ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 4, No. 2, Decmber 2024
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi

Sentiment Analysis for Udemy Reviews with Natural Language Processing and Machine Learning Methods

Sedat SÖNMEZ[1*], Tanju AÇI[2], Hidayet TAKCI[3], Hakan KEKÜL[4]
[1]Sivas Cumhuriyet University, Institute of Science, Department of Artificial Intelligence and Data Science, Sivas, Türkiye; ORCID: 0009-0009-6271-036X
[2]Sivas Cumhuriyet University, Institute of Science, Department of Computer Engineering, Sivas, Türkiye; ORCID: 0009-0008-5822-3393
[3]Sivas Cumhuriyet University, Faculty of Engineering, Department of Computer Engineering, Sivas, Türkiye; ORCID:0000-0002-4448-4284
[4]Sivas Cumhuriyet University, Faculty of Technology, Department of Software Engineering, Sivas, Türkiye; ORCID: 0000-0001-6269-8713
* Corresponding Author: 20239258007@cumhuriyet.edu.tr

Received: 6 August 2024; Accepted: 16 October 2024

Reference/Atıf: S. Sönmez, T. Açi, H. Takcı, and H. Kekül, "Sentiment Analysis for Udemy Reviews with Natural Language Processing and Machine Learning Methods", Researcher, vol. 04, no. 02, pp. 184–191, 2024.

Abstract

Understanding and classifying sentiment content in textual data is an important requirement for many industries. Text-based data such as social media platforms, customer feedback and product reviews are a rich source of human emotions and opinions. Extracting meaningful information from this text data and understanding the emotional content helps businesses make strategic decisions, develop products and improve their services. Machine learning methods are widely used to perform sentiment analysis on large amounts of text data. These methods are used to process text data, extract features, train models and classify emotional content. Natural language processing techniques are used to solve a range of problems such as increasing an application's user satisfaction, improving its services or optimising marketing strategies. In this study, emotional tones are determined by analysing course comments in the Udemy application. Prediction is made by classifying positive or negative comments. Udemy application comments on Google Play were used and sentiment analysis is performed using K-Nearest Neighbour (KNN) and Random Forest Classification (RFC) algorithms. As a result of the analyses, it was observed that the KNN algorithm predicted with 84% accuracy. Accuracy, F1 Score, Recall, Precision metrics were used as performance measures.

Keywords: Sentiment analysis, Udemy, RFC algorithm, KNN algorithm, machine learning

1. Introduction

Sentiment analysis is essentially a text processing task and aims to identify the emotional expression of given text. Initial studies of sentiment analysis focused on emotional polarity, which aims to classify text as positive, negative or neutral[1].

Sentiment Analysis is a scientific inference method that aims to quickly analyze and report individuals' opinions about social, political events, products, services, brands, etc. in large data sets that have emerged as the internet plays an important role in the daily lives of many segments of society at the global level [2].

Technological developments, individuals' desire to access courses at any time and place they want, as well as the freedom to choose courses with the content they want, have increased the demand for educational platforms such as Udemy, Udacity, KhanAcademy. Increasing demand makes customer satisfaction even more important. Analysing the comments about a course and making a positive or negative evaluation quickly require an effective system. When the literature was analysed, it was seen that there were very few studies on the Udemy application. Thanks to sentiment analysis, we can use someone else's recommendation.

In his study on sentiment analysis in 2020, Tuzcu first applied the Multilayer Perceptron (MLP) algorithm. Using RapidMiner software, Support Vector Machines (SVM), Logistic Regression (LR) and

Naive Bayes (NB) algorithms were applied on the dataset. The results of the analyses were compared and the classification success of the MLP algorithm was found to be 89%. The classification success of positive samples was higher than negative samples [3].

