GROUP 9

Sentiment analysis using product reviews

Abstract

Sentiment analysis of client product reviews on e-commerce platforms is the goal of this study. Checking through the reviews left by prior buyers of a certain product is one of the key factors that influence customers' decisions to buy. Verifying the correctness of all these assessments and the product's condition is a difficult process. Instead, by categorizing the reviews and purported user opinions into different groups, the revised platforms let prospective purchasers assess the value of the goods by surfacing the product attributes. The goal is to determine if reviews are good, negative, or neutral and to draw conclusions about how customers feel about the items. In order to categorize the evaluations into several emotion categories, the project entails gathering data from an e-commerce site, preparing the data, and using natural language processing algorithms. The sentiment analysis's findings may aid companies in understanding client feedback, pinpointing areas for development, and making data-driven decisions to improve their goods and services.

1. Introduction

In order to analyze numerous business elements in e-commerce platforms, machine learning is essential. The fact that a corporation does all of its business online makes it difficult to determine why a product isn't in a good enough condition to generate sales, especially at a time when more individuals are accustomed to making purchases online due to improved lives.[3] Nowadays, it's so easy to buy anything and anything online. Platforms for electronic commerce are those that provide products for online purchases. Every day, new platforms with a variety of features and filtering options emerge in an effort to better serve clients.[5]

It might be challenging to use the reviewer's insights to get a comprehensive evaluation of a certain product in a situation where there are many different opinions. When attempting to comprehend a certain product based on a number of criteria, sentiment analysis of customer evaluations is essential. Sentiment analysis is based on the essential tenet that deals with emotions and perspectives.[4]

By addressing their thoughts, one may validate their understanding of the consumers' perspectives. Natural language processing and machine learning techniques together can produce insights.[6].Natural language processing based on supervised learning is utilized to calculate these reviews—known as unstructured data—and then to fine-tune the high-accuracy method.[2]

There hasn't been much advancement for customer satisfaction engines in this specific industry and the idea of the E-Commerce platform. Everything must be done manually and searched using only the filters, not the previously provided reviews, and ratings. Verifying the veracity of all these evaluations and the product's condition is a difficult undertaking. Instead, by classifying customer evaluations and opinions into various categories, the new platforms make it possible to reveal product characteristics and help potential consumers assess the worth of the items. While both customers and company marketers can benefit from these reviews[1][7]. As a result, there isn't a thorough study or research available on how to assess customer reviews, categorize them using different factors like words and emotions, etc., and provide a general attribute on the product's selling in the market based on the users' perspective. It is necessary to conduct further studies on how to examine and categorize all customer feedback. Sentiment analysis of the reviews is thus required to overcome this problem.

So, an intriguing machine learning method called opinion mining, also known as sentiment analysis, may be used to categorize customer evaluations based on their feelings and thoughts. Natural language processing is a tool that may be used for this study.[8]

The specific way in which our brains translate words is beyond our comprehension. Can we teach a machine to learn our language, then? Yes, after much research, several methods that might help robots grasp our languages have been found. The field of study known as "Natural Language Processing" (NLP) focuses on how computers and human language interact.[9] One of the subproblems in natural language processing is the analysis of sentiment, or determining if a statement is positive or bad. Customers may evaluate things on Amazon and give them a good, bad, or even neutral rating. These ratings will be divided into categories that are favorable, neutral, and negative using machine learning and supervised learning techniques.[10]

With the help of this study, it is possible to monitor client perspectives on reviews. The marketing team's examination of any product has as its major goal understanding customer feedback and enhancing product quality to increase sales and purchases. Sorting things into a category requires some effort.

