E-Commerce Site's Fake Review Detection and Sentiment Analysis using ML Technique

J Bharatkumar[1], Kartik M[2], Kiran Shetty[3], K Shreyas Pai[4,] Sunil Kumar S[5]

Student, Computer Science & Engineering, MITE, Moodibidri, India[1-4]

Senior Assistant Professor, Computer Science & Engineering, MITE, Moodibidri, India[5]

Abstract: Most online stores allow their consumers to post reviews of their products and services. These reviews' presence can be used as a source of knowledge. Reviews are becoming a more important source of information for consumers. Unfortunately, phony reviews by certain parties that attempted to produce fake reviews in order to boost the popularity of their product or to disparage the competitor's goods have undermined the significance of the review. The goal of this paper is to identify fake reviews on e-commerce sites using the text, rating properties, and other information from a review. The project also proposes to classify the reviews as positive or negative based on the text used in the reviews, ratings given to the product so on.

Keywords: Supervised Learning, Flask, Framework, Web Application, Naïve Bayes

I. INTRODUCTION

In general, e-commerce websites give users the option to review a product or service. These reviews' presence can be used as a source of knowledge. For instance, businesses can use it to select how to design their goods or services, and prospective customers can use it to decide whether to purchase a product or not. Positive reviews influence customers to buy products and generate revenue, however negative reviews frequently result in sales declines on e-commerce websites. Unfortunately, phony reviews written by some parties to boost the popularity of their product or to disparage a competitor's goods limit the value of the review. As a result, fake reviews are a serious issue for e-commerce sites and other service providers because today's consumers rely heavily on them. The main aim of this project is to assist customers in selecting the best product by identifying phony reviews on E-Commerce sites using the review text, rating**s** given to product and other information. The project also performs the sentiment analysis on customer reviews to classify the reviews as positive or negative based on the text written in the reviews, ratings of the product etc.

II. LITERATURE REVIEW

[1]   P. Chaudhary, A. Tyagi, and S. Mishra, wrote a paper on Supervised Classification for the detection of fake reviews. The goal of this study is to investigate the important review and reviewer-centric elements. Additionally, this work develops and evaluates a few additional criteria that may be beneficial in identifying authentic and fake reviews.

[2]   E.I. Elmurngi and A. Gherbi proposed a paper on Sentiment Analysis and Machine Learning Approaches for Detecting Unfair Reviews. In this study, movie reviews are classified as positive or negative based by performing sentiment analysis using machine learning algorithms.

[3]   Algotor K, Bansal A, written a paper on "Using Opinion Mining to Find Genuine and Useful Consumer Reviews for Products". This paper describes a study that tries to discover and rate true, relevant reviews using an efficient Supervised Learning Method.

[4]   Kumar J, has written a paper on "Using Behavioural and Contextual Features to Detect Fake Reviews". The phony review is recognized in this study by evaluating reviewer deviance as well as other environmental and behavioural factors.

[5]   Ahmed M. Elmogy, Usman Tariq, Atef Ibrahim, published a paper on "Detecting Fake Reviews Using Supervised Machine Learning". This research suggests a machine learning strategy for detecting bogus reviews.

III. TECHNOLOGY USED

A. HTML: HTML which stands for Hyper Text Mark-up Language, is the predominant mark-up   language for web pages. It provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists etc., as well as for links, quotes, and other items. It allows images and objects to be embedded and can be used to create interactive forms.

B. CSS3: CSS3 is the latest evolution of the Cascading Style Sheets language and aims at extending CSS2.1It is used with HTML to create content structure, with CSS3 being used to format structured content. It is used with HTML to create content structure.

C. Flask: Flask is a micro web framework written in Python. Since pre-existing third-party libraries already do common functions, it doesn't have a form validation layer, database abstraction layer, or any other component. However, enabled extensions are used to apply application functionalities exactly as the built-in functionalities of framework.

D: Weka Tool: Weka contains a collection of visualization tools and algorithms for data analysis and predictive modelling, together with graphical user interfaces for easy access to these functions.

