Complexity Matrices in Twitter Sentiment Analysis of Thoughts on Mobile Games Using Machine Learning Algorithms

Erol KINA[1*], Recep ÖZDAĞ[2]

Van Yüzüncü Yıl University, Özalp Vocational School, Van, Turkey, ORCID: 0000-0002-7785-646X

Van Yüzüncü Yıl University, Department of Computer Engineering, Van, Turkey, ORCID: 0000-0001-5247-5591

Corresponding Author: erolkina@yyu.edu.tr

Geliş tarihi:13.12.2021 Kabul tarihi:29.12.2021

Abstract

In modern times, people have started sharing their opinions, thoughts, and feelings with other people through social media. The increasing number of social media users and their shares in social media platforms has naturally drawn the attention of researchers to this field. Twitter is one of the leading data sources in this field. Since Twitter has millions of users from different cultures and classes, it is possible to collect comments in different languages and content. Tweets that people write and share in 280 characters are used for research and analysis. Because not all tweets can be read by people, in this study, sentiment analysis was performed using Naive Bayes (NB) classification algorithm and multilayer artificial neural networks (ML-ANN) based on the content of comments on mobile games. As a result of the analysis, it was found that multilayer artificial neural networks gave better results than the other methods on both training and test data.

Keywords*:* Mobile Games, Sentiment analysis, Twitter, Naive Bayes, Artificial Neural Networks

I. INTRODUCTION

Today people spend a lot of their time on the internet and social media. They can reach a large mass by sharing their feelings and ideas through social media. Personal likes and interests, many comments and shares on certain topics allow us to get a picture of the general mentality and attitude of the society we live in [1]. Sentiment analysis plays an important role in providing richer and more accurate results by allowing shy and introverted people, who are better at expressing themselves on social media platforms, to share their ideas with others. Sentiment analysis on Twitter consists of the steps of organizing, parsing, analyzing, and reporting the tweets sent by users [2]. Certain meanings are extracted and presented to companies or individuals through a special user interface from these results.

Sentiment analysis studies determine whether the data have positive, negative or neutral content. In this study, to process data with machine learning algorithms, training sets are created by splitting the data into positive-negative. The aim of Sentiment analysis is being a part of data science that automatically handles the process of understanding and analyzing textual data as wanted. Thus, certain results are obtained. These results are used to determine people's ideas and thoughts about the subject of the study. Sentiment analysis studies can even be used to determine the general opinion not only of individuals but also of a particular audience. Thus, sentiment analysis can be used as a guide by showing the reaction to a decision to be made for a particular audience and the viewpoint of mobile game players on a particular game based on previous comments. Machine learning is necessary because it is not possible to analyze a large number of data, positive or negative opinions, large data sets one by one and come to a conclusion. Fig. 1 shows the basic machine learning process. First, the data set is obtained. Afterward, the data is passed through the preprocessing stages. The model is divided into a training dataset and a test dataset. The model is developed according to the results obtained. The accuracy rate for the training dataset obtained during the model evaluation process should be high.

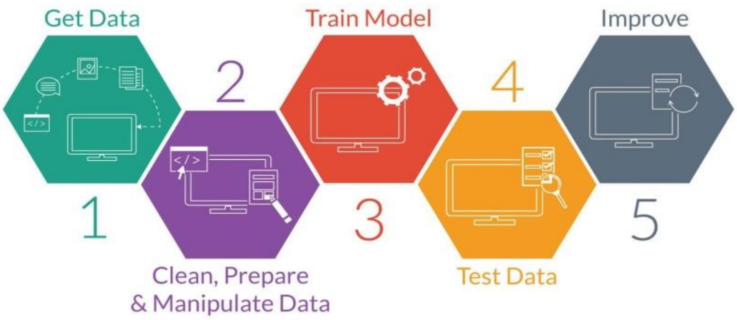

Fig.1. The basic machine learning process

Complexity matrices were used for experimental analysis. Accuracy, Sensitivity (Recall), Specificity, Precision, F-Score metrics were used to calculate complexity matrices and these metrics used to compare classification success rate. ML-ANN and NB machine learning algorithms were preferred in the study. This preference was made because they are suitable for data sets with positive and negative discrimination. For all classification algorithms, 75% of the total data was used to create the training set and 25% was used to create the test data set. 2 hidden neurons were used for the output layer and 10 neurons for the middle layer, and the 3-layer MLP model was preferred.

