DOI: 10.38061/idunas.1596100

IDUNAS	NATURAL & APPLIED SCIENCES JOURNAL	2025 Vol. 8 No. 1 (18-25)

Sentiment Analysis and Automatic Response Generation for E-Commerce Comments

Research Article

Ayşe Macit[1], Seda Postalcıoğlu[2*]

[1] Izmir Demokrasi University, Department of Management Information Systems, İzmir, Türkiye
[2] Izmir Demokrasi University, Department of Computer Engineering, İzmir, Türkiye

Author E-mails:
aysemaciy@gmail.com
seda.postalcioglu@idu.edu.tr
S. Postalcıoğlu ORCID ID: 0000-0002-3188-8116
A. Macit ORCID ID: 0009-0004-6074-2257

*Correspondence to: Seda Postalcıoğlu, Izmir Demokrasi University, Department of Computer Engineering, İzmir, Türkiye
DOI: 10.38061/idunas.1596100

Received: 04.12.2024.; Accepted: 23.12.2024

Abstract

This study addresses the use of machine learning techniques for automatic classification of product reviews on e-commerce sites and generating appropriate responses. It was carried out with approximately 15,000 data labeled as positive, negative and neutral obtained from the "E-Commerce Product Reviews" data set. The TF-IDF vectorization method, which is a text mining technique, was used in the study. Multinomial Naive Bayes, Support Vector Machine, Random Forest, Logistic Regression techniques were used for sentiment analysis. As a result of the studies, the accuracy values of Multinomial Naive Bayes, Support Vector Machine, Random Forest, Logistic Regression algorithms were 87%, 88%, 85% and 88%, respectively. In conclusion, automatic comment analysis tools can significantly improve customer relations for e-commerce sellers.

Keywords: Sentiment Analysis, Text Mining, Automatic Classification

1. INTRODUCTION

Product reviews on e-commerce sites are one of the critical elements that influence customers' shopping decisions. These reviews are of great importance not only for potential buyers but also for sellers in terms of product improvement, marketing and sales strategies, customer relations, credibility and reputation. In this context, studies on categorizing online user reviews with sentiment analysis have made valuable contributions to research in the field of e-commerce.

In particular, the study by Li et al. (2020) made a significant contribution to sentiment analysis in e-commerce reviews. In this study, a new model for sentiment analysis, SLCABG, is proposed based on the real book reviews of "dangdang.com", a Chinese e-commerce website. This model is based on sentiment dictionary. The SLCABG model combines the advantages of sentiment dictionary and deep learning technology and overcomes the shortcomings of the sentiment analysis model of product reviews. In addition, Tuzcu's (2020) applied Multilayer Perceptron (MLP) algorithm using Python programming language for sentiment analysis of online user comments and obtained successful results. There are many studies about this issue.

This study aims to apply machine learning techniques to e-commerce comments, building upon the methods of previous research. However, manually analyzing and responding to comments in large data sets is a time-consuming and challenging task. To alleviate this challenge, this study aims to automatically classify e-commerce comments and generate meaningful responses. After preprocessing steps on text data, a sentiment analysis model was built using TF-IDF vectorization and supervised learning models and then evaluated on test data. The model is used to generate automatic responses for new comments.

2. MATERIALS AND METHODS

Sentiment analysis is an artificial intelligence technology for understanding human emotions from data such as text, audio or image. Text-based sentiment analysis is widely used. This technology is used to identify emotional content in a text document. Sentiment analysis usually classifies emotional content in texts as positive, negative or neutral. It is usually performed with "Natural Language Processing" techniques. Machine learning algorithms are trained to associate certain words, phrases or sentences with certain emotions. Natural Language Processing (NLP) is a machine learning technology that gives computers the ability to interpret, process and understand language. It enables computers to comprehend human language.

