



E COMMERCE WITH PRICE COMPARISON, PRICE ALERT AND FAKE REVIEW DETECTION

¹Nitha C Velayudhan ²Abinav Vijay P, ³Alisha P D, ⁴Chithira Remesh, ⁵Lakshmi C V

¹Assistant Professor, ²Student, ³Student, ⁴Student, ⁵Student

¹Department of Computer Science Engineering,

¹Universal Engineering College Vallivattom, Thrissur, India

Abstract: Acquiring sensitive information from the user in some malicious sites that appears just like the legitimate webpage which they are doing a form of criminal activity that is named as phishing among the electronic world. Academic degree offender can use this kind of phishing or fraud by exploitation such websites, that might be a severe risk to net users for his or her personal and wind. So, among the sphere of e-banking and e-commerce, this act makes a threat for all webpage users. Throughout this paper within the main discerning the varied choices of legitimate, suspicious and phishing websites. These choices are fed to the machine learning algorithms that are constitutional thus are used for comparison and to establish the accuracy of the rule. Algorithms used during this comparison are J48, Naïve Bayes, random forest and provide Model Tree (LMT) are used and they accurately to predict the net website legitimacy is calculated. Also, the foremost effective rule among utterly completely different algorithms are elected. Throughout this paper, we'll compare the finally ends up within the two ways in which. Firstly, we tend to discover the simplest rule by practice the comparison of the assorted attributes like properly Classified Instances, Incorrectly Classified Instances, Mean absolute error and letter of the alphabet statistics. Secondly, the accuracy of these algorithms will analyse with utterly completely different parameters like TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, monster house and People's Republic of China house that is pictured at intervals the chart. The chosen algorithmic rule makes the net website analysing technique machine-controlled. Before making payment on any e-commerce data processor, this prediction model will be used for determinative the legitimacy of that data processor.

Index Terms –CRAWLING,E-SHOPPING,PRODUCT COMPARISON, PRICE ALERT

I. INTRODUCTION

Some malicious folks created these phishing web content that are faux pages that appear as if an equivalent as real web content. These faux web content have extremely visual similarities with real pages to cheat their victims conjointly a number of the faux web content appear as if same as real web content. Variety of careless web users simply cheated by this sort of faux pages. These faux pages can collect the non-public info of their victims this info may be their mastercard range, password, checking account and conjointly some vital info that would be a lot of confidential. In web crimes, phishing web content is searching for a brand new crime alternative web crimes authors see first off like hacking and viruses. And authors see in recent years that the amount of phishing websites increasing apace. Phishing is social engineering and criminal act of stealing victims personal info by surfboarding a faux web content that appear as if an equivalent as real or legitimate web content here the secretarial assistant can raise you to enter your personal info like your account range, username, and parole and wrongdoers will use this info simply to steal a typical persons conjointly those careless persons UN agency can never check the accuracy and demand of web site the web site and apace go and enter their info during this faux website attacker can stick there's an enormous drawback for them conjointly for his or her family conjointly victims suffer from cash, another quite assets and conjointly terribly personal and confidential knowledge. Shortly authors will say that on-line phishing is loosely launched social engineering attack that makes a lot of serious problems in today's on-line e-banking and e-commerce. Phishing is that the word derived from the web site phishing this can be the vibration of the word fishing. The thought behind the phishing web site is same because the fish searching like an attraction is shown to the web users and once user come back and wish to grab this chance then the user can suggest within the lure from the owner of the phishing web site.