Sentiment analysis is trying to solve the text classification problem. When machine learning methods are used in text classification, some of problems appears. Dialect differences, slang words, inverted sentences, allusions in social media comments are some of these difficulties. The TextBlob library was used because of the difficulty in correctly defining irony and sarcasm in texts. Applying sentiment analysis studies, which are applied in almost every field, to the rapidly growing mobile games industry; It is inevitable that software companies producing mobile games will benefit in their strategy and planning. After all, considering how difficult it is to reach every mobile gamer and listen to their complaints and needs, drawing conclusions from the numerous comments on mobile games on Twitter will be a great advantage in a short period of time.

Looking at the studies in the field of sentiment analysis, it can be said that English texts in most studies are examined as positive and negative in two separate classes. Whether the texts have a positive, negative, or neutral meaning is analyzed with different queries. As a result of this analysis, the opinion of the target audience about the study on the topic under study is estimated. In this regard, sentiment analysis studies, preparatory market research on any product to be released for the first time for organizations can evaluate how a decision to be made regarding a large group or individuals, the masses who will play mobile games will react positively or negatively according to the previous comments. The study can be a guide to influence the decision to play a mobile game.

The next sections of the article are planned as follows. Section 2 presents the literature review. Section 3 presents the system modeling of the study and Section 4 presents the methods used NB and ANN. The results are presented and discussed in Section 5 and the final results are explained in Section 6.

II. RELATED WORK

There are many articles in the literature on sentiment analysis research. One of the earliest examples of sentiment analysis is the work of Yi et al [3]. This study, it is clear that it provides ideas for future studies by defining sentiment analysis. The diversification of studies on sentiment analysis gained momentum after this study.

As in many other languages, some sentiment analysis studies have been conducted for Turkish on Twitter. Kaya et al [4] gave an example of Turkish politics. In this study, different machine learning methods were used and their performances were compared. Another Study by Çoban and Özyer [5] deals with the classification of Turkish tweets using LDA (Latent Dirichlet Allocation). In their study, they used the NLP tool Zemberek for language processing [6]. Zemberek is a free, open-source Natural Language Processing framework that can be used for spell checking, inference, word suggestions, etc. Language processing and feature selection are provided with Zemberek.

Pennacchiatti and Popescu [7] classified social media users based on Twitter data in their study. Based on the content of the tweets, they used machine learning methods to determine people's political views and ethnicity by guessing. They examined three different topics such as political affiliation, ethnicity, and the business field they were interested in and found that their results showed differences.

Katz et al [8] proposed the ConSent model, a content-based model of sentiment analysis that differs from other studies. In this model, the keywords in the training set are identified and analyzed. Based on the identified keywords, discriminative features are created and the terms and keywords of each document are examined. Based on the obtained results, the features are extracted and the document is classified. This model was compared with Naive Bayes (NB) and Support Vector Machines (SVM) methods. A total of 2000 data with 24930 positive and 5070 negative polarizations were downloaded from the TripAdvisor website and 30000 comments with negative polarity and 1000 negative and 1000 positive polarizations were downloaded from the IMDB platform. When examining the results, it can be seen that although this new model provides similar results to SVM, the success rate is much lower. Another result of the study is that the ConSent model has a lower success rate for datasets consisting of short texts than for datasets consisting of long texts.

In their sentiment analysis study, Nikfarjam et al [9] investigated sick people's negative thoughts about the side effects of taking medications by examining comments posted on health forums and comments downloaded from the Twitter environment. The application used the ADRMine method. ADRMine, a machine learning-based method, was used to create a dataset of 6279 comments from the health site DailyStrength and 1784 comments downloaded from Twitter. The results show that the ADRMine method applied achieves 82.1% performance. It was found to achieve a much better success rate than the SVM and MetaMap methods.

