Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

GU J Sci, Part C, 12(4): 893-901 (2024)

The Effect of Various Text Representation Methods for Sentiment Analysis on Movie Review Data with Different Machine Learning Methods

Veysel GÖÇ[1] Muhammet Sinan BAŞARSLAN[1*]

[1]*Istanbul Medeniyeti University, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, Istanbul, Turkey*

Article Info

Research article
Received: 09/06/2024
Revision: 22/09/2024
Accepted: 06/10/2024

Keywords

Machine Learning
Movie Review
Sentiment Analysis
Text Representation

Makale Bilgisi

Araştırma makalesi
Başvuru: 09/06/2024
Düzeltme: 22/09/2024
Kabul: 06/10/2024

Anahtar Kelimeler

Makine Öğrenimi
Film Revie
Duygu Analizi
Metin Gösterimi

Graphical/Tabular Abstract (Grafik Özet)

In this study, we investigate the potential of machine learning (ML) models after different text representation methods on the balanced IMDB dataset. After data cleaning and text representation, sentiment analysis classification is performed. In the SVM model, the highest performance was observed with BERT: ACC 0.9033, F1 0.9308, R 0.9015, P 0.9072, AUC 0.9638. The results show that BERT offers high performance in text classification. / Bu çalışmada, dengeli IMDB veri kümesinde farklı metin temsil yöntemleri sonrası makine öğrenmesi (ML) modellerinin potansiyeli incelenmiştir. Veri temizleme ve metin temsil aşamalarının ardından, duygu analizi sınıflandırması yapılmıştır. SVM modelinde, BERT ile elde edilen en yüksek performans gözlemlenmiştir: ACC 0.9033, F1 0.9308, R 0.9015, P 0.9072, AUC 0.9638. Sonuçlar, BERT'in metin sınıflandırmada yüksek performans sunduğunu göstermektedir.

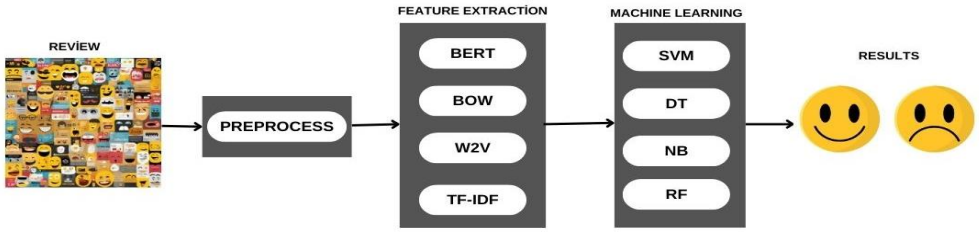

Figure A: *Graphical abstract /* ***Şekil A:*** *Özet grafik*

Highlights (Önemli noktalar)

- *Sentiment analysis was performed for ML models with different text representation methods on the IMDB dataset/ IMDB veri kümesinde farklı metin temsil yöntemleriyle ML modelleri için duygu analizi gerçekleştirilmiştir*
- *BERT-based text extraction provided the highest performance among all the models, with SVM being particularly successful/ BERT tabanlı metin çıkarımı, tüm modeller arasında en yüksek performansı sağlamış; özellikle SVM modelinde başarı elde edilmiştir*
- *Classifications with BERT offer high accuracy and reliability in NLP tasks/ BERT ile yapılan sınıflandırmalar, NLP görevlerinde yüksek doğruluk ve güvenilirlik sunmaktadır*

Aim (Amaç) *The aim of this study is to evaluate the sentiment analysis performance of machine learning models with different text representation methods on the IMDB dataset. In particular, it aims to identify the highest performing models by examining the impact of BERT-based feature extraction. The results aim to demonstrate the effectiveness and reliability of these methods in NLP applications/Bu çalışmanın amacı, IMDB veri kümesinde farklı metin temsil yöntemleriyle makine öğrenmesi modellerinin duygu analizi performansını değerlendirmektir. Özellikle, BERT tabanlı öznitelik çıkarımının etkisini inceleyerek en yüksek performans sağlayan modelleri belirlemeyi hedeflemektedir. Sonuçlar, NLP uygulamalarında bu yöntemlerin etkinliğini ve güvenilirliğini göstermeyi amaçlamaktadır*

Originality (Özgünlük): *It highlights the impact of BERT-based feature extraction in sentiment analysis by comparing the performance of various ML models with different text representation methods on the IMDB dataset. The study offers a new perspective to the field by examining the use of BERT with different ML models in NLP tasks/ IMDB veri kümesinde farklı metin temsil yöntemleriyle çeşitli ML modellerinin performansını karşılaştırarak, BERT tabanlı öznitelik çıkarımının duygu analizindeki etkisini vurgulamasında yatmaktadır. Çalışma, BERT'in NLP görevlerinde farklı ML modelleriyle kullanımını inceleyerek alana yeni bir bakış açısı sunmaktadır*

Results (Bulgular): *It shows that BERT-based feature extraction on the IMDB dataset significantly improves sentiment analysis performance, especially when combined with the SVM model/ IMDB veri kümesinde BERT tabanlı öznitelik çıkarımının özellikle SVM modeliyle birleştirildiğinde duygu analizi performansını önemli ölçüde artırdığını göstermektedir*

Conclusion (Sonuç): *The emphasizes that Transformer models offer high accuracy and reliability in NLP tasks/ Transformer modellerinin NLP görevlerinde yüksek doğruluk ve güvenilirlik sunduğunu vurgulamaktadır*

