ARTIFICIAL INTELLIGENCE STUDIES
2020, 3(2): 1-12

www.aistudies.org

Makine Öğrenmesi Yöntemleri ile Film Yorumlarından Duygu Analizi

Büşra YETGINLER [a], İsmail ATACAK [*,b]

[a] *Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilimdalı,06500 Yenimahalle, Ankara , TÜRKİYE*
[b,*] *Gazi Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, 06500 Teknikokullar, Ankara, TÜRKİYE*

MAKALE BİLGİSİ

Alınma: 14.09.2020
Kabul : 23.11 2020

Anahtar Kelimeler:
Sosyal ağlar, Twitter, Makine öğrenmesi, Film yorumları, Duygu analizi.
[*]Sorumlu Yazar:
e-posta:
iatacak@gazi.edu.tr

ÖZET

Bilgi ve iletişim teknolojilerinin hızlı bir şekilde gelişmesinin bir sonucu olarak, sosyal ağlar günlük yaşantımızın önemli bir parçası haline gelmiştir. Twitter, Instagram, Facebook ve Tumblr gibi sosyal ağ platformları üzerinde seyredilen filmler ve satın alınan ürünler üzerine paylaşılan yorumlar, reklamlar ve kampanyalar bu platformları kullanan hedef kullanıcı kitle üzerinde çok önemli bir etkiye sahiptir. Bununla birlikte kullanıcılar ilgili platformlarda çok büyük miktardaki metinsel veri yığınlarını okuyup anlayabilmek için uzun bir zamana ihtiyaç duyarlar. Bu kullanıcılar açısından önemli bir problemdir. Günümüzde problemin çözümüne yönelik olarak duygu analizi ve yapay zekaya dayanan farklı tipte otomatik metin işleme metot ve algoritmaları uygulanmaktadır. Bu çalışmada, Kaggle üzerinden elde edilen IMDB veri setindeki film yorumlarının olumlu mu yoksa olumsuz mu olduğunu belirleyebilmek için Microsoft Azure Makina Öğrenmesi (MÖ) studyosunda iki sınıflı MÖ temelli 6 farklı model geliştirilerek, karışıklık matrisi metrikleri ve ROC eğrileri yoluyla ilgili modellerin performanslarını test eden bir dizi analizler yapılmıştır. Elde edilen sonuçlar tüm karışıklık matrisi metrikleri açısından tek bir modelin en başarılı sonuca ulaşamadığını, ROC analizleri açısından ise bu sonuca iki sınıflı sinir ağı (SA) modelinin ulaştığını göstermiştir.

DOI: 10.30855/AIS.2020.03.02.01

Sentiment Analyses on Movie Reviews using Machine Learning-Based Methods

ARTICLE INFO

Received: 14.09.2020
Accept : 23.11 2020

Keywords:
Social networks, Twitter, Machine learning, Artificial neural network, Movie reviews, Sentiment analyses.
[*]Corresponding Authors
e-mail:
iatacak@gazi.edu.tr

ABSTRACT

As a result of the rapid development of information and communication technologies, social networks have become an essential part of our daily life. The comments shared on social network platforms such as Twitter, Instagram, Facebook and Tumblr about watched movies and purchased products, advertisements and campaigns have a very important effect on the target users using these platforms. However, the users need a long time to read and understand huge amounts of textual data stacks on the relevant platforms. This is a major problem for the users. Nowadays, different types of automatic text processing methods and algorithms based on sentiment analysis and artificial intelligence are applied to solve the problem. In this study, in order to determine whether the movie reviews in the IMDB dataset obtained on Kaggle are positive or negative, 6 different models based on two-class machine learning (ML) are developed in the Microsoft Azure ML studio, and a series of analyses are performed to test the performance of the relevant models by confusion matrix metrics and ROC curves. The results obtained show that a single model could not reach the most successful result in terms of all confusion matrix metrics, while the two-class neural network (NN) model achieves this result in ROC analyses.

DOI: 10.30855/AIS.2020.03.02.01

1. INTRODUCTION *(GİRİŞ)*

With the technological developments and the widespread use of the internet in recent years, there has been a significant increase in the shares of individuals expressing their views about issues, products and services on the internet. Especially through social networks, people contribute to both producers and other people to have ideas by sharing their views about the movies they watch as well as the places they go, the products they use, and current issues. Users utilize these views to decide whether a movie is worth watching or not. The most important problem related to the issue in social networks is that it requires a serious time to review and to understand a huge amount of data. In addition, when these data are examined, it will be seen as another problem that the majority of them consist of misspelled words, hard-to-understand abbreviations and social media jargon words that are not used in daily speech. Therefore, filtering and processing of the data shared on social networks with natural language processing (NLP) methods emerges as an urgent need for social network users [1].

Sentiment analysis approach is a method that automatically reveals emotional states from textual data by utilizing text analysis, NLP and calculation techniques [2]. A marketing survey for a product to be released, estimation of the viewership for a movie, and measurement of reactions to a decision that concerns everyone can be achieved by using this approach. The study conducted by Lee, Pang and Vaithyanatham in 2002 on the classification of movie reviews in the internet movie database [3] contributed to the literature as the first study in which sentiment analyses were performed. In this study, unigram, bigram and part of speech methods were applied to the relevant data set for feature extraction. Naive Bayes (NB), Maximum Entropy (ME) and Support Vector Machine (SVM) methods were used in the classification process. The best result in the classification related to sentiment analyses was obtained by SVM method with an accuracy rate of 82.9% on the unigram data set. As a different study related to sentiment analyses, Katz et al. [4] proposed ConSent, a content-based sentiment analysis model. They tested the model they had developed on data obtained from TripAdvisor and IMDB web sites. In order to compare the performance of the model, they also used NB and SVM methods in the analyses. Their obtained results showed that the developed model was slightly more successful than SVM. In another study, Rana et al. [5] studied a film data set containing 3000 positive and negative reviews. Besides NB, SVM, logistic regression (LR) and k-nearest neighbors (KNN) methods, they also applied Convolutional Neural Network (CNN), which is a deep learning-based approach, to the related dataset. When comparing the results of the applied methods with each other, they observed that the CNN method gave the most successful results with accuracy of 99.33%, F1-score of 99.43%, precision of 99.67% and sensitivity of 99.02% rates. In the study conducted by Rahman and Hossen [6], different type machine learning (ML)-based methods were applied to a dataset consisting of total 2000 film review among which 1400 are training data and 600 are testing data. As the ML-based methods, they used Bernoulli NB, Decision Tree, SVM, ME and Multinomial NB methods. In terms of performance, they reached the highest value with an accuracy rate of 88.5% over the Multinomial NB classifier.

