E-Commerce with detection of illegitimate web pages and fake reviews

Sai Sarvani Gudluri[1], Sudheer Konijeti[2], Venkatesh Moparthi[3], Salma Banu Shaik[4], Dr.K.M.Rayudu[5]

[5] professor,[1,2,3,4] B.Tech.., Scholors, Department of Computer science and engineering, QIS College of Engineering and Technology.

Abstract:

The act of obtaining confidential information of the user from some dangerous websites, which are known as phishing websites or illigitimate webistes, causes threats to all the users who use the e-commerce website. This will also lead to some financial threats. The main concern of this research work is to detect these types of illigitimate webpages. Here, we observe the similarities in these websites and will subject them to the machine learning algorithms and compare all the algorithms to determine the best. This model helps determine the ligitimacy of webpages before the online payment.

For the product information, customers highly depend on reviews. Taking this as an advantage, the reviews are getting faked by the enterprises for increasing their product's reputation. Therefore, there is a high necessity of detecting the fake reviews. To resolve this issue, we use some machine learning algorithms like K-Nearest Neighbour, Random Forest, Naive Bayes, Support Vector Machine(SVM). These algorithms aid in classifying fake reviews and original ones.

Keywords:

e-commerce, machine learning, fake reviews, K- Nearest Neighbour, Random Forest, Naive Bayes, Support Vector Machine.

Introduction:

During the past decade, the e-commerce has grown rapidly. And as the customers cannot touch the product manually, they are highly depending on the reviews, which made them the most significant aspects to determine the success of a product or a service. A small change in the url may tend the customers to get contact with some fake websites. So it become really important to detect fake reviews and fake websites. We can detect the fake review from number of domains. Fake review is generally a review which misleads the customers[6].

Here we mainly focus on detection of fake reviews and illigitimate webpages. To fullfil this job we use some supervised, semi-supervised and unsupervised classification algorithms such as KNN, Random Forest, Multi layers algorithms to improve the performance of classification. Under semi-supervised classifiers we use label propogation and label spreading algorithms.

Literature Survey:

1) Tittle:

Detection of fraudulent hotel reviews on the internet

Anna V. Sandifer, Casey Wilson, Aspen Olmsted

To decide on available goods and services, people consult online reviews. Businesses and the academic community have become quite interested in identifying false online reviews in recent years. The use of precise algorithms to identify bogus internet reviews helps shield users from spam and false information. We collected filtered and unfiltered online reviews from yelp.com for a number of hotels in the Charleston region. From the data collection, we extracted part-of-speech features, used three classification models, and compared accuracy outcomes to relevant works[1].

2) Tittle:

Ensemble Machine Learning Model for Spam Product Review Classification

M. Fayaz, Atif Khan, Javid Ur Rahman, A. Alharbi, M. I. Uddin, Bader Alouffi less

Online product reviews have become an integral part of how consumers and businesses evaluate products today. They offer suggestions to a business on how to enhance the quality of its products and how to plan and manage its business strategies in order to boost sales and increase profits. Additionally, they assist clients in making the best product choices with less time and effort. Most businesses create spammy product reviews in an effort to boost sales and boost profits. Spam product review detection is a difficult problem in NLP (natural language processing). A variety of machine learning techniques have been used to try and identify and categorise product evaluations as spam or not. However, this study presented an ensemble machine learning approach to raise the classification accuracy[2].

3) Tittle:

For the classification of products in Vietnamese e-commerce, an incremental ensemble learning system

Linh Nguyen Tran Ngoc, Vu Hong Quan, Le Hoang Ngan, Tran Duy Phu, Hoang-Quynh Le less

Text classification algorithms are becoming more and more crucial to businesses as e-commerce platforms flourish. Businesses will likely encounter significant difficulties related to dataset imbalance, continuously added data, language specificity, and product specialisation. In this article, we provide a scalable incremental machine learning system for large-scale implementation in actual business operations. The system also has procedures for enhancing the details of e-commerce products. When compared to models without them, the

proposed tactics—keyword dictionary mapping, sampling, and ensemble learning—brought about improved performance. Additionally, our tests demonstrated that minibatch SVM gave positive outcomes and may be significant in a lighter system[3].

4) Tittle:

Online Transaction Fraud prevention applying Backlogging via E-Commerce portal

Swapnil Malatesh Dharnappa Goudar ,Gyanangshu misra

In today's digital age, credit card fraud is one of the most common crimes. A number of restrictions have been implemented to prevent plutocrats from being gained through illicit means in order to prevent this crime from spiralling out of control. Essentially, credit card fraud occurs when a perpetrator obtains credit card information illegally and then utilises it to carry out his or her unjustifiable intentions. Machine learning techniques have been employed by experimenters to detect this type of credit card fraud. But there are still certain flaws in the delicate posture that need to be fixed in these situations. Therefore, despite this development, fraud detection is still in its infancy, and more effective methods must still be developed in order to entirely reduce the risk of credit card fraud[4].

