9144	0.8439	0.8777	371036	0.9442
bert-base-multilingual-cased	0.9282	0.8898	0.9086	362718	0.9585

Table 7.4: Performance Metrics of SVC in Models for Non-Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9651	0.9779	0.9715	1220051	0.9551
convbert-base-turkish-mc4-cased	0.9587	0.9749	0.9667	1191813	0.9488
electra-base-turkish-cased-discriminator	0.9521	0.9745	0.9631	1210807	0.9422
electra-base-turkish-mc4-cased-discriminator	0.9584	0.9759	0.9671	1190259	0.9494
bert-base-turkish-cased	0.9578	0.9761	0.9668	1216538	0.9479
bert-base-turkish-128k-cased	0.9594	0.9751	0.9672	1200568	0.9492
distilbert-base-turkish-cased	0.9702	0.9810	0.9756	1249180	0.9607
bert-base-cased	0.9526	0.9754	0.9638	1192091	0.9442
bert-base-multilingual-cased	0.9671	0.9792	0.9731	1200409	0.9585

Table 7.5: Performance Metrics of Decision Tree in Models for Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.8160	0.6142	0.7009	343076	0.8849
convbert-base-turkish-mc4-cased	0.8175	0.6469	0.7223	371314	0.8818
electra-base-turkish-cased-discriminator	0.8251	0.6201	0.7081	352320	0.8848
electra-base-turkish-mc4-cased-discriminator	0.8245	0.6666	0.7372	372868	0.8866
bert-base-turkish-cased	0.8284	0.6343	0.7185	346589	0.8898
bert-base-turkish-128k-cased	0.8288	0.6265	0.7136	362559	0.8834
distilbert-base-turkish-cased	0.8194	0.6974	0.7535	313947	0.9084
bert-base-cased	0.8700	0.6720	0.7583	371036	0.8983
bert-base-multilingual-cased	0.8560	0.6208	0.7196	362718	0.8878

Table 7.6: Performance Metrics of Decision Tree in Models for Non-Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.8986	0.9610	0.9288	1220051	0.8849
convbert-base-turkish-mc4-cased	0.8967	0.9550	0.9249	1191813	0.8818
electra-base-turkish-cased-discriminator	0.8969	0.9618	0.9282	1210807	0.8848
electra-base-turkish-mc4-cased-discriminator	0.9015	0.9556	0.9277	1190259	0.8866
bert-base-turkish-cased	0.9023	0.9626	0.9315	1216538	0.8898
bert-base-turkish-128k-cased	0.8950	0.9609	0.9268	1200568	0.8834
distilbert-base-turkish-cased	0.9267	0.9614	0.9437	1249180	0.9084
bert-base-cased	0.9047	0.9688	0.9356	1192091	0.8983
bert-base-multilingual-cased	0.8942	0.9684	0.9298	1200409	0.8878

Table 7.7: Performance Metrics of MNB in Models for Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.8338	0.8268	0.8303	343076	0.9258
convbert-base-turkish-mc4-cased	0.8549	0.8260	0.8260	371314	0.9254
electra-base-turkish-cased-discriminator	0.8306	0.8096	0.8200	352320	0.9199
electra-base-turkish-mc4-cased-discriminator	0.8503	0.8190	0.8343	372868	0.9224
bert-base-turkish-cased	0.8288	0.8285	0.8286	346589	0.9240
bert-base-turkish-128k-cased	0.8386	0.8272	0.8329	362559	0.9230
distilbert-base-turkish-cased	0.8244	0.8303	0.8274	313947	0.9304
bert-base-cased	0.8001	0.7694	0.7845	371036	0.8996
bert-base-multilingual-cased	0.8228	0.8064	0.8145	362718	0.9148

Table 7.8: Performance Metrics of MNB in Models for Non-Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9514	0.9537	0.9525	1220051	0.9258
convbert-base-turkish-mc4-cased	0.9463	0.9563	0.9513	1191813	0.9254
electra-base-turkish-cased-discriminator	0.9450	0.9520	0.9485	1210807	0.9199
electra-base-turkish-mc4-cased-discriminator	0.9439	0.9548	0.9494	1190259	0.9224
bert-base-turkish-cased	0.9511	0.9512	0.9512	1216538	0.9240
bert-base-turkish-128k-cased	0.9480	0.9519	0.9500	1200568	0.9230
distilbert-base-turkish-cased	0.9573	0.9556	0.9564	1249180	0.9304
bert-base-cased	0.9291	0.9402	0.9346	1192091	0.8996
bert-base-multilingual-cased	0.9419	0.9475	0.9447	1200409	0.9148

Table 7.9: Performance Metrics of XGBoost in Models for Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9027	0.7564	0.8231	343076	0.9286
convbert-base-turkish-mc4-cased	0.9075	0.7632	0.8291	371314	0.9253
electra-base-turkish-cased-discriminator	0.9046	0.7541	0.8225	352320	0.9267
electra-base-turkish-mc4-cased-discriminator	0.9116	0.7705	0.8352	372868	0.9274
bert-base-turkish-cased	0.9076	0.7594	0.8270	346589	0.9295
bert-base-turkish-128k-cased	0.9075	0.7633	0.8292	362559	0.9270
distilbert-base-turkish-cased	0.9090	0.7960	0.8487	313947	0.9430
bert-base-cased	0.9136	0.7914	0.8481	371036	0.9327
bert-base-multilingual-cased	0.9159	0.7736	0.8384	362718	0.9310

Table 7.10: Performance Metrics of XGBoost in Models for Non-Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9345	0.9771	0.9553	1220051	0.9286
convbert-base-turkish-mc4-cased	0.9297	0.9758	0.9522	1191813	0.9253
electra-base-turkish-cased-discriminator	0.9318	0.9769	0.9538	1210807	0.9267
electra-base-turkish-mc4-cased-discriminator	0.9314	0.9766	0.9535	1190259	0.9274
bert-base-turkish-cased	0.9345	0.9780	0.9558	1216538	0.9295
bert-base-turkish-128k-cased	0.9318	0.9765	0.9536	1200568	0.9270
distilbert-base-turkish-cased	0.9503	0.9800	0.9649	1249180	0.9430
bert-base-cased	0.9377	0.9767	0.9568	1192091	0.9327
bert-base-multilingual-cased	0.9347	0.9785	0.9561	1200409	0.9310

