

E-Commerce With Price Comparison, Price Alert & Fake Review Using Machine Learning

U. Prasad,¹ K. Amrutha², B. Sowmya³, L. Sippora⁴, M. Anusha⁵

¹ Asst. Professor, Department of Computer Science and engineering

^{2,3,4,5} Student, Department of Computer Science and engineering

QIS Institute of Technology & Science, Ongole, Prakasam (Dt)

Abstract:

Phishing is a kind of criminal behaviour in the internet world that involves obtaining sensitive webpage from users using rogue websites that seem to be reputable websites. AssociateBy utilising such websites, a degree adversary will utilise this form of phishing or fraud, which might be a serious problem.to users of the internet for their personal and sensitive information As a result, when it comes to e-banking and e-commerce, this action poses a risk to all webpage visitors. During this paper, the major focus will be on determining the many valid options. Websites that are questionable or phishing. These options are given into the machine learning algorithm. As a result, constitutional algorithms are employed for comparison and determining the algorithmic rule's accuracy.J48, Nave Bayes, random forest, and supply Model Tree (LMT) are the algorithms employed in this comparison. It is computed using them to properly determine the validity of a website. In addition, the most successful algorithmic rule will be chosen from among a variety of algorithms. We'll compare the outcomes in both approaches in this paper. That. To begin, we mistreat the comparison of the many qualities in order to get the optimum algorithmic rule. Such as successfully classified instances, incorrectly classified instances, mean absolute error, and letter of the alphabetStatistics about the alphabet. Second, the accuracy of such algorithms may be evaluated using a variety of factors such as TP.FP, Rate Rate, Precision, Recall, F-Measure, MCC, mythological creature space, and space that belongs to the People's Republic of China Within the chart, the data is displayed. The algorithmic rule adopted automates the web site analysis procedure. This prediction model may be used to determine the validity of any e-commerce web site before making a payment of that web site.

1. INTRODUCTION

These phishing web pages, which are fake pages that seem exactly like actual web pages, were developed by some evil persons. Some of the fake web pages have a lot of aesthetic similarities with actual web pages in order to deceive their victims. Pages have the same appearance as genuine web pages. This type of fake website easily deceived a number of naive internet users. These bogus websites will acquire personal information from their victims, such as their credit card number. Passwords, bank accounts, and other sensitive information should all be kept private. Phishing is a type of online crime. Other online crimes authors notice first, such as hacking and viruses, when a web page is seeking for a new crime. Authors are aware of this. The number of phishing websites has been dramatically expanding in recent years. Phishing is a social engineering and criminal conduct that involves collecting personal information from victims by visiting a fake website. The receptionist will ask you to enter your personal information, which will seem to be the same as actual or authentic web

pages. Attackers may easily obtain a common password by using information like your account number, username, and password. Persons likewise those thoughtless individuals who will never check the website's accuracy and demand and will rush to visit it. If they enter their information on this fake website, the attacker will cling to them, posing a serious threat to them and their family. Victims lose money, other valuables, and highly sensitive and secret information. In a nutshell, authors might state that online phishing is a widespread social engineering technique that causes increasingly significant problems in today's online e-commerce environment—e-commerce and banking the term phishing comes from the website phishing, which is a vibration of the word fishing. The phishing website's concept is similar to that of a fish hunting attraction that is displayed to internet users and when they click on it, they are taken to a phishing website. If a user comes to take advantage of this chance, the user will fall into the trap set by the phishing site's owner. As authors, we know that in the majority of situations, the attraction will come in the form of an instant messaging site or an email. Lead the user to a phishing website that is actively awful. It may be said that phishing is a unique type of online crime. A network via which the owner of a phishing website may easily get the information he needs from the victims this phishing website is being used. The most prevalent phishing websites impersonating well-known institutions and online traders this work is published under a Creative Commons Attribution 4.0 International License, and credit card companies and others have a variety of faces in e-banking and online shopping. Can easily trap anybody. Phishing websites of this type will have a detrimental influence on the whole economy. Marketing efforts, client connections, and income are all factors to consider. Furthermore, these phishing attempts will cost businesses money. Hundreds of dollars every assault, and this cost will be linked to the damage done to the brand's image and consumer trust. As a type of phishing assaults like these, a lot of people's lives have been ruined. The major goal is to see whether automated data mining can be used in the future. Techniques for recognising the complicated issue of phishing websites Authors utilise prediction in a way that is closely comparable to in this example, the important characteristic is the degree of phishing, which is classified as a classification challenge. Approach for data mining Different features, such as spelling problems, are used in phishing website classification approaches. Prefixes, suffixes, and customisation of lengthy URLs, among other things these features may be found on a variety of websites, as well as from the use of many online tools

