Sentiment Analysis on Remarks/Comments Written on Student's Report Cards using Naïve Bayes Algorithm

Kimverlie G. Abastar
Master in Information Systems
University of Southern Mindanao

OVERVIEW

Each teacher will agree that having original remarks on report cards is necessary since it enables parents to comprehend their child's development and enable them to take the appropriate action by comprehending the comments on the report cards. This information gathering looked on the possible comments, remarks, or feedback given by the teacher or the advisers to their learners.

It was noted that there are various types of comments given by the teachers. This includes the content comment that is to urge students to read their papers critically, to add and remove content, and to rearrange content as necessary. The other sort of comment, known as a surface comment, concentrated on issues with language usage, spelling, grammar, and/or word choice. More likely, comments and remarks were to students in a positive or encouraging manner continue doing their best effort class. There is also the needs improvement/ negative remarks provides corrections and let students know the areas to be improved.

Since the proponent is currently employed as a secondary school teacher. She was able to acquire the required data by interviewing, and collecting information through written notes from her co-teacher or colleagues. On the other, most of the positive and negative or needs improvement remarks were gathered online. Searching on various website to acquire the required data sets needed for the activity. These data were encoded in the Microsoft Excel with columns for the id, sentiments and the remarks. 225 positive comments were identified, and another 225 remarks for negative/ needs improvement with a total of 450 data sets used in training analysis. Moreover, a total of 50 data sets were used in the testing analysis. Lastly, Naïve Bayes Algorithm was the used to determine classification of these sentiments.

OBJECTIVE

The study aimed to:

1. classify the positive and negative sentiments of the remarks and comments written on student's report cards:
2. create a word cloud based on the comments, which is a visual representation of words.
3. generate a confusion matrix based on accuracy score using naïve bayes algorithm.

PROCESS AND METHODS

The study started by gathering remarks and comments available in various websites and through interviewing colleagues and co-teachers. These data are encoded manually, copy, and pasted in Microsoft Excel and saved as CVS file which were divided into training and testing data. The accurate number of data used for training was 450 for and 50 data for testing. The training data was then labeled manually whether it is negative or positive by looking at the sentiment words contained while data testing was not labeled. The frequency of the sentiments was balanced. The data were processed using the python 3 scripts with Natural Language Processing (NLP) techniques to find the division of the words. Stop words package was installed to clear the data from words that have no meaning or influence in the data. Seaborn library was used so that dataset was loaded and plotted into the system showing a bar like graph.

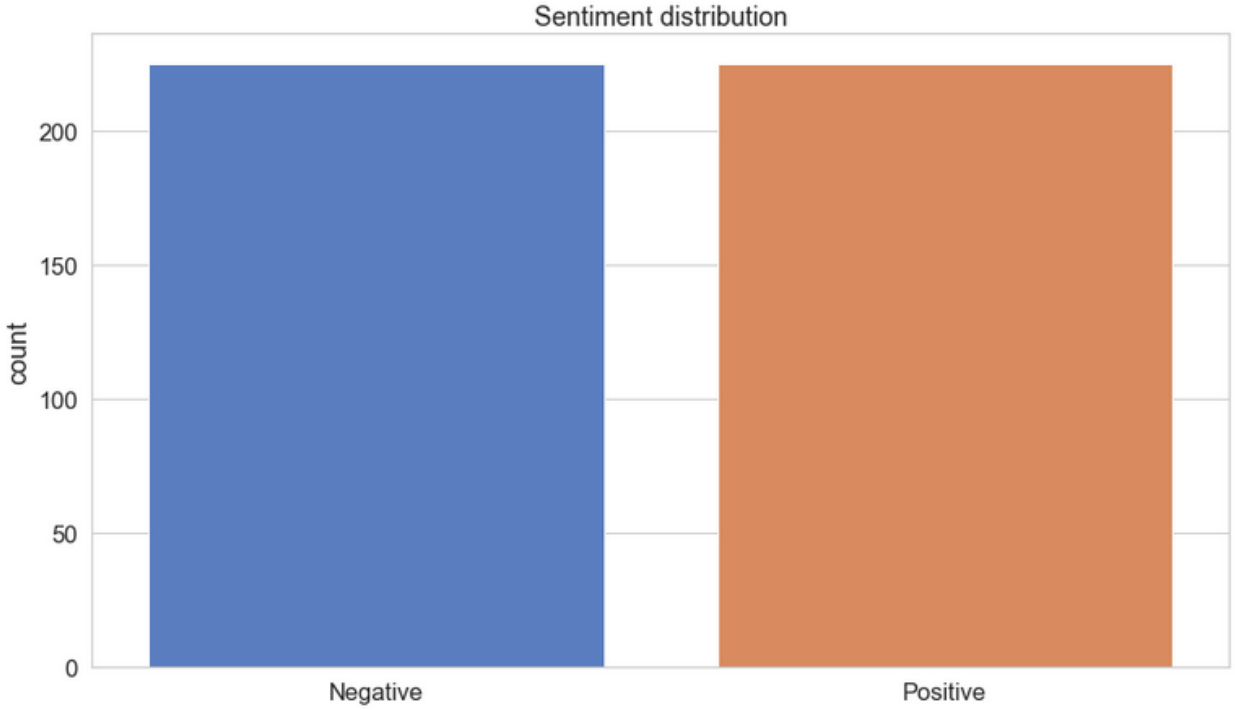

Figure 1. Sentiment Distribution

Figure 1 presents the distribution of sentiments as classified in the train data. The figure shows that the data given are of equal numbers. Then, the next output was the word cloud that was based on the frequency of occurrence of the words used in the remarks and comments. Figure 2 shows the visual representation of words.

