Sentiment Analysis on Student Feedback Using Natural Language Processing

S. Alagukumar
Assistant Professor,
Department of Computer Applications,
Ayya Nadar Janaki Ammal College,
Sivakasi – 626 124
Tamil Nadu, India

R. Vengateshkumar
Associate Professor,
Department of Computer Applications,
Ayya Nadar Janaki Ammal College,
Sivakasi – 626124
Tamil Nadu, India

A. Ananda Gopi
Assistant Professor,
Department of Computer Applications,
Ayya Nadar Janaki Ammal College,
Sivakasi – 626 124
Tamil Nadu, India

ABSTRACT

Text classification is the process of categorizing or labeling text into predefined categories or a category created by the user and has numerous applications in various industries such as sentiment analysis and topic modeling. In this work, it has been analyzed that the student feedback of text data is evaluated with various preprocessing namely, tokenization and word vector matrix and classified using the Multinomial Naïve Bayes algorithm. It can provide valuable insights into how students perceive their learning experience. This work can be used to analyze the emotions and opinions expressed in student feedback, educational institutions can identify areas of improvement and make data-driven decisions to enhance student satisfaction and student support services.

Keywords: Sentiment Analysis, Tokenization, Word Vector Matrix, Preprocessing, Multinomial Naïve Bayes

I.INTRODUCTION

Text classification is the process of categorizing or labeling text into predefined categories or categories created by the user. It is a fundamental task in natural language processing and has numerous applications in various industries such as sentiment analysis, spam detection, and topic modeling. Machine learning algorithms are commonly used for text classification, including support vector machines, Naive Bayes, and Random Forest [1]. Sentiment analysis is a sub-task of natural language processing (NLP) that involves classifying text as positive, negative, or neutral based on its emotional tone. Text classification is the task of categorizing text into predefined categories or labels. Sentiment analysis is a key application of text classification, where the categories are emotions or opinions expressed in the text [2]. Sentiment analysis is a powerful tool used in various applications such as customer feedback analysis, political polling, and social media monitoring. Techniques used for sentiment analysis include machine learning algorithms, and deep learning models. These techniques help to categorize and analyze text data based on its emotional tone, allowing businesses and organizations to make informed decisions. Sentiment analysis of student feedback using machine learning can provide valuable insights into the perspectives and opinions of students. By leveraging natural language processing techniques, these approaches can automatically classify student feedback and can also identify specific topics and themes that are driving the sentiment. It can help instructors and educators to make informed decisions about course design, teaching methods, and student support services, leading to improved learning outcomes and student satisfaction [4]. Section 2, related works are presented. In section 3, Methods are explained. Section 4, describes experimental results and discussion.

II.RELATED WORKS

The most recent innovation in text mining is sentiment analysis. Tarnowska and Ras [5] have analyzed the customers worth to a business is determined not just by their financial impact but also by how satisfied they are. Dissatisfied clients spread negative word-of-mouth whereas happy ones spread positive. Johar and Mubeen [6] have reviewed the sentiment analysis on large scale Amazon product reviews to get the positive and negative comments. Liu [7] has stated that the branch of study known as sentiment analysis and opinion mining examines how people's opinions, sentiments, assessments, attitudes, and emotions are expressed in written language. In addition to being heavily researched in data mining, web mining, and text mining, it is one of the most active

research fields in natural language processing. Singh *et al.* [8] have proposed a method Optimization of sentiment analysis using machine learning classifiers. They have analyzed the four classification methods effectiveness are investigated and contrasted. While OneR appears more promising in producing the accuracy of 91.3% in precision, 97% in F-measure, and 92.34% in correctly classified instances, Nave Bayes was discovered to be fairly quick at learning. Jagdale *et al.* [9] have analyzed sentiment on product reviews using machine learning techniques. They have analyzed and extracted the information from text data from numerous sources, including Facebook, Twitter and Amazon data set. The study shown that the data were collected from Amazon includes product reviews for cameras, laptops, mobile phones and tablets. Hemalatha *et al.* [10] proposed the sentiment analysis on product reviews and results shown that the high accuracy for classifying sentiment on Naive Bayes and maximum entropy classification method. Baid et al. [11] have proposed a sentiment analysis of movie reviews using machine learning techniques using Naïve Bayes, K-Nearest Neighbour and Random Forest and they have obtained the results on Naïve Bayes classifier as 81.45% accuracy, Random Forest classifier as 78.65% accuracy and K-Nearest Neighbour classifier as 55.30% accuracy. Hence from the related works, an effective method is need for the student satisfaction survey and to classify student feedback.

