Journal of Modern Technology and Engineering
Vol.7, No.1, 2022, pp.5-19

Jomard Publishing

# COMPARATIVE ANALYSIS OF COUNT VECTORIZATION VS TF-IDF VECTORIZATION FOR DETECTING CYBERBULLYING IN TURKISH TWITTER MESSAGES

**Mikayıl Sadigzade[1*], Efendi Nasibov[2]**

[1]Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey
[2]Department of Computer Science, Dokuz Eylul University, Izmir, Turkey

**Abstract.** Cyberbullying is a growing problem in Turkey as well as all over the world. According to the findings so far, the probability of being exposed to cyberbullying for those who use social media in Turkey has exceeded 20%. Although there is a lot of cyberbullying detection in English, there are few studies in Turkish. Machine learning is often used to eliminate and detect this problem. In this study, different machine learning algorithms were used to detect cyberbullying on Turkish texts. Our study was carried out using machine learning techniques on a data set consisting of 3000 sentences written in Turkish and collected from social media. Precision, accuracy, recall and F1-score were used to evaluate the performance of the classifiers. In the study, Linear SVM model gave the highest results with 88.35% accuracy and 99.96% F1-score for CountVectorizer. With the same model, 88.69% accuracy and 99.96% F1-score results were achieved for Tf-Idf Vectorizer.

**Keywords**: Cyberbullying detection, Machine learning, N-gram, Tf-Idf, Twitter messages.
**Corresponding author:** Mikayıl Sadigzade, Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir, Turkey, e-mail: *mikayilsadigzade@gmail.com*
*Received: 12 January 2022; Revised: 17 March 2022; Accepted: 2 April 2022; Published: 29 April 2022.*

## 1 Introduction

With the development of the social network, the number of cyberbullying began to increase in the world. Cyberbullying detection is also attracting increasing attention, especially with the use of Machine Learning. The reason for the increase in cyber bullying is that the bullying on the internet cannot be detected or they think that even if it is detected, legal sanctions will not be applied. Such cyberbullying crimes create psychological pressure on people and leave spiritual traces in their lives in the future. It is very difficult to detect and counter cyberbullying in a timely manner. Worldwide, large numbers of children are subjected to sexual discrimination and gender-based violence, corporal punishment, war and other forms of violence. In addition, many are exposed to gang violence, armed attack, rape, harassment, sexual and gender-based violence by their peers in the schoolyard. As a new type of violence, incidents such as cyberbullying, especially through mobile phones, computers, websites and social networking sites, negatively affect children's lives (Altay & Betül, 2017).

The development of information technologies and the rapid introduction of communication tools into the lives of users paved the way for the development and diversity of social media platforms, websites and sharing networks (Altay & Bilal, 2018). There are newly established institutions in many countries to combat cyberbullying. As examples of these institutions, we can show the National Crime Prevention Council in the USA and the Information Technologies and Communications Authority in Turkey (Shukan et al., 2019). The fact that cases of cyberbullying, which can be dangerous for children and adolescents, have become a growing social problem

worldwide, has led educators, academics and law enforcement to examine the risks, prevalence and consequences of cyberbullying and produce solutions in this regard (Fig. 2). Especially in recent years, it has been observed that more and more countries have formed education policies on this issue, supported research and projects, and established partnerships with non-governmental organizations, academics, educators and families working on this issue (Akca & Idil, 2017).

In Figure 1, some statistical results are given as a result of research on cyberbullying. According to these results, LGBTQ individuals are more likely to be cyberbullied on social media (Ravichandran, 2017).

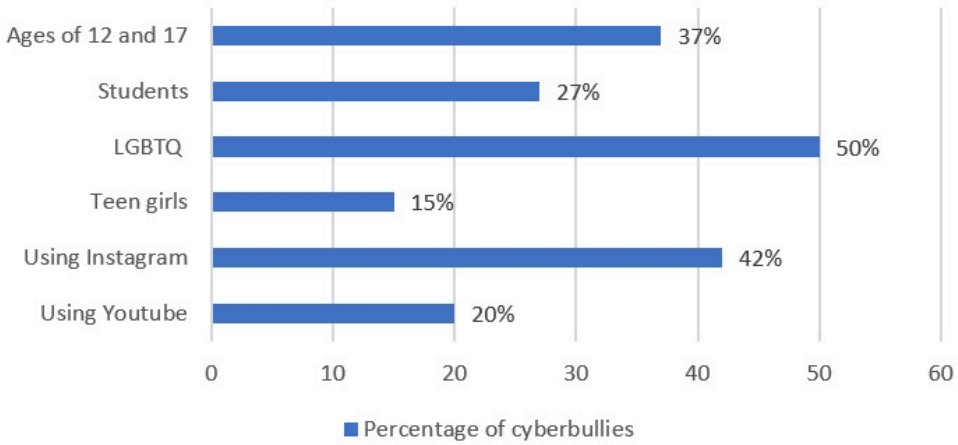


**Figure 1:** The percentage of cyberbullies (Ravichandran, 2017)

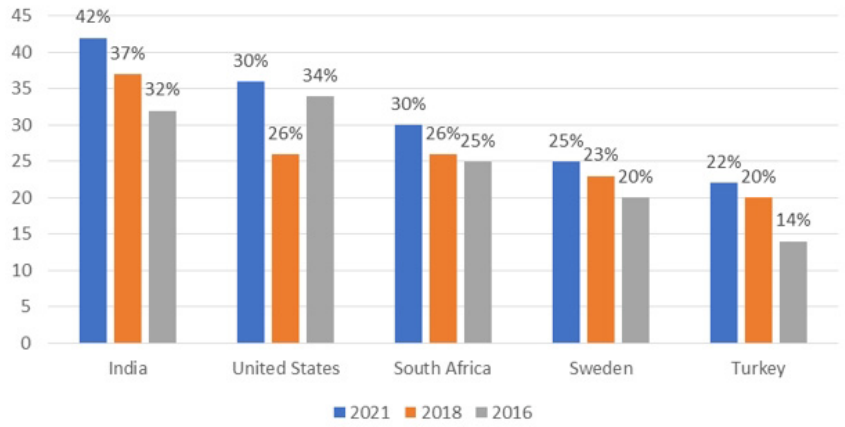


**Figure 2:** The percentage of children being cyberbullied

Academic research in various countries has focused on the perceptions and awareness of children who have been cyberbullied, the effects of cyberbullying on them, and how they try to cope with the problem. Although the nature of cyberbullying, the behaviors of cyberbullying and the results of this phenomenon differ from country to country, it is clear that there are many common points depending on the culture created by the new media (Akca & Idil, 2017).

