ISSN 2063-5346

HYBRID BAG-BOOST ALGORITHM DRIVEN FEATURE ORIENTED SENTIMENTAL ANALYSIS FOR DETECTION OF SMARTPHONE REVIEWS

Nadiya Jabeen[1] Mrs. Sumaiya Shaikh[2]

Article History: Received: 10.05.2023 Revised: 29.05.2023 Accepted: 09.06.2023

Abstract

With the tremendous growth in the Internet technologies and E commerce platforms, many people are showing interest in procuring products and services online. in this scenario recommender systems are a great way to promote new business and get more customers for the relevant products and servicesAnd thereby increase the business. However, it is crucial to be able to predict what a customer would like andneed based on the products he has purchased or he is interested in. In order to achieve this the machine learning model would have to accomplish various tasks such as word segmentation, stop-words, extraction of features and finding similar products other users have purchased etc. In this project we take the example of movie recommendation system and we tried to categorize the movie reviews as positive or negative using sentiment analysis and have built a recommender system using an improved item based collaborative filtering based on the sentiment of users which can suggest movies that a user may like based on the list of movies he has already watched.

Index Terms — Mobile Reviews, Classification Techniques, SVM, Hybrid-Bag-Boost, Naive Bayes.

[1]Research Scholar, Dept. of Computer Science and Engineering, Lords Institute of Engineering & Technology, Hyderabad, Telangana

[2]Asst Professor, Dept. of Computer Science and Engineering, Lords Institute of Engineering & Technology, Hyderabad, Telangana

DOI:10.48047/ecb/2023.12.9.83

I. INTRODUCTION

Nowadays, sales made through e-commerce platforms have increased, and many customers utilize the reviews left by other users to decide whether or not to purchase a particular product. The majority of reviews, however, are split when it comes to buying mobile phones because some specific characteristics of a smartphone might be good, while others might be ok, and a select few other features might be poor. To make an informed choice based on the qualities of one mobile phone, it is crucial to comprehend the review's overall viewpoint. To analyze the overall attitude of the review and categorize it as either positive or negative, we used their smartphone review data set in this research. To select the optimum solution for the situation, the data set is fed into a variety of algorithms, including random forests, decision trees, logistic regression, etc. In this data set, we found that the composite bag boost algorithm performs best at predicting mobile reviews, and the suggested method outperforms the majority of other methods in this situation.

The Random Forest approach sometimes referred to as the bootstrap aggregation technique is used in the Bagging technique. In this method, base learners are arranged in parallel to one another and given sample data from the original dataset as input. Decision Trees are employed as the primary learners in the Random Forest approach. The base learner will use row sampling with replacement to extract the random data from the data collection as input. The outcome will be predicted using a voting classifier.

We employ the ADA Booster and XG Booster algorithms in the boosting strategy. The incorrectly classified classifiers are combined and trained with the base learners in Boosting to make them into strong classifiers that will increase the accuracy of the model. Each base learner takes data from the data set at random and finds the incorrectly classified classifiers or weak classifiers.

II. Literature Survey

"Feature-Based Opinion Mining for Amazon Products using MLT"

One of the biggest challenges facing NLP is sentiment analysis or opinion mining (natural language processing). In the current environment, business analytics is important for those looking to grow their business. This population, in particular, depends on customer reviews of their product to compete in the market and on knowledge analytics, which can provide us a great insight into what to expect in the future. Opinions are frequently mentioned, and future opinions are frequently anticipated. Most of those business owners strive to expand their enterprises and provide high-quality products to maximize profits. Sentiment analysis has therefore received a great deal of interest in recent years in this area. This research work is based on the many methods used to categorize texts according to the opinions they express, i.e., whether a person has a generally positive, neutral, or negative mindset. Along with the experimental results, we also look at the two-advance approaches (feature classification and polarity classification). Finally, we compared three ML classification approaches in this work. 1) SVM, 2) Naive Bayes (NB), and 3) Logistic Regression using Hybrid Algorithm, the latter of which provides greater accuracy than the first three ML algorithms.

"Sentiment analysis from product reviews using SentiWordNet as a lexical resource"

Nowadays, because of the numerous social and technical advancements, consumers ask other customers who have purchased the goods for their thoughts. Additionally, businesses rely on customer input to comprehend how their items are being used. Sentiment analysis and opinion mining are two terms used to describe the method for obtaining sentiments and emotions from data. This opinion mining will assist in understanding the overall consumer reaction to the product. SentiWordNet has been used as a lexicon in this study to complete the sentiment analysis process.