III. METHODOLOGY

The proposed method has three phases, namely, i. tokenization, ii. word vector matrix, and iii. analysis of student feedback using Multinomial Naïve Bayes Classification. The flow of the proposed methodology is explained in the figure 1.

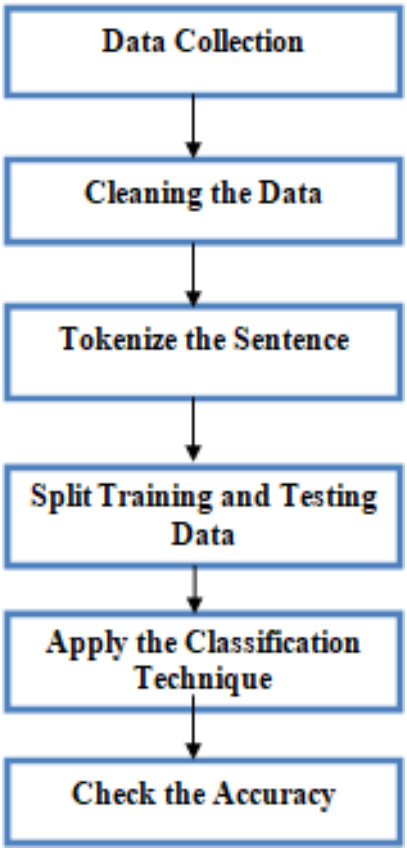


Figure1: Methodology

A. Procedure

Step 1: Read the multiple documents
Step 2: Break the text features into tokens
Step 3: Prepare the word vector matrix of preprocessed data
Step 4: Spilt the data as training and testing
Step 5: Build the model using the Multinomial Naïve Bayes algorithm
Step 6: Identify the student sentiment on feedback information

B. Tokenization

Tokenization [12] refers to the process of breaking down text into smaller units called tokens. This can be done using various techniques, including word-level tokenization, character-level tokenization, and subword-level tokenization. Tokenization is an important step in many natural language processing tasks, such as text classification, sentiment analysis, and machine translation. Word tokenization is the process of breaking down text into individual words or phrases, which can be useful for text classification, sentiment analysis, and language translation. It helps to split the input text into smaller units, enabling the model to understand and analyze the text more effectively.

C. Word Vector Matrix

A word vector matrix [13] is a powerful tool used in natural language processing and machine learning tasks. It represents each word in a vocabulary as a dense vector in a high-dimensional space, capturing the semantic meaning of words and their relationships. This allows for efficient computations, such as word similarities, clustering, and classification, and can be used for various applications like text classification, sentiment analysis, and language modeling. In a word vector matrix, each row represents a word, and each column represents a document or a text sample.

D. Multinomial Naïve Bayes Classification

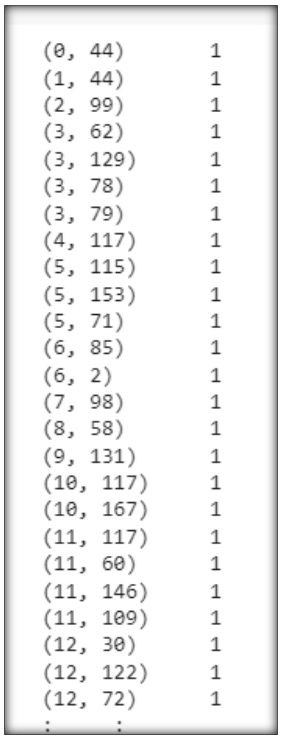
Multinomial Naive Bayes (MNB) classification [14, 15] is a popular machine learning algorithm used for classification tasks when the feature set is large and the class distribution is imbalanced. It is based on Bayes' theorem and assumes that each feature is independent of the others, given the class label. MNB models are simplified versions of the Bayes' theorem, making them computationally efficient and easy to implement. It's a simple and efficient method that assumes independence between features. MNB can also be extended to handle categorical data and missing values. It has the many advantages, such as improved accuracy, robustness to noise and reduced complexity.

IV. RESULTS AND DISCUSSION

In this work, the dataset is collected feedback submitted by the student from the Ayya Nadar Janaki Ammal College, Sivakasi. There are 125 students were submitted the feedback. The work is carried out by using Python. Figure 2 represents the raw data which is submitted by the students.