*Corresponding author, e-mail: muhammet.basarslan@medeniyet.edu.tr
DOI:10.29109/gujsc.1498509

GU J Sci, Part C, 12(4): 893-901 (2024)

Gazi Üniversitesi
Fen Bilimleri Dergisi
PART C: TASARIM VE TEKNOLOJİ

Gazi University
Journal of Science
PART C: DESIGN AND TECHNOLOGY

http://dergipark.gov.tr/gujsc

The Effect of Various Text Representation Methods for Sentiment Analysis on Movie Review Data with Different Machine Learning Methods

Veysel GÖÇ[1] Muhammet Sinan BAŞARSLAN[1*]

[1]*Istanbul Medeniyeti University, Favulty of Engineering and Natural Sciences, Department of Computer Engineering, Istanbul, Turkey*

Article Info

Research article
Received: 09/06/2024
Revision: 22/09/2024
Accepted: 06/10/2024

Keywords

Machine Learning
Movie Review
Sentiment Analysis
Text Representation

Abstract

In this study, we explore the potential of machine learning (ML) models after different text representation methods on the balanced Internet Movie Database (IMDB) dataset, which is widely considered as the gold standard in sentiment analysis, one of the Natural Language Processing (NLP) tasks. On the open-source IMDB movie reviews dataset, we first undertake data cleaning and text representation with data preprocessing steps. Then, we apply sentiment classification using different ML models. To evaluate the models, we used Accuracy (ACC), precision (P), Recall (R), F1-score (F1), and area under curve score (AUC), as well as receiver operating characteristic (ROC) Curve. It is worth noting that text feature extraction with Bidirectional Encoder Representations from Transformers (BERT) provided the highest performance in all models, with the SVM model offering particularly promising results. In this model, we observed the following results: ACC 0.9033, F1 0.9308, R 0.9015, R 0.9015, P 0.9072, AUC 0.9638, and ROC Curve 0.96. These findings suggest that NLP techniques, particularly ML models that employ BERT may offer high levels of ACC and reliability in text classification problems. It would be beneficial for future studies to validate these findings using BERT on different NLP tasks. This would help to evaluate the effectiveness and applicability of the models in practice.

Farklı Makine Öğrenmesi Yöntemleri ile Film Yorumları Üzerine Duygu Analizi için Çeşitli Metin Temsil Yöntemlerinin Etkisi

Makale Bilgisi

Araştırma makalesi
Başvuru: 09/06/2024
Düzeltme: 22/09/2024
Kabul: 06/10/2024

Anahtar Kelimeler

Makine Öğrenimi
Film eleştirisi
Duygu Analizi
Metin Gösterimi

Öz

Bu çalışmada, Doğal Dil İşleme (NLP) görevlerinden biri olan duygu analizinde yaygın olarak altın standart olarak kabul edilen dengeli İnternet Film Veritabanı (IMDB) veri kümesi üzerinde farklı metin temsil yöntemlerinden sonra makine öğrenimi (ML) modellerinin potansiyelini araştırıyoruz. Açık kaynak kodlu IMDB film yorumları veri kümesi üzerinde ilk olarak veri ön işleme adımları ile veri temizleme ve metin gösterimi gerçekleştiriyoruz. Ardından, farklı makine öğrenimi modelleri kullanarak duygu sınıflandırması uyguluyoruz. Modelleri değerlendirmek için Doğruluk (ACC), kesinlik (P), Hatırlama (R), F1-skoru (F1) ve eğri altındaki alan (AUC) skorunun yanı sıra alıcı işletim karakteristiği (ROC) Eğrisini kullandık. Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (BERT) ile metin özelliği çıkarmanın tüm modellerde en yüksek performansı sağladığını ve SVM modelinin özellikle umut verici sonuçlar sunduğunu belirtmek gerekir. Bu modelde aşağıdaki sonuçları gözlemledik: ACC 0.9033, F1 0.9308, R 0.9015, R 0.9015, P 0.9072, AUC 0.9638 ve ROC Curve 0.96. Bu bulgular, NLP tekniklerinin, özellikle de BERT kullanan ML modellerinin metin sınıflandırma problemlerinde yüksek düzeyde ACC ve güvenilirlik sunabileceğini göstermektedir. Gelecekteki çalışmaların bu bulguları farklı NLP görevlerinde BERT kullanarak doğrulaması faydalı olacaktır. Bu, modellerin uygulamadaki etkinliğini ve uygulanabilirliğini değerlendirmeye yardımcı olacaktır.

1. INTRODUCTION (GİRİŞ)

Today, large datasets on the Internet provide access to a variety of information sources, and analyzing this data is becoming an important tool for understanding user trends and preferences. Text data, a subset of this data, is particularly prevalent on social media platforms, online shopping sites, and movie review sites. Analyzing this text data requires the use of NLP techniques to understand users' emotional states and create an overall perception. These techniques are designed to detect, understand, and infer emotional expressions in text. At this point, however, it is important to ask why

*Corresponding author, e-mail: muhammet.basarslan@medeniyet.edu.tr
DOI: 10.29109/gujsc.1498509

businesses need technologies such as artificial intelligence (AI) and NLP. As technology advances, companies are faced with more and more data that needs to be effectively analyzed. In particular, data from consumers' online interactions plays an important role in determining marketing strategies and product development. Therefore, it is inevitable for companies to apply NLP techniques to understand customer feedback, emotional reactions, and overall perceptions.