IV. METHODOLOGY

The fake review detection model and sentiment analyser model are implemented by adopting different Machine Learning Classifiers. The dataset for training the models is collected from Amazon customer review dataset labelled with fake or genuine for target class. To improve the accuracy of the model, the feature selection method is employed, using which some of the important features are selected among the available features by dropping unwanted features. Additionally, the feature extraction method is used to generate the new set of features from original one. The Naive Bayes Algorithm was used to build the fake review detection model, and the accuracy it provided is compared to that of the Random Forest Classifier. For training dataset, the Random Forest classifier performed classification more accurately than the Naive Bayes classifier. In other situations, the Naive Bayes classifier outperformed competing methods. demonstrating how effectively it can predict false reviews and make better generalizations. On the other hand, the sentiment Analyzer model is trained using Support Vector Classifier and the accuracy that the model emitted is above 85 Percent.
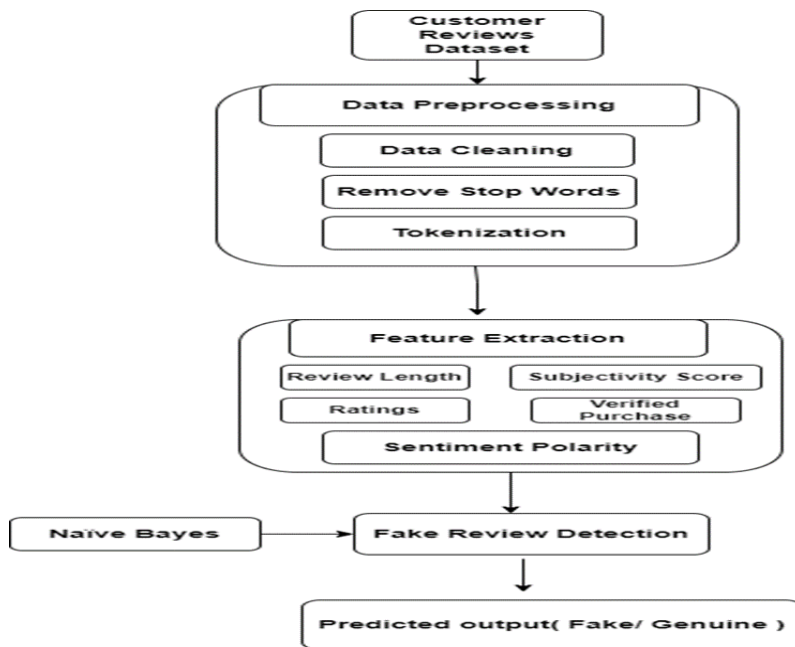