2. Methodology

The dataset is taken from an online source and is based on Amazon products and the reviews given by the customers on various products, the variable attributes which are present in the dataset depict the uniqueness of the data cell. The sample size of the dataset is 34,660, which represents the strength of reviews. The reviews attribute is enough for training the model and drawing insights with the help of the rating attribute setting it as a target variable. The dataset which is in the CSV format is loaded into a variable from the systems' location and it is read by using the Pandas library: data_set = pd.read_csv(data) where data, is the location of the CSV file in the working system, and the data set is the actually retrieved dataset from Pandas. If any warnings occur, the imported Warnings library uses warnings. filter warning ('ignore') can be used to hide them.

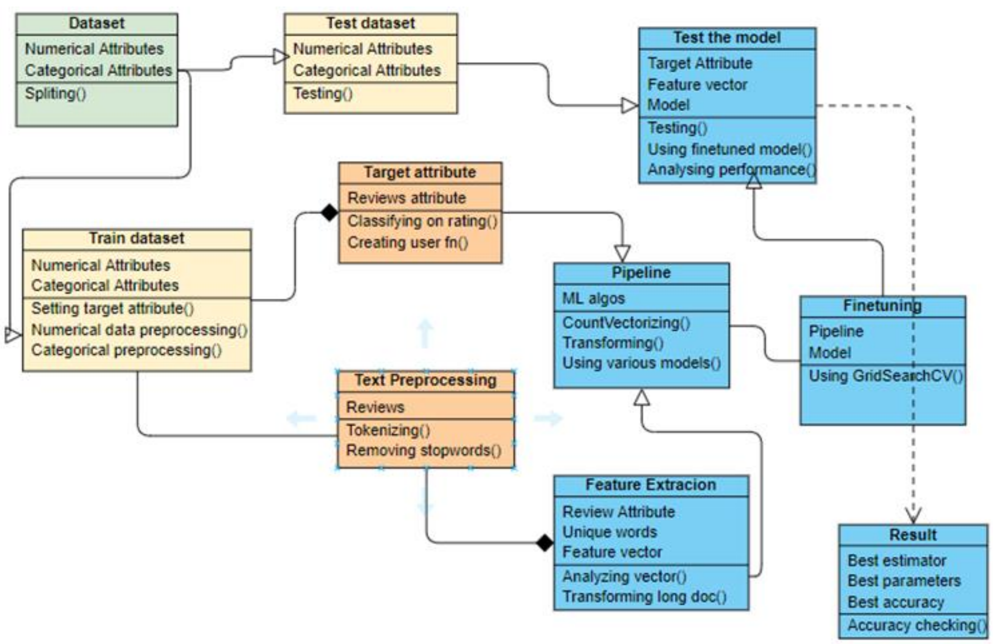

Figure 1: Study Framework

Various machine learning techniques, including Naive Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Trees, and Random Forest, can be utilized for categorization in sentiment analysis on e-commerce platform evaluations. Naive Bayes is a probabilistic technique that works well for classifying texts because it assumes independence between features. For binary classification, logistic regression is a linear model that forecasts the likelihood of the result. SVMs are non-linear models that successfully handle high-dimensional data by attempting to divide classes using a hyperplane. A group of decision trees called a Random Forest combines decision trees to increase accuracy and decrease overfitting. Decision Trees are a tree-like model that recursively divides data into smaller groups depending on characteristics. The data, the issue description, and the performance specifications of the sentiment analysis system all influence the method chosen.

Using Python libraries such as Numpy, Pandas, matplotlib, and Scikit Learn, the numerical and object-type data variables are handled. There are around 21 attributes, 5 of them are numerical and the remaining 16 attributes are categorical, where the absence or not filled data is that reviews.text having 34659 filled data, only 1 tuple is with no review in it. The information of all the attributes is carried with the info() function, indicating an average review score of 4.58 with a minimum standard deviation. By setting the target attribute or variable name as "Sentiment", where values of the attribute can be analyzed using the rating attribute. The user-defined function takes input parameters as ratings, applies the function to the stratified training dataset, and sets up the sentiment attribute to the dataset. As the exploratory data analysis can be done only on the numerical attributes, the text pre-processing can be achieved using the bag of words strategy, this especially follows the natural language processing techniques which are: Tokenization, Stop words, Lower-casing, Occurrence counting: It is by assigning fixed integer to each word occurrence. And then building a feature vector converts the dictionary of text into a feature numerical vector. By building a pipeline to train various models in order to figure out the accuracies of the classification models. The classification algorithms used are LogisticRegression Classifier, SupportVectorMachine Classifier, DecisionTree Classifier, and RandomForest Classifier.