In his study in 2021, Tokcaer reviewed the literature in the field of sentiment analysis in Turkish texts and identified frequently preferred methods, open-source libraries and databases; and also discussed important issues in terms of research. The findings revealed that open-source libraries are widely used in the preprocessing stage of texts. The most frequently used libraries are Zemberek-NLP and ITU NLP Turkish Parser. Research on Turkish texts has shown that binary or ternary word groups, i.e. n-gram approaches, improve the analysis performance. However, it should be noted that punctuation marks and unrelated words remain as attributes that can contribute positively to the analysis performance. It is observed that statistical models such as Chi-Square and Information Gain are widely used in the feature selection phase [4].

In their study conducted in 2020, İlhan and Sağaltıcı performed sentiment analysis using machine learning algorithms on the classified comment data obtained from X. They performed sentiment analysis with the N-gram method and compared the performances of methods such as NB and SVM, which are frequently used in this analysis. As a result of the evaluations, it was determined that the SVM classifier had the highest performance [5].

In their study conducted in 2023, Daşgın and Âdem evaluated the positive and negative reviews of the courses using machine learning and deep learning methods on the Udemy application. They used BayesNet, J48, OneR and SVM among the classical machine learning algorithms and found the highest accuracy rate with 91.57% from the BayesNet algorithm. They also applied Random, GloVe and Word2Vec word embedding techniques to the dataset. They tried GRU and CNN-LSTM hybrid architectures from deep learning architectures and obtained the highest accuracy rate with 95.67% from the combination of GloVe word embedding technique and GRU architecture [6].

Şahinaslan et al. classified 15,082 data into positive, negative and neutral emotion classes using NB classification algorithm in their study conducted in 2022. As a result of their WEKA analysis, they found that the classification success rate was 65.56%. For the analysis, they used comment data from many world languages on the YouTube platform. Their study supports the idea that people's emotions and thoughts can be analysed by sharing their opinions, thoughts and satisfaction on social media platforms [7].

2. Recommended Method

Within supervised machine learning algorithms, there are algorithms that do not take much time and perform well, where predictions are made based on the similarities of observations to each other. The main purpose of classification is to find out which class the objects belong to by looking at their properties. In this study, KNN and RFC algorithms are used because they are simple and effective.

2.1 K-Nearest Neighbour Algorithm

KNN is one of the classification algorithms and is a simple and effective method used to solve classification problems. The algorithm performs classification operations by using the proximity of a given feature to another feature that is closest to it [8].

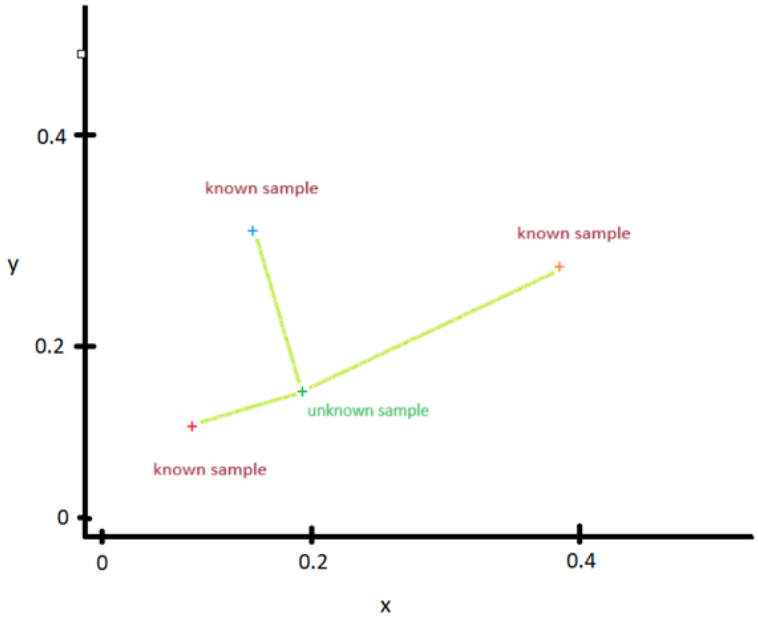

Figure 1: KNN algorithm

If the graph in Figure 1 is interpreted, the distances of the three points close to the unknown sample (green point) are analysed. If the closest proximity of the known red point is confirmed, it is inferred that the unknown sample is also red. It is based on the assumption that your best friend is red, so you are red too. In the KNN algorithm, the number of K nearest neighbours is taken into account instead of one nearest neighbour. Commonly used distance measures are:

Euclidean Distance: As seen in Figure 2, the distances of the points neighbouring the unknown point are calculated by using the Pythagorean theorem used in mathematics. For this, Equation 1 is used according to the x and y coordinates of the points.