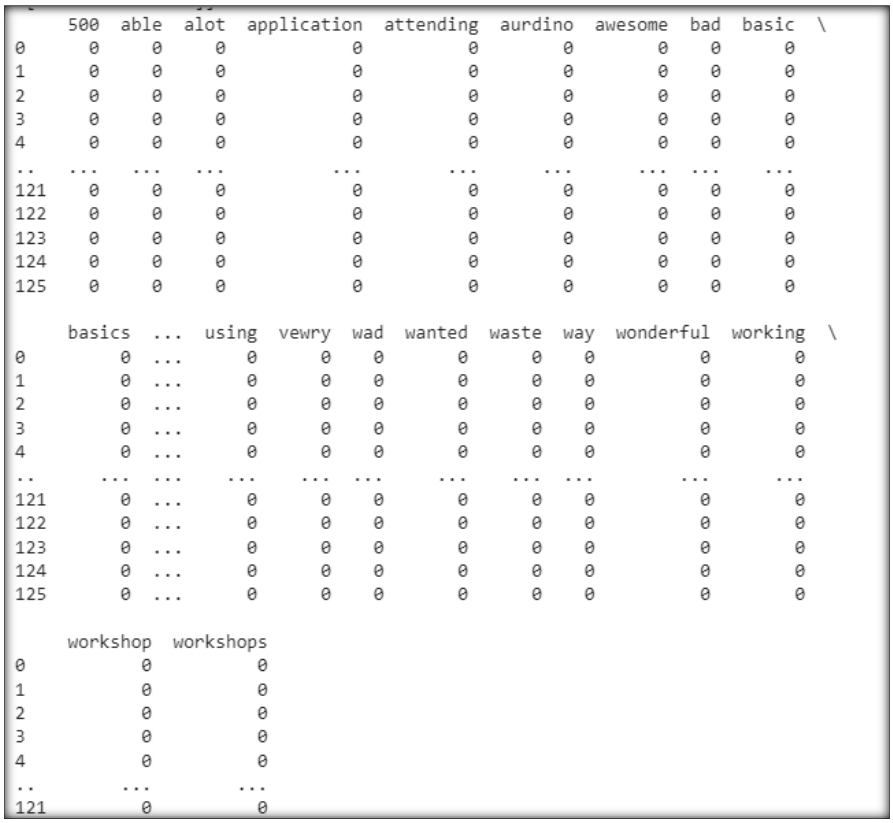
Sentence	Sentiment
It was very interesting to learn	Positive
Usefull	Positive
Thank you so much for the workshop earlier today, so glad I was able to book .	Positive
Time is too short. in this two days only	Negative
Workshop was awesome	Positive
Very useful	Positive
I'm very useful in this session	Positive
Teacher's and students are very kindly	Positive
Very Useful	Positive
Very useful and good teaching Thank	Positive
Very useful	Positive
Workshop wad awesome	Positive
I found the workshop to be very informative. The workshop provided you	Positive
Way of teaching is excellent and I learn	Positive
It is vewry useful for me	Positive

Figure1: Raw Data



```
  (0, 44)       1
  (1, 44)       1
  (2, 99)       1
  (3, 62)       1
  (3, 129)      1
  (3, 78)       1
  (3, 79)       1
  (4, 117)      1
  (5, 115)      1
  (5, 153)      1
  (5, 71)       1
  (6, 85)       1
  (6, 2)        1
  (7, 98)       1
  (8, 58)       1
  (9, 131)      1
  (10, 117)     1
  (10, 167)     1
  (11, 117)     1
  (11, 60)      1
  (11, 146)     1
  (11, 109)     1
  (12, 30)      1
  (12, 122)     1
  (12, 72)      1
  :     :
```

Figure 2: Text Count on Sentiment data



```
     500  able  alot  application  attending  aurdino  awesome  bad  basic  \
0      0     0     0            0          0        0        0    0      0
1      0     0     0            0          0        0        0    0      0
2      0     0     0            0          0        0        0    0      0
3      0     0     0            0          0        0        0    0      0
4      0     0     0            0          0        0        0    0      0
..   ...   ...   ...          ...        ...      ...      ...  ...    ...
121    0     0     0            0          0        0        0    0      0
122    0     0     0            0          0        0        0    0      0
123    0     0     0            0          0        0        0    0      0
124    0     0     0            0          0        0        0    0      0
125    0     0     0            0          0        0        0    0      0

     basics  ...  using  vewry  wad  wanted  waste  way  wonderful  working  \
0         0  ...      0      0    0       0      0    0          0        0
1         0  ...      0      0    0       0      0    0          0        0
2         0  ...      0      0    0       0      0    0          0        0
3         0  ...      0      0    0       0      0    0          0        0
4         0  ...      0      0    0       0      0    0          0        0
..      ...  ...    ...    ...  ...     ...    ...  ...        ...      ...
121       0  ...      0      0    0       0      0    0          0        0
122       0  ...      0      0    0       0      0    0          0        0
123       0  ...      0      0    0       0      0    0          0        0
124       0  ...      0      0    0       0      0    0          0        0
125       0  ...      0      0    0       0      0    0          0        0

     workshop  workshops
0           0          0
1           0          0
2           0          0
3           0          0
4           0          0
..        ...        ...
121         0          0
```