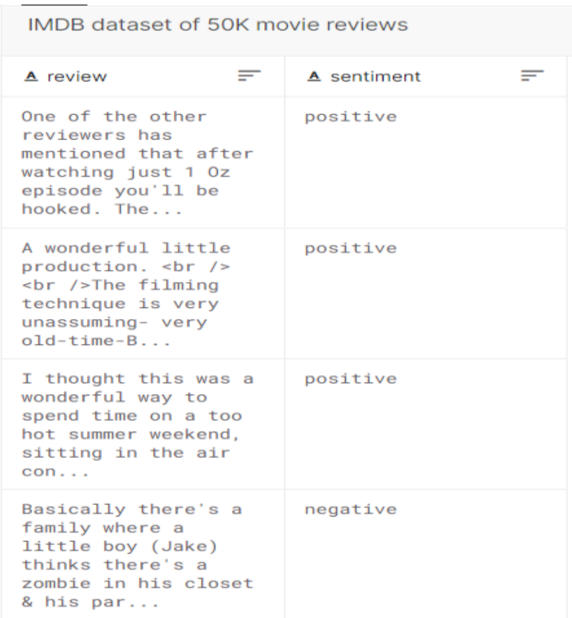
In this study, sentiment analyses are performed with the ML-based models build in Microsoft Azure ML Studio using the dataset containing reviews about IMDB movies and the effects on the performance of the related analyses of these methods are examined. The presentation of the article is organized as follows. In the first section, the importance of NLP in filtering and examining shares made on social networks is explained and the literature review on the classification of movie reviews with sentiment analysis methods is presented. In the second section, we describe in detail the dataset, materials, methods and performance evaluation metrics used in our study. In the third section, the evaluation results of ML-based learning models built in Azure Machine Learning (ML) Studio are given and discussions about the results are presented. In the last section, the inferences obtained from the results are explain.

2. MATERIAL and METHODS *(METERYAL VE METOTLAR)*

In our study, the dataset belonging to IMDB movie reviews is used as material. ML-based methods are applied to the dataset for sentiment analyses. Azure ML Studio is utilized in order to build learning models. Performance of the created models is assessed on the performance metrics such as the accuracy, precision, F1-score and AUC. The details about the material, method and performance assessment are given below in sub-headings.

2.1. Dataset *(Veriseti)*

The dataset used in the study is a collection of 50000 IMDB movie reviews provided by Kaggle for natural language processing and text analytics purposes [7]. Of these comments with binary sentiment, those with an IMDB score below 5 out of 10 are assessed as negative while those with a score of 7 and its above are evaluated as positive. The comments outside this range are neutral and are not included in the data set. The dataset also contains a maximum of 30 reviews per movie. A sample cross-section of the two-column dataset showing these reviews is presented in Figure 1.



IMDB dataset of 50K movie reviews

review	sentiment
One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. The...	positive
A wonderful little production. The filming technique is very unassuming- very old-time-B...	positive
I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air con...	positive
Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his par...	negative

Figure 1. A cross-section of the dataset used in the study *(Çalışmada kullanılan veri setinden bir kesit)*

2.2. Microsoft Azure Machine Learning (ML) Studio *(Microsoft Azure Machine Learning Stüdyo)*

Microsoft Azure ML Studio is a cloud-based environment that allows users to build, manage, train and monitor ML-based models [8]. It provides a working environment where users can easily perform their operations via the interface based on the "drag-and-drop" through web-browser. Azure ML Studio includes a module palette in the visual assembly interface that helps create the end-to-end data science flow. With this palette, users can create new projects, add new data sets, experiment with existing data with the "drag-and-drop" principle, stabilize experiments by accessing training models, and test and evaluate the accuracy of model. Azure ML Studio interface is shown in Figure 2.