2. THEORY

The suggested technique includes all necessary processes and methods for identifying and extracting product records from a database.

Arbitrary e-commerce websites that are;

- (1) Crawling the online store's website
- (2) Identifying and extracting product records from e-commerce pages and
- (3) Identifying and extracting product attributes from product records.

CRAWLING (A)

We examined our collection of e-shops for the presence of product listings before crawling the URIs of their websites.

Within the many tiers of a website's link structure the levels are as follows:

- Level 0: The e-commerce website's front page.
- Level 1: E-commerce web pages that may be accessed straight from the homepage through a link.
- Level 2: The e-web shop's pages, which may be accessed through two links from the homepage.
- Level 3: The e-web shop's pages, which may be accessed through three links from the homepage.

B. EXTRACTION AND IDENTIFICATION OF PRODUCT RECORDS

We employ a unique extraction for identifying and extracting product records from e-commerce websites, which we call Light Extraction. The e-shop URIs crawled in the first stage, as stated in Section IV-A, are fed into the algorithm. Light Extraction displays the real URI's Web page and traverses all components inside the HTML page tree.

C. IDENTIFICATION AND EXTRACTION OF PRODUCT ATTRIBUTES

It is necessary to assign the retrieved attributes to pre-defined products in product to handle the acquired data in subsequent phases. Features to understand the significance of the retrieved attributes as a result, information on the product's format is important. Or a working grasp of the structure of product records, particularly the product attributes included inside them may be used to extract the attributes of a product. The suggested method locates and extracts product attributes. By using existing knowledge about the format and content of the attributes gathered from the analysis of the 50Websites for e-commerce.

3. Related work

[1] This paper contributes a one-of-a-kind approach to automated product worth identification and extraction information from haphazard e-shop websites, which is independent of the e-language shops as well as the product domain. The approach is based on tag path analysis and takes use of the similar structure of product records found on e-commerce websites. Identifying distinctive product records and extracting their attributes the strategy was developed independently from a product domain. Or a method of language the approach's suitability for identifying and extracting product records from e-commerce websites was shown in an experiment in which the attributes of the full product of two completely distinct e-shops were compared. Extractions are made from websites.

This paper [2] Virtual Cart is a smartphone application that facilitates a straightforward and easy shopping experience. In shopping malls, there is a method for purchasing. Virtual Cart, a hybrid programme, will also provide nice customer service. For twenty-four hours a day, seven days a week in regards to identifying, purchasing, and delivering client requests. There are a total of two blessings. Of it: first, there will be no need to stand in line for an extended

period of time in malls just to scan an item; second, there will be no need to queue for an extended period of time in malls just to scan an item; third, there will be no need to queue for an extended period of time in malls There is no scope for the frauds that happen while searching on a mobile device. The transactions that will often appear in the outlets cloud are as follows:

Will be created in a secure manner.

The purpose of the paper [3] is to create the 'modified SentiWordNet algorithm,' which is developed mostly on machine learning. Used to persuade options to buy the merchandise these features will be kept and perhaps created for the future. Users when they're on the lookout for a merchandise this approach provides a rating at the tip for each feature of the product. This work is licenced under a Creative Commons Attribution 4.0 International License merchandise and therefore advises customers to travel with the one that has the highest product rating in a very specific manner. Feature. When the options are prioritised, the products with the greatest positive ratings in those categories are selected. Would be recommended to the user, assisting them in finding merchandise that meets their needs. Thither approach enables the product's creators to comprehend their customers' input and, as a result, to improve the product. The merchandise's real options and design products that are in merchandise with the client's requirements

This paper [4] proposes a completely alternative model-based technique for identifying and isolating abnormalities. Rather than typical points, the profile takes a different approach. The Isolation Forest, or forest, creates a group of trees for a specific purpose. Anomalies are cases with short average route lengths on the trees in a given information collection.