As a result of many studies on cyberbullying, it has been shown that cyberbullying and victimization have an extremely bad effect on people's social, academic and emotional lives (Eroğu & Neşe, 2015). In their study Polat & Seda (2016), conducted using the relational screening model, data on internet use, anger and expression styles of anger and stress management variables were used, as well as socio-demographic variables believed to affect adolescents' cyberbullying and victimization. Behavior variables were collected, evaluated and analyzed systematically and statistically. The data obtained were explained with a descriptive method.

Ünver & Zihni (2017) study was conducted to determine how much cyberbullying by secondary school students is related to problematic internet use and risky online behaviors. In addition, different factors such as age, gender, daily internet use and preferred websites have an effect on this.

Cyberbullying is a growing problem in Turkey as well as all over the world. According to the researches, with the increase in the number of people using social media recently, the probability of being exposed to cyberbullying in Turkey has exceeded 20%. In this context, although there are many studies on the detection of cyberbullying in English texts in the literature, there are few studies on Turkish texts. In addition, only a limited number of algorithms and methods are used in Turkish studies.

In the recent study, the Turkish dataset prepared in the article Bozyiğit et al. (2018) was used. First, various preprocessing steps were applied to the respective dataset. After the preprocessing steps, stop word, n-gram and Tf-Idf approaches were applied as attributes on the texts in the data set. Later, different machine learning algorithms developed were tested to detect cyberbullying Twitter messages. Finally, the results of the developed models were compared in terms of prediction success. As a result, Linear SVM model achieved the highest success with 88.35% accuracy and 99.96% F1-score for Count Vectorizer. With the same model, 88.69% accuracy and 99.96% F1-score results were achieved for Tf-Idf Vectorizer.

## 2 Related Work

In the study Reynolds et al. (2011) a data set was created from the questionnaire from the content of Formspring.me, a question-answer site where bullying occurs at a much higher level. The labeling process of the large amount of data obtained in this study was carried out using the Amazon Mechanical Turk web service. The application of the C4.5 decision tree method to the created dataset revealed that texts containing cyberbullying were detected with an accuracy of 78.5%.

Arroyo-Fernández et al. (2018) tested a number of vector space modeling techniques together with several well-known classifiers to separately predict each of the aggression levels. Vector space modeling techniques include latent semantic analysis (LSA) of Tf-Idf vectors, calculated for n-gram features of characters or words according to both Tf-Idf and LSA.

Sugandhi et al. (2016) tested different classifiers on the dataset to see which classifier gave the best accuracy. To do this, they split the labeled data into two sets, using 80% of the data for training and 20% for testing. In the study, labeled data of 393 bullying and 2886 non-bullying messages were compared with other studies using Linear SVM, Multinomial Naive Bayes and KNN classifiers.

Novalita et al. (2019) addressed both cyberbullying and non-cyberbullying tweets. Looking at the results of the random forest classification test, it was determined that the system successfully found the cyberbullying tweets with the best F1-score of 90%. Isa & Ashianti, (2017) searched for the most appropriate SVM kernel for cyberbullying classification. They found that the best model was a multicore model with an accuracy of 97.11%, since the data used were not linearly separable.

Artificial neural network (ANN) models such as LSTM, D-CNN, Random Forest (RF) were used in the paper dealing with Turkish texts (Pervan & Keleş, 2019). The studies have been carried out on the data collected from the N11.com website. In the study Agrawal & Awekar (2018), artificial neural network models such as ANN, CNN, RNN, LSTM were used and among them, LSTM showed the highest accuracy. Studies have been carried out on the data collected from Twitter, a social media tool. In the study Yuvaraj et al. (2021), only the CNN algorithm was used in this article examining the English texts and studies were conducted on the data collected from Formspring.me. In the study, the highest score was obtained in the F1-score.

Zhao et al. (2020) and Rezvani et al. (2020) studied on the LSTM algorithm by considering

different data in English collected from Twitter. In both studies, the LSTM showed its highest score in the F1-score.

Rachid et al. (2020) used the CNN technique on the data generated from English texts. Research was conducted on the data collected from Twitter, and the study found the highest F1-score for CNN. In the study Pawar & Raje (2019), the data were created from English texts and NB, LR (Logistic Regression), SGD (Stochastic Gradient Descent) machine learning techniques were used. Data collected from Formspring.me was studied and in the study, LR showed the highest accuracy and F1-score results. The GHSOM (Growing Hierarchical Self-Organizing Map) technique was used in the paper on the English text (Di Capua et al., 2016). Studies were carried out on the data collected from Formspring.me, Youtube, Twitter.

In the paper Aind et al. (2020), the Reinforcement Learning technique, which is an area of machine learning, was used. Studies have been made on the data collected from the comments, and reinforcement learning has shown the highest accuracy result in the study. In the study Soni & Singh (2018), LR technique was used. Data collected from posts have been studied, and LR showed its highest result in AUROC.