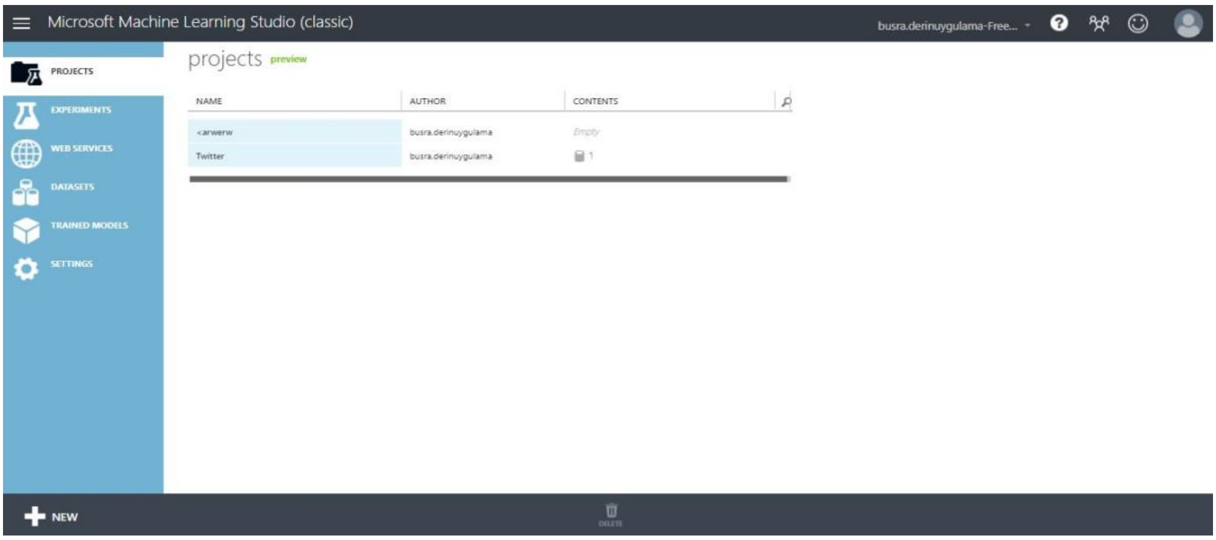


Figure 2. Azure machine learning (ML) studio interface *(Azure makine öğrenmesi (MÖ) stüdyosu arayüzü)*

ISSN: 2651-5350

2.3. Machine Learning-Based Methods Used in the Study *(Çalışmada Kullanılan Makine Öğrenmesi Temelli Metotlar)*

2.3.1. Neural networks *(Sinir ağları)*

Artificial neural networks (ANNs), which was introduced by McCulloch and Pitts in 1943 [9], is a mathematical model of the biological neural networks which consists of a series of layers interconnected by weighted connections. In other words, it is a computer algorithm that produces solutions to problems that require the ability to think and observe by using the brain's learning ability. In practice, it is possible to encounter different type neural network models including feedforward networks, back propagation networks, self-organizing maps, holdfield networks and adaptive resonance theory networks. Back propagation networks, also known as the multilayer perceptron model, are the most widely used among them [10]. As shown in Figure 3, this model includes one input layer, one or more hidden layers and one output layer, and each layer except the input layer is composed of nodes named as the neuron that represents a biological nerve cell.

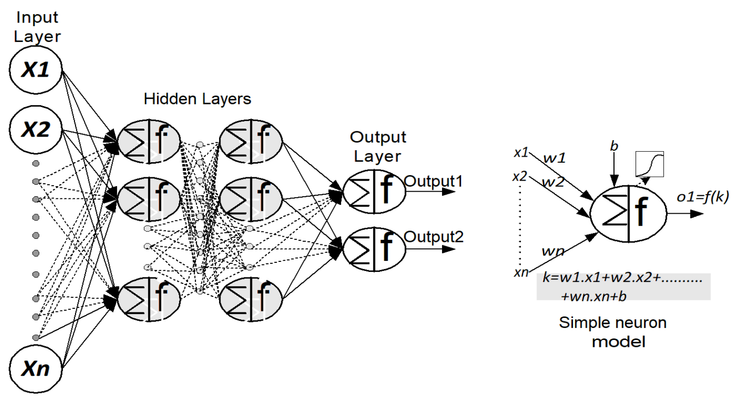


Figure 3. A basic multilayer neural network *(Temel çok katmanlı bir sinir ağı)*

As shown in Figure 3, a simple neuron is a processing element that adds up the incoming entries by multiplying their connection weights, and then obtains its own output by passing this sum through a function called the activation (transfer) function. In simple form, an ANN model can be formulated as given in Eq.1.

$$O_k = f\left(\sum_{i=1}^{n} x_i * w_i + b_k\right) \quad (1)$$

where $x_1, x_2, \ldots . x_n$ is the inputs applied externally to the multilayer neural network. $w_{k1}, w_{k2} \ldots \ldots . w_{kn}$ are the connection weights of the inputs. f(.) is the activation function that produces the neuron output depending on the sum of the product of the incoming input and the weights. o_k denotes the output of neuron [11].

2.3.2. Support vector machine *(Destek vektör makinesi)*

Support vector machine (SVM), the foundations of which is based on a study conducted by Bernhard E. Boser, Isabelle Guyon, and Vladimir N. Vapnik in 1992 [12], has been successfully applied in many areas such as handwriting recognition, disease diagnosis, face recognition and time series analysis from past to present. The basic idea of this method is based on finding the hyper-plane that maximizes the margin between the two classes. Figure 4 represents the hyper-plane representation for the SVM. The training set for classification problems with SVM is represented as { (x_1, y_1),….., (x_n, y_n) }, where $x_i \in R^P$ gives the input vector while $y_i \in \{-1,1\}$ is the output vector. Suppose that w denotes the weight vector to be optimized, and b denotes the bias term. Accordingly, the equation of the hyper-plane separating the two classes can be given as follows.

$$w^T x + b = 0 \quad (2)$$

In order to solve the equation given above, that is, to find an optimal hyper-plane, the optimization problem in Eq. 3 needs to be solved.

$$\begin{aligned} & Min\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ & s.t.\quad y_i(w^T x_i + b) \geq 1 - \xi_i,\ i = 1, \dots, n \\ & \qquad\qquad \xi_i \geq 0,\ \ i = 1, \dots, n \end{aligned} \tag{3}$$

where $\boldsymbol{\xi_i}$ is the slack variable, and it gives a degree of misclassification in the margin. C is expressed as the penalty parameter, and the greater this value corresponds to the higher penalty for error terms [13].