In this context, this study on the IMDB movie reviews dataset [1] is also very important. This dataset contains user reviews of different movies and these reviews may reflect positive or negative emotions. Therefore, this study aims to analyze these reviews using NLP techniques and develop a method to understand emotional reactions to movies. The findings can be extremely valuable not only in the film industry, but also for companies looking to improve their marketing strategies. This analysis can help filmmakers evaluate the feedback and provide an important guide to understanding the preferences of potential viewers and shape their future projects.

In conclusion, this study provides a framework for how NLP can be used in sentiment analysis, illustrating why companies need technologies such as AI and NLP and why this study is important. The results of this study can help companies make better decisions and develop customer-centric strategies.

The main contribution of the study is to investigate the contribution of text representation to model performance in sentiment analysis, an NLP task. For this purpose, frequency-based representation methods such as Bag of Words (BoW), Term Frequency-Inverse Document Frequency (TF-IDF), and the embedding method Word2Vec (W2V) were used. In addition, a groundbreaking transformer architecture in NLP such as BERT was also used. These methods were trained on a publicly available dataset using Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF), which are the first methods that come to mind when it comes to ML methods. The best model is also compared with the literature in the conclusion and discussion section. Figure 1 shows a visualization of the process of the study.

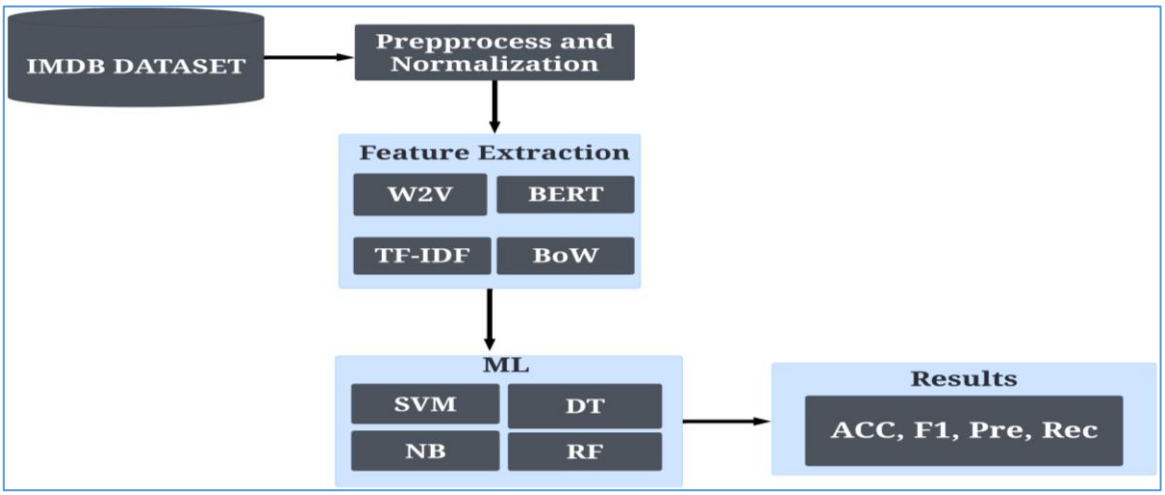

Figure 1. Process of the study (Çalışma süreci)

2. LITERATURE REVIEW (LİTERATÜR ARAŞTIRMASI)

Studies on sentiment analysis on IMDB movie reviews [1] using NLP techniques have attracted great interest in recent years. In this section, we will describe what kind of work has been done in the literature to perform sentiment analysis from IMDB reviews.

Shaukat et al. obtained 0.8667 ACC rate with a multi-layer perceptron (MLP) model using BoW approach on IMDB data. This study shows that the combination of BoW and MLP can be effective for sentiment analysis [2]. Kaynar et al. achieved an ACC of 0.8610 using TF-IDF. [3]. Similarly, Amulya et al. achieved an ACC of 0.88 using TF-IDF and Recurrent Neural Network (RNN) on IMDB data [4]. These studies show that text features extracted using TF-IDF are highly suitable for sentiment analysis. Misini et al. achieved an ACC of 0.8667 with MLP using BoW [5]. This result confirms that the combination of BoW and MLP is a frequently preferred and effective method in the literature. Basarslan and Kayaalp obtained an ACC of 0.8821 using Keras embedding representation and Bidirectional Long Short-Term Memory (BiLSTM) [6]. This study shows that deep learning methods are effective in improving the performance of sentiment analysis. Mohaiminul and

Sultana applied TF-IDF followed by NB, SVM, RF, and Stochastic Gradient Descent (SGD) algorithms and obtained 0.8366 ACC with RF [7]. These results provide important data for comparing the sentiment analysis performance of different ML algorithms. Başa and Basarslan preprocessed the data before performing word vectorization with TF-IDF on IMDB data, and then achieved 0.90 ACC with the SVM algorithm [8]. This study shows that the combination of SVM and TF-IDF can provide strong performance. Pang et al. proposed an approach for sentiment analysis on IMDB comments and classified positive and negative emotions using text mining techniques. This work provided a basic framework for assessing the emotional tone of IMDB comments [9]. Subsequent work has suggested that deep learning techniques can improve the performance of sentiment analysis. Kim et al. performed sentiment analysis on IMDB movie reviews using deep neural networks and achieved higher ACC than traditional methods. This study demonstrated that deep learning models are effective in sentiment analysis [10]. Similarly, the use of word embedding can also improve sentiment analysis performance Maas et al. [1] investigated how distributional representations such as W2V can be used for sentiment analysis on IMDB reviews.