As Çelik and Koç (2021) mention that the TF-IDF vectorization method is a feature engineering technique frequently used in text mining. For machine learning or deep learning, verbal expressions need to be made meaningful and expressed numerically. TF-IDF is a statistical measure that indicates how representative a word is within a document. "TF" stands for term frequency. It shows the frequency of the word in the document, i.e. the number of times the word occurs in the document. "DF" stands for document frequency, i.e. the number of documents divided by the number of occurrences of the word. "IDF" is the inverse document frequency, which is the logarithm of DF.

Supervised learning is a type of machine learning technique. Supervised learning, the data set used to train the algorithm contains labelled data (Onan and Korukoğlu, 2016). It is based on a machine learning model trying to learn the input and output relationships in the training data set. For each example, the training dataset contains inputs (features) and target outputs (labels or responses) corresponding to these inputs. Supervised learning uses various algorithms to model the relationship between inputs and outputs.

Multinomial Naive Bayes is a text classification algorithm in which features are classified and correlated with word counts or frequencies in documents (Calis et al., 2013). It aims to determine the class and category of the data presented. It is a supervised learning algorithm used in text classification. It is used in areas such as spam filtering and sentiment analysis. It uses Bayes theorem. Figure 1 shows the naive bayes classification.

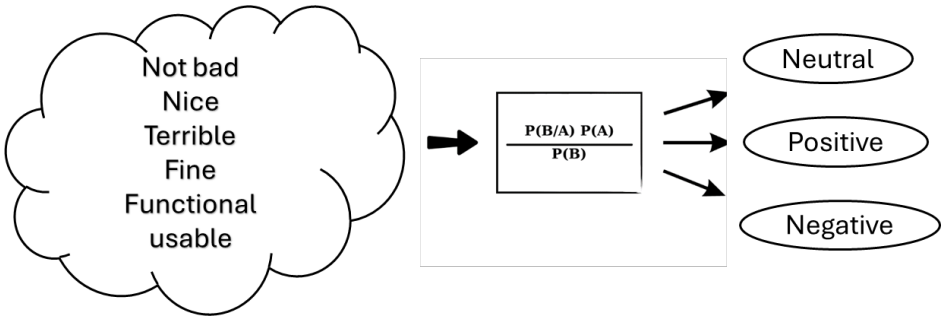

Figure 1. Naive Bayes Classification

Support Vector Machine is a supervised learning method used in classification problems (Ülgen et al., 2017). It draws a line to separate points placed on a plane. It aims for this line to be at the maximum distance for the points of both classes. It is a suitable algorithm for small and medium data sets. Kernel is an important parameter for Support Vector Machine and can take values such as 'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'.

Random Forest is an ensemble learning method that generates a large number of decision trees in the training phase and predicts classes or numbers according to the problem. It is effective on complex and variable data sets. It is based on the principle that many decision trees come together to form a forest. Each tree is trained on data subsets generated by random sampling. Since the trees are trained independently on each of these subsets, they can focus on different features, which increases the generalization ability of the model.

Logistic Regression is a supervised learning algorithm used to predict the probability of a categorical dependent variable. It is a statistical model commonly used to solve classification problems. It focuses on predicting the probability of a dependent variable using a combination of independent variables.

3. DATA SET CHARACTERISTICS

For this study, which aims to automatically classify e-commerce comments and generate meaningful responses using artificial intelligence techniques, the "E-Commerce Product Comments" dataset was used (Çabuk, 2022).

As can be seen in Figure 2, the dataset consists of data classified as negative, positive and neutral. The label distribution helps us to better understand the success rates of the model. This figure shows that there is a balanced distribution of positive and negative labels, but there are not enough examples of neutral data, which may affect the performance of the model.