[5] In this paper, they use an internet data processing system to find hidden patterns and business opportunities in e-commerce. Using information from their clients and the internet, they provide a new framework for data processing that is backed by technology. Constructing a Web-page recommender system, and demonstrating how data processing technologies may be used efficiently in Associate ecommerce environment This paper describes a framework that seeks to find the hidden value in e-commerce. Information about their business and on the internet

[6] In this paper, they collected filtered and unfiltered online reviews for a variety of hotels in the Charleston area from a variety of sources. Yelp.com took the information set and extracted part-of-speech options, then applied three categorization algorithms and compared the results. The accuracy of the findings is based on similar studies. Yelp.com is one of the most popular online review sites. It employs an algorithmic filtration system. When it comes to detecting fake reviews, there is a rule to follow. The algorithm, on the other hand, may be kept a secret. During this work, they gathered feedback from a variety of reviews.

Yelp.com for a hundred random hotels in the Charleston area, then categorised filtered and unfiltered reviews accordingly. Reviewing reviews as if they were real, extracting part-of-speech options, training and testing the information set, creating a model, and so on. Compared the findings to relate work Part-of-speech features were retrieved from the data and 3D modelling was performed. To detect phoney online reviews, researchers used a

variety of classification methods. They found that the greatest level of accuracy was our dataset was classified using the Multinomial Nave mathematician classification model.

This paper [7] contributes a very unique approach to automated product value identification and extraction Information from impulsive e-shop websites, which is freelance from the language of the e-shops, and then the merchandise domain. The approach's novelty as compared to other similar methodologies and tools for net data extraction, as well as its price. The straightforward of the e-shops' independence from the language and hence the product area, is a comparison.

Setup of the extraction activities, as well as the ability to recognise merchandise records automatically among a comprehensive e-commerce site Tag path analysis is used in this approach. The strategy was developed independently from a product. A language or a domain

[8] In this paper, specifically, distinguishing between authentic, questionable, and phishing websites. These are your options. Area unit input into the Maori hen's built-in machine learning algorithms, which are then utilised for comparison and analysis. The accuracy of the algorithmic programme the main goal is to examine into the possibilities of using machine-generated data. In order to discover the sophisticated flaw in phishing websites, processing techniques are used. Four algorithms were utilised by the authors.i.e. For our experiments, we used Nave Thomas Bayes, J48, Random Forest, and Logistic Model Tree. Then there are these algorithms. TheMaori hen data processing approach was compelled to be mistreated in order to explore algorithmic programme accuracy. When these algorithms are executed inside the output window, the following results are achieved.

[9] In this paper, discuss the various information extraction techniques used by various authors that takes the the user required information from a number of different sites. Extracting user information from the internet might be an approach. Information from websites is essential. Using a web crawler, the extraction approach indexes information on the internet theonline thenet. Crawler. The paper presents different internet information extraction techniques that are currently available on the market. A number of each approach has its own recall and precision rates. This shows how effective the extraction procedure is.

In his paper [10], he says: PathSim and SimRank are two similarity search algorithms introduced. Currently, the majority of the networks engaged are networks that are diverse. Learning data mining algorithms has become more important as the trend of data analytics has grown. Andprogramming languages for machine learning many applications use data mining techniques to find the order of similarity between two objects. In such difficult settings, similarity search has become more common. Instead of employing standard methods, similarity search has become more common.SimRank and passim are two new methods that may be employed. PathSim may be superior to PathSim in terms of time complexity.SimRank, however the conclusion is that the algorithm should be chosen based on many circumstances. PathSim may be used in a variety of ways.SimRank is more efficient than SimRank.

4. Proposed System:

In the proposed system, each review goes through tokenization process first. Then, unnecessary words are removed and candidate feature words are generated.