Haidar et al. (2019) a meta-algorithm technique used to improve the stability and accuracy of machine learning algorithms. Research has been done on data collected from Twitter, and in the study, SVM showed the highest score in the F1-score. Paul & Saha (2020) used BERT machine learning technique in English. Studies have been done on data collected from Wikipedia, Formspring.me and Twitter. In the study, BERT showed its highest result in the F1-score.

Algorithms such as BOW, Linear SVM, LR were used by Karcioğlu and Aydin, (2019), which was studied on Turkish texts. Studies have been made on the data collected from Twitter, and in the study, SVM has shown the highest result in accuracy. When we look at the study Altay & Alatas (2018), techniques such as BLR, RO, Multi-layer Perceptron, J48, DVM, which are used relatively less, were used. Studies were conducted on data collected from the Formspring.me website, and the study showed the highest results in F1-score and ROC.

In Kumari et al. (2021), Random Forest was used and studies were conducted on the data collected from the comments written in the video conversations. In the study, Random Forest scored highest in the F1-score. Hani et al. (2019) used SVM, Neural Network techniques. Studies have been done on data collected from e-mail and additionally instant messaging, and also in the study, SVM showed its highest result in F1-score. Song & Song, (2021) the CRT algorithm was used and studies were conducted on the data collected from buzzes (phone calls), and in the study, CRT showed the highest accuracy.

## 3 Preliminaries

In general, many types of cyberbullying are discussed in the literature, as cyberbullying differs according to its effect on people (Weru et al., 2017). The most common types of cyberbullying are:

- Harassment

Harassment is the transmission of insulting and profanity messages from the other party of the cyberbully (Akca & Sayımer, 2017; Öztürk, 2017; Yazğılı & Baykara 2021).

- Outing

Outing is the sharing of the personal information of the person who has been cyberbullied, the information that should not be seen by anyone, or the photos, in a way that everyone can see, on the phone and social media (Akca & Sayımer, 2017; Öztürk, 2017; Yazğılı & Baykara, 2021).

- Exclusion

It is the deliberate removal or restriction of the person exposed to cyberbullying from groups such as whatsapp and instagram (Öztürk, 2017; Yazğılı & Baykara, 2021).

**Table 1**: The summary of relevant studies.

| Works | Methods used | Datasets | Results | Supported language |
|---|---|---|---|---|
| Aind et al., 2020 | Reinforcement Learning | Comments | Reinforcement Learning Accuracy=%89 | English |
| Agrawal and Awekar, 2018 | CNN | Formspring. me | CNN F1-Score = %93 | English |
| Altay and Alatas, 2018 | BLR, RO, J48, DVM | Formspring. me | RO F1-Score=%83,2; ROC=%92 | English |
| Bozyiğit et al., 2021 | SVM, MNB, RF, LR, AdaBoost | Twitter | AdaBoost F1-Score = %87.6 | English |
| Di Capua et al., 2016 | GHSOM Network Algorithm | Formspring. me,Youtube, Twitter | GHSOM Accuracy = %73, F1-Score = %74 | English |
| Haidar et al., 2019 | SVM | Twitter | SVM F1-Score = %90,5 | English |
| Hani et al., 2019 | SVM, Neural Network | Email and instant messenger | SVM F1-Score = %89,8 | English |
| Paul and Saha, 2020 | BERT | Twitter Wikipedia Formspring. me | BERT F1-Score = %94 | English |
| Pawar and Raje, 2019 | NB, LR, SGD | Formspring. me | LR Accuracy=%90,24, F1-Score = %90,56 | English |
| Pervan and Keleş, 2019 | LSTM, D-CNN, Random Forest (RF) | N11 website | LSTM Accuracy =%94,21 | Turkish |
| Rachid et al., 2020 | RNN, CNN | Comments | RNN F1-Score = %84 | English |
| Reynolds et al., 2011 | J48, JRIP, IBK and SMO | Formspring. me | J48 Accuracy = %81,7 | English |
| Rezvani et al., 2020 | LSTM | Instagram, Twitter | LSTM F1-Score = %92,8 | English |
| Karcioğlu and Aydin 2019 | Lineer SVM, Lojistik Regresyon | Twitter | Lineer SVM Accuracy =%65,62, Logistic Regression Rated average= =%60,99 | English |
| Kumari et al., 2021 | Random Forest | Images with comments | Random Forest F1-Score = %74 | English |
| Yuvaraj et al., 2021 | ANN, DRL | Twitter | ANN Accuracy=%80,7 | English |
| Song and Song, 2021 | CRT | Buzzes | CRT Accuracy = %74,5 | English |
| Soni and Singh, 2018 | LR | Posts | LR AUROC = %83,4 | English |
| Van et al., 2020 | CNN | Posts | CNN F1-Score = %88,5 | English |
| Zhao et al., 2020 | LSTM with attention network | Twitter | LSTM F1-Score = %86,3 | English |

- Imitation

Impersonation or phishing is to send or share various messages by disguising the cyberbullied person or someone else in order to hurt the person or people who are being bullied and to make them look bad in the society and in the presence of their friends (Yazğılı & Baykara, 2021).

- Attack

An attack is the sending of various lies, bad and gossip phrases or information about the person who has been cyberbullied to other third and fourth parties (Öztürk, 2017; Yazğılı & Baykara, 2021).

- Anger

This action, also known as flaming incitement, generally creates low demonstrative actions and a hostile effect in the relations between people who use social media (Akca & Sayımer, 2017;

Öztürk, 2017; Yazğılı & Baykara, 2021).

- Fake profiling

Create fake profile is an action taken to embarrass people who are cyberbullied through fake platforms such as websites and online groups, and to reduce their security with malicious news (Öztürk, 2017; Yazğılı & Baykara, 2021).

- Peer cyber tracking

Peer cyberharassment is a more severe form of harassment, which is another form of cyberbullying. The difference of this type from harassment is that the victim of cyberbullying is constantly followed, threatened, intimidated, and constantly feeling in danger through different communication technologies (Öztürk, 2017; Yazğılı & Baykara, 2021).