Similarly, the use of word embedding can also improve sentiment analysis performance Maas et al [1] investigated how distributional representations such as W2V can be used for sentiment analysis on IMDB reviews. This study showed that word-level representations can improve sentiment analysis performance. In this context, this study aims to investigate how NLP techniques can be used for sentiment analysis on IMDB movie reviews and how they can contribute to the existing literature. This study can extend the results of previous studies, introduce new methods, and fill existing gaps in the field of sentiment analysis.

Sentiment analysis studies on the IMDB dataset [1] have been conducted using different text feature extraction methods and ML algorithms. These studies compared the performance of models built using different text representation methods.

3. MATERIALS AND METHODS (MATERYAL VE METOD)

This section is a description of the text representation, the dataset, and the ML algorithms used in this study.

3.1 Text Representation (Kelime Temsilleri)

It aims to convert text data into a digital format and put it into a format that the computer can understand. This is usually realized as word vectors or document vectors. BeautifulSoup is a popular HTML and XML parser library in Python [11]. It is widely used in text mining projects to extract and preprocess data from web pages.

In this study, frequency-based (TF-IDF, BoW), embedding (W2V), and transformer (BERT) methods were used to investigate their contribution to classification models. These methods will be described in this section. W2V is one of the word embedding techniques and is used to compute word vectors [12]. In this method, a vector representing the meaning of a word is obtained. W2V methods are shown in Figure 2.

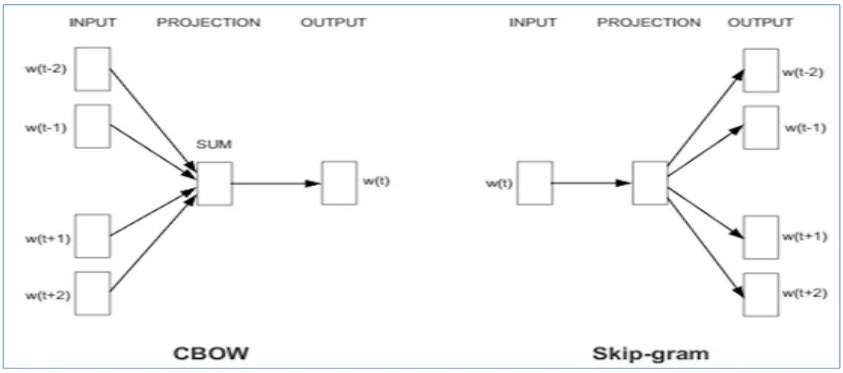

Figure 2. W2V demonstration (W2V gösterimi) [12]

TF-IDF is a statistical measure used to determine the importance of a term in a document. It calculates the ratio between the frequency of a term in a document (TF) and its rarity in all documents (IDF). TF is given in equation (1), IDF in equation (2), and TF-IDF in equation (3) [8].

$$TF(t,d)=\frac{number\ of\ t\ terms\ in\ d\ documents}{total\ number\ of\ terms\ in\ document\ d} \quad (1)$$

$$\text{IDF}=\log\left(\frac{N}{DF(t)}\right) \quad (2)$$

$$\text{TF-IDF}=TF(t,d)*IDF(t) \quad (3)$$

BoW is a model for representing the frequency of words in a document. In this model, each word in a

document is put into a bag and only its presence is considered, regardless of its frequency in the document. That is, the content of the document is not considered as ordered, but only as a set [13]. The BoW model is widely used in natural language processing applications such as text classification and sentiment analysis. This model can be used to represent the content of documents in a numerical vector format and can be used as input data for ML algorithms.

BERT is an NLP model created by Google. BERT works by examining the context of sentences bidirectionally (both left-to-right and right-to-left) to better understand the meaning of language. This allows it to build more accurate and meaningful representations of language, taking into account information from words on both sides of the text. A representation of the BERT architecture is shown in Figure 3 [14].

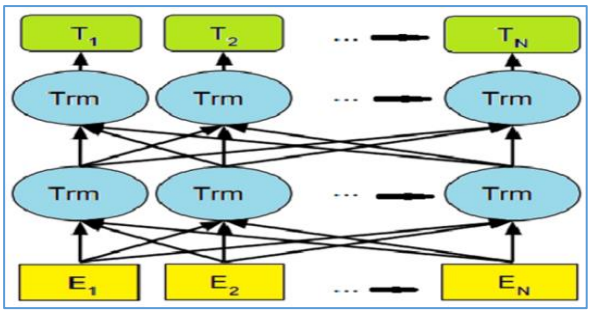

Figure 3. BERT demonstration (BERT gösterimi) [13]

3.2 Dataset (Veri Seti)

In this section, we describe the details of the dataset collected by Maas et al [1]. The dataset consists of two attributes: review and sentiment. The sentiment attribute contains a total of 50.000 comments in two classes, positive and negative.

4. MACHINE LEARNING (MAKİNE ÖĞRENMESİ)

ML is a branch of AI that enables computer systems to improve themselves by analyzing data and using algorithms to perform specific tasks [14]. This type of algorithm allows computers to solve complex problems through the process of learning from data. ML models often gain their experience through data and make future decisions using the knowledge learned from that data. It can be used in a variety of tasks such as classification, regression, clustering, and pattern recognition. ML is increasingly used to solve complex problems involving large amounts of data and computing power. Its main goal is to make computer systems capable of learning from data to produce solutions without human intervention [15].

