

Brand Review Prediction using Sentiments: Machine learning Algorithms

¹Dr.A.Bamini, ²Miss.A.Shanmugapriya

¹M.Phil, Ph.D., ²MCA

¹Head, ²Student

Department of Computer Applications

The Standard Fireworks Rajaratnam College for Women, Sivakasi, India.

Abstract: In today's competitive Digital market Rating and Review of various brands makes the customers to understand the quality about the product. In this paper the ratings of the users submit only the rating rather than dropping the feedback review. To analyze this concept we propose a new deep learning algorithm where we can get 98% accuracy by analyzing the original dataset given from flip kart. Also the accuracy of different machine learning algorithm were compared. Due to the vivid reviews provide by the customers, there is a feedback environment being industrial for helping customers to buy the right product and guiding enterprise to enhance their features of product which will suit the consumer's demand. The customer finds it difficult to exactly find the review for a particular feature of a product that she/he intends to buy. Also, there is a mixture of positive and negative reviews thereby making it problematic for consumer to find a clear response. Also these reviews have been affected severely from spammed reviews from unauthenticated users. In order to avoid this confusion and make this review system more transparent and user friendly we propose a technique to extract feature based opinion from a diverse pool of reviews and giving out it further to separate it with respect to the aspect of the product and further categorize it into positive and negative reviews using machine learning based approach. The Paper Focus on "Brand Review Prediction using Sentiments: Machine learning Algorithms" is developed using Python as front end and by using a Spyder tool.

Keywords: Machine Learning Algorithms, Data Mining Techniques

I. Introduction

Rapid development in technology and internet brought us to the era of ecommerce. Ecommerce is not anything but selling and buying goods or services online. Many customers share their good or bad opinion about products or services online nowadays. These opinion become a part of decision making process to make impact on business model. Also understanding and taking into consideration reviews, will help to gain trust of customer which will help to expand business. Thus, we need to study and examine customer reviews. Usually ratings are given in star format. Single product has number of ratings. It is hardly possible to read each ratings in detail. Many researches shown pictorial representation is more effective and can be memorize, understood easily rather than textual representations. So if we are going to convert textual reviews into visual format, it will enhance consistency in decision making. Existing work shows that various approach are used for sentiment analysis like machine learning, mass based, NLP based or even based on clustering. Also few research consider neutral reviews for analysis. Many of them do not have diagram representations for end results or multipart visual representation which are not user oriented.

The possibility of research work does not include:

- a. Consideration of emotions
- b. Work aims only to calculate expressions about product not the quality of product.

Thus, the paper work will provide novel approach based on hadoop environment to provide visual representation of sentiment analysis results apply to online product rating given by customer.

II. Literature Review

[1] Gurneet Kaur and Abhinash Singla, **Sentimental Analysis of Flip kart reviews using Naive Bayes and Decision Tree algorithm, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)**. Product users review comments about product and review about retailers from Flip kart are taken as data set and then the dataset were classified as subjectivity/objectivity and negative/positive based on the attitude of buyer. The empirical study of efficacy of classifying product review by semantic meaning. Different approach were used as well as spelling correction in review text, and then classify comments employ hybrid algorithm combine Decision Trees and Naive Bayes algorithm. The spelling correction is done to make the most reasonable comment for meaningful the polarity of words using Word Net dictionary. Then stemming is performed to remove the stop words.

[2] Krutika Wase, Pranali Ramteke, Rushabh Bandewar, Nadim Badole and Bhuvneshwar Kumar, **Sentiment analysis of product review, International Journal of Innovations in Engineering and Science**. In the field of sentiment analysis there are

many algorithms have to begin NLP problems to recognize the positive and negative reviews of the user's for your products on online market. Data used in this were collected from Amazon.com, Redif.com, Flip kart.com. In this product is divided into positive, negative and neutral categories which indicate the sentiment of an product. Negative sentiment product were used for the evidence generation for cybercrime cases. The experiment were used for filtering the negative sentiment products. The advantage are the analysis method for the classification of product, which was done by feature extraction followed by a classification step. Many of the applications of Opinion Mining were base on bag-of-words, which do not hold context which is necessary for Sentiment Analysis.

[3]Janhavi N L, Santhosh Kumar K L and Jharna Majumdar, **Sentiment Analysis of Customer Reviews on Laptop Products for Flipkart, International Research Journal of Engineering and Technology (IRJET)**. Analyze the fundamentals of opinion mining. It consist of different approaches including Extraction, Clustering and Classification. Extract reviews from the website using flip kart product API, using product API here they were attractive the brand name, reviews, rating and other related things for product, clustering using ROCK and using CART algorithm to order reviews as positive and negative words from the comments and finally they come to know which product have more percentage of positive reviews. The input were classified as positive and negative words from reviews and it calculates percentage of positive and negative words. Therefore the result analysis of review percentage it helps the user to end based on the positive review percentage of the product.