Table 7.11: Performance Metrics of SGD in Models for Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9192	0.8472	0.8817	343076	0.9501
convbert-base-turkish-mc4-cased	0.9161	0.8438	0.8785	371314	0.9445
electra-base-turkish-cased-discriminator	0.9070	0.8152	0.8586	352320	0.9395
electra-base-turkish-mc4-cased-discriminator	0.9213	0.8440	0.8810	372868	0.9456
bert-base-turkish-cased	0.9127	0.8316	0.8703	346589	0.9450
bert-base-turkish-128k-cased	0.9167	0.8416	0.8776	362559	0.9455
distilbert-base-turkish-cased	0.9226	0.8615	0.8910	313947	0.9577
bert-base-cased	0.9174	0.8309	0.8720	371036	0.9421
bert-base-multilingual-cased	0.9361	0.8606	0.8968	362718	0.9540

Table 7.12: Performance Metrics of SGD in Models for Non-Rescue Class

Classification Models	Precision	Recall	F1 Score	Support	Accuracy
convbert-base-turkish-cased	0.9580	0.9791	0.9684	1220051	0.9501
convbert-base-turkish-mc4-cased	0.9525	0.9759	0.9641	1191813	0.9445
electra-base-turkish-cased-discriminator	0.9478	0.9757	0.9615	1210807	0.9395
electra-base-turkish-mc4-cased-discriminator	0.9524	0.9774	0.9647	1190259	0.9456
bert-base-turkish-cased	0.9532	0.9773	0.9651	1216538	0.9450
bert-base-turkish-128k-cased	0.9533	0.9769	0.9650	1200568	0.9455
distilbert-base-turkish-cased	0.9658	0.9818	0.9737	1249180	0.9577
bert-base-cased	0.9489	0.9767	0.9626	1192091	0.9421
bert-base-multilingual-cased	0.9589	0.9823	0.9704	1200409	0.9540

We examined the evaluation metrics of all BERT models and found that the most successful BERT model was distilbert-base-turkish-cased. In almost all classifications, the distilbert-base-turkish-cased obtained higher success rates for both rescue and non-rescue classes. The DistilBERT model achieved higher success than other models for rescue class in Decision Tree, Multinomial Naïve Bayes, XGBoost, and SGD classifications with recall values of 0.6974, 0.8303, 0.7960, and 0.8615, respectively. This implies that DistilBERT is the most effective in distinguishing tweets involved in rescue response and those not related to the rescue class. These evaluation metrics also suggest that DistilBERT models generally demonstrate great performance in identifying rescue-related data points accurately.

The outcome of the performance of all the deep learning models in rescue and non-rescue categories is shown in Table 7.13 and Table 7.14. In terms of recall, which indicates the ability of a model to correctly identify features of a class, the “Only DNN” model obtained the highest recall for both classes, with values of 0.8972 for the rescue class and 0.9751 for the non-rescue class. This shows that the DNN model is the most successful in accurately identifying of both classes within our data set. The “Only LSTM” model also performs well, especially for the rescue class, but lower than the “Only DNN” model according to recall. The “LSTM + CNN” model indicates a weaker performance in identifying features of both classes because it has lower recall values compared to the “Only LSTM” and “Only DNN” models. This examination shows that deep learning is efficient in classifying social media data of disaster response. The high accuracy of DNNs results from their capability to memorize complex interactions in the data and to learn efficiently from the significant set of characteristics acquired during the training process.

Table 7.13: Performance Metrics of Deep Learning Algorithms for Rescue class

Deep Learning Algorithms	Precision	Recall	F1 Score	Support	Accuracy
Only CNN	0.8664	0.8325	0.8491	62899	0.9405
Only DNN	0.9006	0.8972	0.8989	62899	0.9594
Only LSTM	0.9013	0.8871	0.8941	62899	0.9577
LSTM + CNN	0.8605	0.8524	0.8564	62899	0.9425

Table 7.14: Performance Metrics of Deep Learning Algorithms for Non-Rescue class

Deep Learning Algorithms	Precision	Recall	F1 Score	Support	Accuracy
Only CNN	0.9582	0.9677	0.9629	249727	0.9405
Only DNN	0.9741	0.9751	0.9746	249727	0.9594
Only LSTM	0.9717	0.9755	0.9736	249727	0.9577
LSTM + CNN	0.9629	0.9652	0.9641	249727	0.9425

In the topic modeling part, Table 7.15 presents an illustration of the data acquired from clustering the dataset based on Global Positioning System (GPS) coordinate information and identifying significant terms. GPS coordinates are used to cluster messages originating from locations near to each other. Each cluster is a rectangular area with a 1-kilometer width. For instance, this table displays exemplars from three distinct groups. Data within a cluster exhibits identical and significant keywords, originating in positions near to each other and in a period of 30 minutes. The goal is to quickly recognize emergencies and requirements by identifying key terms.

Table 7.15: Keywords in Texts into the Distinct Clusters

Text	Longitude	Latitude	Cluster	Words
deprem anından itibaren haber alamıyoruz hayrullah mahallesi kuddusi baba bulvarı altınköşk apt b blok kat daire kahramanmaraş necdet amp gülten erdoğan arslan erdoğan	29.024374	37.7216173	60	deprem, hatay, yardım, acil, ol, bölge, insan, kurtar, yap, yer
maraş şubat ilçesi sokak bahar çiçeği apartmanı onu nurcan alacacı arkadaşımın ailesi yardım ulaşmamış ilk günden beri lütfen yardımcı olur musunuz	32.905111	36.069849	270	deprem, aile, duy, ulaşma, halk, sokak, kurtar, kahramanmaraş, arkadaş, yardım
abi sahada maraş ilçesinde larva çadır temin edip dağıtıyor cebinden son montunu çadırını vermiş pazarcık ilçesinde çadır sorunu insanlar diyor	28.6321043	40.8027337	2	deprem, hatay, ol, bölge, yardım, insan, yap, kurtar, afad, acil

A total of three heat maps are created in this study: clusters-words heat map, which consists of tweets divided into clusters according to location information, cluster-time-words heat map, containing important words by grouping each cluster according to time, and city-words heat map, which contains significant words according to cities. Each cluster is internally organized based on location, and significant key terms a obtained. The cluster-words heat map is presented in Figure 7.1. A total of 590 clusters is obtained using tweets with coordinate information. We extracted a total of 10 important words for each cluster. Some of the clusters and keywords are displayed in this heat map.