In this study, NB, DT, RF, SVM were used. These algorithms will be described in this section.

NB is a statistical classification algorithm. It is often used in classification tasks and is based on Bayes' theorem. It assumes the independence of each feature. Equation (4) shows the equation of the Naive Bayes theorem [16].

$$P(A|B) = \frac{P(B|A) * P(A|B)}{P(B)} \quad (4)$$

SVM is a supervised ML algorithm used for classification and regression analysis. The main goal of SVM is to find a hyperplane that provides the best discrimination when classifying data. Figure 4 shows a visualization of SVM [13].

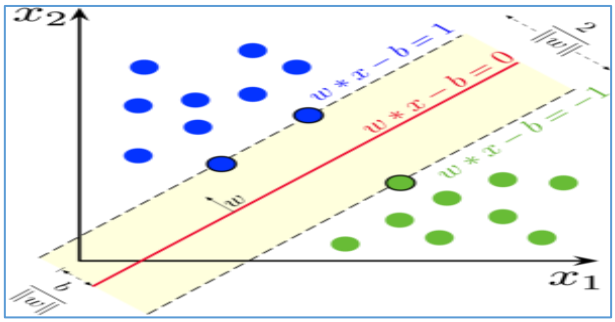

Figure 4. Demonstration of the SVM method (SVM yönteminin gösterimi) [14]

LR is a classification algorithm and makes probability estimates using the logistic function. It is widely used in binary classification problems. Figure 5 shows a visualization of LR [17].

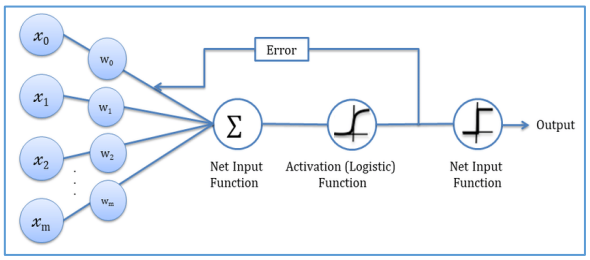

Figure 5. Demonstration of the LR method (LR yönteminin gösterimi) [17]

DT is a model used in classification and regression problems. It organizes the dataset into a tree-like structure based on features and uses this structure to perform classification or prediction [18]. Figure 6 shows a visualization of DT.

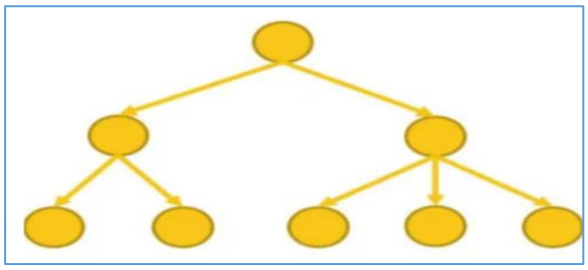

Figure 6. Demonstration of the DT method (DT yönteminin gösterimi) [19]

RF is a decision tree-based classification and regression method. It is created by combining multiple decision trees and is used to provide more accurate and balanced results [20] . Figure 6 shows a visualization of RF.

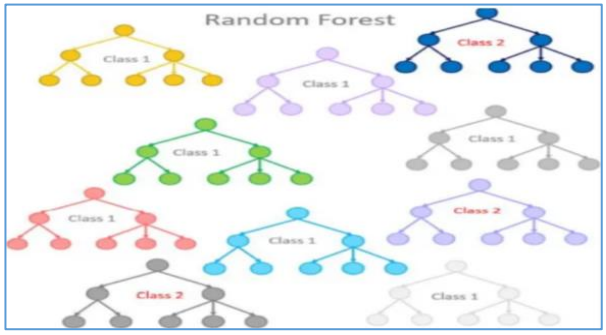

Figure 7. Demonstration of the RF method (RF yönteminin gösterimi) [19]

5. EXPERIMENT SETUP (DENEYSEL KURULUM)

In the study, text cleaning and normalization operations were applied to the comments in the IMDB dataset [1]. Then, an 80%-20% holdout training test separation was performed and text representation was extracted using wor2vec, TF-IDF, BoW, and BERT. Then the models were built with ML models SVM, RF, DT, NB. The whole study was written on Google Colab using Python libraries.

5.1. Performance Metrics (Performans Metrikleri)

ACC, P, R, F1, which are used to calculate the performance metrics of ML models for sensitivity analysis, are given in equation (5) and equation (8), respectively [21-23]. These metrics are obtained from the complexity matrix. This matrix contains the true values and the model predicts correctly or incorrectly and the false values and the model predicts correctly or incorrectly. These are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

$$ACC = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \quad (5)$$

$$P = \frac{Tp}{Tp+Fp} \quad (6)$$

$$R = \frac{Tp}{Tp+Fn} \quad (7)$$

$$F1 = \frac{P*R}{R+R} \quad (8)$$

The metrics given in equations (5) to (8) are obtained from the complexity matrix. The other performance metric used in the study is ROC Curve. ROC Curve is a graphical representation used to evaluate the performance of classification models. The area under the ROC graph is called the ROC Curve. This curve shows the relationship between the true positive rate and the false positive rate of the model. These two rates help us understand how the model performs at different thresholds. The area under the ROC Curve (AUC) summarizes the overall performance of the model. The AUC value is between 0 and 1 [20]:

- AUC = 0.5: indicates that the model predicts at random. The ROC Curve is a line with a 45-degree angle [23].
- AUC > 0.5: Indicates that the model is better than random prediction. The model can discriminate between classes.
- AUC = 1: Indicates that the model performs perfectly and can discriminate classes without error.

