OFFLINE PRODUCT DETECTION AND RECOMMENDATION SYSTEM

Aishwarya Kelgandre, Computer Engineering, Trinity Academy of Engineering, Pune, India
Kishor Dhaygude , Computer Engineering, Trinity Academy of Engineering, Pune, India
Aditya Battul , Computer Engineering, Trinity Academy of Engineering, Pune, India
Atharva Tikone, Computer Engineering, Trinity Academy of Engineering, Pune, India

Abstract
E-commerce is currently growing in the technological industry. E-commerce platforms have a lot of advantages, but they also have risks and disadvantages. There are risks, such as writing false product reviews, not rating products correctly, using shady websites, purchasing subpar products, etc. Due to the abundance of e-commerce websites, customers frequently struggle to choose one to use while making purchases. We have developed a mechanism that can help people with these problems. Our technology employs artificial intelligence to rank and assess websites based on user reviews. Customers can now utilise the internet to buy any thing from anywhere thanks to technological advances. Based on the product's quality, the customer might comment, rate, or write a review. Giving users a grade based on elements like product quality and other criteria is the project's main objective. Online commerce platforms are many. On phoney websites, fraudulent activities could occur often. The suggested approach assigns ratings to online stores based on the feedback or opinions of prior users. Performance of the product is condensed to a collection of variables. These elements are used to rate that particular product, which the consumer can use to decide which one to purchase.
Keywords–:Product, Machine Learning, Mining.

I. Introduction

Since the process of evaluating technology began, it has developed swiftly and is still doing so. People's everyday routines are evolving today. They want everything delivered quickly to their door. As a result, the technology sector is rapidly expanding. Everything has benefits and drawbacks. Technology has greatly streamlined and facilitated our lives. On sometimes, people misuse technology.

Technology has made it feasible for the e-commerce sector, which is growing every day. With only a few clicks, consumers may purchase any kind of thing online, saving themselves the extra time and effort of having to go to a store to make their purchases. This is just one of the many services it offers.
Fraudulent users emerged as the user base grew. These frauds are affecting a sizable number of victims. The use of false accounts, the provision of inferior goods, the recommendation of inappropriate goods, wrong pricing, etc. are just a few of the many ways they have found to carry out these scams. Before making a purchase, consumers should check the product reviews provided by previous buyers. However, not all reviews can be trusted. These reviews could divert the audience. Large-scale occurrences like this should be stopped. We have created a technique that can assist users in determining a reliable website and high-quality merchandise. And this can aid in reducing the chance of falling victim to these scams.

I. RELATED WORK

Detailed To identify group colluders in online product review spam campaigns, the pair wise features are first explicitly used. This can expose collusions in spam campaigns at a more granular level. In [1] article, spam campaigns that have been discovered on well-known product review websites (like

amazon.com) have drawn increasing interest from both industry and academics. In these campaigns, a team of online posters is employed to work together to create false evaluations for some target products. The objective is to influence the targets' perceived reputations in a way that serves their best interests.

Online product reviews have become into a significant source of customer opinions, according to [2] study. Imposters have been creating false or misleading reviews to promote and/or denigrate some specific businesses or services due to money or fame. Review spammers are those who act as imposters. Several strategies have been put out in recent years to address the issue. Use a different strategy in this work that takes advantage of reviews' burrstones nature to spot review spammers.

Online reviews of goods and services can be highly helpful to clients, but they must be safeguarded from manipulation, according to [3] study. The majority of studies to far have mainly examined online reviews from a single hosting site. How could information from various review hosting sites be utilised? Our main research question is this one. Create a methodical approach to combine, contrast, and assess evaluations from various hosting sites in response. leverage more than 15 million reviews from more than 3.5 million users across three major travel websites with a focus on hotel reviews.

Users increasingly rely on user-generated content, such as comments on Yelp and Amazon reviews and liked post and ad content on Facebook, according to [4] study. This has created a market for underhanded marketing strategies like collusion networks, bogus accounts (like Sybil), and compromised accounts. The majority of current methods for detecting this behaviour rely on supervised (or semi-supervised) learning over actual or potential attacks. They cannot identify attacks that the operator overlooked when labelling or when the adversary modifies their tactics.

Online reviews have grown in importance as a resource for decision-making and product creation, according to [5] study. However, opinion spam is a common target of review systems. Although supervised learning has been used by researchers to study fake review detection for years, the ground truth of huge datasets is still unknown, and the majority of supervised learning techniques currently in use are based on phoney reviews that aren't actually fake. Present the first study on fake review detection in Chinese using filtered reviews from Damping's fake review detection system in collaboration with Dianping1, the largest Chinese review hosting website.