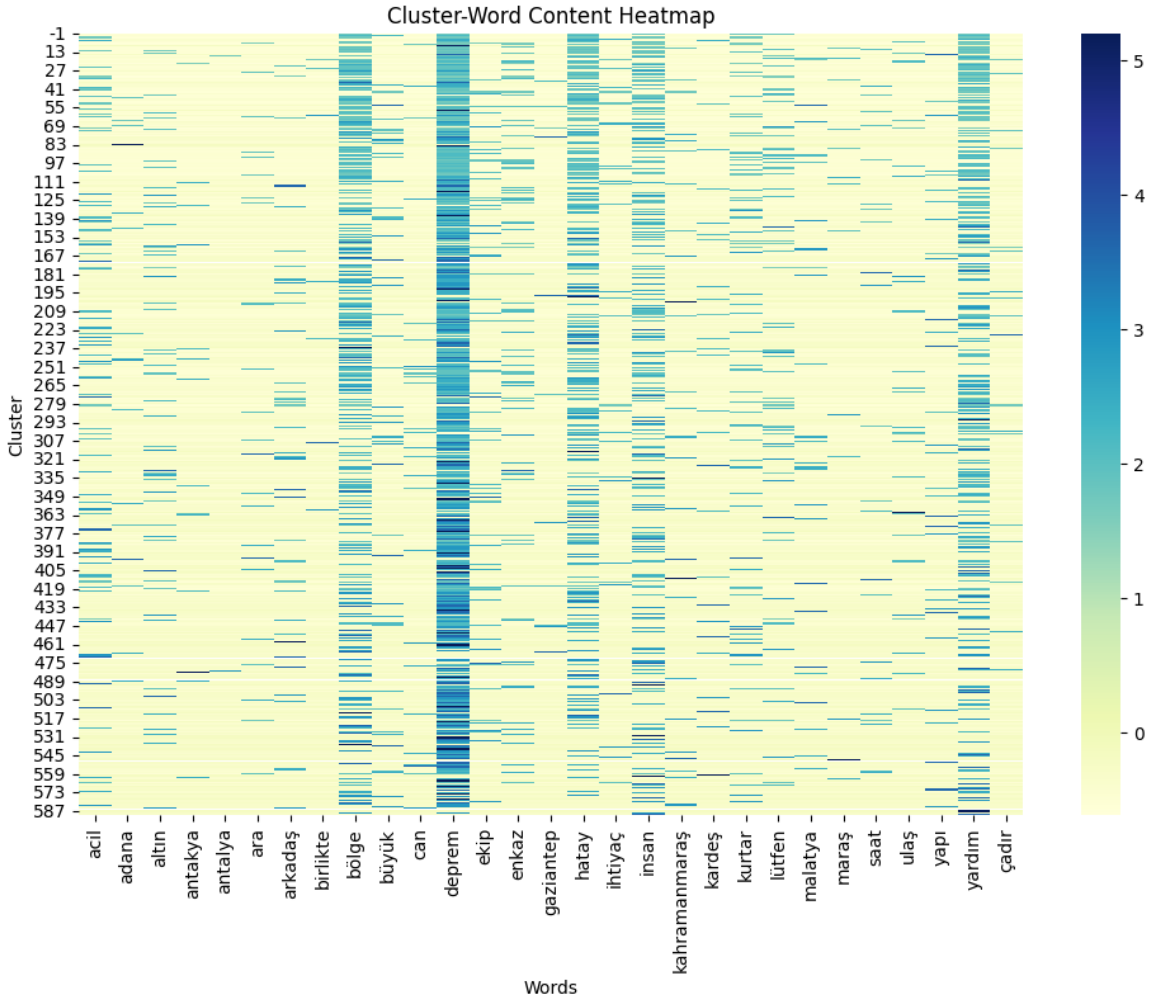

Figure 7.1: Heat Map of the Clusters and Keywords

The clusters are grouped according to different time periods, and important words are extracted again. The heat map in Figure 7.2 shows word examples found according to some time periods.

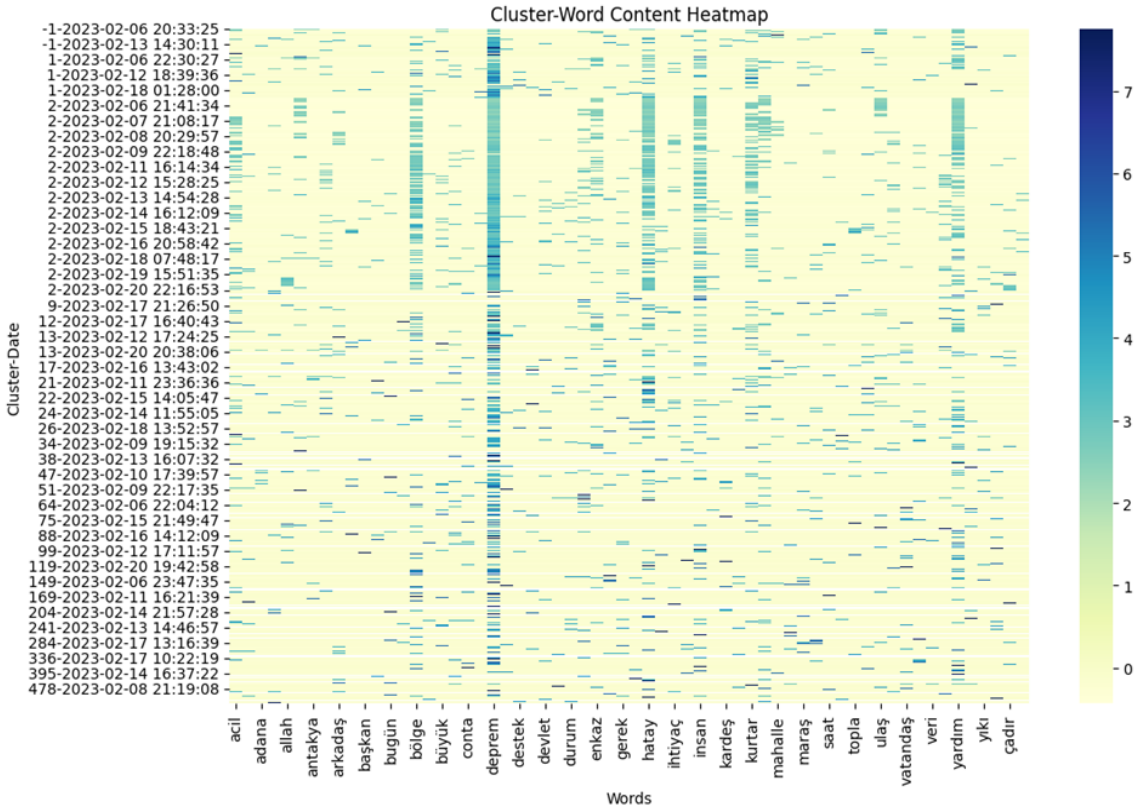

Figure 7.2: Heat Map of the Clusters-Times and Keywords

Cities in Turkiye are determined using the coordination information of the data. The city-based representation of the heat map is as shown in Figure 7.3. From this chart, it is evident that the word “deprem” is the most utilized term. Furthermore, the keywords “yardım” , “acil” , “ekip” , “enkaz” , and “hatay” have been recognized as significant vocabulary for the rescue class.