# 4 Feature Extraction

At this stage, the texts shared on social media are handled, the features that machine learning can work on are converted into vector format and certain studies are carried out. The reason why it is used in n-gram applications, which is a widely used method in many applications, is to capture the properties of the contents of the data. Packet load analysis is a technique used in many areas such as text analysis. Especially in network load analysis, this concept is used in various ways with various approaches (Hadžiosmanović et al., 2012).

Unlike other Turkish studies on the detection of cyberbullying, the document is expressed numerically in the term matrix by using the n-gram method (n=1,2,3,..) in the data set. Thus, it is aimed to evaluate the slang word sequences as a whole in the classification process. As an example, the following terms are taken from the message "çok gerizekalı birisin" ("You are such a retard").

- 1- gram terms: "çok", "gerizekalı", "birisin",
- 2- gram terms: "çok gerizekaı", "gerizekalı birisin",
- 3- gram terms: "ç gerizekalı birisin"

CountVectorizer is a widely used technique provided by the scikit-learn library in the Python programming language. The purpose of using this technique is to transform a text under consideration into a vector based on the frequency, or count, of each word in the entire text. In this method, each word in the documents is vectorized using the frequency of words in the corpus.

Tf-Idf is the weight factor calculated with the statistical method that shows the importance of a term in the document. The weight of a Tf-Idf term is obtained by multiplying the term frequency (the number of times the term occurs in a document) by the inverse term frequency (the log ratio of the number of documents in the dataset to the number of documents containing the term).

More specifically, the Tf-Idf value for the word t in document d is calculated from the corpus C (documents set) as follows:

$$tf\_df\,(t,d,C) = tf\,(t,d) * idf\,(t,C)\,, \tag{1}$$

where

$$tf\,(t,d) = log\,(1 + freq\,(t,d)) \tag{2}$$

and

$$idf(t,d,C) = log(\frac{N}{count(d \in C : t \in d)}) \tag{3}$$

## 5 Evaluation Metrics

Since cyberbullying detection is a classification task, the accuracy score is also considered as the first rating scale. Accuracy, which is a metric, is mostly used to measure the success of the model, but unfortunately it is insufficient on its own. The value of accuracy is calculated by the ratio of the predicted areas in the model to the total dataset. One of the biggest reasons for using F1-score instead of it due to lack of accuracy is the uneven distribution of the considered dataset. In addition, another reason that makes the F1-score important for us is that it is a measurement metric that includes not only False Negative or False Positive but also all error costs. Recall indicates how well positive situations were predicted, and precision indicates success in a positively predicted situation. Learning performance was calculated using the confusion matrix in Table 2. The F1-score, accuracy, precision, and recall are given in the equations below.

**Table 2**: Confusion Matrix

| | **Actual Value (confirmed by experiment)** | | |
|---|---|---|---|
| **Predicted by the model** | | **positives** | **negatives** |
| | **positives** | **TP**(True Positive) | **FP**(False Positive) |
| | **negatives** | **FN**(False Negative) | **TN**(True Negative) |

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \tag{4}$$

$$Recall = \frac{TP}{TP + FN}, \tag{5}$$

$$Precision = \frac{TP}{TP + FP}, \tag{6}$$

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}. \tag{7}$$

## 6 Machine Learning Methods

Machine learning is a standalone computing solution based on data and experience. Decisions to be made in machine learning are usually oriented towards prediction or classification (Hutter et al., 2019; Sarkar, 2019). There are some training methods for labeling training data. By tagging here, it is meant that the data is first classified by humans and then used for machine learning (Aggarwal, 2018). Also, supervised learning of a model depends on labeled data, and unsupervised learning is when the learning data is unlabeled. In unsupervised learning, the machine learns to classify itself according to the similarities and differences between the data. Semi-supervised learning is learning applied to a combination of labeled and unlabeled data (Rosa et al., 2019).

Depending on the labeling of the training data with different training methods, the algorithms used in this study are listed below:

- Multinomial Naive Bayes
- Decision Tress
- Random Forest
- Linear SVM

# 7 Application

## 7.1 Dataset

In this study, it is aimed to detect Turkish cyberbullying messages by using different machine learning algorithms. The proposed work consists of machine learning methods such as preprocessing and feature extraction. The steps performed in the study and the flow between them are visualized with Figure 3.

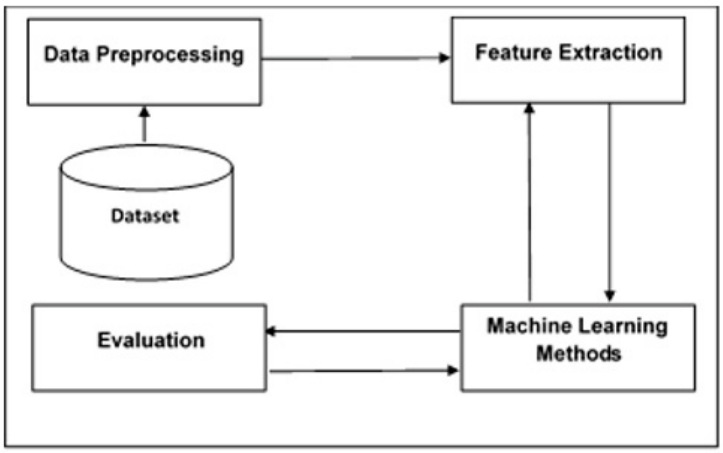


**Figure 3:** Flow of proposed work

In the recent study, the data set consisting of 3,000 Twitter shares used in the article (Bozyiğit et al., 2019) was used, and half of the messages in this set were labeled positively (containing cyberbullying) and the other half were labeled negatively (not containing cyberbullying). In order to obtain better results for the data set used in the study, some preprocessing steps were carried out. In the first step, numeric characters, punctuation, and web links were removed from the shared folders in the dataset and converted to lowercase. It is noteworthy that the data set used in this direction has a fairly balanced distribution. Some common messages in the dataset and their labels are shown in Table 3 as an example.