Online reviews are cited in [6] article as one of the most significant sources of information for customers on a variety of goods and services. Due to their growing significance, there is a greater chance for spammers or unscrupulous business owners to fabricate evaluations in an effort to promote their products and services or disparage those of their rivals. There have been numerous research on the best approaches to identify review spam using different machine learning algorithms in response to this rising issue. The conversion of reviews to word vectors, which has the ability to produce hundreds of thousands of features, is a recurring theme in the majority of these investigations.

Based on the two premises that people are more likely to trust reviews from those who are connected to them and that review spammers are less likely to maintain a significant network of relationships with regular users, the [7] paper presents an efficient and effective method to identify review spammers by incorporating social relations. This paper makes two contributions: (1) it describes how social relationships can be taken into account when predicting review ratings, and it suggests a trust-based rating prediction model that uses proximity as a measure of trust; and (2) it develops a trust-aware detection model based on rating variance that iteratively calculates user-specific overall trustworthiness scores as a spam city indicator.

Using the content and rating properties from a review, [8] article attempts to identify phoney product reviews. In essence, the suggested system (ICF++) will evaluate the reliability of a product, the trustworthiness of the reviewers, and the honesty of a review. The text mining and opinion

mining methodologies will be used to determine a review's honesty value. The experiment's findings demonstrate that the proposed system is more accurate than the outcomes of the iterative computation framework (ICF) method.

Online social networks (OSNs), according to [9] article, are used for a variety of purposes, including commerce, education, telemarketing, medical care, and entertainment. OSNs capture the structure and dynamics of person-to-person and person-to-technology interaction. Additionally, this technology makes illegal actions possible. In this new view of social life that articulates and reflects off-line connections, spotting anomalies is crucial because they can indicate a serious issue or contain information that would be helpful to the analyst.

In [10] paper, they put forth a fresh, all-encompassing strategy they call SpEagle that gathers information from all metadata (text, timestamp, and rating) and relational data (network) and uses it collectively to identify dubious users and reviews as well as spam-targeted products. The review-network-based classification task used by SpEagle allows prior knowledge regarding the nodes' class distribution, which is estimated from metadata. Positives include: When labelled data is available, it facilitates smooth integration of that data. It is quite effective.

Mangoes are divided into four categories in this research [11] based on machine learning, including Green Mango, Yellow Mango, and Red Mango. The size and form of the mangoes are taken into account using the RGB values. The analysis that follows is used to get good probabilities. This aids in training the system to recognise mangoes at the proper age. Naive Byes and SVM (Support Vector Machine) are two machine learning techniques used in this study.
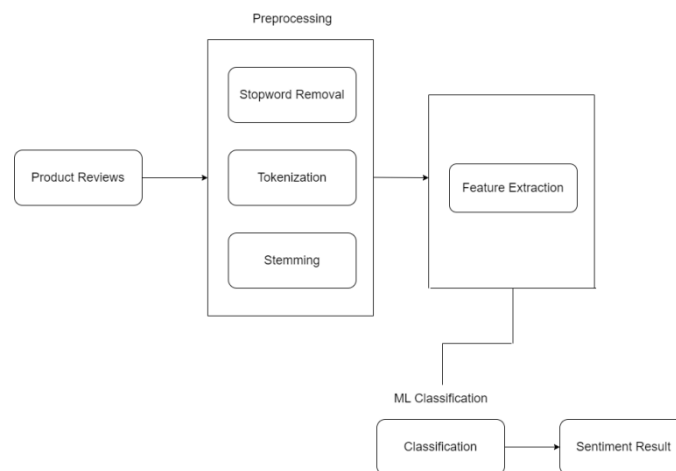
II. PROPOSED SYSTEM

Figure 1. Proposed System Architecture

Algorithm:
Naive Bayes:
Step 1: Convert the data set into a frequency.
Step 2: Create Likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.
Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.
For example:
Problem: Players will play if weather is sunny. Is this statement is correct?
We can solve it using above discussed method of posterior probability.
P(Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)
Here we have P (Sunny |Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36, P(Yes)= 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

Naive Bayes uses a similar method to predict the probability of different class based on various attributes. This algorithm is mostly used in text classification and with problems

4) Performance Evaluation: The evaluation can be done based on following factors:

i) Performance matrices such as TPR FPR Precision Recall etc.