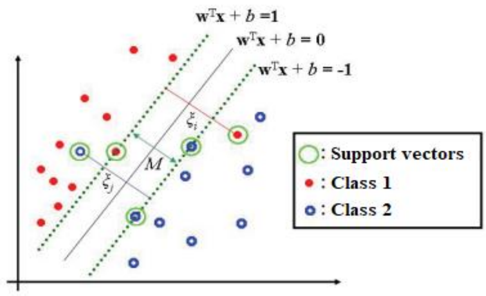


Figure 4. Hyper-plane representation for support vector machine (SVM) *(Destek vektör makinesi (DVM) için süper düzlem gösterimi)*

2.3.3. Logistic regression *(Lojistik regresyon)*

The term "logistic model" was first introduced by Joseph Berkson in 1944 [14]. The study that was the beginning of the binary Logistic Regression (LR) analysis was conducted by Gordon and Kannel on cardiologic diseases in 1968. The purpose of LR is to establish an acceptable coherent model which defines the relationship between result variable and descriptive variables by using minimal variables. The most important feature of this method is that it uses the categorical variable as the result variable [15].

Let Y be a dependent variable, X be an independent variable and β0, β1…. βp be regression coefficients. Accordingly, the LR model with two variables can be formulated as in Eq. 4.

$$P(Y) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} = \frac{1}{1+e^{-(\beta_0+\beta_1 X)}} \tag{4}$$

2.3.4. Decision forest *(Karar ormanı)*

The decision forest (DF) is a type regression model which consists of an ensemble of decision trees. DFs come across to us as fast and supervised ensemble models designed for classification tasks. The bases of such ensemble methods is based on building a more generic model by combining multiple models instead of relying on a single model in order to achieve better results [16]. Figure 5 shows the decision forest structure formed by N decision trees.

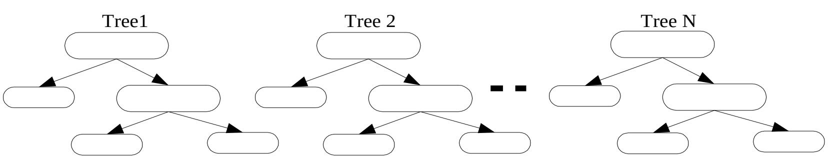


Figure 5. An ensemble model with N trees for the decision forest *(Karar ormanı için N ağaçlı bir topluluk modeli)*

In this structure, decision trees act as nonparametric models that perform a series of tests for each sample and transfer the binary tree structure end-to-end until a leaf node is reached. Each tree in the decision forest gives appropriate outputs to the Gaussian distribution by estimation. The decision forest achieves its output by finding

ISSN: 2651-5350© 2020 Gazi Akademik Yayıncılık

the gaus distribution closest to the combined distribution for all tree structures. For this purpose, it implements a number of aggregating methods such as Bagging and Replication in the model [17].

2.3.5. Boosted decision tree *(Artırılmış karar ağacı)*

Boosted decision tree (BDT) is an algorithm in which an ensemble of regression trees is created according to the previous trees of the trees given in the model. This algorithm corrects the errors of previously created decision trees. In other words, it follows a correction process that continues until the N^{th} tree, in which the second tree corrects the errors of the first tree, the third tree corrects the errors of the first tree and the second tree. The correction process is performed through a loss function that will minimize the error, and it continues until the optimum tree is obtained [18-19]. An ensemble model with N trees for the boosted decision tree is shown in Figure 6.

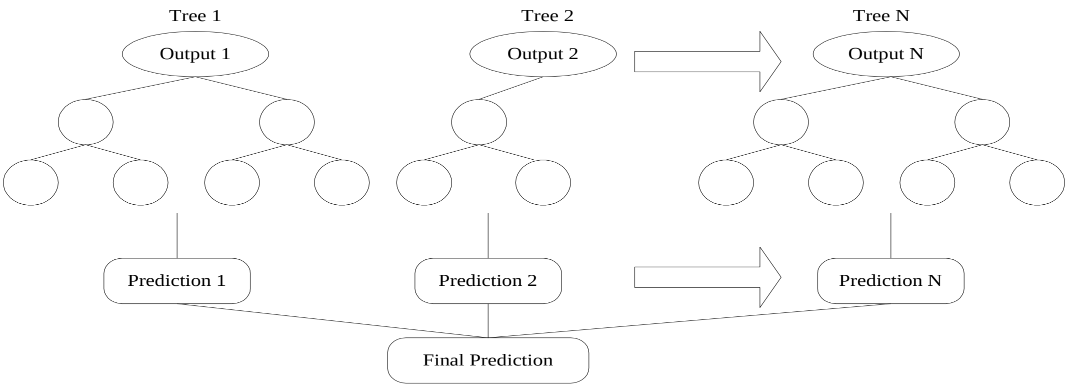


Figure 6. An ensemble model with N trees for the boosted decision tree *(Güçlendirilmiş karar ormanı için N ağaçlı bir topluluk modeli)*

BDTs are quite successful in dealing with tabular data. The most important advantages of this algorithm are that it is strong against missing data and can allocate feature importance scores [19].