II. THEORY

The appearance of the net has created new avenues for the merchants to sell their product and has additionally reduced the barriers for getting in retail business. This, along with the expanded reach of the net marketing and wide mentioned 'long-tail' phenomenon, has generated intense competition within the on-line retail sector. Within the on-line domain, the oligopolistic nature of the brick and mortar (B&M) marketing has nearly become open-for-all style of marketing. This has raised the amount of accessible choices for the shopper's getting selections and merchant's selling selections. Because the optimum decision-making involves thorough comparison and analysis of all alternatives accessible, the rise in number of accessible choices has created users' call issues a lot of complicated. Moreover, such complexness is combined by the actual fact that as compared to their B&M counterparts, online merchants square measure a lot of various. They're not solely heterogeneous in terms of the services that they supply, however additionally disagree in different areas like the kind of product that they stock (e.g. regular, refurbished, used, outdated, or long-tail) and also the channel mixture of their operations (e.g. pure-play, click and mortar (C&M), and manufacturer-owned). It's no marvel that in the days of the net, intermediaries have occupied a outstanding area in facilitating the decision-making method of each the merchants and also the shoppers. Majority of those Web-based intermediaries bring merchants and shoppers along and facilitate a victorious sales transaction. One amongst the foremost standard kinds of the Web-based intermediaries is that the comparison-shopping agents (CSA) or shopbots. CSAs give call support tools to shoppers for examination their purchase alternatives supported each the value and non-price (e.g. product, businessperson reputation) primarily based factors. The CSA-based call support systems square measure also proactively integrated by some merchants like get.com. The focus of this paper is to review the up to date literature concerning CSAs to investigate them in the context of call support systems (DSS). So as to supply comprehensive call support, a typical DSS ought to have four components: knowledge, models, interfaces, and user specific customization. During this paper, the four part DSS framework is employed to survey the current CSA-based analysis and study the choice support role of the up to date CSAs. The paper provides suggestions for up the choice support facet of the CSAs and proposes a look agenda for the CSA-based call support systems. Our survey of the literature synthesizes this CSA-based tutorial analysis supported the four major components of the DSS: knowledge, model, user interfaces, and users. The CSAs square measure Web-based agents that receive requests from users for the product merchant data and so provide a hierarchic list of the relevant data like value, delivery choices, availableness, and pledge. Although, CSAs' role in finding call issues of the consumers and merchants wide mentioned, they're not analyzed extensively within the context of DSS. The DSS style literature identifies four major purposeful parts necessary for a DSS –data, models, interfaces, and user-specific personalization. Previous researchers have wide discussed the importance of exploit and organizing relevant knowledge, developing and managing quantitative models and rule-based algorithms, simple and economical interactive user interfaces, and user specific focus and customization.

I. PRICE COMPARISON

Price comparison sites can collect data directly from merchants. Retailers who want to list their products on the website then supply their own lists of products and prices, and these are matched against the original database. This is done by a mixture of information extraction, fuzzy logic and human labour. Comparison sites can also collect data through a data feed file. Merchants provide information electronically in a set format. This data is then imported by the comparison website. Some third party businesses are providing consolidation of data feeds so that comparison sites do not have to import from many different merchants. Affiliate networks aggregate data feeds from many merchants and provide them to the price comparison sites. Many of the popular shopping websites provide direct affiliation to the customer who wants to become affiliate partner. They provide their own API to the affiliate partner to show their products with specifications to the affiliate partner's website. This enables price comparison sites to monetize the products contained in the feeds by earning commissions on click through traffic.¹ Other price comparison sites have deals with merchants and aggregate feeds using their own technology.

III. USER REVIEW

A user review is a review conducted by any person who has access to the internet and publishes their experience to a review site or social media platform following product testing or the evaluation of a service.^[1] User reviews are commonly provided by consumers who volunteer to write the review, rather than professionals who are paid to evaluate the product or service. User reviews might be compared to professional nonprofit reviews from a consumer organization, or to promotional reviews from an advertiser or company marketing a product. Growth of social media platforms has enabled the facilitation of interaction between consumers after a review has been placed on online communities such as blogs, internet forums or other popular platforms

IV. NOTIFICATION ALERT

In information technology, a notification system is a combination of software and hardware that provides a means of delivering a message to a set of recipients. It commonly shows activity related to an account.^[2] Such systems constitute an important aspect of modern Web applications. For example, a notification system can send an e-mail announcing when a computer network will be down for a scheduled maintenance. The complexity of the notification system may vary. Complicated notification systems are used by businesses to reach critical employees. Emergency notification systems may take advantage of modern information technologies. Governments use them to inform people of upcoming danger. In mobile phones and smartphones, dedicated hardware such as a notification LED is sometimes included to deliver messages or notify users.

III. RELATED WORKS

This paper ^[1] contributes a very distinctive approach for the automatic identification and extraction of product value info from capricious e-shop websites that freelance from the e-shops' language and additionally the merchandise domain. The approach uses tag path analysis and it exploits the common structure of product records within e-shop websites for characteristic product records and extracting their attributes. The approach was created freelance from a product domain or a language. The adequacy of the approach for the identification and extraction of product records from e-shop websites was shown in academic degree experiment where the attributes of the complete product of two altogether completely different e-shop websites square measure extracted.

This paper ^[2] proposes smartphone application named Virtual Cart for facilitating a simple and convenient technique for buying in wanting malls. The hybrid application Virtual Cart can even extend friendly client service for 24 hours daily with relevance locating, buying and shipping shopper needs. There square measure two blessings of it: first no got to exchange the queue for a drawn-out time in malls merely for scanning the item, second there will be no scope for the frauds that happen in mobile wanting. The transactions which will present itself oftentimes with the retailers cloud area unit progressing to be created secure.