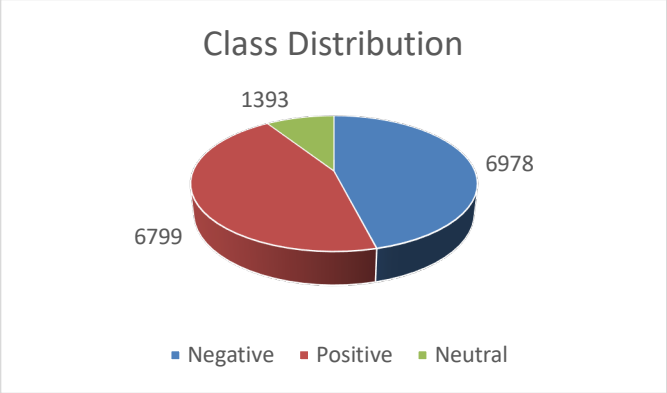

Figure 2. Data Set Class Distribution

In addition to the numerical distribution of the dataset, the textual content is also very important for the study. The structure of the text data may impact the results; therefore, the dataset was preprocessed to improve text processing and make it suitable for the model, ensuring accurate results.

The following steps were applied to the dataset (Ulusoy, 2022) ; Converting Text Data to Lowercase, all text data is converted to lower case, removing special characters, punctuation and special symbols (such as punctuation marks) have been removed from the text. Removing Numbers, removing stop words, extracting the Word root with stemmer.

4. RESULTS AND DISCUSSION

After modifying the dataset, the modeling and testing phase began, using the TF-IDF vectorization method and classifiers such as Multinomial Naive Bayes, Support Vector Machine, Random Forest, and Logistic Regression. The data set was divided into 20% test set and 80% training set. Figure 3 shows the accuracy values of the algorithms used in the project. The Support Vector Machine and Logistic Regression algorithms produced similar results.

Accuracy is calculated with the mathematical expression is shown in equation 1. TP means instances correctly predicted as positive, FP means instances incorrectly predicted as positive, FN means instances incorrectly predicted as negative, TN means instances correctly predicted as negative [9].

$$(TP+TN) / (FP+TP+TN+FN) \quad (1)$$

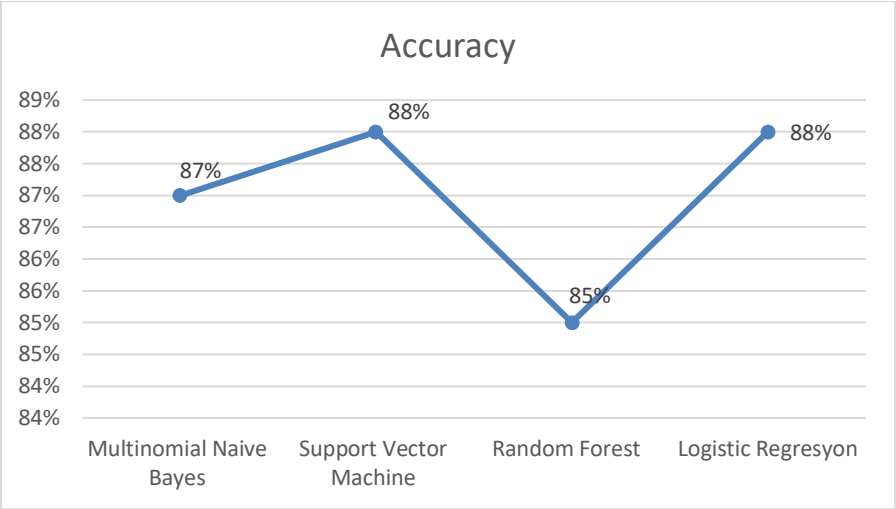

Figure 3. Accuracy rates

Figure 4 shows the precision values of the algorithms used in the study. It is calculated with the mathematical expression is shown in equation (2).