2.3.6. Bayes point machine *(Bayes nokta makinesi)*

Bayes point machine (BPM) developed by Herbrich et al. [20-21] is a method used the Bayesian principle in order to successfully classify samples given in a kernel space. BPM conducts the classification process through a hypothesis, also called a classifier. The hypothesis is a function in the form of *h(x)* that assigns an input vector *x* to an output label y, and using the training data, it creates a version space *V*, each point of which represents a possible classifier [22]. The version space *V(z)* formed in a BPM with training data as $z = (x,y) = (\{x_1,y_1\},\ldots,\{x_m,y_m\})$ is as follows.

$$V(z) := \{h \in H \mid \forall i \in \{1,\ldots,m\} : h(x_i) = y_i\} \quad (5)$$

Classification of the test data *h(xi)* within the Bayesian theory is made based on its posterior probability $P_{H|Z^m=z}(h)$ with minimal loss.

$$Bayes_z(x) := \underset{y \in Y}{argmin} E_{H|Z^m=z}[l(H(x),y)] \quad (6)$$

where the loss and posterior probability values are obtained using the formulas given below, respectively.

$$l(y,\acute{y}) = \begin{cases} 0 & y = \acute{y} \\ 1 & y \neq \acute{y} \end{cases} \quad (7)$$

$$P_{H|Z^m=z}(h) = \begin{cases} \frac{PH(h)}{PH(V(z))} & h \in V(z) \\ 0 & h \notin V(z) \end{cases} \quad (8)$$

ISSN: 2651-5350© 2020 Gazi Akademik Yayıncılık

2.4. Performance Metrics *(Performans Metrikleri)*

Performance assessment metrics are defined as the measurement parameters that reveal how successful the algorithms based on artificial intelligence are. The assessment metrics including classification accuracy, precision, recall and F1-Score are commonly used in measuring the performance of a binary classification problem, and they calculate assessment results using the confusion matrix. Table 1 shows the confusion matrix for a binary classification problem.

Tablo 1. Confusion matrix for a binary classification problem *(İkili bir sınıflandırma problemi için karışıklık matrisi)*

		Predicted values	
Actual values		**Positive**	**Negative**
	Positive	True positive (TP)	False negative (FN)
	Negative	False positive(FP)	True negative (TN)

The confusion matrix produces 4 outputs which are used in the calculation of performance metrics: true positive (TP), false positive (FP), false negative (FN), and true negative (TN) [23]. Here, the first output TP is the case in which data with a positive actual class value is correctly tagged as positive by the model. The second output FP corresponds to the case where data with a negative actual class value is mislabeled as positive by the model. The third output FN represents the case in which the predicted class value of data is negative when the actual class value is positive. The last output TN defines the case where data with a negative actual class value is correctly labeled as negative by the model.

Accuracy is the ratio of results that a model successfully estimates trues to all predicted results and is formulated as follows.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \qquad (9)$$

As given in Eq. 10, precision is defined as the ratio of results that a model correctly estimates actual positives to all results that are predicted as positive by the model.

$$Precision = \frac{TP}{TP+FP} \qquad (10)$$

Recall is a metric that gives us how much ratio of all actual true positives our model is able to capture. It is calculated by Eq. 11.

$$Recall = \frac{TP}{TP+FN} \qquad (11)$$

F1-score metric has a very important role in determining the entire error cost of the model since it is based on the harmonic mean of the precision and sensitivity values. In order not to make an incorrect model selection, it would be more accurate to use this metric instead of the accuracy metric, especially in the classification of unbalanced datasets. Its formulation is as given in Eq. 12.

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision+Recall} \qquad (12)$$

Another approach used to evaluate model performance is the ROC curve and the AUC values expressing the area under it. The ROC curve is constructed using the false positive ratio on the x axis and the true positive ratio on the y axis. The fact that the curve is close to the top-left corner on the y-axis increases the ability of model to successfully distinguished positives and negatives.

3. RESULTS and DISCUSSIONS

Figure 7 shows the model built with the help of modules in Microsoft Azure ML Studio for this study, where positive or negative sentiment analyses are performed by applying the different ML-based methods to the reviews about IMDB movies. As seen from the block diagram of the model, after the dataset is uploaded to the system, the pre-processing text module is first processed to perform sentence detection, case normalization, removal of URL addresses, numbers and e-mail addresses, removal of special characters, removal of repetitive character sequences, expansion of verb abbreviations, conversion of words to dictionary form and conversion of backslashes into slashes. Then, the rates for the training data and test data are determined using the split data module as 60% and 40% respectively, resulting in 30000 training data and 20000 test data. With the edit metadata module, it is ensured that the data column and the data type to be processed are selected and the column containing the class tag is marked. In the next process, the words in the sentence are extracted with the Extract N-Gram Features from Text module from the training data according to the weighting ratios that indicate the importance weight in the dataset. In this module, N-Gram size is selected as 1 while the weighting function is chosen as TF-IDF. Other settings for performing the process in the module are made as follows: the minimum word length is 3, the maximum word length is 15, the minimum N gram document absolute frequency is set to 4, and the maximum N gram document rate is set to 1. Using the Chi Squared feature scoring method, 500 words are extracted according to the features determined in the module and these words are selected on the Select column dataset module. After that, the training models that will allow the classification process to predict the sentiment are obtained from the output of the train model module by applying respectively ML-based algorithms that represent the model with the desired result tags and variables to the input of the corresponding module. After the training is completed, the test of the model is performed through the score model module. This module obtains the test score by using two inputs, one from the trained model and the other from the select column in dataset module. Finally, the outputs produced by the score model module is applied to the evaluate model module, and the performance results of the trained model are obtained from the output of the relevant module.