6. EXPERIMENT RESULTS (DENEY SONUÇLARI)

The study was written in Python sci-kit learn library. The parameter details of all the models are given in Table 1. These parameters and models were created using Python sci-kit learn library.

Table 1. Model parameter details (Model parametre detayları)

Model	Parameters	Parameter Details
SVM	Kernel	Linear
	C	1
NB	alpha	1.0
	fit_prior	True
	class_prior	None
DT	max_depth	None
RF	n_estimators	100

	random_state	0

The parameters listed in Table 1 are used as reported in the literature or with default values. Parameter optimization will be discussed in detail and applied in future studies. Table 2 shows the results of the models built with SVM, NB, DT, RF according to the text representation.

Table 2. Results of the ML model (ML model sonuçları)

	SVM				**NB**			
	BoW	TF-IDF	W2V	BERT	BoW	TF-IDF	W2V	BERT
Acc	0.8194	0.8434	0.8793	0.9033	0.8631	0.7474	0.8422	0.8458
F1	0.8220	0.8436	0.8798	0.9308	0.8641	0.7502	0.8457	0.8483
R	0.8101	0.8423	0.8765	0.9015	0.8578	0.7421	0.8273	0.8346
P	0.8342	0.8449	0.8831	0.9072	0.88706	0.7585	0.8649	0.8625
AUC	0.8897	0.9185	0.9496	0.9638	0.9383	0.8256	0.9038	0.9256
ROC	0.89	0.92	0.95	0.96	0.94	0.83	0.90	0.93
	DT				**RF**			
	BoW	TF-IDF	W2V	BERT	BoW	TF-IDF	W2V	BERT
Acc	0.7254	0.7115	0.7177	0.7797	0.8466	0.8168	0.8515	0.8626
F1	0.7263	0.7130	0.7194	0.7789	0.8479	0.8166	0.8531	0.8625
R	0.7238	0.7094	0.7152	0.7818	0.8408	0.8174	0.8440	0.8635
P	0.7288	0.7166	0.7237	0.7760	0.8552	0.8157	0.8624	0.8614
AUC	0.7253	0.7115	0.7177	0.7796	0.92	0.89	0.92	0.93
ROC	0.73	0.71	072	0.78	0.92	0.90	0.93	0.94

Table 2 According to the results, the SVM model performed the best, especially when BERT feature extraction was used. The AUC value obtained with BERT-based feature extraction is 0.9638, which shows that the model performs very well. The W2V and TF-IDF methods also performed well. On the other hand, the BoW method underperformed compared to the other methods.

According to Table 2, the NB model performed best when BoW feature extraction was used. The AUC value obtained with BoW is 0.9383. In contrast, the TF-IDF method performed poorly compared to the other feature extraction methods. This shows that TF-IDF is not as effective as BoW in NB models.

The DT model presented in Table 2 showed the highest performance with BERT-based feature extraction. The AUC value obtained with this feature extraction method is 0.7796, which shows that BERT is a highly effective feature extraction method for DT models. The lower performance of other feature extraction methods compared to BERT shows the superiority of BERT for this type of model.

The RF model presented in Table 2 performs best with the BERT feature extraction method. The AUC value obtained with BERT is 0.93. In addition, the BoW and W2V feature extraction methods also performed quite well. This shows that RF models can be flexible and perform well with different feature extraction methods.

Considering the results of the models in all the tables together, the SVM model performed best when BERT feature extraction was used, and the AUC obtained was quite high at 0.9638. W2V and TF-IDF methods also provide high performance in SVM models, while BoW method shows lower performance compared to other methods. NB models achieved the best performance with BoW feature extraction and the AUC value was 0.9383. However, the TF-IDF method showed lower performance in NB models. DT models also performed best with BERT feature extraction and the AUC was 0.7796. RF models performed best with an AUC of 0.93 using BERT. BoW and W2V methods also performed well on RF models.

In general, BERT and SVM achieved better results compared to the other methods used in the project. These results show that feature extraction methods play an important role in model performance. The ROC Curve of the generated models are shown in Figure 8.