III. SYSTEM ANALYSIS

Existing system:

It has become crucial to examine product attributes via mining customer reviews. The

classification of smartphone reviews has been the subject of numerous studies, which differ in the datasets used and the methods used. Numerous research has employed a variety of unbalanced datasets with various labels to classify the reviews using supervised learning approaches such as naive Bayes, SVM, and decision trees, but these strategies fall short of being truly accurate.

Disadvantages:

- Not accurate
- Cannot be deployed in real-time scenarios

Proposed System:

We suggest creating and implementing a web application that forecasts consumer reviews based on characteristics. With the help of Natural Language Processing, a kind of machine learning, we can evaluate text to determine if consumer evaluations are favorable or negative. Sentiment analysis is yet another name for this. We suggest employing hybrid machine learning techniques to analyze user attitudes based on mobile features and then display the findings.

Advantages:

- Accurate review prediction

IV. PROPOSED MODULAR IMPLEMENTATION

The project's planned modular implementation is shown below. There are two modules in it:

1. Admin
2. Consumer

Admin Module:

The system administrator is in charge of tasks like:

1. Uploading the dataset
2. Sentiment analysis of mobile review data.
3. Create a sentiment analysis model using the hybrid bag boost technique, SVM, and logistic regression.
4. Examine how well the algorithms performed on the provided dataset.
5. Check out expected customer feedback

Consumer Module:

The system's user can make use of the following machine-learning services:

1. logging in to the system,
2. reviewing features, and
3. Get forecasts for the total rating and reviews

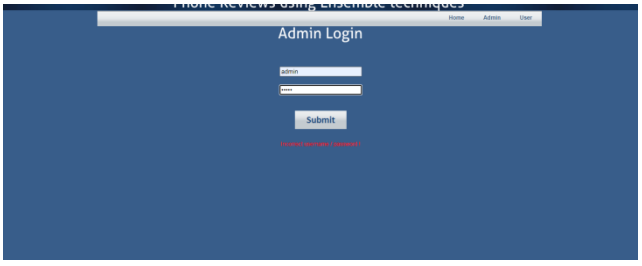
V. PROJECT EXECUTION

Home page:

This is the starting page of the application when the application is executed on Pycharm, the application is hosted on a web server and a URL is generated to access the application once the user clicks on the URL the below page is opened on the browser.

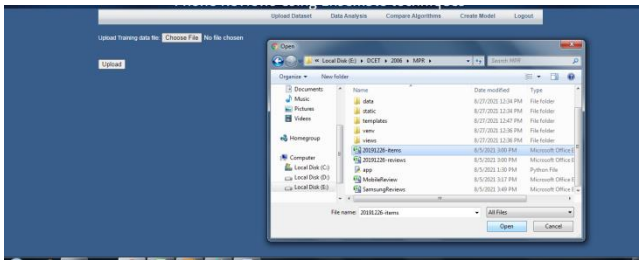
Admin Login:

This is the login page for the admin module. The admin need to login into the system with his credentials to perform operations like uploading the dataset, Training the dataset, Exploratory Data Analysis of the dataset, and Feeding the dataset to different Machine learning Algorithms to find the Algorithm that can meet the best accuracy and Create a model that can be hosted on the Flask Application to be used by the users.



Upload Dataset:

On this page, the administrator of the system can upload datasets that are used for training the machine learning models. The admin has to select the file by clicking on the Choose file button and clicking on the upload button to upload the file to the server. Once the upload is complete, a success message would be displayed that the file is successfully uploaded. For this project, we are using Train_3.csv as a dataset.

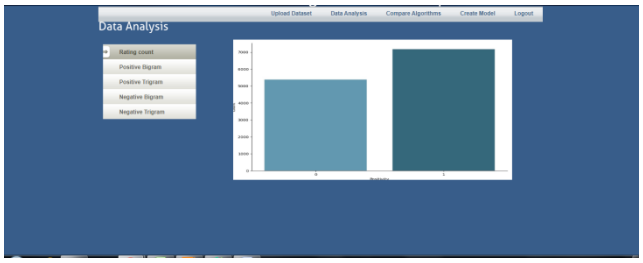


Data Analysis:

Exploratory Data Analysis is performed on the dataset to clean the dataset for any missing data, identify patterns, and identify the relationships of various parameters of the outputs with the help of graphs, statistics, etc.

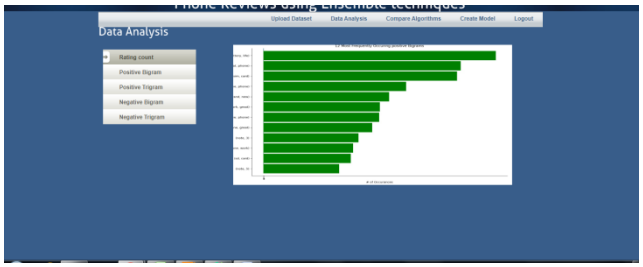
Rate Count:

The below graph shows the Rate count Analysis.