ii) Impact of Different Sampling method

iii) Investigation of time related data
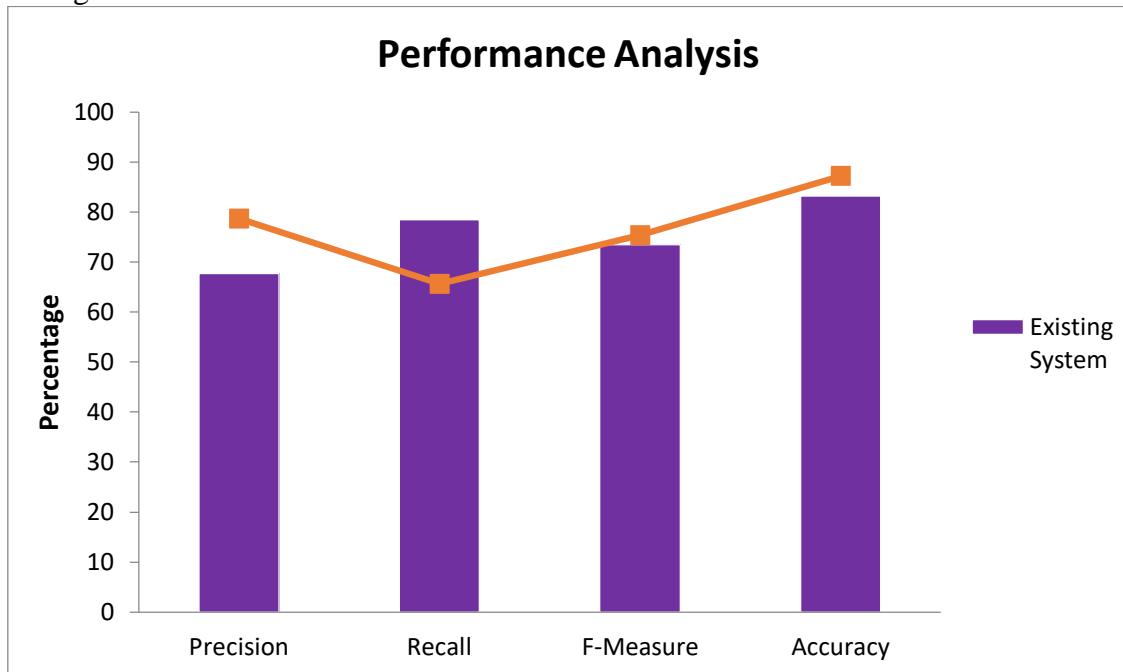
	Existing System	Proposed System
Precision	67.78	78.70
Recall	79844	65.64
F-Measure	73.11	74.31
Accuracy	83.29	87.26

Conclusion

The customer gains a better grasp of e-commerce websites thanks to our method. It shields the client from falling victim to fraud. They will receive a thorough survey about the quality of the service or customer comments regarding those online stores. Our system might serve as a critic for the shopping sites, pushing them to fix their flaws. Depending on the data's accessibility, our solution can be extended to many industries. Additionally, it is available to customer opinions based on their regular online buying experiences. To overcome the cyber frauds. To make the customers aware of fraudulent websites using customers reviews. This Application shows customer to check the product neary by store whether the Product is Avalable or not at that store. with the help of reviews and available

References

[1]. Ch. Xu and J. Zhang," Combating product review spam campaigns via multiple heterogeneous pairwise features", In SIAM International Conference on Data Mining, 2014.

[2]. G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting bustiness in reviews for review spammer detection", In ICWSM, 2013.

[3]. A. j. Minnich, N. Chavoshi, A. Mueen, S. Luan, and M. Faloutsos, "True view: Harnessing the

power of multiple review sites",In ACM WWW, 2015.

[4]. B. Viswanath, M. Ahmad Bashir, M. Crovella, S. Guah, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards detecting anomalous user behavior in online social networks", In USENIX, 2014.

[5]. H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao," Spotting fakereviews via collective PU learning", In ICDM, 2014.

[6]. M. Crawford, T. M. Khoshgoftaar, and J. D. Prusa," Reducing Feature Set Explosion to Faciliate Real-World Review Sapm Detection", In Proceeding of 29th International Florida Artificial Intelligence Research Society Conference, 2016.

 [7]. H. Xue, F. Li, H. Seo, and R. Pluretti," Trust-Aware Review Spam Detection",IEEE Trustcom/ISPA.,2015.

[8]. E. D. Wahyuni , A. Djunaidy," Fake Review Detection From a ProductReview Using Modified Method of Iterative Computation Framework", In Proceeding MATEC Web of Conferences, 2016.

[9]. R. Hassanzadeh," Anomaly Detection in Online Social Networks: Using Datamining Techniques and Fuzzy Logic", Queensland University of Technology, Nov, 2014.

[10] R. Shebuti, L. Akoglu," Collective opinion spam detection: bridging review networks and metadata", In ACM KDD, 2015.

[11] G.D. Upadhye, D.Pise,"Grading of Harvested Mangoes Quality and Maturity Based on Machine Learning Techniques",IEEE International conference on smart city and Emerging Technology,2018.