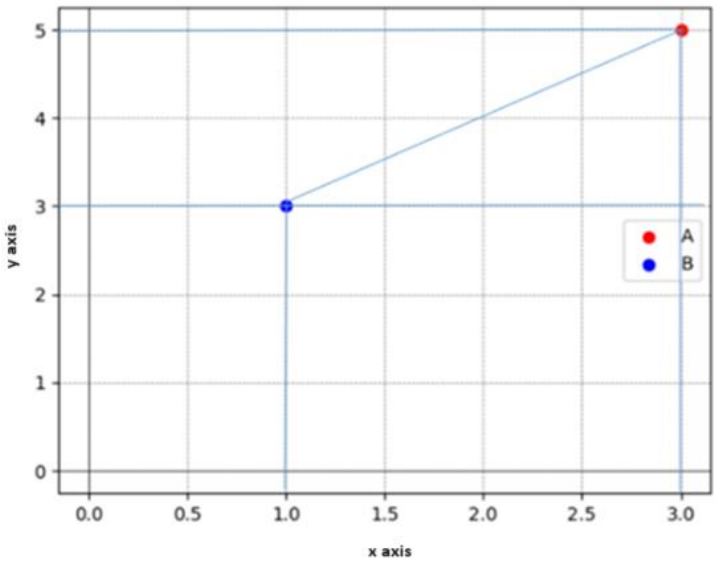

Figure 2: Calculation of Euclidean distance

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Manhattan (City-block) Distance: It is obtained by taking the absolute values of the differences of the coordinates of the sample points in the data set and summing these values. Equation 2 is used in this measurement method.

$$\boldsymbol{d(i,j) = \sum_{n=1}^{m} |x_{in} - x_{jn}|} \quad (2)$$

Minkowski Distance: The Minkowski distance is a measurement in Euclidean space. It is created by generalising the Euclidean distance and Manhattan distance. When p=2, the Minkowski distance is equal to the Manhattan distance. When p=1, the Minkowski distance is equal to the Euclidean distance as shown in Equation 3.

$$\boldsymbol{d(i,j) = \left(\sum_{n=1}^{m} |x_{in} - x_{jn}|^{\rho} \right)^{1/p}} \quad (3)$$

Compliance with some standards determined during the determination of the K value is beneficial for the model to give healthy results. At the same time, the maximum value of K should not be greater than the square root of the number of samples (n) in the data set.

In our study, 3, 5 and 7 values for the K parameter were tested. The most successful K value was found to be 3.

2.2 Random Forest Classification Algorithm

RFC algorithm is one of the supervised classification algorithms and can be used in regression and classification problems. This algorithm aims to increase the accuracy in the classification process by creating multiple decision trees. The RFC algorithm works in the form of multiple decision trees that work independently of each other, coming together and selecting the value with the highest score. It is possible to show this process in Figure 3. As the number of trees increases, the probability of obtaining the correct result increases. The main difference with the decision tree (DT) algorithm is that the root node is found and the division of nodes is randomised [9].