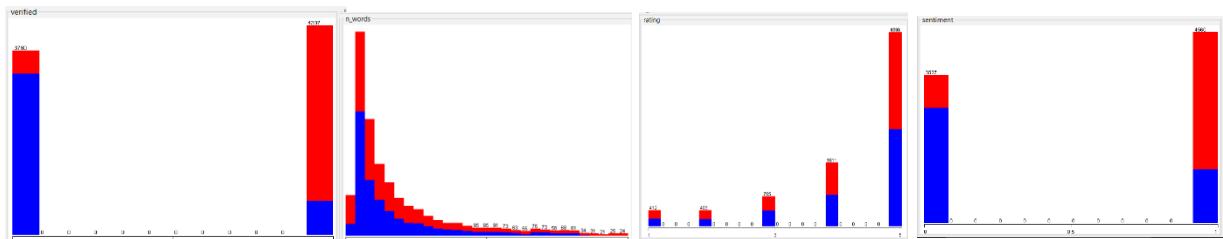Fig 1. Proposed system architecture

Fig 2. Verified Purchase Vs Target class   Fig 3. Word Counts Vs Target class   Fig 4. Rating Vs Target class
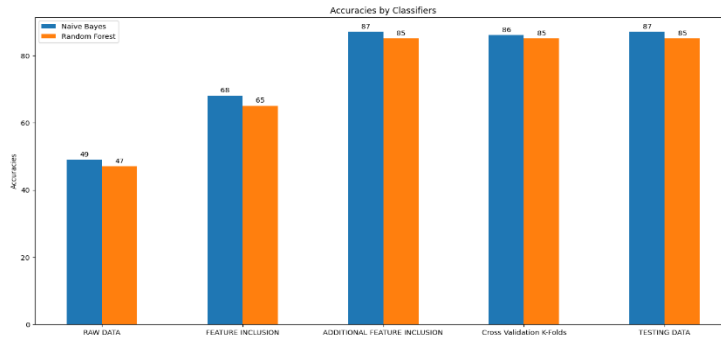Fig 5. Sentiment Polarity Vs Target

Fig 6. Model Accuracy graph

V.   RESULTS

With the help of the Python Flask framework and frontend technologies like HTML and CSS, a web interface is created to display the outcome. Using the web interface users can enter the product reviews to identify whether the review is Fake or Genuine.
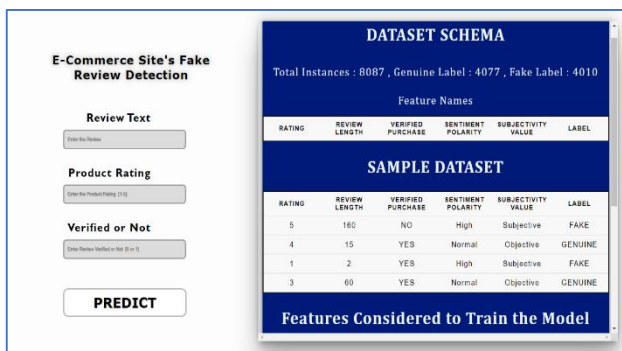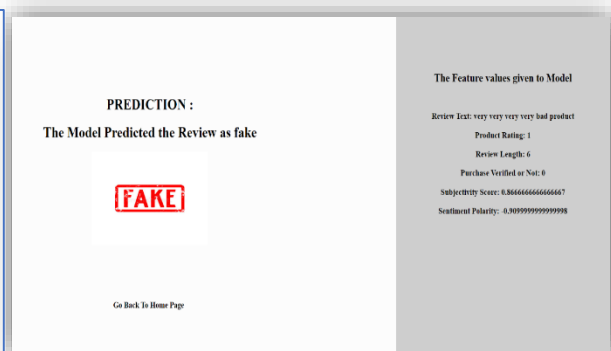

Fig 7. Fake review Detector

Fig 8. Predicted review as Fake

VI.   CONCLUSION

The Project will detect the fake reviews and will help the potential customer to identify and distinguish between Real and Fake reviews. So that the customer can take right decisions while making online purchase through E-Commerce sites. The predicted result is be represented in the web interface which will be useful for the users to interact.

REFERENCES

[1] E. I. Elmurngi and A.Gherbi, "Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques," Journal of Computer Science, vol. 14, no. 5, pp. 714– 726, June 2018.
[2] Alimuddin Melleng, Anna-Jurek Loughrey, Deepak P, Text Representation Based on Sentiment and Emotion for Detection of Fake Reviews.
[3] B. Liu and M. Hu, "Opinion Mining, Sentiment Analysis and Opinion Spam Detection," [Online]. Available: https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon [Accessed: January 2019].
[4] Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-gyo Jung, Bernard J. Jansen, Creating and detecting phony online product reviews.
[5] Hassan Raza, M. Faizan, Ahsan Hamza, Ahmed Mushtaq, Naeem Akhtar, Machine Learning Techniques for Scientific Text Sentiment Analysis.
[6] Wu, Xian, Dong, Yuxiao, Tao, Jun, Huang, Chao, Nitesh, V., Chawla, 2017. Reliable fake review detection via modeling temporal and behavioral patterns. In: 2017 IEEE International Conference on Big Data. Big Data), pp. 494–499. https://doi.org/ 10.1109/BigData.2017.8257963.