**Table 3**: Some messages in the dataset

| Sharing# | Sharing | Tag |
|---|---|---|
| 1 | rabbim kalan ömrünü geçen ömründen hayırlı eylesin<br>(may my lord make the rest of your life better than your past life) | Negative |
| 2 | bende biliyorum benden bı bok olmicak<br>(I know it won't be shit from me too) | Positive |
| 3 | dogruyu soyleyince kadro verince adalet yerini bulacak<br>(when you tell the truth and give a staff, justice will be served) | Negative |
| 4 | hava gavur şeyi gibi yanıyor diyorlar ama o konuda hiç tecrübem yok bilemiyorum<br>(they say the air burns like a gavur thing, but I don't know, I have no experience in that) | Positive |
| 5 | eğlenceli geceye devam<br>(keep up the fun night) | Negative |

The 3,000 Twitter messages, of which 1,500 are related to cyberbullying (cyberbullying=1) and 1,500 are not related to cyberbullying (cyberbullying=0).

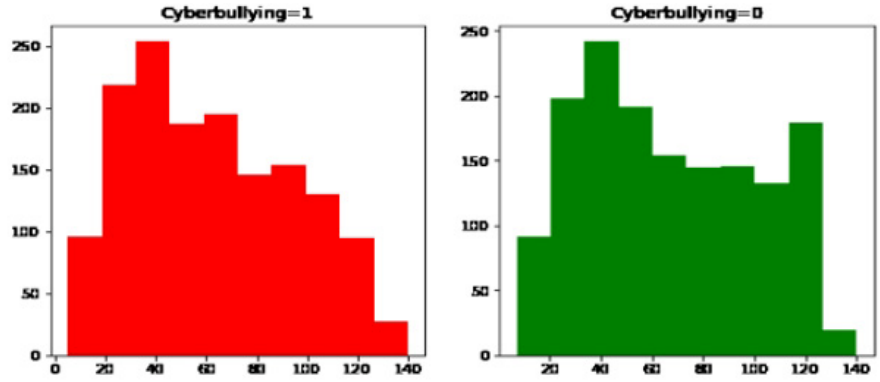


**Figure 4:** Characters' distribution in tweets

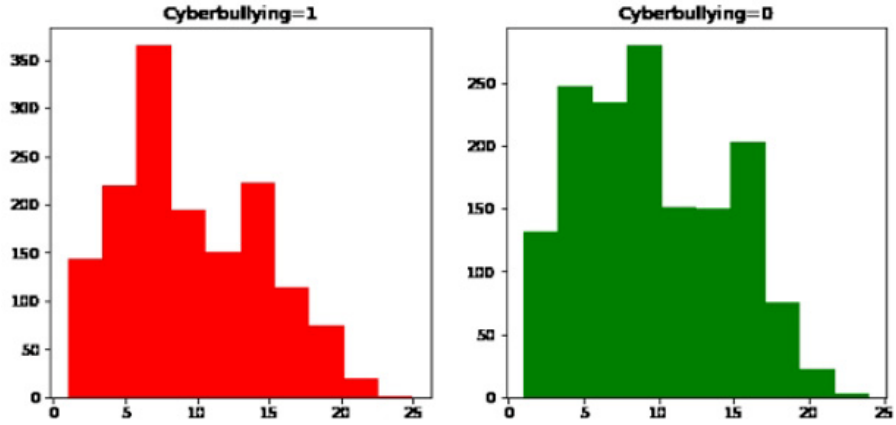


**Figure 5:** Words' distribution in tweets

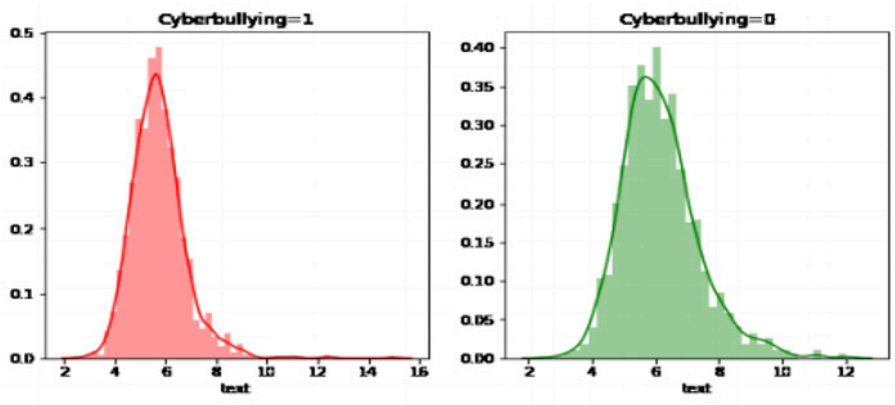


**Figure 6:** Average word length of each tweet

Also, classification with machine learning is used as it is important to have a balanced distribution and an even number of labels. An exploratory analysis of the word, character, and average word length of each tweet from Twitter with and without cyberbullying was performed, and the results are shown in Figures 4-6. As a result of the study, formal similarities were found between Turkish messages with and without cyberbullying.