Nizam and Akın [10] investigated how the distribution of data across classes affects the success rate of the classification algorithm, depending on whether they are balanced or unbalanced. 2 separate datasets were used, which were taken from Twitter comments of different product groups from companies in the food industry. For two different datasets, the data were categorized as positive, negative, and neutral. For the unbalanced dataset, a total of 2000 Twitter comments were examined, consisting of 1113 positive, 277 negatives, and 610 neutral data. For the balanced dataset, a total of 824 Twitter comments were examined, consisting of 257 positive, 277 negatives, and 299 neutral data. 2 datasets using the classification algorithms Random Forest (RF), NB, SVM, and k-nearest neighbors, decision tree (J48) Precision, Accuracy, Sensitivity (Recall)) and the Kappa coefficient was evaluated based on the results obtained. After this evaluation, the statistical results were examined using the performance results and the Kappa coefficient and it was found that the balanced dataset was more powerful than the unbalanced dataset. The SVM classification algorithm achieved the best performance with an accuracy of 72.33%.

Turkmen et al [11] used the SVM and RF -classifier approach to predict the political orientation of individuals based on tweets posted during the Gezi Park protests. A total of 1351 Twitter messages were classified into three groups and determined as "pro-protest", "anti-protest" and "neutral" messages. Furthermore, the chi-square method was used to determine the characteristics. In addition to the criterion of performance accuracy, they determined the F-measure by calculating precision and sensitivity. While they achieved more efficient results in estimating whether individuals have political connections or not using the SVM algorithm, they achieved more than 80% success in estimating individuals' political tendencies using the RF algorithm.

Literature studies on sentiment analysis have been conducted in many areas, such as determining customer opinions about a product [12,13], predicting election results [14], evaluating movie and book reviews [15], and some other types of data analysis [16].

In their study, Topal and Uçar [17] analyzed TripAdvisor data using artificial neural networks to determine the preference potential of Chinese consumers in Turkey. They used a scaled conjugate gradient (SCG) as the training algorithm. The data of 544 cities and provinces visited by Chinese citizens in their home countries were used. Each city is a characteristic of tourists, which created a "zero matrix" and a "frequency matrix". After this step, the F criterion was applied step by step and the classification was done with MLFFNN.

Olgun and Özdemir [18] used NB and ANN classifiers to work on pattern recognition in control charts and proposed SCG as a learning algorithm because it gives faster results [19]. When they compared the classification performance for ANN and NB with other classifiers using raw data, they found that they achieved higher recognition performance with ANN. Our aim in this study is to solicit positive or negative opinions about any game genre to get ideas about mobile game developers and people's attitudes towards mobile games and to create a roadmap for the future. For this purpose, the following methods were used to preprocess the data based on the downloaded tweets:

Linguistic labeling (noun, verb, adjective, etc.), n-Grams (number of repetitions), reduction to word stems (stemming and lemmatization), extraction of ineffective words (stop words), removal of negative words that generate positive meaning (treatment of negations and but phrases). Unlike the literature, the results were classified as negative or positive by using the library "TextBlob" to calculate the polarity. In addition, the dataset was split into 75% training and 25% test datasets using the algorithm NB, and a model was built using the library Scikit-learn and the class GaussianNB to make predictions for the test data. By comparing the target variables of the obtained training set with the test data, intermediate processes were applied to the data between the input layer and the output layer in the hidden layer by multilayer artificial neural networks (ML-ANN).

III. SYSTEM MODEL

To process data with machine learning, training sets are needed, which are created by splitting the data into positive-negative or positive-negative-neutral. Also, this data is preprocessed to make it suitable for classification. Preprocessing is done by steps such as string fragmentation, removing stop words, searching for word stems, computing term frequencies, and inverting document frequencies. String fragmentation involves breaking up data into chunks to obtain meaningful information. Stop words are lists of words that do not make sense on their own.