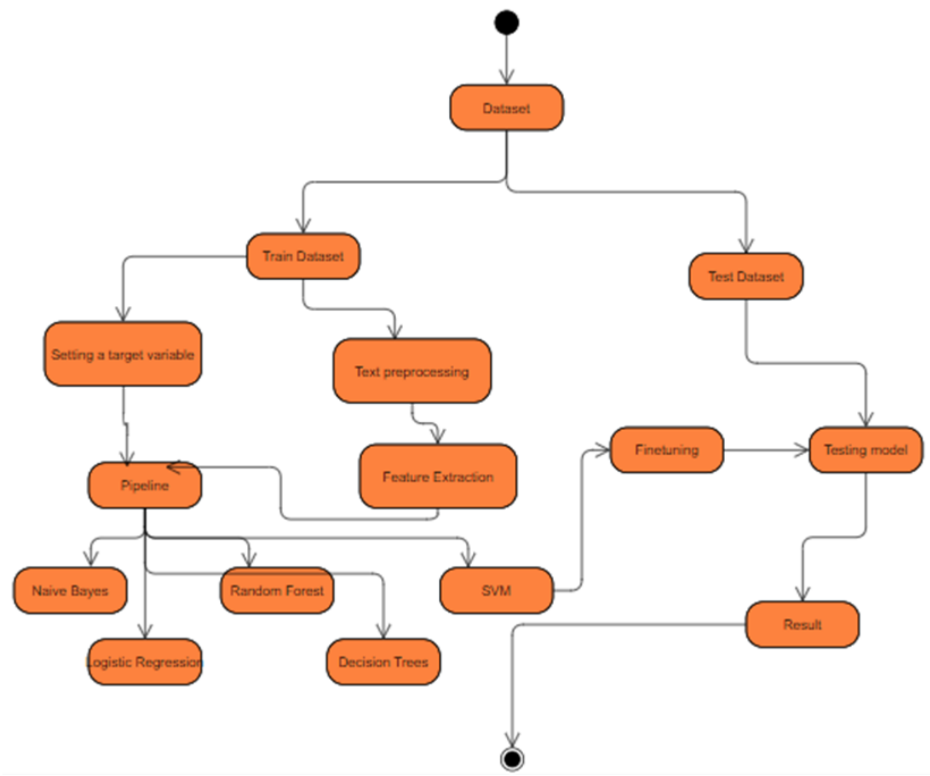

Figure 2: Flowchart of the analysis

Splitting the dataset to build a model with 80% of the training and 20% of the testing datasets. Proceeding with the correlation analysis shows that there is no significant correlation among any variables. By training and testing using various classification models, the accuracy is compared and visualized using the matplotlib library to figure out the algorithm which can be proceeded to fine-tuning.[3] SVM, the plot is pointed by using data items in the n-featured space. Then the classification is performed by using the hyperplanes that are used to differentiate the classes into categories. Here, classify the reviews (decisions) using decision boundaries. By using the GridSearchCV of the best boundaries on a grid of potential qualities, rather than tweaking the boundaries of different segments of the chain fine-tuning is done. At that point, they were fitting the grid search into the preparation informational index. Next, utilize the last classifier (after adjusting) to test some subjective audits, then test the precision of the last classifier (after tweaking). Model evaluation can be done using the performance analysis and classification report. Basically, using the performance analysis best score, best estimator, and best parameters can be found. applying classification report on target test variable and variable consisting of the fine-tuned model.