### 7.2 Data Preprocessing

Before the dataset is processed by the written program, some preprocessing processes are done. To make these operations understandable, the pseudocode is shown in Figure 7.

```
Text Preprocessing (List of Cyberbullying sentences)
{
    Remove numerical characters, punctuation marks, web links
    from input.
    Apply lowercase conversation input.
    For each word in sentences
        Using the Turkish stemmer, words were separated into
        their roots zeyrek. Fixed incorrect words using
        MorphAnalyzer lemmatization.
    End loop
}
```

**Figure 7:** Pseudo code of the text preprocessing method

## 8 Results and Discussion

The computational results using CountVectorizer and Tf-Idf Vectorizer approaches are given in the tables 4 and 5. In the study, when CountVectorizer was used, Linear SVM model had the highest results with 88.35% accuracy and 99.96% F1-score, and Random Forest model had the lowest results with 86.36% accuracy and 99.21% F1-score (Fig. 8). When Tf-Idf Vectorizer was used, Linear SVM model had the highest results with 88.69% accuracy and 99.96% F1-score, and Decision Tree model had the lowest results with 85.69% accuracy and 99.16% F1-score (Fig. 9). Therefore, in the context of the classification developed for the detection of cyberbullying in Turkish texts:

- The F1-score of 99.96%, obtained with the Linear SVM algorithm, is the most successful result known in Turkish literature, according to studies using classical machine learning models.
- It can be said that these results are also successful when compared with the results of the research conducted in different languages.

**Table 4**: Evaluation Results of Machine Learning Models in Train datasets using CountVectorizer

| Models | F1-score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| Linear SVM | 99.96 | 100 | 99.92 | 88.35 |
| Multinomial Naive Bayes | 99.88 | 99.83 | 99.92 | 87.02 |
| Decision Tree | 99.96 | 100 | 99.92 | 87.19 |
| Random Forest | 99.21 | 99.66 | 98.75 | 86.36 |

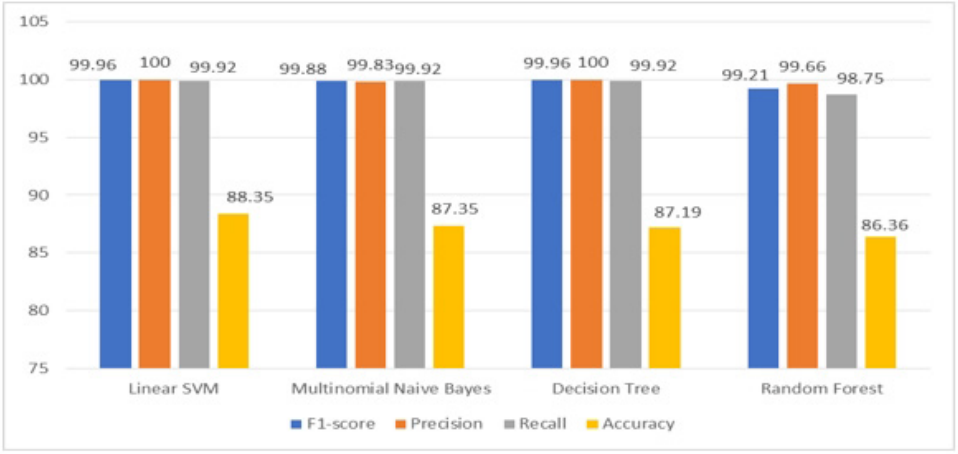


**Figure 8:** Evaluation Results graphics of Machine Learning Models in Train datasets using Count Vectoizer

**Table 5**: Evaluation Results of Machine Learning Models for Train datasets using Tf-Idf Vectorizer

| Models | F1-score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| Linear SVM | 99.96 | 100 | 99.92 | 88.69 |
| Multinomial Naive Bayes | 99.92 | 99.92 | 99.92 | 87.35 |
| Decision Tree | 99.16 | 99.16 | 98.42 | 85.69 |
| Random Forest | 99.16 | 99.66 | 98.67 | 86.86 |

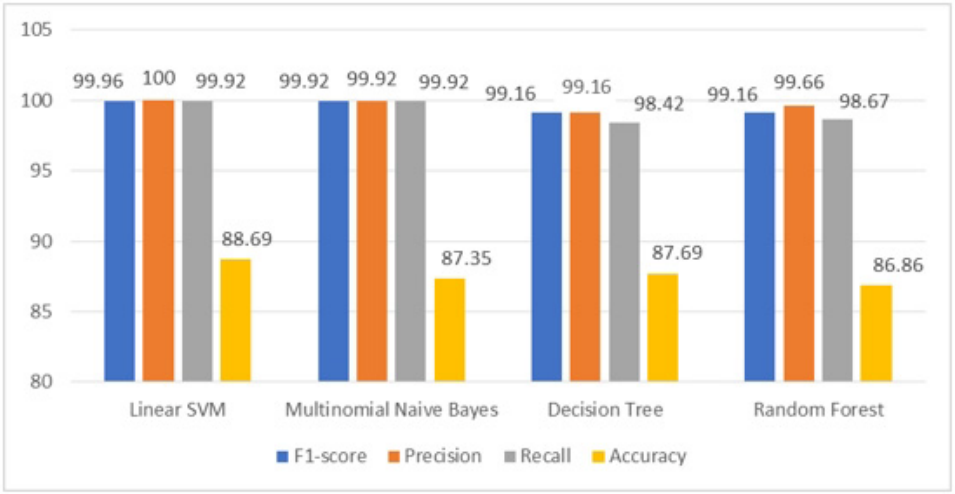


**Figure 9:** Evaluation Results graphics of Machine Learning Models for Train datasets using Tf-Idf Vectorizer

In the study Bozyiğit et al. (2021), three same and three different algorithms were used on similar data, the results are shown in detail in Table 6 and Figure 10. The results of the Linear SVM, Multinomial Naive Bayes, Random Forest and KNN algorithms used in the aforementioned study were compared with the results of this study. According to the comparisons, it has been observed that there are noticeable differences between the results obtained. The highest results in the article Bozyiğit et al. (2021) were found to be 85.4% F1-score and 86.8% accuracy in Linear SVM. It was observed that the values obtained according to the results of both articles were higher, especially the F1-scores. In this study, it is thought that the results are better

thanks to the n-gram, stop word, word of bag and text preprocessing techniques used.