$$TP / (TP + FP) \quad (2)$$

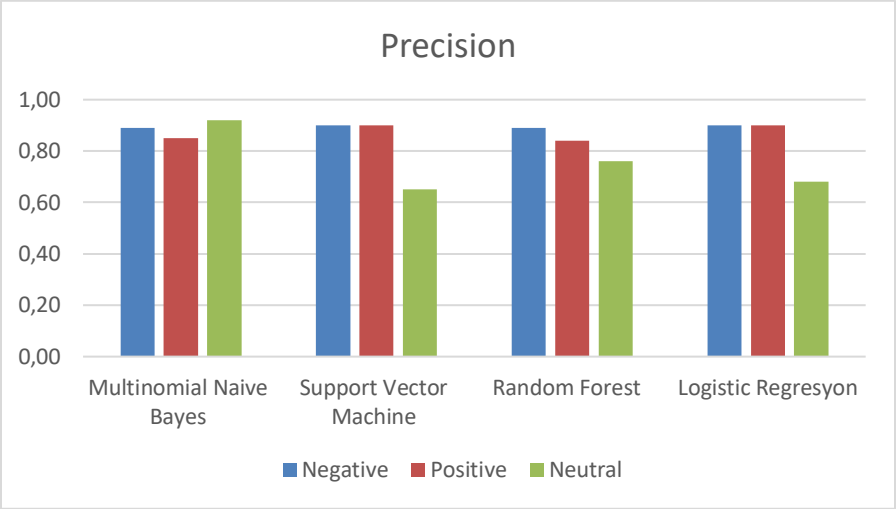

Figure 4. Precision values of the algorithms

Figure 5 shows the recall values of the algorithms used in the study. Mathematical expression of the recall is shown in equation (3).

$$TP/(TP+FN) \quad (3)$$

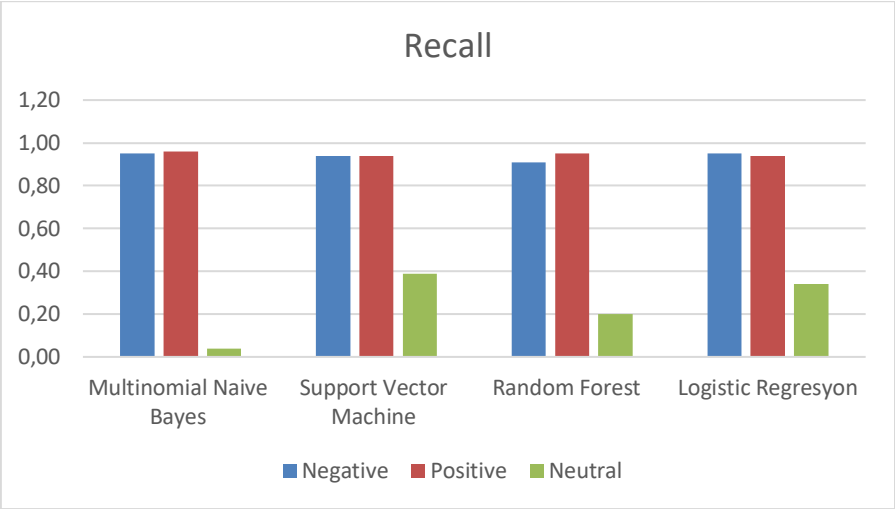

Figure 5. Recall Values of algorithms

The F1 score is the harmonic mean of Precision and Recall metrics. It is calculated with the formula is shown in equation (4) [9].

$$2*(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Figure 6 shows the F1 score values of the algorithms used in the study.