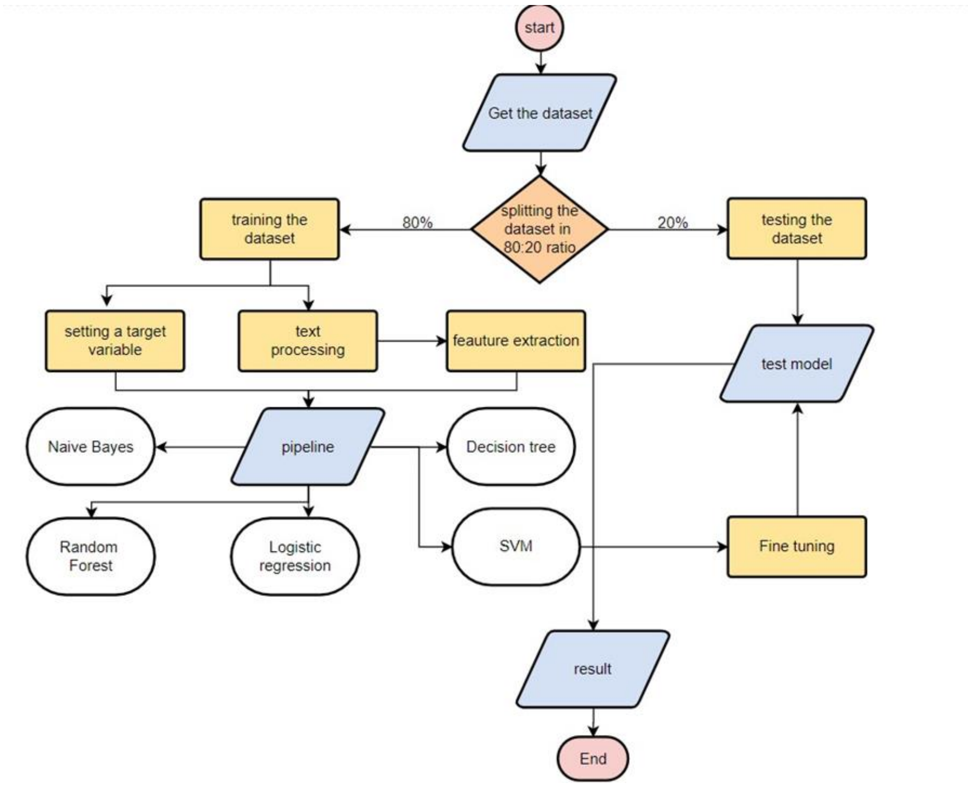

Figure 3: Models and Analysis

Precision: Decides the number of items chosen was right. (Percent of predictions are correct). Recall: Discloses the number of the items that ought to have been chosen were really chosen. (Percent of positive cases chosen).

F1-score: estimates loads of review and exactness (1 method's precision and recall are similarly significant, 0 in any case). (Percent of positive predictions are chosen).

Support: is the number of events of each class.

$$\begin{aligned} precision &= \frac{TP}{TP+FP} \\ recall &= \frac{TP}{TP+FN} \\ F1 &= \frac{2 \times precision \times recall}{precision + recall} \\ accuracy &= \frac{TP+TN}{TP+FN+TN+FP} \end{aligned}$$

Figure 4: Calculations of precision, recall, f1, accuracy

3. Results and Discussion

With SVM and Random Forest having the greatest accuracy ratings, it is clear from the provided accuracy scores that all algorithms perform well in sentiment analysis. The final method to choose, nevertheless, will rely on the particular specifications of the sentiment analysis project.

It is commonly known that SVM performs well with high-dimensional data and is a powerful technique for classification jobs. It operates by identifying the optimum hyperplane that efficiently divides the data into distinct classes. This demonstrates that SVM has high generalization capabilities and is successful in handling complicated datasets. In this situation, SVM's high accuracy score of 93.8% demonstrates that it can successfully categorize the sentiment of reviews on e-commerce platforms.