In this study, the list of stop words for Turkish generated by the Lucene application was used. If there is more than one repeating letter for each word, the dimensions are reduced by normalization to reduce these repetitions to one. For example, the word "gaaaaaame" was changed to "game". Also, the increase in the length of tweets from 140 characters to 280 characters has led to a decrease in the use of abbreviations. While people used to write "Pls" instead of "Please" or "Thx" instead of "Thanks" when characters were limited, this usage has decreased with the transition to 280 characters, and so words are used long and correctly. Zemberek, a Turkish library, was preferred to convert words from their attached form to the root word.

The criteria used when comparing classification algorithms with each other; the model is modeled under 3 headings: Performance Criteria, Term Frequency, and Reverse Document Frequency.

A. Model Performance Criteria

The criteria used to compare classification performance are listed below.

Precision: indicates the degree of precision of the result obtained for the classifier. The ratio between the number of positively identified and labeled samples and the total number of positively classified samples is calculated using the following equation.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

Recall: The ratio of positively marked samples to the total number of true positive samples is calculated using the following equation.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

Accuracy: It is a measure often used in classification procedures. The ratio of correctly classified and separated samples to the total number of samples is calculated using the following equation.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\varepsilon} \tag{3}$$

The abbreviations used in Eqs. (1) to (3) are explained below.

TP (True Positive Rate) refers to comments that are actually positive and correctly classified as positive by the classifier.

FP (False Positive Rate) refers to comments that are positive but were not classified as positive by the classifier.

TN (True Negative Rate) refers to comments that are negative and were correctly classified as negative by the classifier.

FN (False Negative Rate) refers to comments that are negative but have not been classified as negative by the classifier.

The ε symbol denotes the sums TP, FP, TN, and FN.

F-measure: Calculated using the values of the sensitivity and precision measures. It ensures that the system is correctly optimized in terms of sensitivity or precision and is given by equation (4).

$$\mathrm{F-measure} = 2 * \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{4}$$

Using model performance criteria, annotations containing text are converted into numerical matrices that machine learning algorithms can work with. The method used for this conversion, called CountVectorizer, involves an operation called tokenization, which decomposes sentences into a series of tokens. The CountVectorizer operation converts documents that contain text into numeric matrices. This creates a sparse matrix representing all the words in the document (see Fig. 2) [20].

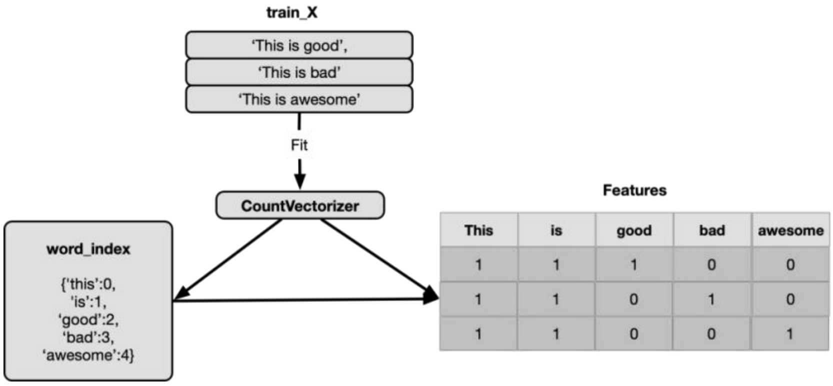

Fig.2. CountVectorizer Transformation

B. Term Frequency

This method is used to calculate the weight of terms in a document. Term frequency (TF) is calculated by dividing the number of occurrences of the term in the document by the total number of words.