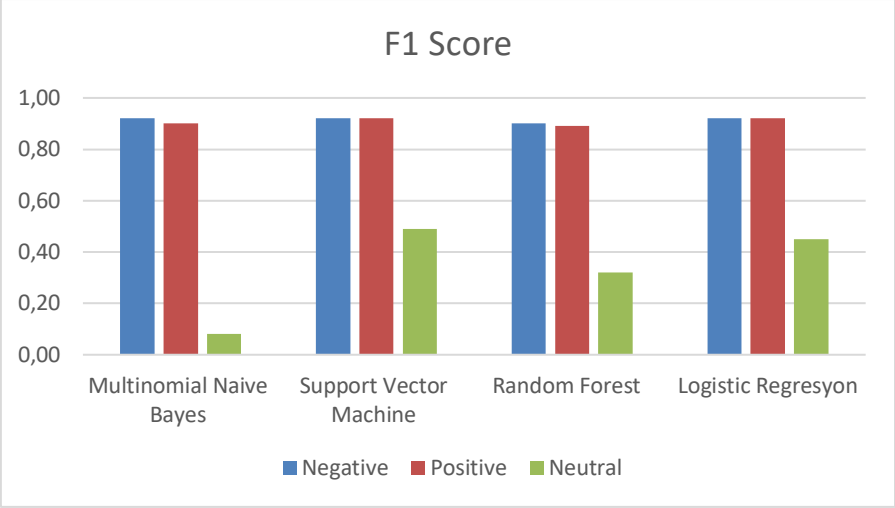

Figure 6. F1 scores of algorithms

After training the model, a randomly selected comment from the dataset was classified, and the automatically generated response corresponding to the predicted sentiment was displayed. The same process was applied to a comment received from the user and the result was printed on the screen.

Figure 7 shows the comments returned from the dataset, the comments received from the user and the automatic responses. “0” represents Negative, ‘1’ Positive, ‘2’ Neutral data.