Several ways are employed to experimentally explore possible directions for the improvement of the suggested methodology.

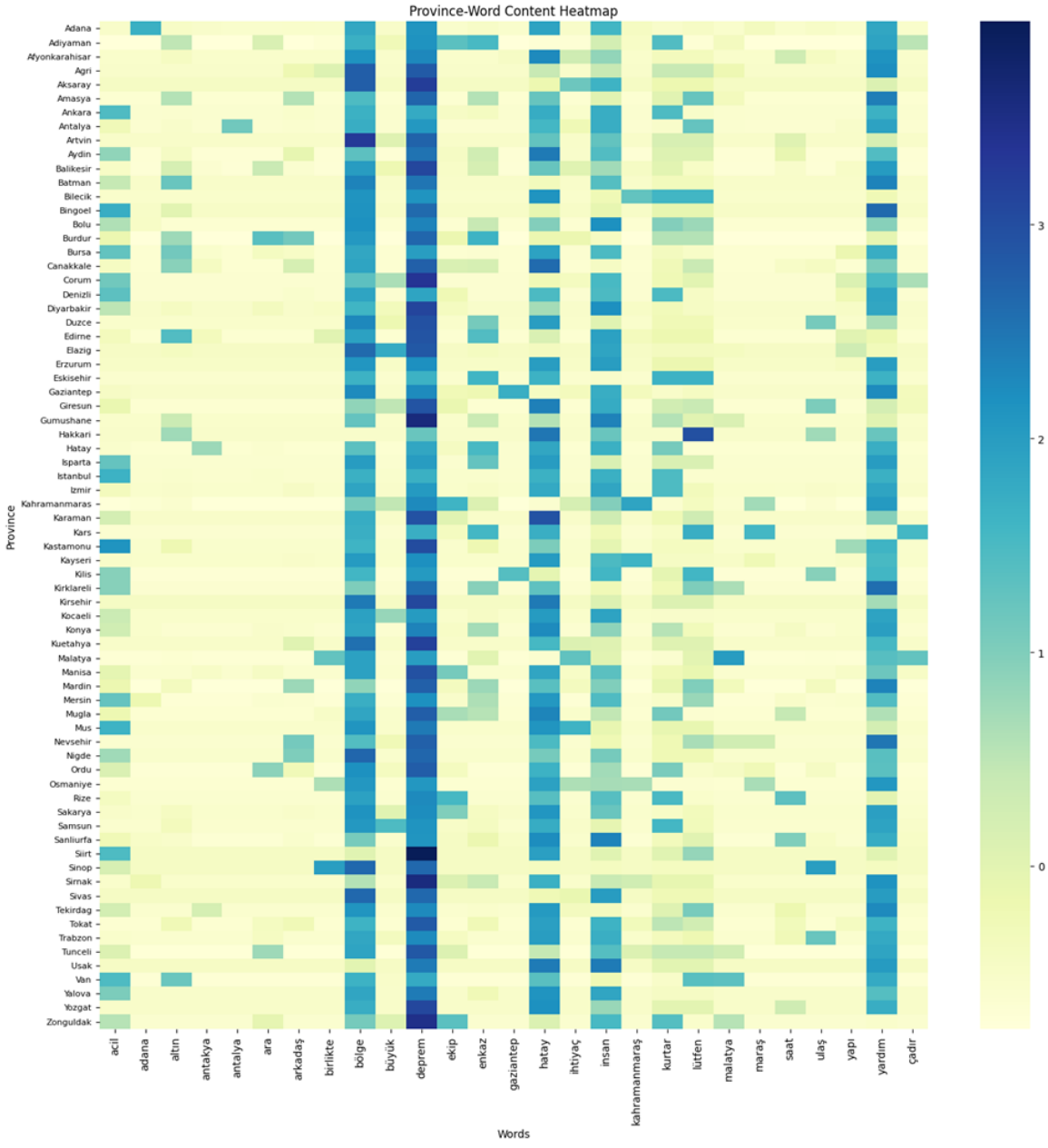

Figure 7.3: Heat Map of Cities and Keywords

Studies have also been carried out to show the visualization of clusters on the map of Turkiye. In the Figure 7.4, clusters on the map of Turkiye can be seen. The numbers inside the colored circles indicate the total coordination number. These coordination numbers are determined through tweets with known location information.