**Table 6**: The detailed results of (Bozyiğit et al., 2021) with experimented classifiers. using Tf-Idf Vectorizer

| Models | F1-score | Precision | Recall | Accuracy |
|---|---|---|---|---|
| Linear SVM | 85.4 | 88.1 | 82.8 | 86.8 |
| Multinomial Naive Bayes | 78.4 | 72.8 | 85.3 | 83.2 |
| Random Forest | 85.2 | 85.7 | 82.0 | 85.2 |
| Logistic regression | 86.4 | 86.3 | 86.6 | 87.4 |
| AdaBoost | 87.6 | 89.0 | 83.6 | 87.6 |
| KNN | 75.8 | 87.6 | 66.8 | 80.2 |

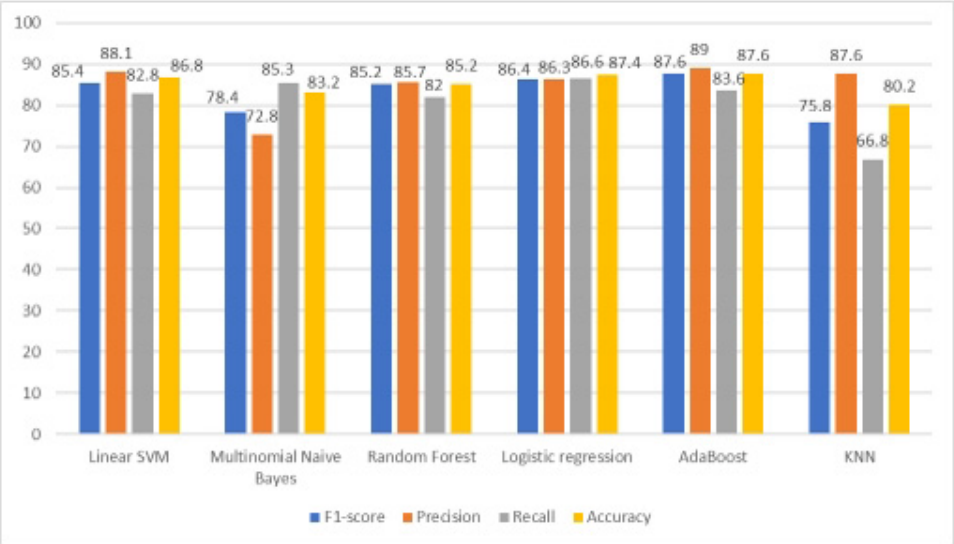


**Figure 10:** Graphical representation of Bozyiğit et al. (2021) results

## 9 Conclusion

Recently, our social life has helped people to express themselves on many issues. But unfortunately, some users do not use social media with cyberbullying-like actions in good faith, and they use some abusive and threatening words against users they know or do not know. For this, machine learning models were used in our study to prevent such malicious actions. In this study, considering Turkish social media messages such as twitter, four different machine learning algorithms were used and cyberbullying was detected. In this context, according to future studies, it is aimed to increase the performance of the recognition model for bad words used in the Azerbaijani language by including the shares made in Azerbaijani language from Youtube or Twitter, and to evaluate various machine learning methods at the same time.

## References

Aggarwal, C.C. (2018). Information retrieval and search engines. In *Machine Learning for Text* (pp. 259-304). Springer, Cham.

Agrawal, S., Awekar, A. (2018). Deep learning for detecting cyberbullying across multiple social media platforms. In *European conference on information retrieval* (pp. 141-153). Springer, Cham.

Aind, A.T., Ramnaney, A., & Sethia, D. (2020). Q-bully: a reinforcement learning based cyberbullying detection framework. In 2*020 International Conference for Emerging Technology* (INCET) (pp. 1-6). IEEE.

Akca, E.B., & Sayımer, I. (2017). Cyberbullying, it's tyeps and related factors: an evaluation through the existing studies. *AJIT*, *8*(30), 7.

Altay, E.V., Alatas, B. (2018). Detection of Cyberbullying in Social Networks Using Machine Learning Methods. *2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism* (IBIGDELFT), 87-91.

Arroyo-Fernández, I., Forest, D., Torres-Moreno, J. M., Carrasco-Ruiz, M., Legeleux, T., & Joannette, K. (2018). Cyberbullying detection task: the ebsi-lia-unam system (elu) at coling'18 trac-1. In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying* (TRAC-2018) (pp. 140-149).

Aslan, A., Doğan, B.Ö. (2017). Online violence: the case of "potinss" as a cyber-bullying space. M*armara Journal of Communication,* 27, 95-119 (in Turkish).

Bozyiğit, A., Utku, S., & Nasiboğlu, E. (2018). Detection of social media messages containing cyberbullying. In 3*rd International Conference on Computer Sciences and Engineering* UBMK (in Turkish).

Bozyiğit, A., Utku, S., & Nasiboğlu, E. (2019). Cyberbullying detection by using artificial neural network models. In *2019 4th International Conference on Computer Science and Engineering* (UBMK) (pp. 520-524). IEEE.

Bozyiğit, A., Utku, S., & Nasibov, E. (2021). Cyberbullying detection: Utilizing social media features. *Expert Systems with Applications*, 179, 115001.

Di Capua, M., Di Nardo, E., & Petrosino, A. (2016). Unsupervised cyber bullying detection in social networks. In *2016 23rd International Conference on Pattern Recognition* (ICPR) (pp. 432-437). IEEE.

Eroğlu, Y. (2014). Holistic Model Determining Risk Factors Which Predict Cyber Bullying and Victimization in Adolescents (Doctoral dissertation, Bursa Uludag University (Turkey)).

Eroğlu, Y., & Güler, N. (2015). Contingent self-worth, risky internet behavior and cyberbullying/ examining the relationship between victimization. *Sakarya University Journal of Education, 5*(3), 118-129.