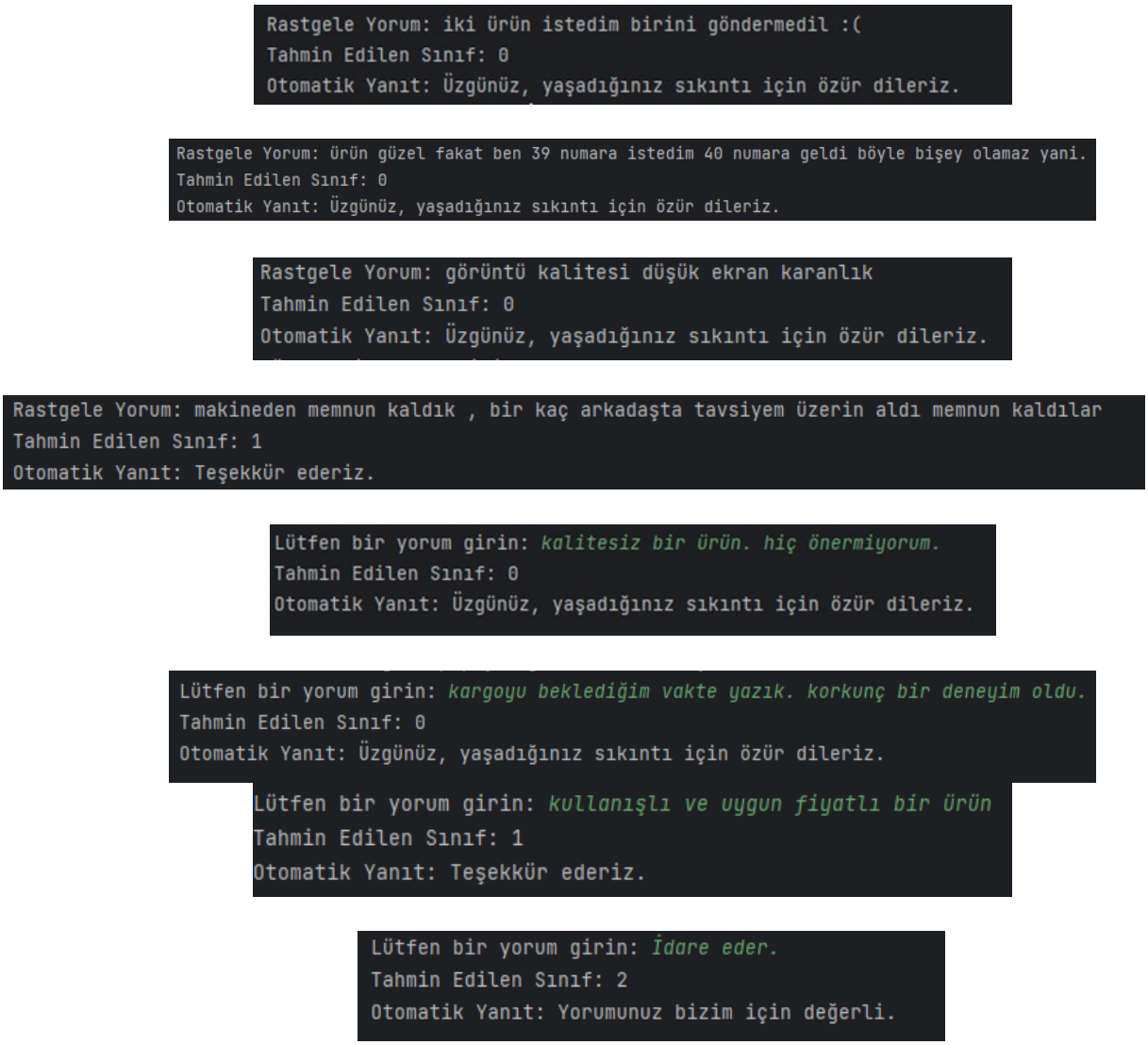

Figure 7. The comments results

CONCLUSION

Multinomial Naive Bayes, Support Vector Machine, Random Forest, Logistic Regression algorithms and the accuracy values are 87%, 88%, 85% and 88% respectively. For the neutral class, where sample sizes were insufficient, the Support Vector Machine algorithm produced the best results. The performance of the study can be improved by using larger and more diverse datasets, improving the data preprocessing steps and trying more complex algorithms. In this way, the study can be further developed and used as an industrial product. E-commerce sellers can enhance customer relationships and satisfaction by using such automatic analysis tools.

REFERENCES

1. Ulusoy, B. (2022). Ürün Yorumları Duygu Analizi. [Proje].Kaggle.https://www.kaggle.com/code/barisulusoy/r-n-yorumlar-duygu-analizi (Erişim Tarihi: 30 Ekim 2023)

2. Çabuk, M. (2022). E-Ticaret Ürün Yorumları [Veri seti]. Kaggle.https://www.kaggle.com/datasets/mujdatcabuk/eticaret-urun-yorumlari (Erişim Tarihi: 30 Ekim 2023)

3. Çalış, K., Gazdağı, O., & Yıldız, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. Bilişim Teknolojileri Dergisi, 6(1), 1-10. https://dergipark.org.tr/en/download/article-file/75317

4. Çelik, Ö., & Koç, B. C. (2021). TF-IDF, Word2vec ve Fasttext Vektör Model Yöntemleri ile Türkçe Haber Metinlerinin Sınıflandırılması. Dergi Adı (DEUFMD), 23(67), 121-127. https://dergipark.org.tr/en/download/article-file/940060

5. Onan, A., & Korukoğlu, S. (2016). Makine Öğrenmesi Yöntemlerinin Görüş Madenciliğinde Kullanılması Üzerine Bir Literatür Araştırması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi / *Pamukkale University Journal of Engineering Sciences*, 22(2), 111-122. https://jag.journalagent.com/pajes/pdfs/PAJES_22_2_111_122.pdf

6. Tuzcu, S. (2020). Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. ESTUDAM Bilişim, 1(2), 1–5.

7. Ülgen Oğul, İ., Özcan, C., & Hakdağlı, Ö. (2017). Spark Destek Vektör Makinesi ile Metin Sınıflandırma. https://www.researchgate.net/profile/Caner-Ozcan-2/publication/321579721_Text_Classification_with_Spark_Support_Vector_Machine/links/5a27b7bba6fdcc8e866e7e35/Text-Classification-with-Spark-Support-Vector-Machine.pdf

8. Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access*, 8, 23522-23530. https://doi.org/10.1109/ACCESS.2020.2969854

9. Postalcioglu, S., (2022). Design of Automatic Tool for Diagnosis of Pneumonia Using Boosting Techniques, Brazilian Archives of Biology and Technology. 2022, v. 65, pp:1-16. doi.org/10.1590/1678-4324-2022210322.