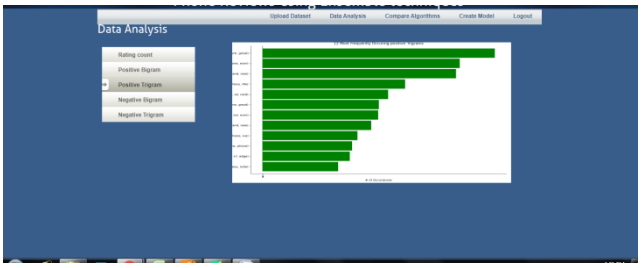
Positive Bigram:

The below graph shows the Positive Bigram Analysis.



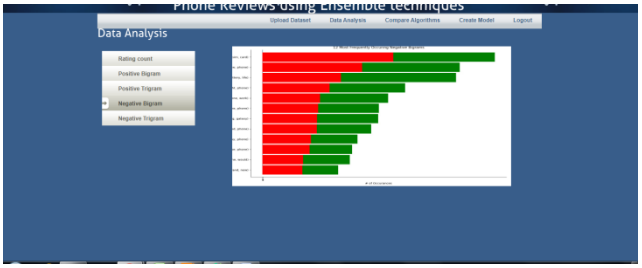
Positive Trigram:

The below graph shows the Positive Trigram Analysis.



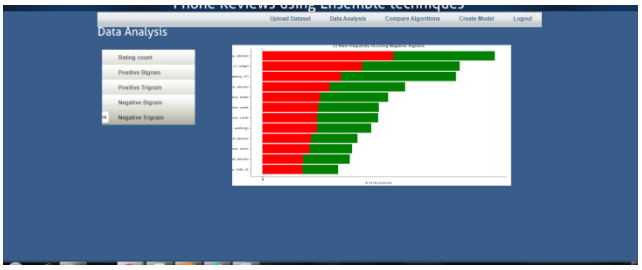
Negative Bigram:

The below graph shows the Negative Bigram Analysis.



Negative Trigram:

The below graph shows the Negative Trigram Analysis.

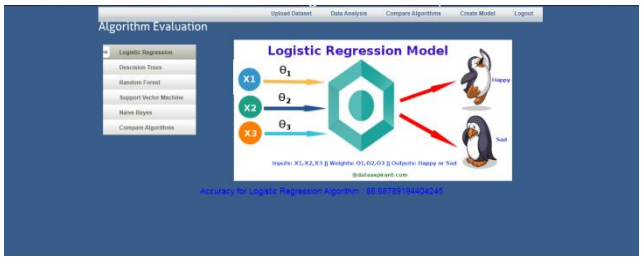


Compare Algorithms:

On this page, the admin can feed the dataset to various Algorithms to train them and get the test accuracy for each algorithm.

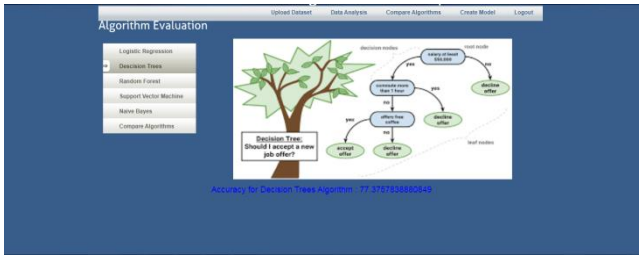
Logistic Regression:

When the dataset is fed to Logistic regression algorithm we observe that the test accuracy is 88.68%.



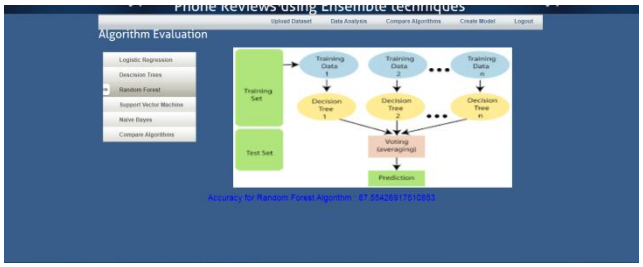
Decision Trees:

When the dataset is fed to Decision Trees algorithm we observe that the test accuracy is 77.37%.



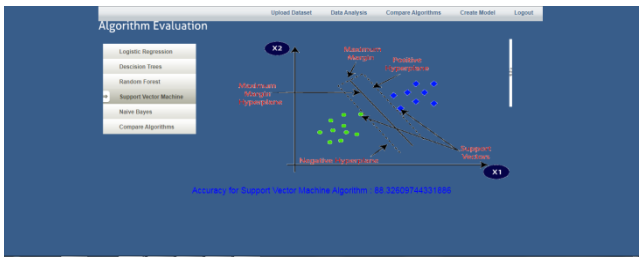
Random Forest:

When the dataset is fed to Random Forest algorithm we observe that the test accuracy is 87.56%.