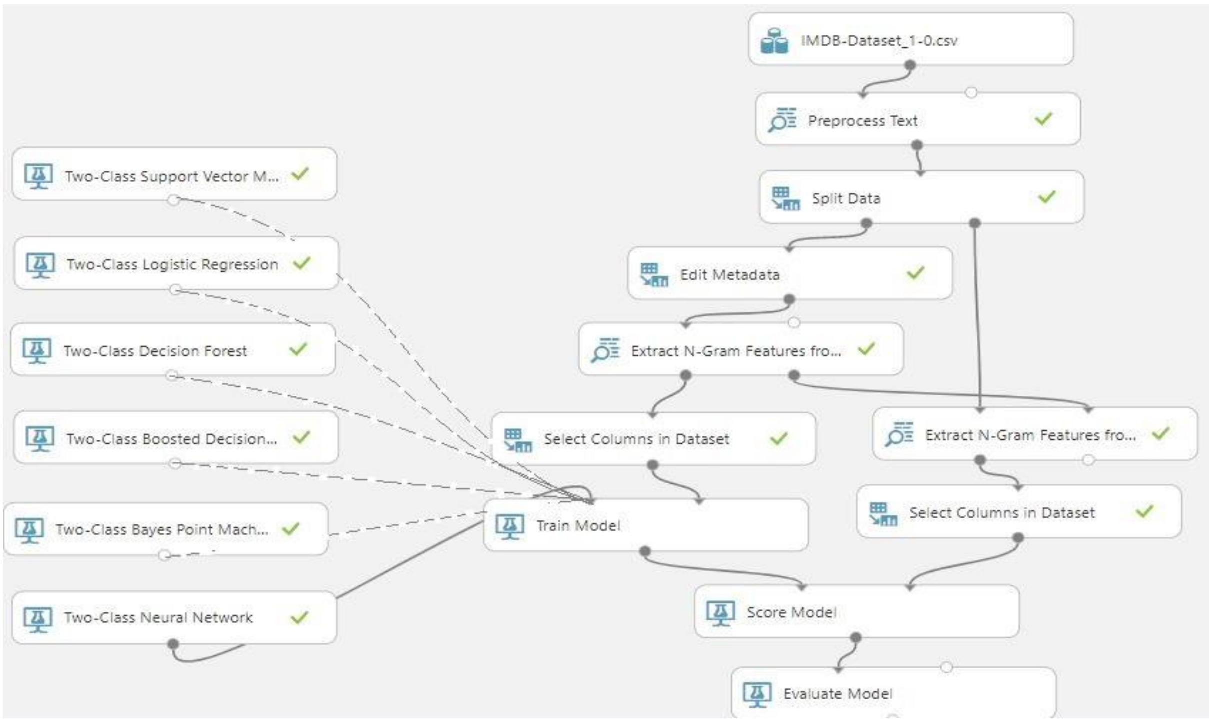


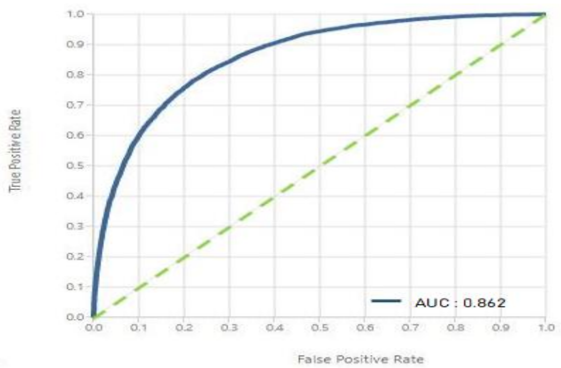
Figure 7. Model built in Microsoft Azure ML studio *(Microsoft Azure ML stüdyo' da oluşturulan model)*

Of the 20000 test data obtained at the end of the split data process, 10055 is positives and 9945 is negatives. When the confusion matrix results given in Table 2 are evaluated in the context of these values, it is seen that the two-class NN model produces the closest result to the true positive value of 10055 with the value 8072 while the two-class BPM model produces the closest result to the true negative value of 9945 with the value of 7849. When considering the performances of these models in terms of proportional accuracies, the rate of predicting true positives (Recall) and predicting true negatives (Specificity) of two-class NN and two-class BPM models are 80.3% and 78.9%, respectively. Two-class DF give the lowest performance for both metrics with 75.5% and 76.5% values.

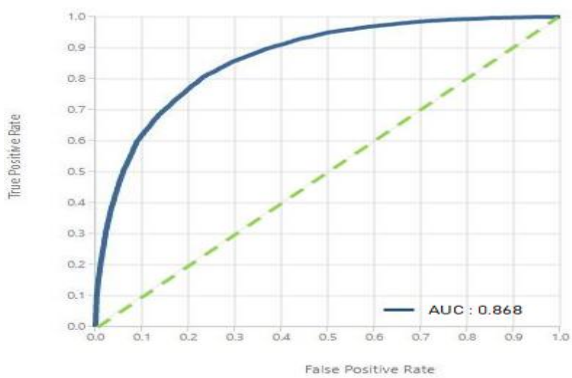
Table 2. Confusion matrix results for ML-based models *(ML temelli modeller için karışıklık matrisi sonuçları)*

ML-Based Model	True/False	Positive	Negative	Recall	Specificity
Two-class SVM	True	7930	7660	78,9%	77,0%
	False	2285	2125		
Two-Class LR	True	8025	7705	79,8%	77,5%
	False	2240	2030		
Two-class DF	True	7588	7611	75,5%	76,5%
	False	2334	2467		
Two-class BDT	True	7884	7724	78,4%	77,7%
	False	2221	2171		
Two-class BPM	True	7674	7849	76,3%	78,9%
	False	2096	2381		
Two-class NN	True	8072	7715	80,3%	77,6%
	False	2230	1983		

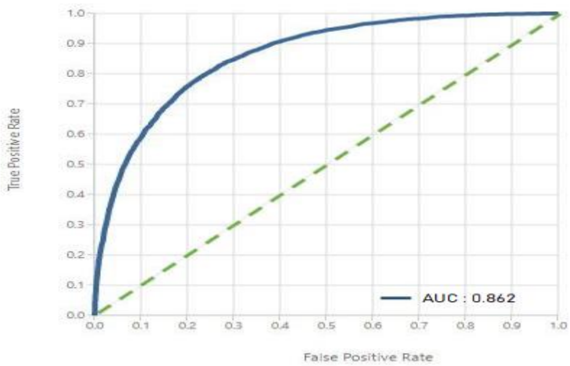
ROC curves and area under the curve (AUC) values of ML-based models are shown in Figure 8. In ROC curve analyses, the closer the curve is to the top-left corner or the closer the AUC value is to 1, the better the discrimination power of the model. In this context, when the ROC curves and AUC values of the proposed models are examined, it is understood that the two-class NN model, whose ROC curve is shown in Figure 8 f, gives the most successful result with an AUC value of 0.87. The lowest performance in terms of AUC value, with a value of 0.779, is obtained from the two-class BPM model whose ROC curve is given in Figure 8 e.