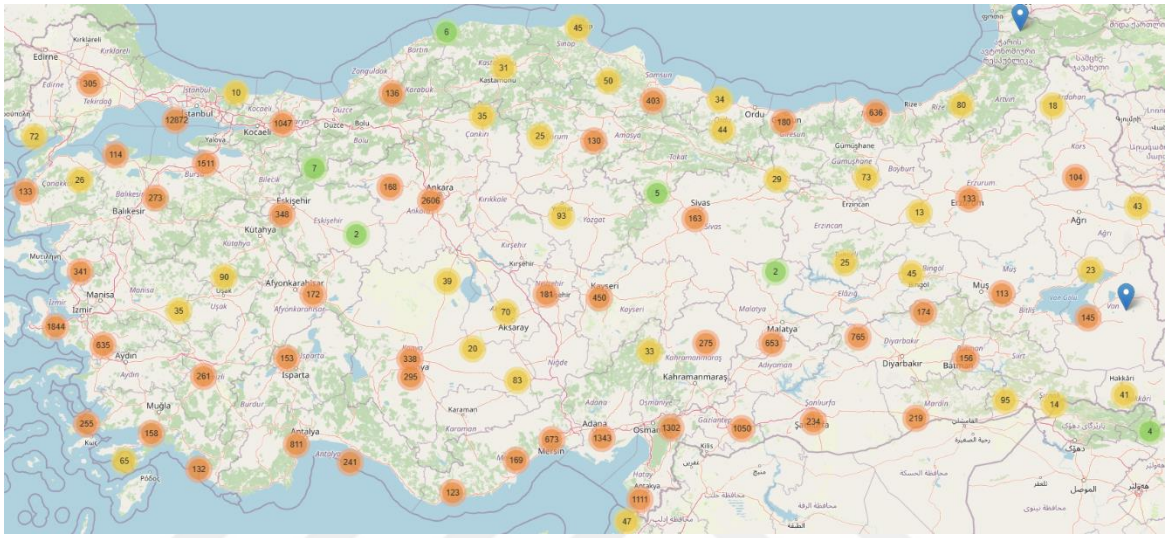

Figure 7.4: Number of Coordinates in the Map of Turkiye

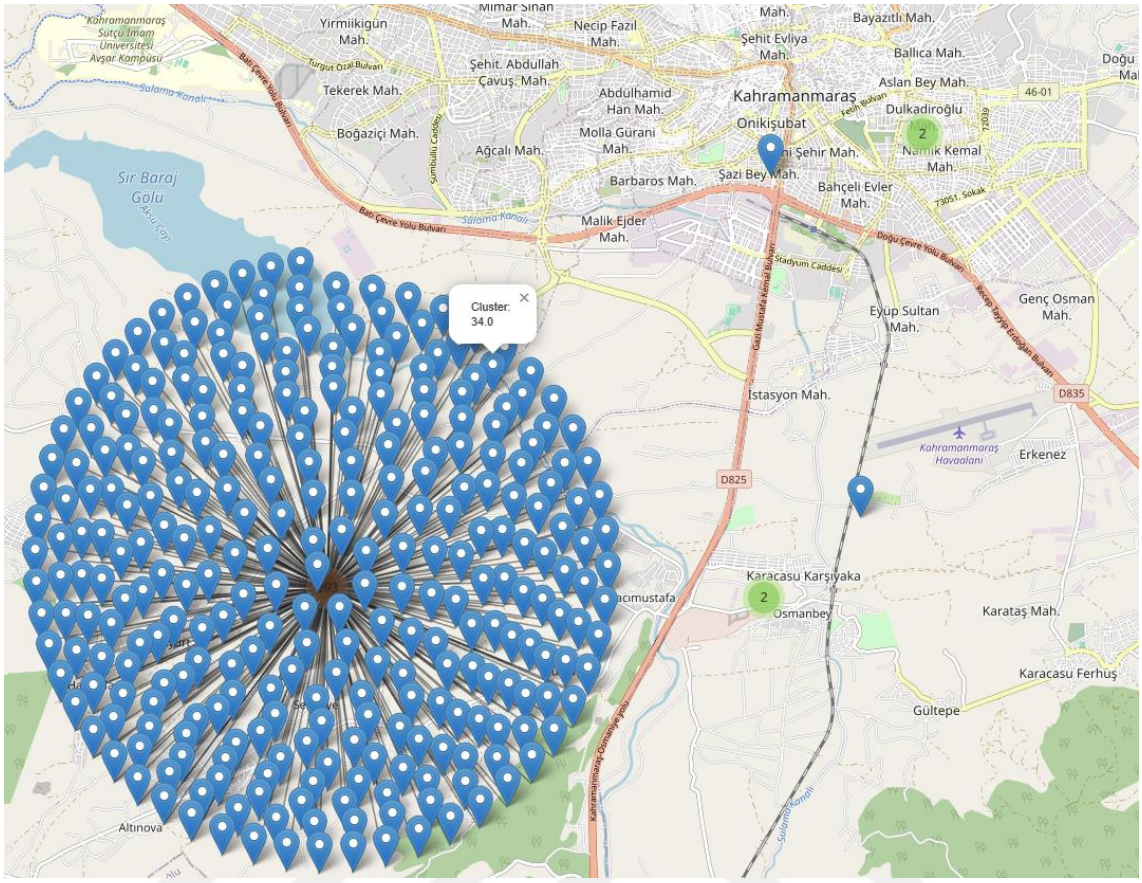

Figure 7.5: Example Map from the Clusters in Kahramanmaraş City

In Figure 7.5, an example map can be seen from the city of Kahramanmaraş, which is the center of earthquakes. For example, it seems that the posts sent from the location in the figure are grouped in cluster 34.