Figure 3: Word Vector Matrix

```
MNB = MultinomialNB()
MNB.fit(X_train, Y_train)

#Caluclating the accuracy score of the model
predicted = MNB.predict(X_test)
accuracy_score = metrics.accuracy_score(predicted, Y_test)
print("Accuracuy Score: ",metrics.accuracy_score(predicted, Y_test))


Accuracuy Score:  1.0
```

Figure 3: Accuracy

First, the raw data are preprocessed by the converting tokens to identify keywords and phrases, of the text. Then the tokenized text data are counted to create the word vector matrix. Figure 3 represents the text count on of the preprocessed data. The process of creating a word vector matrix follows tokenization and converting each word into a vector representation using techniques such as bag-of-words. Figure 4 represents word vector matrix. The resulting word vector matrix can be used for Multinomial Naïve Bayes classification. The result of sentiment analysis is typically such as positive or negative. The proposed method provides 100 % accuracy on Multinomial Naïve Bayes classification.

V.CONCLUSION

In this work, it has been proposed a method to analyze the student feedback of text data is evaluated with various preprocessing namely, tokenization and word vector matrix and classified using the Multinomial Naïve Bayes algorithm. The student feedback data is workshop conducted data can provide valuable insights into the effectiveness of the workshop and the level of satisfaction among participants. From this work, it has been help workshop organizers identify areas of improvement and make necessary changes to enhance the overall learning experience. Proposed method is used to analyze the feedback and organizers can gain a better understanding of what worked well and what didn't. This information is used to improve future workshops for the student’s benefits.

REFERENCES

1. Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150.
2. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.
3. Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.
4. Nasim, Z., Rajput, Q., & Haider, S. (2017, July). Sentiment analysis of student feedback using machine learning and lexicon based approaches. In 2017 international conference on research and innovation in information systems (ICRIIS) (pp. 1-6). IEEE.
5. Tarnowska, K. A., & Ras, Z. W. (2019, January). Sentiment analysis of customer data. In Web Intelligence (Vol. 17, No. 4, pp. 343-363). IOS Press.
6. Johar, S., & Mubeen, S. (2020). Sentiment analysis on large scale Amazon product reviews. IJSRCSE, 8(1), 7-15.
7. Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. Human-centric Computing and information Sciences, 7, 1-12.
8. Liu, B. (2022). Sentiment analysis and opinion mining. Springer Nature.
9. Jagdale, R. S., Shirsat, V. S., & Deshmukh, S. N. (2019). Sentiment analysis on product reviews using machine learning techniques. In Cognitive Informatics and Soft Computing: Proceeding of CISC 2017 (pp. 639-647). Springer Singapore.
10. Hemalatha, I., Varma, G. S., & Govardhan, A. (2013). Sentiment analysis tool using machine learning algorithms. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), 2(2), 105-109.
11. Baid, P., Gupta, A., & Chaplot, N. (2017). Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications, 179(7), 45-49.
12. Ahuja, S., & Dubey, G. (2017, August). Clustering and sentiment analysis on Twitter data. In 2017 2nd International Conference on Telecommunication and Networks (TEL-NET) (pp. 1-5). IEEE.
13. Kasri, M., Birjali, M., & Beni-Hssane, A. (2021). Word2Sent: A new learning sentiment-embedding model with low dimension for sentence level sentiment classification. Concurrency and Computation: Practice and Experience, 33(9), e6149.
14. Raschka, S. (2014). Naive bayes and text classification i-introduction and theory. arXiv preprint arXiv:1410.5329.
15. Xu, S., Li, Y., & Wang, Z. (2017). Bayesian multinomial Naïve Bayes classifier to text classification. In Advanced Multimedia and Ubiquitous Engineering: MUE/FutureTech 2017 11 (pp. 347-352). Springer Singapore.