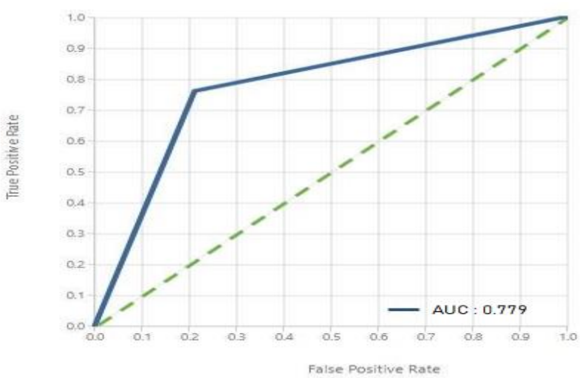
(a) Two-class SVM



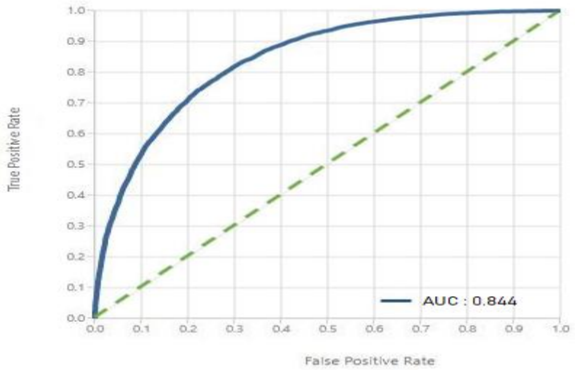
(b) Two-class LR



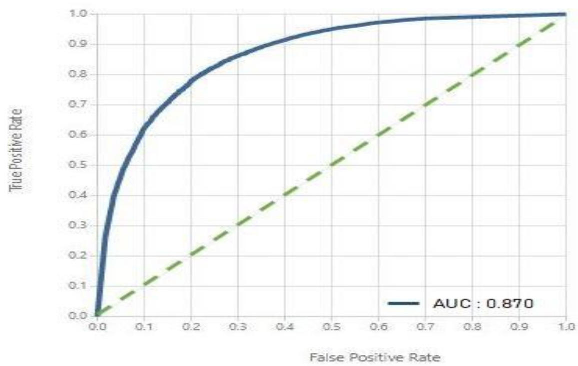
(d) Two-class BDT



(e) Two-class BPM



(c) Two-class DF



(f) Two-class NN

Figure 8. ROC Curves of ML-based models (*Makine öğrenmesi modellerine ait ROC eğrileri*)

Table 3 presents the comparison of ML-based models in terms of accuracy, precision and F1-score performance metrics. When the results in the table are evaluated in terms of three basic metrics, it can be seen that the two class NN model provides the best results on accuracy and F1-score metrics with values of 78.9% and 79.3%, respectively, while the two classes BPM model achieves the best result on the precision metric. However, the two class NN produces quite close to that of the two class BPM in terms of respective metric performance. The lowest performance related to three metrics is obtained from the two class DF model.

Table 3. Comparison of ML-based models in terms of accuracy, precision, and F1-score performance metrics *(Makine öğrenmesi modellerinin doğruluk, kesinlik ve F1-skor performans metrikleri açısından karşılaştırılması)*

ML-Based Model	Accuracy	Precision	F1-score
Two-class SVM	77.9%	77.6%	78.2%
Two-class LR	78.6%	78.2%	79.0%
Two-class DF	76.0%	76.5%	76.0%
Two-class BDT	78.0%	78.0%	78.2%
Two-class BPM	77.6%	78.5%	77.4%
Two-class NN	78.9%	78.4%	79.3%

4. CONCLUSION

Sentiment analyses have become very important NLP tools for both scientists and companies working in this field in converting the big data stacks that has emerged with the widespread use of the internet and social media nowadays into easily interpretable and understandable information. In this context, different models in which artificial intelligence-based methods are integrated have been developed to make understandable the relevant data stacks. In this study, sentiment analyses were performed by applying ML-based methods to the IMDB movie reviews provided by Kaggle. The performances of the proposed models were compared with each other in the study, where the movie reviews were labeled as two classes: positive and negative. In analyses, the performance measurements of the models were performed using the accuracy, precision, recall, specificity and F1-score metrics based on the confusion matrix and the ROC curves with AUC values. The confusion matrix analyses showed that no ML-based model alone could most successfully predict both positive and negative movie reviews together. While the two-class NN model provided the most successful result regarding the prediction of positive movie reviews, it was seen that the two-class BPM model obtained this result in the prediction of negative reviews. When comparing the performance of the models in terms of confusion matrix-based performance metrics, it was observed that the two-class NN model achieved a higher score according to the accuracy, recall and F1-score metrics, and the two-class BPM model produced a higher score according to the specificity and precision metrics. When the results were evaluated in terms of ROC curves, it was understood that the two-class NN model achieved the best performance with the highest AUC value. In our future studies, we plan to propose a hybrid model based on fuzzy logic and deep learning algorithms that we believe will provide higher performance in such datasets.