C. Reverse Document Frequency

To determine if the word is a term (conjunction, etc.), the number of occurrences of the word in more than one document is calculated using the following equation and the Inverse Document Frequency (IDF) is determined.

$$tf_{ij} * \mathrm{Idf}_{ij} = f_{ij} * \log\left(\frac{N}{df_i}\right) \tag{5}$$

In the given equation, tf_{ij} indicates the number of times term *i* occurs in document *j*, Idf_i indicates the number of documents in which term *i* occurs at least once, and *N* represents the total number of documents.

For example, if the word "And" occurs in every document under study, it is calculated as log (10/10) = 0. If the word "And" is mentioned in only one document, it is calculated as log (1/10) = 1. If the word "And" occurs 1 time in 100 documents, the IDF value is calculated as 2. Thus, if the number of documents in which the term occurs repeatedly decreases, the IDF value increases inversely [21].

The calculation of reverse document frequency is shown in equation (5).

IV. MATERIAL AND METHOD

In this study, the classification algorithms NB and ML-ANN, machine learning methods, were used to evaluate the performance of machine learning models performing classification using performance measures such as precision, accuracy, and F-measure.

A. Naive Bayes

Bayes theorem is a probability theory that shows the relationship between conditional probabilities and exceptional probabilities within the probability distribution for a random variable. In turn, NB is one of the machine learning algorithms commonly used in the field of text classification, created based on Bayes' theorem. For example, the method NB can be used to find out whether a particular attribute of a record is included in which class value. NB algorithms are widely used in sentiment analysis and spam filtering and recommendation systems. It is known as a lazy learning algorithm. It is an algorithm that works with datasets that are unstable in terms of data variety. The working logic of the algorithm is explained by the fact that it calculates the probability of each case for an element and classifies it according to the one with the highest probability value. It is the algorithm that can achieve accurate results even with a small amount of data in the training dataset [9].

NB is mathematically modeled by the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{6}$$

In the given equation, P(A|B), P(B|A), P(A), and P(B) give respectively the probability of A occurring if B is known to be true, and the probability of B occurring if A is known to be true. They indicate the probability of A and B occurring [22].

B. Multilayer Artificial Networks

ANN is a supervised machine learning method that mimics the working structure of the human brain and nervous system and aims to learn. Our brain is made up of special information-transmitting cells called neurons. Neurons connect to communicate with each other. In this way, a process called learning takes place. This coupling is divided into two types: forward and backward [23].

Feedforward artificial neural networks: in feedforward ANN the arrangement of neurons starts at the input and progresses in regular layers to the output. One layer is only connected to the next layer. The information received at ANN is processed by the intermediate layers and then by the output layer, just after the input layer, and the output of the network is provided. The multilayer feedforward network ANN is shown graphically in Fig. 3.

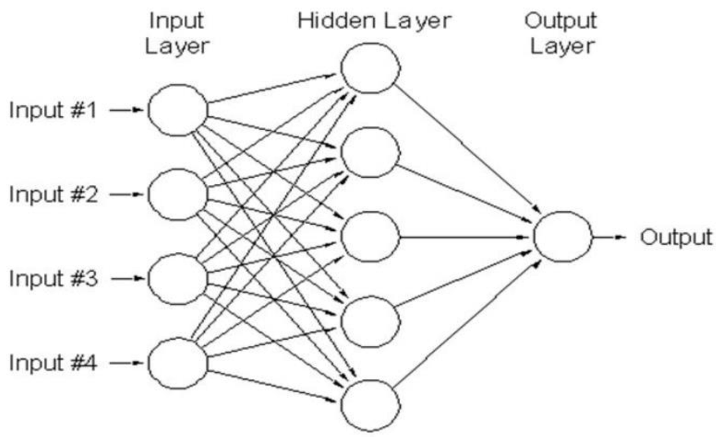

Fig.3. Multilayer Feedforward ANN

Artificial Neural Networks with Feedback: In the feedforward layer ANN, the output of a neuron serves only as input to the following neuron layer, while in the feedback layer ANN it can serve as input to any neuron before it or in its layer. Therefore, the feedback ANN has a dynamic form that does not contain linearity. Depending on how the connections providing the feedback feature are connected to the neurons, feedback ANNs with different behavior and structure can be obtained from the existing ANNs.