Hadžiosmanoviç, D., Simionato, L., Bolzoni, D., Zambon, E., & Etalle, S. (2012). N-gram against the machine: On the feasibility of the n-gram network analysis for binary protocols. In *International Workshop on Recent Advances in Intrusion Detection* (pp. 354-373). Springer, Berlin, Heidelberg.

Haidar, B., Chamoun, M., & Serhrouchni, A. (2019). Arabic cyberbullying detection: Enhancing performance by using ensemble machine learning. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)* (pp. 323-327). IEEE.

Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E., & Mohammed, A. (2019). Social media cyberbullying detection using machine learning. *Int. J. Adv. Comput. Sci. Appl, 10*(5), 703-707.

Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). *Automated Machine Learning: Methods, Systems, Challenges* (p. 219). Springer Nature.

Isa, S.M., Ashianti, L. (2017). Cyberbullying classification using text mining. *1st International Conference on Informatics and Computational Sciences* (ICICoS) 241-246.

Karcioğlu, A.A., & Aydin, T. (2019). Sentiment analysis of Turkish and english twitter feeds using Word2Vec model. In *2019 27th Signal Processing and Communications Applications Conference* (SIU) (pp. 1-4). IEEE.

Kumari, K., Singh, J.P., Dwivedi, Y.K., & Rana, N.P. (2021). Multi-modal aggression identification using convolutional neural network and binary particle swarm optimization. *Future Generation Computer Systems*, 118, 187-197.

Novalita, N., Herdiani, A., Lukmana, I., & Puspandari, D. (2019). Cyberbullying identification on twitter using random forest classifier. *Journal of Physics: Conference Series*, *1192*(1), 012029.

Öztük, C. (2017). Examining the Cyberbullying Tendencies of 8th Grade Students in Terms of Various Variables: The Case of Adapazari, Sakarya Province. *Journal of Interdisciplinary Education: Theory and Practice, 1*(1), 42-58.

Paul, S., Saha, S. (2020). CyberBERT: BERT for cyberbullying identification. *Multimedia Systems*, 1-8.

Pawar, R., Raje, R.R. (2019). Multilingual cyberbullying detection system. In *2019 IEEE International Conference on Electro Information Technology* (EIT) (pp. 040-044). IEEE.

Pervan, N., Keleş, Y. (2019). *Making Semantic Inferences from Turkish Texts Using Deep Learning Approaches*. Ankara Universitesi, Ankara, Turkey.

Polat, Z.D., Bayraktar, S. (2016). Cyber Bullying and Cyber Victimization in Adolescents. *Mediterranean Journal of Humanities, 5*(1), 115-132.

Rachid, B.A., Azza, H., & Ghezala, H.H.B. (2020). Classification of cyberbullying text in arabic. In *2020 International Joint Conference on Neural Networks* (IJCNN) (pp. 1-7). IEEE.

Ravichandran, K., Arulchelvan, S. (2017). The model of multilayer perceptron analysed the crime news awareness in India (quantitative analysis method). In *2017 4th International Conference on Advanced Computing and Communication Systems* (ICACCS) (pp. 1-6). IEEE.

Reynolds, K., Kontostathis, A., & Edwards, L. (2011). Using machine learning to detect cyberbullying. In *2011 10th International Conference on Machine Learning and Applications and Workshops* (Vol. 2, pp. 241-244). IEEE.

Rezvani, N., Beheshti, A., & Tabebordbar, A. (2020). Linking textual and contextual features for intelligent cyberbullying detection in social media. In *Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia* (pp. 3-10).

Rosa, H., Pereira, N., Ribeiro, R., Ferreira, P.C., Carvalho, J.P., Oliveira, S., ... & Trancoso, I. (2019). Automatic cyberbullying detection: A systematic review. *Computers in Human Behavior*, 93, 333-345.

Sarkar, D. (2019). *Text analytics with Python: a practitioner's guide to natural language processing*. Bangalore: Apress.

Shukan, A., Abdizhami, A., Ospanova, G., & Abdakimova, D. (2019). Issues of information technology crime control in the republic of Turkey. *Informatologia, 52*(1-2), 65-73.

Song, T.M., & Song, J. (2021). Prediction of risk factors of cyberbullying-related words in Korea: Application of data mining using social big data. *Telematics and Informatics*, 58, 101524.

Soni, D., Singh, V.K. (2018). See no evil, hear no evil: Audio-visual-textual cyberbullying detection. *Proceedings of the ACM on Human-Computer Interaction, 2*(CSCW), 1-26.

Sugandhi, R., Pande, A., Agrawal, A., & Bhagat, H. (2016). Automatic monitoring and prevention of cyberbullying. *International Journal of Computer Applications*, 8, 17-19.

Ünver, H., Zihni, K.O.Ç. (2017). Examining the relationship between cyberbullying, problematic internet use and risky internet behavior. *Turkish Journal of Educational Sciences, 15*(2), 117-140.

Weru, T., Sevilla, J., Olukuru, J., Mutegi, L., & Mberi, T. (2017). Cyber-smart children, cyber-safe teenagers: Enhancing internet safety for children. In *2017 IST-Africa Week Conference* (IST-Africa) (pp. 1-8). IEEE.

Yazğılı, E., Baykara, M. (2021). Cyberbullying detection methods potential application areas and challenges. *Dicle University Engineering Faculty Journal of Engineering, 12*(1), 23-35.

Yuvaraj, N., Srihari, K., Dhiman, G., Somasundaram, K., Sharma, A., Rajeskannan, S., ... & Masud, M. (2021). Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking. *Mathematical Problems in Engineering*, 2021.

Zhao, Z., Gao, M., Luo, F., Zhang, Y., & Xiong, Q. (2020). LSHWE: improving similarity-based word embedding with locality sensitive hashing for cyberbullying detection. In *2020 International Joint Conference on Neural Networks* (IJCNN) (pp. 1-8). IEEE.