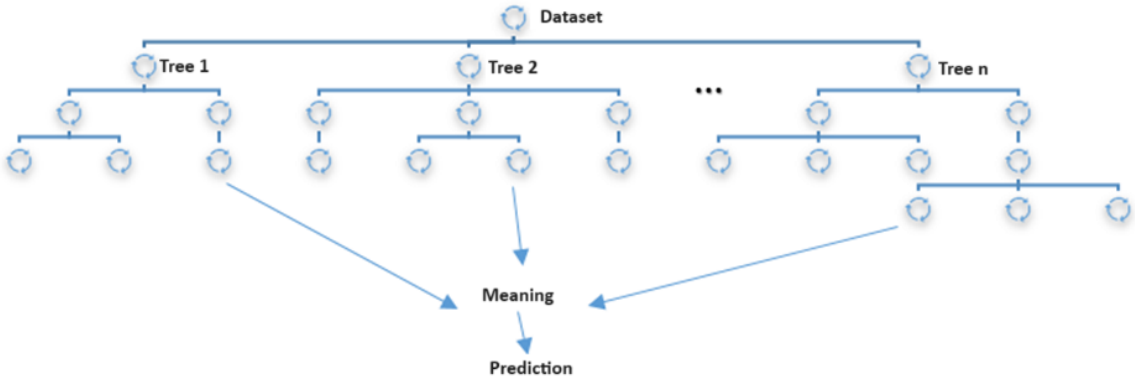

Figure 3: Finding the root node in the RFC algorithm

In our study, the value expressing the depth of decision tree generation on the algorithm was set as 2 and the results were modelled accordingly.

No feature extraction was performed on the original data set. Empty values and repetitive rows were not encountered during the data preparation phase.

3. Results and Experimental Study

In this study, an original dataset was created by extracting user comments from a training platform application, and after some preprocessing, emotion classifications were made with many classification algorithms and the results were compared. Python programming language and Spyder software were used to implement the applications.

3.1 Dataset

The original dataset was collected automatically by Web Scraping method using Python's selenium library from the application named Udemy. It contains comments of 880 users. 880 rows consist of 2 columns. A balanced dataset consisting of 400 positive and 400 negative comments was created by emotion classification on the comments taken from the dataset. An example from the dataset is given in Table 2.

Table 1: A sample from the original data set

Sn.	Comment	Score	Content
0	Don't use the app!! Use your browser. If you….	0	use app use browser use app good luck use control use
1	I wish that I could be in control of the search. The filter is disabled for …	0	wish could control search filter disable odd…
2	The content is great for when it works, however I've only been able…	0	content great work however able watch video...
3	Has it all. Just wish there was a. option to go strictly audio…	1	wish option strictly audio applicable course…
4	Mostly good. It doesn't work in split screen mode on my table though…	0	mostly good work split screen mode table though…
5	This app has degraded itself to be designed to sell…	0	app degrade design sell course rather medium actually watch..

3.2 Text Preprocessing

For text analysis, the interpretations to be used should undergo certain pre-processing steps. In this study, the following were done as data pre-processing steps;

- ✓ Empty lines were deleted.
- ✓ All characters other than the letters between A-Z (instead of punctuation marks) were replaced with the space character.
- ✓ English words that do not make sense on their own (the, a, or, and, etc.) were separated from the data set.
- ✓ Words have been reduced to their roots.

TF-IDF method was preferred as the term weighting method. TF-IDF is obtained by multiplying the term frequency (TF) and inverse document frequency (IDF) values of the terms. In this way, the text data were digitised. As a result of the digitisation process, the data were normalised.

3.3. Classification

In this study, model applications on the original dataset with KNN and RFC algorithms were performed using Pandas, NumPy and Scikit-learn (sklearn) libraries in Python programming language. Pandas was used for data storage and processing, NumPy library for mathematical operations and Scikit-learn (sklearn) library for creating, training, testing and evaluating machine learning models. These libraries were used to cover all stages of the machine learning model starting from the data processing process.

The algorithm performed the training process by dividing the original data set into 10 parts. This approach aimed to avoid memorisation by training the data set in parts instead of the whole data set. In this way, it is aimed to achieve a more generalised learning and a better understanding of the patterns in the data set.

In the training process, the model was trained on a dataset created by scanning the 300 most frequently used words in each comment fragment. This approach prevented the inclusion of meaningless or unnecessary words in the learning process.