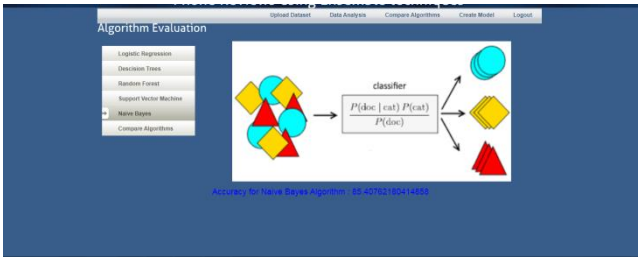
Support Vector Machine:

When the dataset is fed to Support Vector Machine algorithm we observe that the test accuracy is 88.32%.



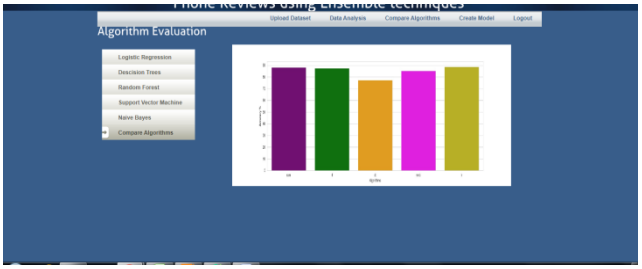
Naive Bayes:

When the dataset is fed to Naive Bayes algorithm we observe that the test accuracy is 85.40%.



Compare Algorithms:

This screen shows the comparison of various test accuracies of the Algorithms.



VI. CONCLUSION

In this project, we have put forth a brand-new, hybrid method for categorizing mobile evaluations according to their polarity and features. We used the Amazon mobile reviews data set and fed it to many algorithms, including the Hybrid Bag Boost method, SVM, Naive Bayes, and logistic regression. We have made the program available as a web application, and users may now determine whether the sentiment or opinion of a review is favorable or unfavorable. We see that the hybrid bag boost algorithm outperforms other algorithms in terms of accuracy.

REFERENCES

[1] Siva Kumar Pathuri, N.Anbazhagan, “Feature-Based Opinion Mining For Amazon Product’s Using MLT”,IJITEE,Issn:2278-3075,vol8,issue-11,sep-19.

[2] V. Keselj, F. Peng, N. Cercone, and C. Thomas, "N-gram-based
author profiles for authorship attribution," in Proceedings of the
conference pacific association for computational linguistics,
PACLING, vol. 3, pp. 255–264, 2003.
[3] PeriakaruppanSudhakaran, Joan Lu, "Classifying product reviews
from balanced datasets of Sentiment Analysis & Opinion
Mining", *Published in 2014 6th International Conference on*
Multimedia. Computer Granhics and Broadcasting IEEE.
[4] Alexandra Cernian, Valentin Sgarciu, Bogdan Martin" Sentiment
analysis from product reviews using SentiWordNet as lexical
resource"ECAI 2015 - International Conference – 7th Edition
Electronics, Computers and Artificial Intelligence 25 June -27 June,
2015.
[5] Long-Sheng Chen, Hui-Ju Chiu, "Developing a Neural Network
based Index for Sentiment Classification", Published in International
MultiConference of Engineers and Computer Scientists, vol. 1, 2009.
[6] Siva Kumar Pathuri, N.Anbazhagan, "Prediction of cardiovascular
Disease UsingClassification Techniques With High
Accuracy",Jardcs,Vol-12 issue-02,Mar-20
[7] A. Ejaz, Z. Turabee, M. Rahim and S. Khoja, "Opinion mining
approaches on Amazon product reviews: A comparative study," *2017*
International Conference on Information and Communication
Technologies (ICICT), Karachi, 2017, pp. 173-179.
[8] Z. Singla, S. Randhawa and S. Jain, "Sentiment analysis of customer
product reviews using machine learning," *2017 International*
Conference on Intelligent Computing and Control (I2C2), Coimbatore,
2017, pp. 1-5.
[9] Anagnostopoulos, Theodoros&Anagnostopoulos, Christos
&Hadjiefthymiades, Stathes &Kyriakakos, Miltos&Kalousis,
Alexandros. (2009). Predicting the location of mobile users: A
machine learning approach. 65-72. 10.1145/1568199.1568210.
[10] Cho, Eunah& Amazon, He &Lalor, John & Kumar, Varun &
Campbell, William. (2019). Efficient Semi-Supervised Learning for
Natural Language Understanding by Optimizing Diversity.1077-1084.
10.1109/ASRU46091.2019.9003747.