Figure 2. Data Set Word Cloud

After, a stopword was also installed in the program to do the filtering of words which have no meaning. A word cloud was also installed to do the tokenization process and calculate the frequency of occurrence of words. To categorize the dataset, the Naive Bayes algorithm was utilized. The information that was applied to anticipate brand-new data classes that have never been is represented by this model. After that, train and test data were assigned. The sentence classification stage in the test data may then be known by applying the results of the word probability of the training data after conducting the learning process of the Nave Bayes classifier in the training data using the Multinomial Nave Bayes technique. After that, predict the test data's attitude and calculate the prediction's accuracy score using the algorithm.

	precision	recall	f1-score	support
0	0.67	0.86	0.76	43
1	0.83	0.62	0.71	47
accuracy			0.73	90
macro avg	0.75	0.74	0.73	90
weighted avg	0.75	0.73	0.73	90

Figure 3. Classification Report of Naive Bayes Algorithm

Finally, obtain the classification report to assess the accuracy of the classification algorithm's predictions (see Figure 3). The performance of a classification model is then evaluated by computing and plotting the confusion matrix.

RESULTS AND DISCUSSION

Using the Naive Bayes method, sentiment classification was carried out in two stages: the learning process stage and the classification stage. The main three terms associated with the remarks and comments, namely **work**, **class**, and **time**, which are clearly displayed in Figure 2. The algorithm's accuracy score is 73%, which translates to a true positive accuracy rate of 73% and a false positive rate of 27%. It implies a very good rate of accuracy.

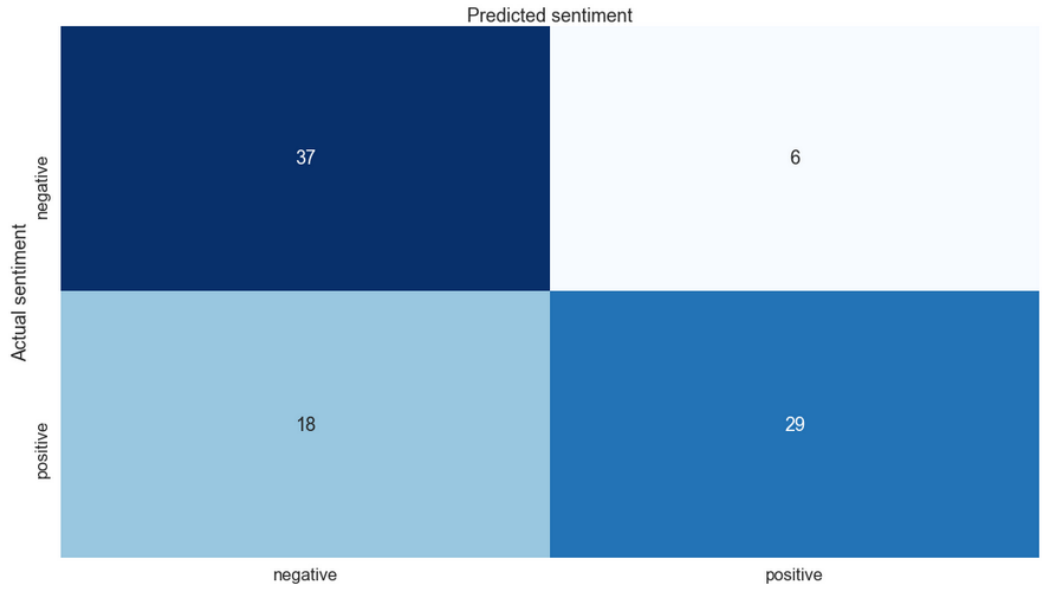

Figure 4. Confusion Matrix

The confusion matrix is depicted in Figure 4 and contains the following values: True Negative = 37, False Negative = 18, True Positive = 29, and False Positive = 6. The findings indicate that the majority of comments and remarks are negative in nature.

References

Evan-Moor. (2020, April 7). *100 Sample Report Card Comments and 5 Strategies for Simplifying Report Cards*. Retrieved from Education Blog Evan-Moor: https://teacherblog.evan-moor.com/2020/04/07/100-sample-report-card-comments-and-5-strategies-for-simplifying-report-cards/

Mary. (2022, April 4). *325 amazing report card comments and remarks to save your time!* Retrieved from EduCloud: https://educloud.app/lms/blog/report-card-comments

Raudys, J. (2021, September 1). *105 Report Card Comments to Use and Adapt*. Retrieved from Prodigy: https://www.prodigygame.com/main-en/blog/report-card-comments/

Rish*, I. (2001, January). *An empirical study of the naive Bayes classifier*. Retrieved from ResearchGate: https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier

Suresh, A. (2020, November 17). *What is a confusion matrix?* Retrieved from Medium: https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5