REFERENCES ***(KAYNAKLAR)***

[1] Y. Yamak, A. Güneş and E. Sunar, "Makine öğrenmesiyle metinde duygu analizi," Bitirme ödevi, Erciyes Üniversitesi, Bilgisayar Mühendisliği, Kayseri, Haziran 2018.

[2] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges," Journal of King Saud University - Engineering Sciences, vol. 30, no. 4, pp. 330–338, 2018. doi:10.1016/j.jksues.2016.04.002

[3] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques, " *In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP-2002*, Philadelphia, Pennsylvania, 2002. pp. 79–86.

[4]G. Katz, N. Ofek and B. Shapira, "ConSent: Context-based sentiment analysis," *Knowledge-Based Systems*, vol. 84, pp. 162-178, 2015. doi: 10.1016/j.knosys.2015.04.009

[5] T. Dholpuria, Y. K. Rana and C. Agrawal, "A Sentiment analysis approach through deep learning for a movie review," *2018 8th International Conference on Communication Systems and Network Technologies, CSNT-2018*, Bhopal, India, 2018. pp. 173-181.

[6] A. Rahman and M. S. Hossen, "Sentiment analysis on movie review data using machine learning approach, " *in 2019 International Conference on Bangla Speech and Language Processing, ICBSLP-2019*, Sylhet, Bangladesh, 2019. pp. 1–4.

[7] A.L. Maas, R.E. Daly, P.T. Pham, Dan. Huang, A.Y. Ng and C. Potts, "Learning word vectors for sentiment analysis, " *in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, ACL-2011*, Portland, Oregon, USA, 2011. pp. 142–150.

[8] Microsoft, "What is Azure Machine Learning?" microsoft.com, 2019. [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-ml. [Accessed: June 21, 2020].

[9] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity, " *Bulletin of Mathematical Biology*, vol. 5, no. 4, pp. 115–133, 1943. doi: 10.1007/BF02478259

[10] R. Barga, V. Fontama, W. H. Tok and L. Cabrera-Cordon, Predictive Analytics with Microsoft Azure Machine Learning, Berkely, CA:Apress, 2015.

[11] H. Simon, Neural networks: *a comprehensive foundation*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[12] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A training algorithm for optimal margin classifiers," *In D. Haussler, editor, Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, ACM Press, Pittsburgh, PA, 1992. pp. 144–152.

[13] B. Yetginler, "Rahim ağzı kanserinin veri madenciliği yöntemleri ile sınıflandırılması," YL Tezi, Kırıkkale Üniversitesi, Fen Bilimleri Enstitüsü, Kırıkkale, Haziran 2019.

[14] J. S. Cramer, "The Orijin and development of the logit model," Cambridge University Press, pp. 149–157, 2003.

[15] C. Aktaş, "Lojistik regresyon analizi: öğrencilerin sigara içme alışkanlığı üzerine bir uygulama," *Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, vol. 1, no. 26, pp.107–121, 2009.

[16] Microsoft, "Azure Machine Learning Two-Class Decision Forest," microsoft.com, 2020. [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-forest. [Accessed: July 03, 2020].

[17] A. A. Çolakoğlu, "Makine öğrenmesi algoritmaları ile Avrupa havalimanları analizi ," YL Tezi, Pamukkale Üniversitesi, Sosyal Bilimler Enstitüsü, Denizli, Haziran 2020.

[18] E. Jumin, N. Zaini, A. N. Ahmed, S. Abdullah, M. Ismail, M. Sherif, A. Sefelnasr and A. El-Shafie, "Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction," *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 713–725, 2020. doi: 10.1080/19942060.2020.1758792

[19] J.-P. Lai, Y.-M. Chang, C.-H. Chen, P.-F. Pai, "A survey of machine learning models in renewable energy predictions," *Applied Sciences*, vol. 10, pp. 1-20, 2020. doi:10.3390/app10175975.

[20] R. Herbrich, T. Graepel and C. Campbell, "Bayes point machines," *Journal of Machine Learning Research*, vol. 1, pp. 245–279, 2001.

[21] R. Herbrich, T. Graepel and C. Campbell, "Bayes Point Machines: Estimating the Bayes Point in Kernel Space, " *In: Proceedings of International Joint Conference on Artificial Intelligence Work-shop on Support Vector Machines*, Stockholm, Sweden, 1999. pp. 23–27.

[22] J. Li, "Texture classification of landsat TM imagery using Bayes point machine, " *In Proceedings of the 51st ACM Southeast Conference*, ACMSE 2013, Savannah, GA, 2013. pp. 16.1-16.6.

[23] J. Han, M. Kamber and J. Pei, Data Mining:*Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2012. [Online]. Available: http://ieeexplore.ieee.org/document/6918822/

RESUME (*ÖZGEÇMİŞ*)

Büşra YETGINLER [a]

Büşra YETGINLER was born in Malatya, Turkey, on 13 January 1993. She graduated from the Computer Engineering Department in Kırıkkale University, Kırıkkale, Turkey, in 2016. In 2016, she attended the MSc programme at the Computer Engineering Department in Kırıkkale University and graduated in 2019. In 2020, she attended the Ph.D. programme at the Computer Engineering Department in Gazi University and still continues. His current fields of research are bioinformatics, machine learning based algorithms and engineering education.

İsmail ATACAK [*,b]

İsmail ATACAK was born in Konya, Turkey in 1972. He received the BS., MSc and Ph.D. degrees from Department of Electronics and Computer Education, Gazi University, in 1994, 1998, and 2005, respectively. From 2007 to 2012, he worked as an Assistance Professor at the Department of Electronics and Computer Education, Faculty of Technical Education, Ankara, Turkey. He is currently working as an Assistance Professor at the Department of Computer Engineering, Faculty of Technology, Ankara, Turkey. His research interests include power systems, artificial intelligent based algorithms, optimization based algorithms and engineering education.