Accuracy, F1 Score, Recall and Precision metrics were used as performance measures. The performance evaluation metric results obtained in this direction are shown in Table 2 and Table 3.

Table 2: KNN Algorithm performance evaluation results

Metrics	**K values**		
	3	**7**	**10**
Accuracy Score	**0.840**	0.803	0.768
F1 Score	**0.842**	0.796	0.751
Recall Score	**0.818**	0.745	0.672
Precision Score	**0.868**	0.857	0.852

Table 3: RFC Algorithm performance evaluation results

Metrics	**Accuracy Score**	**F1 Score**	**Recall Score**	**Precision Score**
Results	0.758	0.736	0.650	0.848

The Receiver Operating Characteristic Curve (ROC curve) was also used to evaluate the performance of the learning algorithm. The ROC curve, which is a preferred evaluation tool especially in classification problems, is a graph showing the relationship between sensitivity and specificity. This graph is used to evaluate the accuracy performance of the model at different thresholds and is an important visual tool to understand the success of the classification model [10].

Sensitivity: As seen in Equation 4, it refers to the true positive rate. That is, the ratio of true positives to the total number of positive samples. Sensitivity measures how good a model is for not missing true positives.

$$Sensitivity = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Negatives + Number\ of\ False\ Positives} \quad (4)$$

Specificity: As seen in Equation 5, it refers to the true negative rate. That is, the ratio of true negatives to the total number of negative samples. Specificity measures how good a model is for not missing true negatives.

$$Specificity = \frac{Number\ of\ True\ Negatives}{Number\ of\ True\ Negatives + Number\ of\ False\ Positives} \quad (5)$$

ROC Curve: X axis represents specificity. Y axis represents sensitivity. The ROC curve shows the sensitivity and specificity performance of the model at different probability thresholds. As the curve moves from the upper left corner to the lower right corner, it shows that the performance of the model improves.

Area Under the Curve (AUC): It refers to the area under the ROC curve. The AUC value is used as a metric that measures the classification performance of the model with a single number. The closer the AUC value is to 1, the better the model performs.

The ROC curve and AUC are widely used to evaluate model performance, especially in imbalanced classification problems. A large area under the curve indicates a good classification ability of the model. The ROC curve graph of the KNN and RFC algorithms models is given in Figure 4.

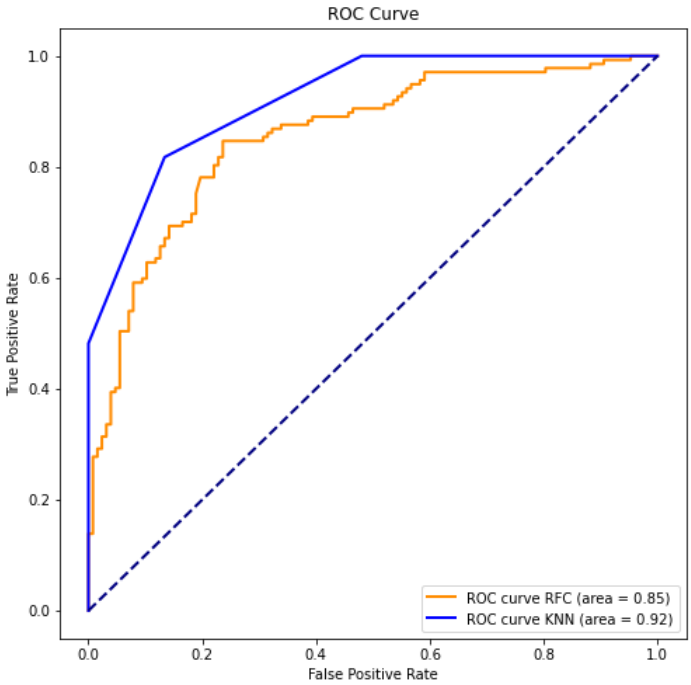

Figure 4: ROC curves of KNN-RFC algorithms

According to the ROC curve metric of the KNN algorithm, the area under ROC was calculated as 0.92.