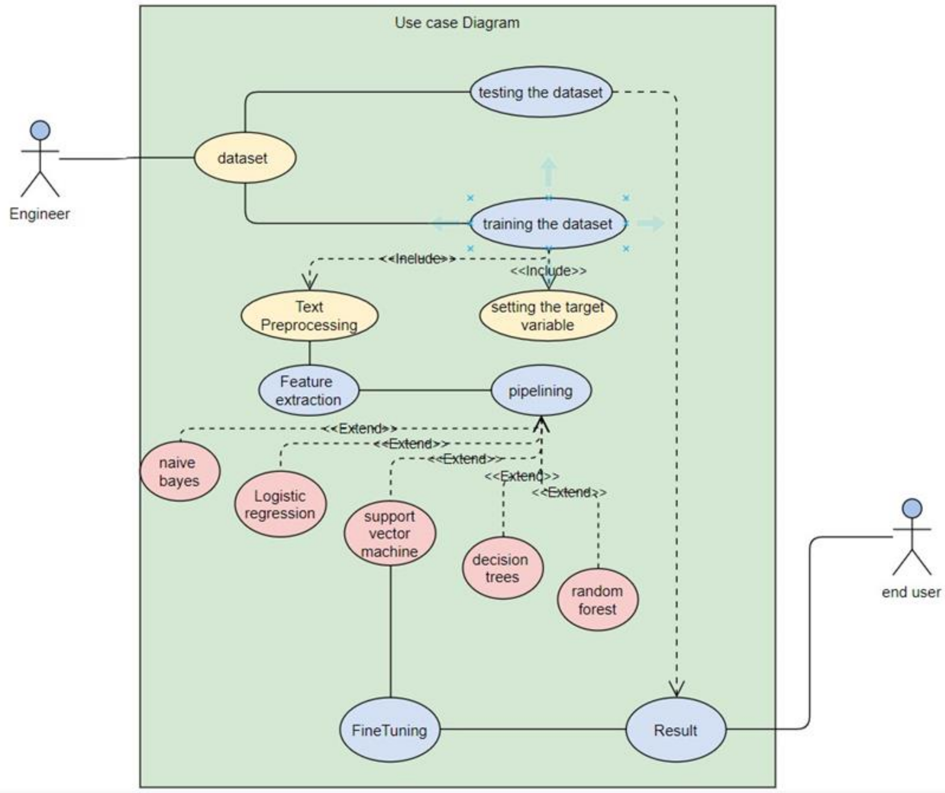

Figure 5: Result from User Perspective

Precision is named as the measurement of quality and recall is named as the measurement of quantity. They can be figured as inversely proportional to each other. It is clear that out of all the algorithms chosen this SVM produces high precision indicating it is returning more relevant outcomes. Low recall results in producing the results correctly.

It's crucial to remember, though, that accuracy alone is not the only factor to take into account when choosing an algorithm for sentiment analysis. It's also important to consider other aspects including the model's interpretability, computational complexity, and training time. To guarantee the optimum performance and fit for the particular sentiment analysis project, the choice of SVM should be based on a thorough examination of all these variables.

Classifier Name	Accuracy
NaiveBayes Classifier	93.4%
LogisticRegression Classifier	93.7%
SupportVectorMachine Classifier	93.8%
DecisionTree Classifier	89.99%
RandomForest Classifier	93.5%

Table 1: Comparing various models

The classification report presents the performance of a machine learning model for sentiment analysis. The report includes measurements for each class including precision, recall, F1-score, and accuracy.

When examining the given classification report, it is clear that the model obtained an accuracy of 0.939, demonstrating that it is able to accurately categorize the sentiment of user reviews of e-commerce platforms. The accuracy scores were 0.71, 0.39, and 0.95 for the Negative, Neutral, and Positive groups, respectively. This indicates that Positive reviews had a higher percentage of accurate predictions from the algorithm than Negative reviews or Neutral reviews.