Each candidate feature words are checked against the dictionary and if its entry is available in the dictionary then its frequency is counted and added to the column in the feature vector that corresponds the numeric map of the word. Alongside with counting frequency, the length of the review is measured and added to the feature vector. Finally, sentiment score which is available in the data set is added in the feature vector. We have assigned negative sentiment as zero valued and positive sentiment as some positive valued in the feature vector.

As we have used 'gold standard' dataset prepared by Ottet al. [3], we did not require the steps like handling missing values, removing inconsistency, removing redundancy etc. Instead we needed to merge the texts, create a dictionary and map the texts to numeric value as the tasks of preprocessing. We have used word frequency count, sentiment polarity and length of the review as our features. We have taken 2000 words as features. Hence the size of our feature vector is 160×2002 . We have not taken n-gram or parts of speech as features because these are the derived features from bag of words and may cause over-fitting. The process of feature extraction is summarized

Advantages of proposed system:

The system is very fast and effective due to semi-supervised and supervised learning. Focused on the content of the review based approaches. As feature we have used word frequency count, sentiment polarity and length of review.

5. CONCLUSION

As the need for the mobile looking is increasing the want for a lot of secure, safe and trustworthy dealing is of essential demand. Smartphones, which have become an integral part of modern life, have significantly decreased the amount of work required. That's what you'll need if you're searching. It has two advantages: first, it eliminates the need to replace the queue for a line. Time spent in malls just for the purpose of scanning an item, and second, there will be no room for the frauds that occur with mobile devices. Shopping. The transactions that will frequently take place with the retailer's cloud will be created secure. The trend the popularity of online shopping has returned. Online stores are available 24 hours a day, 7 days a week, and can be accessed from anywhere in the world. From any location with a web connection the advantages and convenience of shopping online can invariably entice a customer. This work is licenced under the Creative Commons Attribution 4.0 International License, which has a large lot of customers. Consumers, on the other hand, must be aware of and attentive to the hazards involved. When you're searching for anything on the internet, be extra cautious. Because of the net market's openness and competition, most Businesses always strive to maintain the highest level of security, but as a user-centered website, they may spice things up a little business.

REFERENCES

- (1) Andrea Hoch, Holger Kett and Anette Weisbecker, "Mining E-Commerce Data from E-Shop Websites" Fraunhofer Institute for Industrial Engineering IAO Stuttgart, Germany, 2015
- (2) K P Vidya, K M Swathi, D Chaitra, S H Jayalakshmi, M V Manoj Kumar, H R Sneha, Likewin Thomas and B H Puneetha, "Virtual Cart: Novel Approach for Revamping Smart Shopping Experience" 2018 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)
- (3) Priyanka M B, Subham Ankit Karan, Pruthvi M, Manushree M R and Sharath Kumar S," Recommender System Using Machine Learning" 2019
- (4) Fei Tony Liu, Kai Ming Ting and Zhi-Hua Zhou ," Isolation Forest" 2008 Eighth IEEE International Conference on Data Mining
- (5) Jinhua Sun and Yanqi Xie," A Web Data Mining Framework for E-commerce Recommender Systems" 2009 International Conference on Computational Intelligence and Software Engineering
- (6) Anna V. Sandifer, Casey Wilson and Aspen Olmsted," Detection of fake online hotel reviews" 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)
- (7) Dao-Quan Li, Hua Yang and Li-li Li," Research and Application of Data Mining Technique in E-commerce" 2010 International Conference on E-Business and E-Government
- (8) Rana M. Amir Latif*, Muhammad Umer, Tayyaba Tariq, Muhammad Farhan, Osama Rizwan and Ghazanfar Ali," A Smart Methodology for Analyzing Secure EBanking and E-Commerce Websites" Proceedings of 2019 16th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 8th – 12th January, 2019
- (9) Miss. N. V. Kamanwar and Prof. S. G. Kale," Web Data Extraction Techniques: A Review" 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR'16)
- (10) SriUsha, K. RasmithaChoudary, Kavitha C. R. and T. Sasikala," Data Mining Techniques used in the Recommendation of E-commerce services"