C. Recommended Method

In this study, 75% of the available dataset is reserved for the training set and 25% for the test set. How this was coded in the Python environment can be seen in Fig. 4. Also, the model was created using the scikit-learn library and the classifier object in the GaussianNB class in the NB module. The Python coding used for this model is shown in Fig. 5. To compare the test data with Python, the dataset obtained by estimation (y_pred) was compared with the dataset reserved for testing (X_test). The obtained result was compared with our test dataset and the accuracy of the data was measured. The coding used to compare the data is shown in Fig. 6.

```
from sklearn.cross_validation import train_test_split
X_train, X_test, y_train, y_test = train_test_split
(X, y, test_size = 0.25, random_state = 0)
```

Fig.4. Separation of data set with Python code

```
from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)
```

Fig.5. Model creation with Python code

```
@author: erolk
"""
y_pred = classifier.predict(X_test)
```

Fig.6. Comparison of data with Python code

In this study, as many neurons were included as class types were used ANN. The factors that determine the quality of the data to be trained are the number of hidden layers and the number of unknown neurons in that layer. The input layer and the output layer are in the hidden layer. The data between these two layers is subjected to intermediate processing. Before training and classifying the obtained data set, the preprocessing steps were completed. During the preprocessing of the data, stop words and words with less than three letters were removed. For all classification algorithms, 75% of the total data was used to create the training set and 25% was used to create the test data set. 2 hidden neurons were used for the output layer and 10 neurons for the middle layer, and the 3-layer MLP model was preferred.

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, the simulations were performed in Matlab and Python environments. In Turkey, comments on mobile games based on tweets sent by users in 2019-2020 were used. The dataset was used to try to distinguish between positive and negative emotions. The Turkish dataset was used, of which 13000 were classified into positive classes and 13000 into negative classes. Using the NLTK library prepared in a Python environment, the datasets were prepared for the training and classification processes. The preprocessing steps mentioned in the introduction were performed. After the preprocessing steps, we obtained a dataset consisting of 20868 tweets, 10323 positive, and 10545 negatives. The total dataset used was reduced to 20000 tweets, 10000 negatives, and 10000 positives. The distribution of tweets that make up the dataset can be found in Table 1.

TABLE I. Distribution Of Downloaded And Used Tweets According To Polarity Calculations

Tweet Definition	Number	Distribution
Downloaded	26000	Positive + Negative + Neutral
Used	20000	Positive + Negative
Used Negative	10000	Negative
Used Positive	10000	Positive
Allocated for the Training Set	15000	Positive + Negative
Reserved for the Test Set	5000	Positive + Negative

The cross-complexity matrices obtained with the values "TP", "FP", "TN", and "FN" depend on the dataset reserved for training, the dataset reserved for testing, and the whole dataset according to the classification methods NB and ML-ANN used in Table 2 are shown. The P-value explains the number of positive tweet comments and the "N" value explains the number of negative tweet comments.

From Table 2, it can be seen that the predictions of the machine learning algorithm ANN are more accurate than those of the machine learning algorithm NB in terms of true negative and true positive values for the training set. It was found that the machine learning algorithm NB gives more accurate results for the true-negative prediction for the test data and the machine learning algorithm ANN for the true-positive prediction. For false-negative and false-positive detection, the machine learning algorithm NB provided more false results than the machine learning algorithm ANN in both the test and training datasets.

TABLE 2. Observed Complexity Matrices In Terms Of Training And Test Data For Nb And Ml-Ann As A Result Of Classification

		Training		**Test**		**All**	
		N	**P**	**N**	**P**	**N**	**P**
NB	**N**	4630	1925	2280	1010	8520	4655
	P	2750	5695	340	1370	1480	5345
ML-ANN	**N**	6530	1100	1870	632	8346	1695
	P	930	6440	670	1828	1654	8305

It was found that the machine learning algorithm ANN and the machine learning algorithm NB provide overall similar accuracy results in estimating the accuracy of true negative values, but the machine learning algorithm ANN provides significantly more accurate results in estimating the accuracy of true positive values.