III. Proposed Work

The Proposed work focus on a new approach of Machine Learning algorithm called Support Vector Machine and Random Forest algorithm is proposed to classify the dataset taken from the e-commerce websites as positive negative and neutral depending on the star ratings provided by the user for the product which helps the upcoming buyers to buy the product without any difficulties, the both machine learning algorithm is combined and simulated by using a tool called SPYDER in Python.

IV. Methodology

The Methodology focus on the Machine Learning algorithms called SVM Classifier and Random Forest algorithm to analyze the dataset taken from the flip kart websites as positive, negative and neutral depending on the star ratings provided by the user for the product. A Confusion Matrix provides the highest accuracy in deep forest algorithm so that the buyers can able to purchase the product without any difficulties, the both machine learning algorithms were integrated and simulated by using a tool called ANACONDA SPYDER in Python.

The data set consists of 20,000 records out of which 50% were taken as training data and 50% were taken as testing data. Also the irrelevant dataset were removed from the samples as there is no usage of those data. The results were calculated depending on the measures such as Precision, Recall and Overall Accuracy.

❖ Detection Process:

After Training the next process is to predict the output of the model on the testing dataset, and then a confusion matrix is generated which determines the reviews as either positive or negative. The results involve the following attributes:

True Positive - accurate Positive Reviews in the testing data.

Forged Positive - Fake Positive Reviews in the testing data.

Correct Negative - Actual Negative Reviews in the testing data.

Copied Negative - Counterfeit Negative Reviews in the testing data.

The reviews from datasets whether they are fake or real are analyzed and framed as a Confusion matrix.

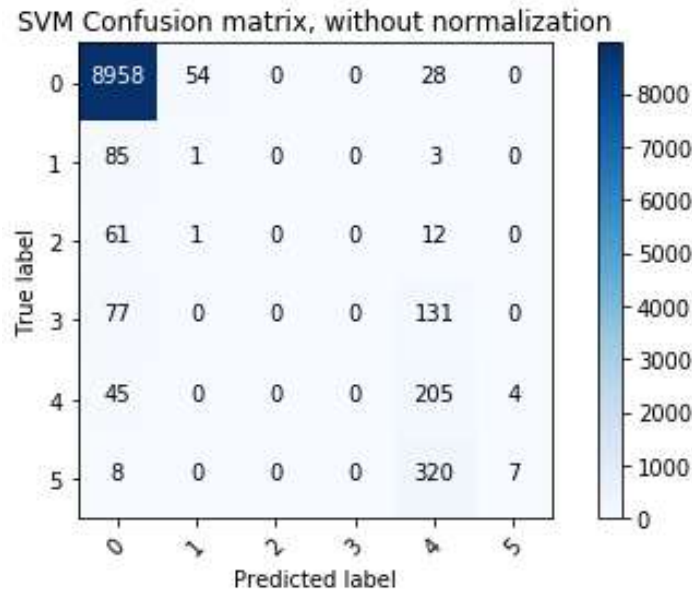
	Real	Fake
Real	True Negative Reviews (TN)	False Positive Reviews (FP)
Fake	False Negative Reviews (FN)	True Positive Reviews (TP)

❖ Sentiment Classification algorithms:

In this module, the sentiment classification algorithms are used and they have been applied in recent technologies. There are many different techniques in classification method like NB, DT-J48, SVM, K-NN, Neural Networks, and Genetic Algorithm, Deep forest algorithm. In this study, the two popular supervised classifiers are used.

❖ Support Vector Machine (SVM)

SVM in machine learning is a supervised learning model with the related learning algorithm, which studies data and identifies patterns, which is used for regression and classification analysis. Newly, many classification algorithms have been proposed, but SVM is still one of the most widely and most current used classifiers.



❖ **Random Forest Algorithm**

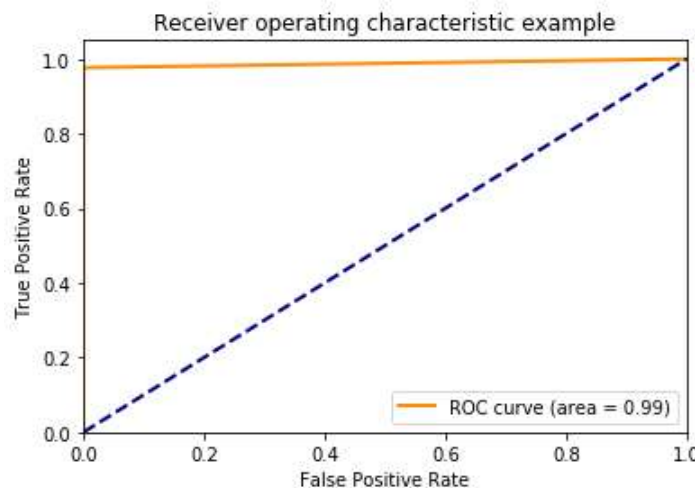
This algorithm is defined as a highly accurate and robust method when compared to other algorithms in sentimental analysis because of the number of decision trees participating in the process. Random forests is a set of many decision trees. Deep decision trees may smart from over fitting, but random forests prevents over fitting by making trees on random subsets.