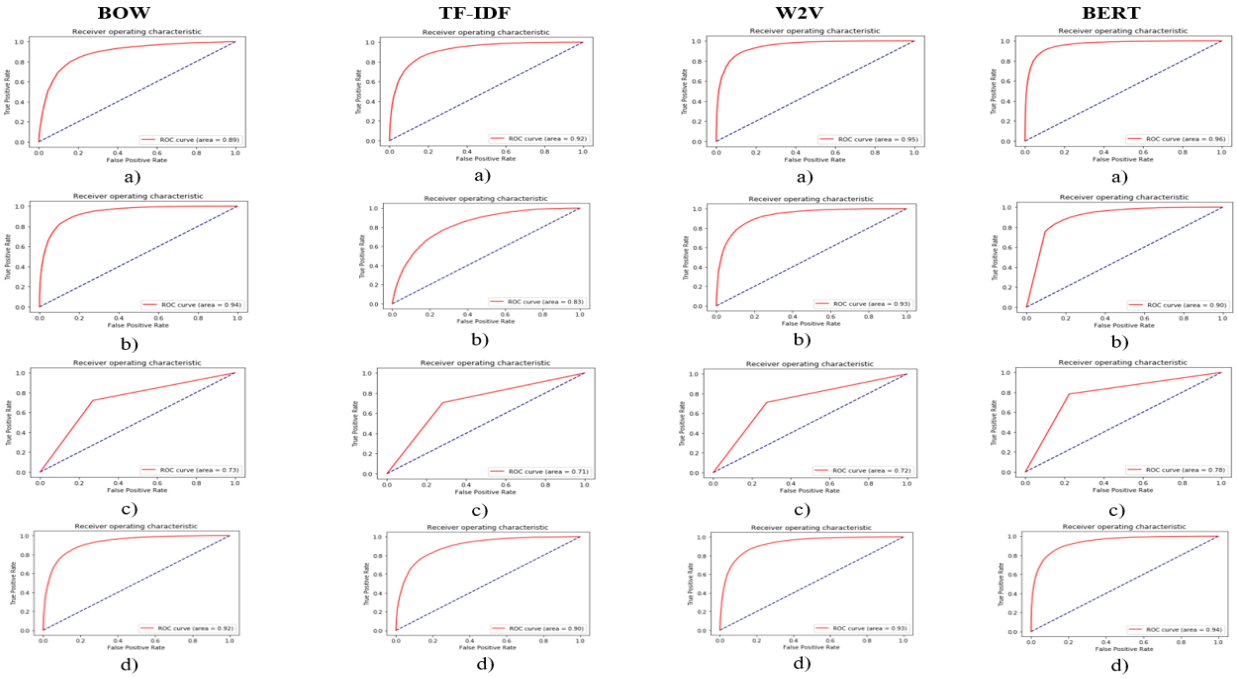

Figure 8. ROC Results of a) SVM b) NB c) DT d) RF Models (a) SVM b) NB c) DT d) RF Modellerinin ROC Sonuçları)

As illustrated in Figure 8, the area under the curve (AUC) values for all models fall within the range of 0.71 to 0.96. This indicates that the models exhibit excellent performance.

7. CONCLUSION AND DISCUSSION (SONUÇ VE TARTIŞMA)

In this study, we applied NLP techniques to perform emotional analysis on the comments of a balanced dataset collected from IMDB. The comments were classified as positive and negative. The results of this study show that BERT is the best performing text feature extraction method for text classification problems. Although BoW and W2V showed competitive results in some models, TF-IDF showed lower overall performance. Similar studies on word representation with similar content and ML-generated models are shown in Table 3.

Table 3. Similar studies on the IMDB dataset (IMDB veri kümesi üzerinde benzer çalışmalar)

	ML Methods	Text Representation	ACC
[2]	MLP	BoW	0.8667
[3]	MLP	TF-IDF	0.8610
[4]	RNN	TF-IDF	0.88
[5]	MLP	TF-IDF	0.8867
[6]	BiLSTM	Embedding	0.8821
[7]	RF	TF-IDF	0.8366
[8]	SVM	TF-IDF	0.90
The present	SVM	BERT	**0.9033**

As can be seen in Table 3, the BERT method used in our study achieved the highest result with an accuracy rate of 0.9033, which is higher than other studies, when compared to studies with similar content in the literature. This success was especially achieved with the SVM model. Other text feature extraction methods such as BoW, TF-IDF, and W2V also showed high performance, but were not as effective as this method. These results show that the method provides superior performance in sentiment analysis problems and is more successful in text classification problems compared to other methods.

The results obtained show that the methods used are highly effective. Text feature extraction with BERT gave the highest performance in all models and especially the SVM model gave the best results (AUC = 0.9638). This shows that NLP techniques and especially ML models used in combination with BERT provide high accuracy and reliability in text classification problems.

However, using larger datasets for similar studies in the future and experimenting with different NLP techniques may help to further improve the results. In particular, methods such as BoW, TF-IDF, and W2V were found to be effective in certain scenarios. To increase the generalizability of the model, it may be useful to investigate different datasets, for example, datasets specific to different movie genres or language groups.

This study highlights ML techniques and shows that NLP techniques can be used effectively in emotion analysis. These techniques have the potential to be a valuable tool in industrial applications. At the same time, the results of this study provide practical implications that can be used in academic research and commercial applications.

In addition to the superior performance of BERT, it was observed that other feature extraction methods can be effective in certain models. This suggests that each model and method combination should be carefully evaluated in specific application areas. For example, the better performance of BoW in NB models proves that simpler methods can also work in certain situations.

In conclusion, this study shows that BERT and SVM perform well in text classification and that NLP techniques can be very powerful tools when combined with the right model and feature extraction methods. These results lay the foundation for future work and pave the way for more in-depth research in the field of text analysis.

It is crucial for future studies to validate these findings using different NLP models and more diverse datasets. It would also be useful to study how these methods perform in real-world applications, such as customer feedback analysis or social media sentiment analysis. This will be crucial in assessing the effectiveness and applicability of the models in practice.

DECLARATION OF ETHICAL STANDARDS (ETİK STANDARTLARIN BEYANI)

The author of this article declares that the materials and methods they use in their work do not require ethical committee approval and/or legal-specific permission.