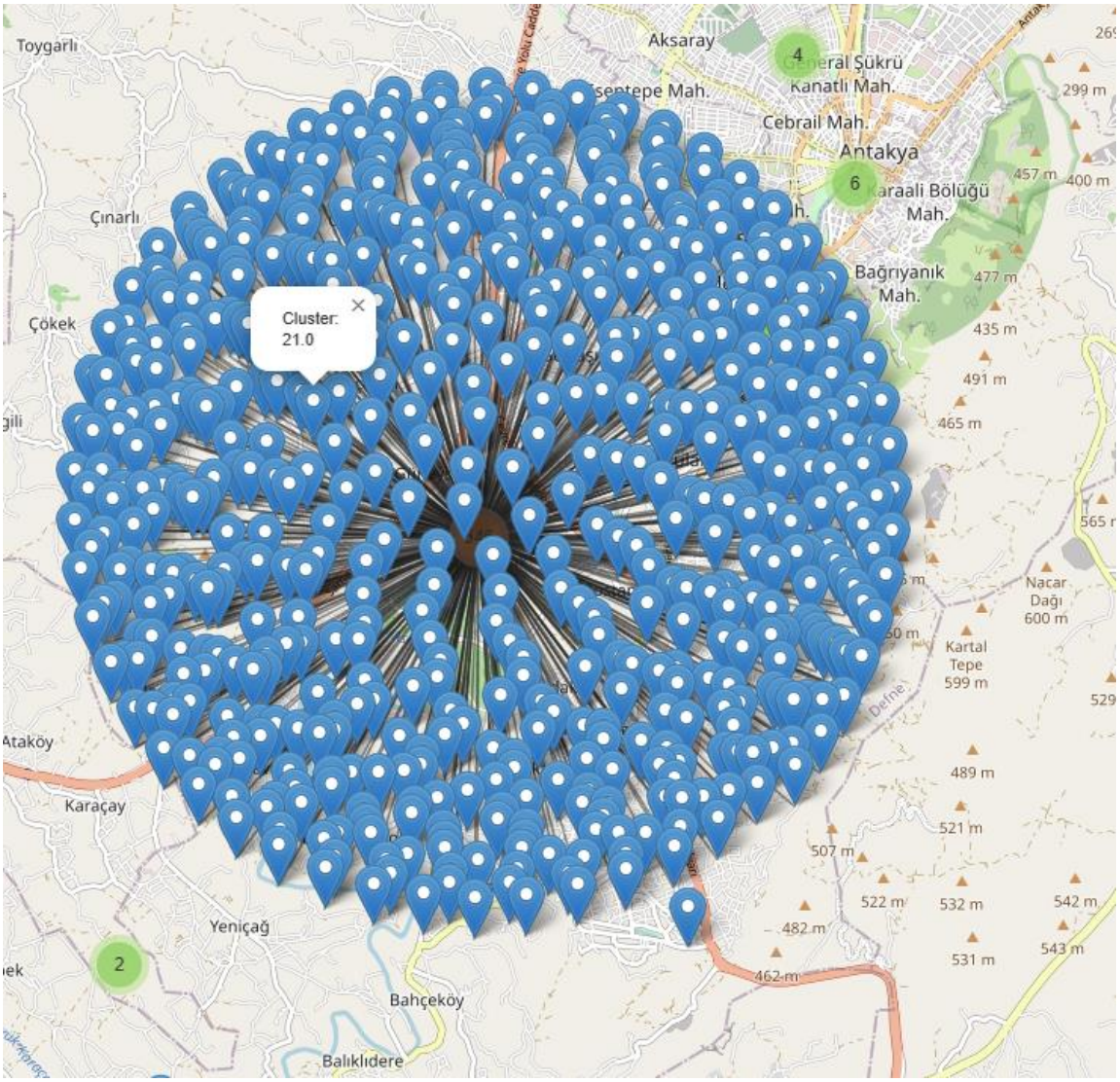

Figure 7.6: Example Map from the Cluster in Hatay City

An example map can also be seen from the city of Hatay in Figure 7.6, which is the city that was greatly affected by earthquakes.

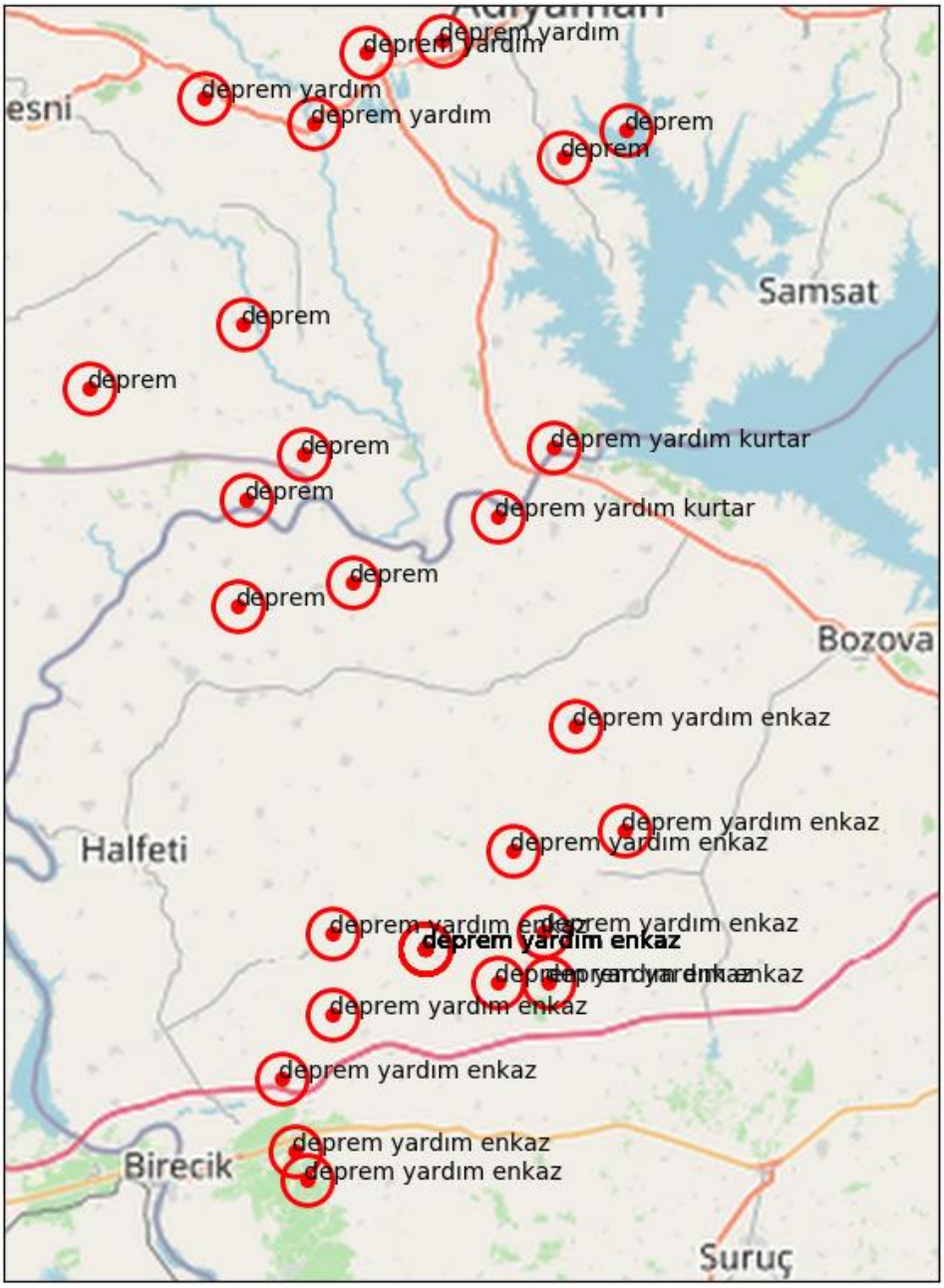

Figure 7.7: Representation of Words on the Map of Turkiye

Figure 7.7 presents a map showing the most important words found according to tweets sent from Şanlıurfa, one of the cities greatly affected by the earthquakes in Turkiye. The 3 most important keywords for this location are “deprem” , “yardım” , and “enkaz”.

8. CONCLUSIONS AND FUTURE WORK

Changes caused by natural disasters can negatively affect people's lives. Today, the influence of social media is becoming increasingly important in taking emergency measures against natural disasters. Following or amidst catastrophic events, many individuals turn to social media platforms for news because it is the fastest and most effective way to obtain critical updates from people directly involved in the event. An extremely important concern revolves around the utilization of social media to help individuals locate their lost or detained loved ones during disasters.