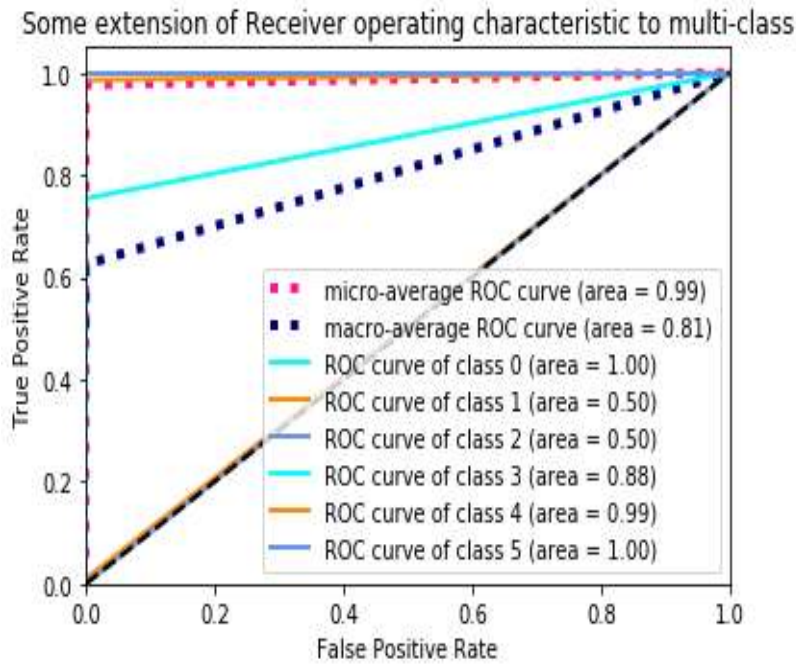
❖ **Evaluation parameters and accuracy for all methods**

In this module five main performance evaluation measures have been introduced for Classification algorithms. These include Fake Positive Reviews predictive value, False Negative Reviews analytical value, Real Optimistic Reviews projecting value, Actual Undesirable Reviews prognostic value, accuracy and Precision. The results of evaluation parameters for all methods creates a summary of recordings obtained from the experiment. Random Forest surpasses as the best accuracy among the other classification algorithms with 96.66%. The tabular column list the readings as well as accuracies obtained for a specific supervised learning algorithm on a dataset of a product ratings.

❖ **Receiver Operating Characteristic (ROC) metric to estimate classifier output value using cross-validation.**

ROC curves normally feature exact hopeful rate on the Y bloc, and artificial positive velocity on the X axis. This means that the top left corner of the plot is the model point - artificial positive rate of zero, and a factual positive rate of one. This is not very practical, but it does mean that a larger area under the curve (AUC) is usually better. This module shows the ROC reaction of different datasets, created from K-fold cross-validation. Taking all of these curves, it is feasible to analyze the mean area under curve, and see the variance of the curve when the training set is split into unlike subsets. This unevenly shows how the classifier output is affected by changes in the training data, and how different the splits generate by K-fold cross-validation are from one another.





V. Conclusion

Due to a large number of real-world applications discovering people's opinion is important in better decision making, therefore, there is exciting new research in the field of sentiment analysis. Recently people have started to express their opinion on the web that increases the need for analyzing opinion online content for the various real-world applications. There is a enormous scope of development of these existing sentiment analysis model and to describe an investigation to find automatically the polarization of textual review given by the user as positive or negative. Because rankings and stars are becoming crucial in assisting potential clients to take decision or buying items, there is a definite need for such a research. We have used flip kart data set for performing experiments and then to apply a machine learning algorithm to predict the review. Our system has achieved recognition rate of 98.5% by using MLP which is better than recognition rate of existing system.

References:

- [1]Thakkar, Harsh & Patel, Dhiren. (2015). Proceed for Sentiment examination on Twitter: A State-of-Art study.
- [2]Arti, K. P. Dubey and S. Agrawal,(2019) "An Opinion Mining for Indian ruler association Using Machine Learning Techniques".
- [3]A. Razia Sulthana, A. K. Jaithunbi, and L. Sai Ramesh,(2018) "Sentiment analysis in twitter data with data logical techniques for analytical model".
- [4]L. Kao and Y. Huang,(2017) "Predicting give for meaning according to fan page users' sentiment".
- [5]F. Ren and Y. Wu,(2013) "Predicting User-Topic Opinions in Twitter with Social and Topical Context." in IEEE communication on sentimental Computing.
- [6]R. Chakraborty, M. Bhavsar, S. K. Dandapat and J. Chandra,(2019) "Tweet Summarization of News Articles: An purpose Ordering-Based perception," in IEEE Transactions on Computational Social Systems.
- [7]J. Zhu, H. Wang, M. Zhu, B. K. Tsou and M. Ma,(2011) "Aspect-Based Opinion poll from Customer Reviews," in IEEE Transactions on emotional Computing.