Bu makalenin yazarı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

AUTHORS' CONTRIBUTIONS (YAZARLARIN KATKILARI)

***Veysel GÖÇ*:** Preprocessing the dataset, data analysis, experiments and evaluations, manuscript draft preparation.

Veri setinin ön işlenmesi, veri analizi, deneyler ve değerlendirmeler, makale taslağının hazırlanması.

***Muhammet Sinan BAŞARSLAN*:** Defining the methodology, conceptualization, evaluations of the results and draft editing.

Metodolojinin tanımlanması, kavramsallaştırma, sonuçların değerlendirilmesi ve taslak düzenleme.

CONFLICT OF INTEREST (ÇIKAR ÇATIŞMASI)

There is no conflict of interest in this study.

Bu çalışmada herhangi bir çıkar çatışması yoktur.

REFERENCES (KAYNAKLAR)

[1] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 142–150, 2011.

[2] Z. Shaukat, A. A. Zulfiqar, C. Xiao, M. Azeem, and T. Mahmood, "Sentiment analysis on IMDB using lexicon and neural networks," SN Appl Sci, vol. 2, no. 2, p. 148, Feb. 2020, doi: 10.1007/s42452-019-1926-x.

[3] O. Kaynar, Y. Görmez, M. Yldz, and A. Albayrak, "Makine öğrenmesi yöntemleri ile Duygu Analizi," in International Artificial Intelligence and Data Processing Symposium (IDAP'16), 2016, pp. 17–18.

[4] K. Amulya, S. B. Swathi, P. Kamakshi, and Y. Bhavani, "Sentiment Analysis on IMDB Movie Reviews using Machine Learning and Deep Learning Algorithms," in 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), IEEE, Jan. 2022, pp. 814–819. doi: 10.1109/ICSSIT53264.2022.9716550.

[5] A. Misini, A. Kadriu, and E. Canhasi, "Albanian Authorship Attribution Model," in 2023 12th Mediterranean Conference on Embedded Computing (MECO), IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/MECO58584.2023.10155046.

[6] M. S. Basarslan and F. Kayaalp, "Sentiment Analysis with Various Deep Learning Models on Movie Reviews," in 2022 International Conference on Artificial Intelligence of Things (ICAIoT), IEEE, Dec. 2022, pp. 1–5. doi: 10.1109/ICAIoT57170.2022.10121745.

[7] M. Mohaiminul and N. Sultana, "Comparative Study on Machine Learning Algorithms for Sentiment Classification," Int J Comput Appl, vol. 182, no. 21, pp. 1–7, Oct. 2018, doi: 10.5120/ijca2018917961.

[8] S. N. Başa and M. S. Basarslan, "Sentiment Analysis Using Machine Learning Techniques on IMDB Dataset," in 2023 7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/ISMSIT58785.2023.10304923.

[9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," May 2002.

[10] Y. Kim and O. Zhang, "Credibility Adjusted Term Frequency: A Supervised Term Weighting Scheme for Sentiment Analysis and Text Classification," May 2014.

[11] L. Richardson, "Beautiful Soup Documentation Release 4.4.0," 2019.

[12] J. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, "Distributed representations of words and phrases and their compositionality," in Advances in Neural Information Processing Systems, 2013. [Online]. Available: https://proceedings.neurips.cc/paper

[13] M. S. Başarslan and F. Kayaalp, "Sentiment analysis of coronavirus data with ensemble and machine learning methods," Turkish Journal of Engineering, vol. 8, no. 2, pp. 175–185, Apr. 2024, doi: 10.31127/tuje.1352481.

[14] M. B. Çakı and M. S. Başarslan, "Classification of fake news using machine learning and deep learning", Journal of Artificial Intelligence and Data Science, vol. 4, no. 1, pp. 22–32, 2024.

[15] P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, Aug. 2018, pp. 1–6. doi: 10.1109/ICCUBEA.2018.8697857.

[16] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, Y. Fauziah, and N. H. Cahyana, "Automated Text Annotation Using a Semi-Supervised Approach with Meta Vectorizer and Machine Learning Algorithms for Hate Speech Detection," Applied Sciences, vol. 14, no. 3, p. 1078, Jan. 2024, doi: 10.3390/app14031078.

[17] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), IEEE, May 2017, pp. 900–903. doi: 10.1109/UKRCON.2017.8100379.

[18] M. P. LaValley, "Logistic Regression," Circulation, vol. 117, no. 18, pp. 2395–2399, May 2008, doi: 10.1161/circulationaha.106.682658.

[19] M. H. L. Louk and B. A. Tama, "Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system," Expert Syst Appl, vol. 213, p. 119030, Mar. 2023, doi: 10.1016/j.eswa.2022.119030.

[20] L. Breiman, "Random Forests," Mach Learn, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[21] M. Z. Khaliki and M. S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN," Sci Rep, vol. 14, no. 1, p. 2664, Feb. 2024, doi: 10.1038/s41598-024-52823-9.

[22] T. Öztürk, Z. Turgut, G. Akgün, and C. Köse, "Machine learning-based intrusion detection for SCADA systems in healthcare," Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 11, no. 1, p. 47, Dec. 2022, doi: 10.1007/s13721-022-00390-2.

[23] H. A. Ardaç and P. Erdoğmuş, "Question answering system with text mining and deep networks," Evolving Systems, May 2024, doi: 10.1007/s12530-024-09592-7.