In previous research, there was a noticeable gap in effectively using machine learning, deep learning, natural language processing techniques together to improve response and rescue operations during natural disasters. Instead, existing studies generally focus on one of these issues to achieve the goal. Furthermore, existing studies often do not emphasize how machine learning algorithms can help improve resource allocation, decision-making processes, and overall efficiency in disaster management. By addressing these gaps, we aimed to significantly increase the effectiveness and timeliness of response efforts resulting in more efficient rescue operations and better outcomes in crisis situations. We conducted a study that was generally more successful and more detailed, using more methods than the studies we reviewed in the literature review.

This study investigated the use of machine learning, deep learning, and natural language processing techniques to evaluate social media data from X, formerly known as Twitter, users who were following the devastating earthquakes in Turkiye. The analysis is extended to extract information on a spatio-temporal basis. Through intensive data collection and preprocessing, as well as the utilization of advanced algorithms such as BERT, and classification algorithms the research successfully achieved automatic lassification of tweets into two separate categories: rescue and non-rescue. We detected

significant keywords by analyzing the data from a spatio-temporal viewpoint, which facilitated the identification of essential information that indicates urgency. This would streamline the process of providing assistance and support following natural disasters.
Our data set contains a huge amount of data. However, all this data may not fully represent the different perspectives and experiences of individuals affected by the earthquake. Furthermore, taking a subset of our dataset and labeling it manually was a subjective process, which could lead to potential inconsistencies in the labeled data. These were the limitations we saw in our study. Moreover, the most challenging aspect of our work was the data collection part, as we encountered restrictions in using data extraction through X, but we worked hard to collect the data and managed it successfully. It is determined that the performance could be better utilizing fine-tuning of machine learning model hyper parameters utilizing appropriate regularization and optimization algorithms to prevent overfitting.

This study enhances the existing research on utilizing machine learning and deep learning techniques for immediate crisis management and highlights the significance of social media data as a useful resource in disaster response efforts. The models and algorithms can be modified to suit additional emergency scenarios in the future. Enhanced and precise forecasts can be achieved by enhancing current models and algorithms. Through persistent innovation and refinement of approaches, we can strive to deliver increasingly efficient and prompt reactions in forthcoming emergencies. Moreover, by broadening the analysis to include social media platforms beyond Twitter, we can gain a more comprehensive understanding of public opinion and the dissemination of information during the crisis. Furthermore, while social media data is extremely helpful for disaster management, it is important to remember that a significant amount of false information can also be disseminated through social media platforms. Incorrect or misleading information can be spread either accidentally or purposefully. Hence, it is crucial to focus on developing dependable, efficient, and high-performing applications. This study also assists other specialists in identifying non-emergency data.

REFERENCES

Akın, A. A., & Akın, M. D. (2007). Zemberek, an open-source NLP framework for Turkic languages. Structure, 10(2007), 1-5.

Albelwi, S., & Mahmood, A. (2017). A framework for designing the architectures of deep convolutional neural networks. Entropy, 19(6), 242.

Algiriyage, N., Sampath, R., Prasanna, R., Stock, K., Hudson-Doyle, E., & Johnston, D. (2021). Identifying Disaster-related Tweets: A Large-Scale Detection Model Comparison (pp. 731-743).

Autelitano, A., Pernici, B., & Scalia, G. (2019). Spatio-temporal mining of keywords for social media cross-social crawling of emergency events. GeoInformatica, 23(3), 425-447.

Carpenter, K. A., Cohen, D. S., Jarrell, J. T., & Huang, X. (2018). Deep learning and virtual drug screening. Future medicinal chemistry, 10(21), 2557-2567.

Contreras, D., Wilkinson, S., Alterman, E., & Hervás, J. (2022). Accuracy of a pre-trained sentiment analysis (SA) classification model on tweets related to emergency response and early recovery assessment: the case of 2019 Albanian earthquake. Natural Hazards, 113(1), 403-421.

Cvetković, V., Nikolić, A., & Ivanov, A. (2023). The role of social media in the process of informing the public about disaster risks. Journal of Liberty and International Affairs, 9(2), 104-119.

Dagaeva, M., Garaeva, A., Anikin, I., Makhmutova, A., & Minnikhanov, R. (2019). Big spatio-temporal data mining for emergency management information systems. IET Intelligent Transport Systems, 13(11), 1649-1657.

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

Dwarakanath, L., Kamsin, A., Rasheed, R. A., Anandhan, A., & Shuib, L. (2021). Automated machine learning approaches for emergency response and coordination via social media in the aftermath of a disaster: A review.

Eligüzel, N., Çetinkaya, C., & Dereli, T. (2020). Comparison of different machine learning techniques on location extraction by utilizing geo-tagged tweets: A case study. Advanced Engineering Informatics, 46, 101151.

Elroy, O., & Yosipof, A. (2023, September). Semi-Supervised Learning Classifier for Misinformation Related to Earthquakes Prediction on Social Media. International Conference on Artificial Neural Networks (pp. 256-267).

Fuzzy Topic Modeling - methods derived from Fuzzy Latent Semantic Analysis **URL:** https://pypi.org/project/FuzzyTM/

GeeksforGeeks (2023) Semi-supervised learning in ML, GeeksforGeeks. **URL:** https://www.geeksforgeeks.org/ml-semi-supervised-learning/

GeeksforGeeks (2024) ML: Stochastic gradient descent (SGD), GeeksforGeeks. **URL:** https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd

Huang, X., Li, Z., Wang, C., & Ning, H. (2019). Identifying disaster related social media for rapid response: a visual-textual fused CNN architecture. International Journal of Digital Earth.

Jia, J., & Ye, W. (2023). Deep Learning for Earthquake Disaster Assessment: Objects, Data, Models, Stages, Challenges, and Opportunities. Remote Sensing, 15(16), 4098.

Jiang, Z. H., Yu, W., Zhou, D., Chen, Y., Feng, J., & Yan, S. (2020). Convbert: Improving bert with span-based dynamic convolution. Advances in Neural Information Processing Systems, 33, 12837-12848.

Joseph, J. K., Dev, K. A., Pradeepkumar, A. P., & Mohan, M. (2018). Big data analytics and social media in disaster management. In Integrating disaster science and management (pp. 287-294). Elsevier.

Kanade, V. (2022, September 29). All you need to know about support vector machines Spiceworks. **URL:** https://www.spiceworks.com/tech/big-data/articles/what-is-support-vector-machine/

Koshy, R., & Elango, S. (2023). Multimodal tweet classification in disaster response systems using transformer-based bidirectional attention model. Neural Computing and Applications, 35(2), 1607-1627.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150.