Considering the performance criteria in Table 3, the best performance was obtained in classification using ANN. Comparing the values for accuracy, precision, and sensitivity, which are part of the performance criteria of the model used to determine the result values, it can be seen that the values for accuracy and precision for the machine learning algorithm ANN provide significantly more accurate results than the machine learning algorithm NB and both machine learning algorithms provide similar accuracy for the sensitivity values.

The results obtained using the F-measure were very accurate for the test data, while the machine learning algorithm ANN produced much more accurate results for the training dataset. Since 1 is taken as the ideal value for the AUC, it can be seen that the machine learning algorithm ANN covers a much wider range than the covered range and has a high success rate of 86.20%. The average success of the machine learning algorithm NB was calculated to be 69.25%. It can be said that the success rates of the two machine learning algorithms for the test data sets are close to each other.

TABLE 3. F-Measure (F) And Auc Result Values According To Accuracy (A), Precision (P), And Recall (R) Results

	Method	**A**	**P**	**R**	**F**	**AUC**
Training Data	NB	0,6947	0,6542	0,8241	0,7105	0,6953
	ML-ANN	0,8982	0,8981	0,8951	0,8956	0,8982
Test Data	NB	0,6812	0,6441	0,8312	0,7241	0,6794
	ML-ANN	0,7532	0,7354	0,7952	0,7638	0,7527
All Data	NB	0,6925	0,6480	0,8350	0,7308	0,6925
	ML-ANN	0,8620	0,8561	0,8697	0,8629	0,8620

VI. CONCLUSION

In this study, an analysis of sentiment and thought was performed by applying the algorithms NB and ML-ANN to classify according to the content of comments on mobile games. There are two major limitations in this study that could be addressed in future research. First, the study focused on only 2 machine learning algorithms, second, we used A, P, R, F and AUC metrics for Training Data, Test Data and All Data sets. Voting classifier models and hybrid models may give better results. ROC curves can be drawn with the metrics used in the study. When evaluating the training dataset, ML-ANN showed much better success with an 89.82% correct classification rate, while it was even more successful with a 75.27% correct classification rate on the test data. Considering the whole dataset, since 1 is accepted as the ideal value for the AUC, we can see that the machine learning algorithm ML-ANN covers a much wider range and achieves a high success rate of 86.20%. The best classification in terms of performance was obtained with ML-ANN. In a similar study by Basarslan and Kayaalp [24], they obtained 4500 Tweets by using the Twitter API. Tweets consist of English tweets about health in 2019. They used NB and ML-ANN machine learning algorithms. They obtained an A value of 0.72 for NB, and an A value of 0.86 for ML-ANN. At the same time, they obtained a P-value of 0.73 and an F value of 0.76 for NB, while a P value of 0.87 and an F value of 0.86 for ML-ANN. The low number of tweets they used in their studies and the use of a tagged data set increased their success.

The results that can be obtained by working with a larger dataset or by increasing the percentage of test data can provide more efficient results by using different algorithms. It is planned to obtain results with higher accuracy in future studies by using different classification algorithms.