5)Tittle:

Big Data Analysis of the Impact of Emotional Factors (Positive and Negative) on the Utility of Product Reviews

Lianzhuang Qu School of Information and Business Management, Dalian Neusoft University of Information

Product reviews are crucial in the era of successful online shopping since they influence consumers' decisions to buy things. Additionally, it can assist companies in identifying the corresponding flaws and implementing targeted fixes. However, user comments are rife with emotion, and remarks that exhibit strong emotional inclinations have a bigger influence on consumers than common comments, particularly those that express negative feelings. At the moment, the majority of text sentiment analysis is focused on chapters and sentences, and there aren't many examples of sentiment analysis that has been honed to product-specific features. Based on this, this article aims to investigate the elements that affect the emotional tendency of product evaluations, analyse a significant number of reviews of a certain mobile phone device, and extract the keywords of these reviews[5].

Existing System:

Existing system is a completely unique approach for the automated identification and extraction of product worth information from e-shop websites that is freelance from the e-shops' language and also the product domain.

The method was independently developed from a language or a product domain.

Disadvantages:

1. In this system Accuracy is low
2. Fake reviews also pose additional problems in that they mislead consumers to spend money with a business or on a product they may not have otherwise chosen.
3. While almost all (88%) shoppers use ratings and reviews to evaluate or learn more about products, fake reviews can affect their ability to confidently turn to reviews as a trusted source
4. Damaged Reputation
5. Public Danger – Fake reviews of physicians, attorneys, accountants or auto repair shops could potentially harm consumers

Proposed System:

Each review initially goes through the tokenization procedure in the proposed system. After that, extraneous words are eliminated, and potential feature words are created.

Each potential feature word is compared to the dictionary, and if it has an entry, its frequency is calculated and added to the column in the feature vector that corresponds to the word's numeric map.

The length of the review is calculated and added to the feature vector together with counting frequency. The feature vector is then updated to include the sentiment score that is included in the data set. Positive sentiment has some positive values while negative sentiment has zero values in the feature vector.

This kind of phishing websites will have a detrimental effect on businesses' entire operations, marketing initiatives, client relationships, and revenue.

Additionally, these phishing attempts will cost businesses hundreds of dollars every attack, and this cost will be linked to the damage done to the brand image and customer confidence as a result of these attacks.

The major goal is to assess if automated data mining approaches can effectively identify the complex issue with phishing websites.

Ottet al"gold .'s standard" dataset, which we used, was created.

Advantages:

1. Due to semi-supervised and supervised learning, the system is incredibly quick and efficient. Centred on the review-based techniques' substance. In addition to word frequency and emotion polarity, review length was utilized as features.
2. Helps Customers know about the ligitamacy before payment.
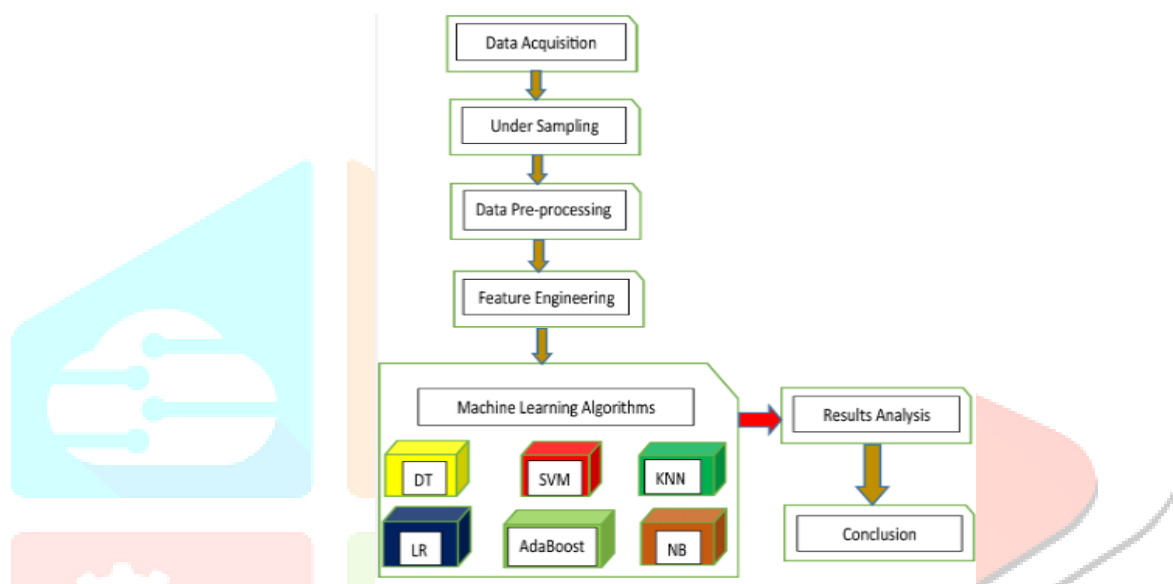
Working Principle:

fig: Steps involved in fake review detection[7]

There are three main steps in identifing or detecting a fake review. These include Data Pre-processing, Feature Extraction and Classification[8].

Data Pre-processing: It is an act of cleaning and organizing the raw data such that it is well suited to train the machine learning algorithms to get the desired output. Here we remove noise, duplicates, errors, incomplete and inconsistent data.

Feature Extraction: In our research work we have extracted the features manually based on the URL because in some cases by looking into the URL we can identify the maliciousness of web pages to some extent or by querying the information associated with the referenced host, it's safety can be detected. And for the reviews we have considered the keywords and classified the reviews as highly negative, negative, positive and highly positive. It helps increase the performance by detecting the accurate pattern[9].

Appling Classifiers: Here we classify the review or webpage based on some supervised, semi-supervised and un-supervised classifiers from machine learning such as Logistic regression, Gaussian Naïve Bayes, Random Forest, Support Vector Machine Training algorithms.
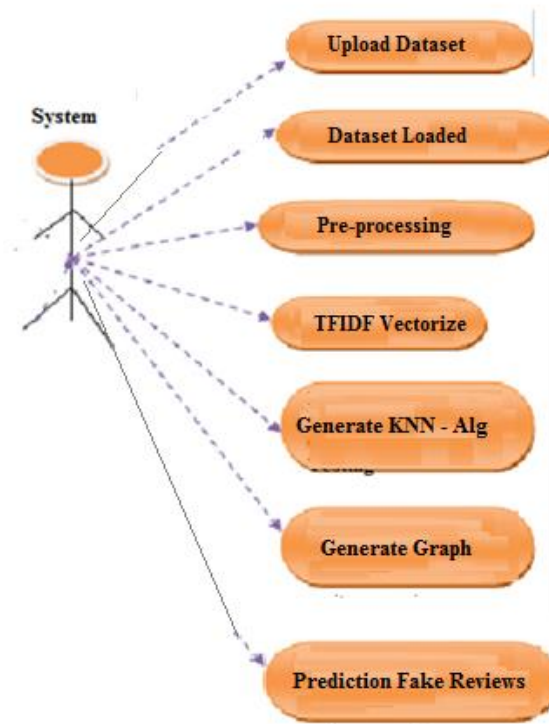
Conclusion:

We have considered several supervised, semi supervised and unsupervised classification techniques to detect illigitimate webpages and fake reviews. And we considered the classification algorithms which are not used on previous research works. We have compared all the accuracy of all the classification algorithms used to determine the best classification algorithm. During this research work we have found that Naïve Bayes Classifier gives the highest accuracy, which is greater than the accuracy of algorithms used in previous research works.

References:

1)https://www.semanticscholar.org/paper/Detection-of-fake-online-hotel-reviews-Sandifer-Wilson/db586a5e5bc3be06482e544665c1494940ed6a75

2) https://www.semanticscholar.org/paper/Ensemble-Machine-Learning-Model-for-Classification-Fayaz-Khan/4450b179e19d312c01b1076624ada4d46223da47

3)https://www.semanticscholar.org/paper/An-incremental-ensemble-learning-system-for-product-Ngoc-Quan/680a563dc60cd0813649cdb1f3f13fc6fbee946c

4) https://www.ijert.org/online-transaction-fraud-detection-using-backlogging-on-e-commerce-website

5) https://www.semanticscholar.org/paper/Influence-of-Emotional-Factors-(Positive%2C-Negative)-Qu/15555732a6fa5181c15ac2a928d9357f44d758b6

6) https://utu.io/blog/12-things-you-can-do-right-now-to-fight-fake-reviews/

7) https://images.app.goo.gl/PsbXLpj1UvjTUZcX9

8) https://ijarcce.com/wp-content/uploads/2022/07/IJARCCE.2022.11733.pdf

9) https://thesai.org/Downloads/Volume12No1/Paper_69-Fake_Reviews_Detection_using_Supervised_Machine.pdf