Kumar, A., Singh, J. P., Dwivedi, Y. K., & Rana, N. P. (2022). A deep multi-modal neural network for informative Twitter content classification during emergencies. Annals of Operations Research, 1-32.

Lemos, A.R. (2022) Cross-validation, Medium.
URL: https://towardsdatascience.com/cross-validation-705644663568

Li, L., Zhang, Q., Tian, J., & Wang, H. (2018). Characterizing information propagation patterns in emergencies: A case study with Yiliang Earthquake. International Journal of Information Management, 38(1), 34-41.

Li, L., Bensi, M., & Baecher, G. (2023). Exploring the potential of social media crowdsourcing for post-earthquake damage assessment. International Journal of Disaster Risk Reduction, 98, 104062.

Linardos, V., Drakaki, M., Tzionas, P., & Karnavas, Y. L. (2022). Machine learning in disaster management: recent developments in methods and applications. Machine Learning and Knowledge Extraction, 4(2).

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. V. (2020). Classifying earthquake damage to buildings using machine learning. Earthquake Spectra, 36(1), 183-208.

Manimegalai, R., Sri Rajiv Jegan, G. V., Arawind, S., & Gomathi, B. (2023, April). Dtweet: Disaster Tweet Analysis Using Deep Learning Techniques. In International Conference on Soft Computing for Security Applications (pp. 331-343). Singapore: Springer Nature Singapore.

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.

Ofli, F., Alam, F., & Imran, M. (2020). Analysis of social media data using multimodal deep learning for disaster response. arXiv preprint arXiv:2004.11838.

Powers, C. J., Devaraj, A., Ashqeen, K., Dontula, A., Joshi, A., Shenoy, J., & Murthy, D. (2023). Using artificial intelligence to identify emergency messages on social media during a natural disaster: A deep learning approach. International Journal of Information Management Data Insights, 3(1), 100164.

Pykes, K. (2023, October 19). What is Topic Modeling? An Introduction with Examples. **URL:** https://www.datacamp.com/tutorial/what-is-topic-modeling

Resch, Bernd & Usländer, Florian & Havas, Clemens. (2017). Combining machine-learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment. Cartography and Geographic Information Science. 45. 1-15. 10.1080/15230406.2017.1356242.

Rijcken, E., Mosteiro, P., Zervanou, K., Spruit, M., Scheepers, F., & Kaymak, U. (2022, July). FuzzyTM: a software package for fuzzy topic modeling. In 2022 IEEE international conference on fuzzy systems (FUZZ-IEEE) (pp. 1-8). IEEE.

Rijcken, E. (2022, May 20). FuzzyTM: A Python package for Fuzzy Topic Models - Towards Data Science. Medium. **URL:** https://towardsdatascience.com/fuzzytm-a-python-package-for-fuzzy-topic-models-fd3c3f0ae060

Saad, O. M., Chen, Y., Trugman, D., Soliman, M. S., Samy, L., Savvaidis, A., ... & Chen, Y. (2022). Machine learning for fast and reliable source-location estimation in earthquake early warning. IEEE Geoscience and Remote Sensing Letters, 19, 1-5.

Sanchez, G., Marzban, E. (2020) All Models Are Wrong: Concepts of Statistical Learning. **URL:** https://allmodelsarewrong.github.io

Savci, P., & Das, B. (2023). Comparison of pre-trained language models in terms of carbon emissions, time and accuracy in multi-label text classification using AutoML. Heliyon, 9(5).

Scheele, C., Yu, M., & Huang, Q. (2021). Geographic context-aware text mining: Enhance social media message classification for situational awareness by integrating spatial and temporal features. International Journal of Digital Earth, 14(11), 1721-1743.

Shehzad, F., Rashid, M., Sinky, M. H., Alotaibi, S. S., & Zia, M. Y. I. (2021). A scalable system-on-chip acceleration for deep neural networks. IEEE Access, 9, 95412-95426.

Shenfield, A., & Howarth, M. (2020). A novel deep learning model for the detection and identification of rolling element-bearing faults. Sensors, 20(18), 5112.

Simple Transformers (2020, April 24).

URL: https://pypi.org/project/simpletransformers/0.26.0/

Wang, Q., Guo, Y., Yu, L., & Li, P. (2017). Earthquake prediction based on spatio-temporal data mining: an LSTM network approach. IEEE Transactions on Emerging Topics in Computing, 8(1), 148-158.

Wikipedia contributors. (2024, April 12). 2023 Turkey–Syria earthquakes - Wikipedia. **URL :** https://en.wikipedia.org/wiki/2023_Turkey–Syria_earthquakes

Xiao, Yu & Huang, Qunying & Wu, Kai. (2015). Understanding Social Media Data for Disaster Management. Natural Hazards. 79. 10.1007/s11069-015-1918-0.

Yu, M., Huang, Q., Qin, H., Scheele, C., & Yang, C. (2020). Deep learning for real-time social media text classification for situation awareness–using Hurricanes Sandy, Harvey, and Irma as case studies. In Social Sensing and Big Data Computing for Disaster Management (pp. 33-50). Routledge.

Zhou, B., Zou, L., Mostafavi, A., Lin, B., Yang, M., Gharaibeh, N., Mandal, D. (2022). VictimFinder: Harvesting rescue requests in disaster response from social media with BERT. Computers, Environment and Urban Systems, 95, 101824.

APPENDICES

Appendix A. ENGLISH TRANSLATION OF SOME TURKISH WORDS

Turkish Language	**English Language**
Acil	Urgent
Adres	Address
Apartman	Apartment
Arama	To search
Arkadaşlar	Friends
Battaniye	Blanket
Bağış	Donation
Birlikte	Together
Can	Life
Daire	Flat
Deprem	Earthquake
Destek	Support
Enkaz	Debris
Etkilemek	To affect
Göndermek	To send
Güvenli	Secure
İhtiyaç	Need
İletişim	Contact
İnternet	Internet
Kan	Blood
Kayıp	Loss
Kişi	Person

Turkish Language	**English Language**
Kurtarma	Rescue
Lütfen	Please
Mahalle	Neighbourhood
Önemli	Important
SMS	SMS
Sokak	Street
Telefon	Phone
Toplanma Alanı	Assembly Area
Yardım	Help
Yayalım	Let’s spread
Yıkılan	Collapsed