REFERENCES

1. Go, A., Huang, L., & Bhayani, R., Twitter sentiment analysis. Entropy, 17, 252-269,2009.
2. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113, 2014.
3. Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W., Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Third IEEE international conference on data mining* (pp. 427-434), 2003.
4. Kaya, M., Fidan, G., & Toroslu, I. H., Sentiment analysis of Turkish political news. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology* (Vol. 1, pp. 174-180), 2012.
5. Çoban, Ö., & Özyer, G. T., Sentiment classification for Turkish Twitter feeds using LDA. In *2016 24th Signal Processing and Communication Application Conference (SIU)* (pp. 129-132), 2016.
6. Akın, A. A., & Akın, M. D. , Zemberek, an open source nlp framework for turkic languages. *Structure*, *10*(2007), 1-5, 2007.
7. Pennacchiotti, M., & Popescu, A. M. , A machine learning approach to twitter user classification. In *Fifth international AAAI conference on weblogs and social media*, 2011.
8. Katz, G., Ofek, N., & Shapira, B., ConSent: Context-based sentiment analysis. *Knowledge-Based Systems*, *84*, 162-178, 2015.
9. Nikfarjam, A., Sarker, A., O'connor, K., Ginn, R., & Gonzalez, G., Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, *22*(3), 671-681, 2015.
10. Nizam, H., & Akın, S. S., Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması. *XIX. Türkiye'de İnternet Konferansı*, 1-6, 2014.
11. Türkmen, A. C., & Cemgil, A. T., Political interest and tendency prediction from microblog data. In *2014 22nd Signal Processing and Communications Applications Conference (SIU)* (pp. 1327-1330), 2014.
12. Wang, W., Sentiment analysis of online product reviews with Semi-supervised topic sentiment mixture model. In *2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery,* Vol. 5, pp. 2385-2389, 2010.
13. Santos, R. L. D. S., de Sousa, R. F., Rabelo, R. A., & Moura, R. S., An experimental study based on fuzzy systems and artificial neural networks to estimate the importance of reviews about product and services. In *2016 International Joint Conference on Neural Networks (IJCNN),* pp. 647-653, 2016.
14. Anjaria, M., & Guddeti, R. M. R., Influence factor-based opinion mining of Twitter data using supervised learning. In *2014 Sixth International Conference on Communication Systems and Networks (COMSNETS),* pp. 1-8, 2014.
15. Neethu, M. S., & Rajasree, R., Sentiment analysis in twitter using machine learning techniques. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT),* pp. 1-5, 2013.
16. Nurwidyantoro, A., Sentiment analysis of economic news in Bahasa Indonesia using majority vote classifier. In *2016 International Conference on Data and Software Engineering (ICoDSE),* pp. 1-6, 2016.
17. Uçar, M. K., & Topal, İ. Yapay Sinir Ağları ile Reklam Sektöründe Kullanıcı Profili Çıkarma Uygulaması: Çin–Türkiye Örneği. *Veri Bilimi*, *1*(1), 20-28, 2018.
18. Olgun, M., & Özdemir, G., Istatistiksel Özellik Temelli Bayes Siniflandirici Kullanarak Kontrol Grafiklerinde Örüntü Tanima. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, *27*(2), 2012.
19. Bustami, R., Bessaih, N., Bong, C., & Suhaili, S.,Artificial Neural Network for Precipitation and Water Level Predictions of Bedup River. *IAENG International Journal of computer science*, *34*(2), 2007.

20. Raza, G. M., Butt, Z. S., Latif, S., & Wahid, A., Sentiment Analysis on COVID Tweets: An Experimental Analysis on the Impact of Count Vectorizer and TF-IDF on Sentiment Predictions using Deep Learning Models. In *2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, pp. 1-6, 2021.
21. Tripathy, A., Agrawal, A., & Rath, S. K., Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, *57*, 117-126, 2016.
22. O. Kaynar ve S. Yıldız. "İstenmeyen Elektronik Postaların Centroid Tabanlı Sınıflayıcı, Yapay Sinir Ağları ve Destek Vektör Makinaları Yardımıyla Tespit Edilmesi,", 2. Ulusal Yönetim Bilimleri Kongresi, Türkiye, cilt.1, ss.144-153, 8-10 Ekim 2015.
23. Ulusoy, T., İmkb Endeks Öngörüsü İçin İleri Beslemeli Ağ Mimarisine Sahip Yapay Sinir Aği Modellemesi. Uluslararası İktisadi ve İdari İncelemeler Dergisi, (5), 2010.
24. Basarslan, M. S., & Kayaalp, F., Sentiment Analysis with Machine Learning Methods on Social Media. **DOI:** https://doi.org/10.14201/ADCAIJ202093515, 2020.