According to the RFC algorithm ROC curve metric, the area under ROC is calculated as 0.85.

4. Conclusion

In this study, sentiment analysis was performed with various machine learning techniques on an original dataset containing course reviews on Udemy. Since our dataset is original and no previous work has been done, it has not been compared with previous studies. RFC and KNN classifiers, which are commonly used classification algorithms, were used. The performance of these classifiers was compared using various model performance measures. In the study, 3, 5 and 7 values for the K parameter were tested and the most successful K value was determined as 3. According to the performance criteria analysed, the highest performance was obtained from the KNN algorithm with a rate of 84%. With piecewise training, the model gained better generalisation ability and became more successful on real world data.

Contribution of Researchers

All researchers have contributed equally to writing this paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] S. E. Şeker, 'Duygu Analizi (Sentimental Analysis)', *YBS Ansiklopedi*, vol. 3, no. 3, pp. 21–36, 2016.

[2] O. Kaynar, Y. Görmez, M. Yildiz, and A. Albayrak, 'Makine Öğrenmesi Yöntemleri ile Duygu Analizi', *Int. Artif. Intell. Data Process. Symp.*, pp. 234–241, 2016, [Online]. Available: https://www.researchgate.net/publication/311136507

[3] S. Tuzcu, 'Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması', *Eskişehir Türk Dünyası Uygul. ve Araştırma Merk. Bilişim Derg.*, vol. 1, no. 2, pp. 1–5, 2020, Accessed: Nov. 26, 2023. [Online]. Available: https://dergipark.org.tr/tr/pub/estudambilisim/issue/53654/676052

[4] S. Tokcaer, 'Türkçe Metinlerde Duygu Analizi', *J. Yaşar Univ.*, vol. 16, no. 63, pp. 1516–1536, 2021, doi: 10.19168/jyasar.928843.

[5] N. İlhan and D. Sağaltıcı, 'Twitter'da Duygu Analizi', *Harran Üniversitesi Mühendislik Derg.*, vol. 5, no. 2, pp. 146–156, Aug. 2020, doi: 10.46578/humder.772929.

[6] R. Daşgın and K. Adem, 'Kitlesel Çevrimiçi Ders Platformlarında Kurslara Yapılan Yorumların Metin Madenciliği Kullanılarak Duygu Analizi', *Uluslararası Muhendis. Arastirma ve Gelistirme Derg.*, vol. 15, no. 2023, pp. 636–646, 2023, doi: 10.29137/umagd.1266187.

[7] Ö. Şahinaslan, H. Dalyan, and E. Şahinaslan, 'Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi', *Bilişim Teknol. Derg.*, vol. 15, no. 2, pp. 221–229, Apr. 2022, doi: 10.17671/gazibtd.999960.

[8] D. Kılınç, E. Borandağ, F. Yücalar, V. Tunalı, M. Şimşek, and A. Özçift, 'KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi', *Marmara Fen Bilim. Derg.*, vol. 28, no. 3, 2016, doi: 10.7240/mufbed.69674.

[9] D. Akmaz, E. Elektronik Mühendisliği, M. Fakültesi, and M. Üniversitesi, 'Stockwell Dönüşümü, ONE-R Özellik Seçme Yöntemi ve Rastgele Orman Algoritması ile Güç Kalitesi Bozulumu Sinyallerinin Sınıflandırılması', *Fırat Univ. J. Eng. Sci.*, vol. 34, no. 1, pp. 267–276, Mar. 2022, doi: 10.35234/FUMBD.1018023.

[10] F. TAŞDEMİR, 'ROC Analizi ve R Yazılımı ile Verilerin Sınıflama Doğruluklarının Karşılaştırılması', *Ulus. Eğitim Akad. Derg.*, vol. 6, no. 2, pp. 230–241, Oct. 2022, doi: 10.32960/uead.1166987.