Recall scores were 0.29, 0.07, and 1.00 for the Negative, Neutral, and Positive classes, respectively. This indicates that a larger percentage of Positive reviews than Negative or Neutral reviews were accurately detected by the algorithm. The harmonic mean of accuracy and recall is known as the F1-score, which assesses the model's overall efficacy. The Positive class had a high F1 score of 0.97, whereas the Neutral and Negative classes had lower values of 0.41 and 0.11, respectively.

In conclusion, the model scored highly for accuracy, although there is still potential for development in terms of accurately detecting Negative and Neutral evaluations.

```
In [106]:  from sklearn.metrics import classification_report
           from sklearn.metrics import accuracy_score

           print(classification_report(X_test_target, predictedGS_clf_LinearSVC_pipe))
           print('Accuracy: {}'. format(accuracy_score(X_test_target, predictedGS_clf_LinearSVC_pipe)))

                 precision    recall  f1-score   support

                      0.00      0.00      0.00         7
        Negative      0.71      0.29      0.41       161
         Neutral      0.39      0.07      0.11       286
        Positive      0.95      1.00      0.97      6472

        accuracy                          0.94      6926
       macro avg      0.51      0.34      0.37      6926
    weighted avg      0.92      0.94      0.92      6926

Accuracy: 0.9395033208200981
```

Figure 6: Classification report

References

[1] A. Alrehili and K. Albalawi, "Sentiment Analysis of Customer Reviews Using Ensemble Method," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-6, doi: 10.1109/ICCISci.2019.8716454.

[2] A. Abdul Aziz and A. Starkey, "Predicting Supervise Machine Learning Performances for Sentiment Analysis Using Contextual-Based Approaches," in IEEE Access, vol. 8, pp. 17722-17733, 2020, doi: 10.1109/ACCESS.2019.2958702.

[3] R. S. Jagdale and S. S. Deshmukh, "Sentiment Classification on Twitter and Zomato Dataset Using Supervised Learning Algorithms," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), Aurangabad, India, 2020, pp. 330-334, doi: 10.1109/ICSIDEMPC49020.2020.9299582.

[4] N. U. Pannala, C. P. Nawarathna, J. T. K. Jayakody, L. Rupasinghe and K. Krishnadeva, "Supervised Learning Based Approach to Aspect Based Sentiment Analysis," 2016 IEEE International Conference on Computer and Information Technology (CIT), Nadi, Fiji, 2016, pp. 662-666, doi: 10.1109/CIT.2016.107.

[5] S. Wladislav, Z. Johannes, W. Christian, K. André and F. Madjid, "Sentilyzer: Aspect-Oriented Sentiment Analysis of Product Reviews," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2018, pp. 270-273, doi: 10.1109/CSCI46756.2018.00059.

[6] Fang, X., Zhan, J. Sentiment analysis using product review data. *Journal of Big Data* 2, 5 (2015). https://doi.org/10.1186/s40537-015-0015-2

[7] C. Chauhan and S. Sehgal, "Sentiment analysis on product reviews," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, pp. 26-31, doi: 10.1109/CCAA.2017.8229825.

[8] Purohit, Amit and Patheja, Pushpinder Singh. 'Product Review Opinion Based on Sentiment Analysis'. 1 Jan. 2023 : 3153 – 3169.

[9] Leung, C.W. (2009). Sentiment Analysis of Product Reviews. *Encyclopedia of Data Warehousing and Mining*.

[10] Nguyen, Heidi; Veluchamy, Aravind; Diop, Mamadou; and Iqbal, Rashed (2018) "Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches,"SMU Data Science Review: Vol. 1: No. 4, Article 7. Available at: https://scholar.smu.edu/datasciencereview/vol1/iss4/7