The aim of the paper ^[3] is to developed the 'modified SentiWordNet algorithm' that in the main is on machine learning is employed to urge the alternatives of the merchandise. These choices will then be keep and can be created out there for the users whereas they are finding out the merchandise. This approach offers the rating at the tip for every feature of the merchandise so recommends the patrons to travel with the one that has the foremost effective rated product terribly} terribly specific feature. Whenever the priority of the alternatives is chosen, then the merchandise that have highest positive scores during this choices would be supported for the user so helps in shopping for the merchandise in line with their wants. The system helps the makers of the merchandise to understand their feedback of the merchandise so helps them improve the actual decisions of the merchandise and develop merchandise that ar in line with the requirements of the buyer.

In this paper^[4], they gathered filtered and unfiltered on-line reviews for many hotels within the Charleston space from yelp.com, extracted part-of-speech options from the information set, applied three classification models, and compared accuracy results to related works. Yelp.com is one in all the largest on-line review sites. It uses a filtering algorithmic rule to sight pretend reviews. However, the algorithm may be a secret. during this work, they collected reviews from yelp.com for one hundred random hotels within the Charleston space, and labeled filtered reviews as real and unfiltered reviews as pretend, extracted part-of-speech options, trained and tested the information set, engineered a model and compared results to connected work. From the collected information, extracted part-of-speech features, and applied 3 completely different classification models to identify pretend on-line reviews. They found that highest accuracy was achieved by applying the Multinomial Naïve mathematician classification model to our dataset.

In this paper ^[5], they introduce web{an online|a web} processing answer to e-commerce to search out hidden patterns and business strategies from their consumer and internet data, propose a brand new framework supported processing technology for building a Web-page recommender system, and demonstrate however processing technology is effectively applied in associate ecommerce atmosphere. This paper describe a framework that aims at answer to e-commerce to search out the hidden insight of their business and net data.

IV. IMPLEMENTATION

The proposed approach comprises all required steps and methods for identifying and extracting product records from arbitrary e-shop websites which are

- (1) crawling the e-shop website,
- (2) identifying and extracting the product records within the e-shop pages and
- (3) identifying and extracting the product attributes within the product records.

The approach was designed to be independent from a specific website language or a product domain and to be fast and easy configurable. Its functionality is described in the following subsections. For creating the approach we have analysed a set of 50 e-shops. These e-shops have different sizes from small e-shops over medium e-shops to very large eshops. The e-shops are a mixture of online shops from the United States of America (USA), Greece, Germany, Spain and the United Kingdom (UK) having various product domains (e.g. clothes, home electronics or books). The approach takes the Uniform Resource Identifier (URI) of a special product or of any other page of the e-shop website as input and returns attribute-value pairs for pre-defined product features of all products within the e-shop website as output.

A. CRAWLING

For crawling the URIs of the e-shop websites we have analysed our set of e-shops for the occurrence of product lists within the different link levels of the websites. We could determine that no website provides product lists on a deeper level than level three. In this paper the levels of the e-shop websites are defined as follows:

- Level 0: Homepage of the e-shop website.
- Level 1: Web pages of the e-shop which can be directly accessed by a link on the Homepage.
- Level 2: Web pages of the e-shop which can be accessed by following two links from Homepage.
- Level 3: Web pages of the e-shop which can be accessed by following three links from Homepage.

The analysis of the 50 e-shops revealed the following results: 76% of the analysed e-shop websites have product lists on level 0, 96% have product lists on level 1, 74% show lists of products on level 2 and 12% of the e-shops even have product lists on level 3. Level 0 mostly includes special offers of the e-shop which are on sale as well as the category links to the product pages available on level 1 where the product list of each category is presented. Level 2 pages include all product list pages from level 2, that means level 2 links in the most cases are page links on the first product page including the following product list pages. In some cases level 1 or 2 includes an overview of the product categories or an overview of subcategories of a special category. In such cases the product lists move for one level. Following the results of the analysis of the distribution of the product lists within the e-shop websites the approach crawls all website links for reaching all pages of the website until level 3.

B. PRODUCT RECORD IDENTIFICATION AND EXTRACTION

For the identification and extraction of product records within e-shop websites we use a special approach called LightExtraction. The algorithm takes the e-shop URIs crawled in the first step as described in Section IV-A as input and identifies and extracts the product records on each page as shown in Algorithm 1. LightExtraction renders the Web page of the actual URI and runs through all elements inside the HTML page tree. As shown in line 4 of Algorithm 1 all script tags are ignored. Next, LightExtraction uses a filter to check if the actual element is probably a product record. The filter expects that a product record

- (1) Contains at least five child nodes,
- (2) Includes some text (product name and short description)
- (3) Implies an image tag (product image) and
- (4) An anchor tag (link to detail page). Only elements matching the filter are considered for the clustering.

The first part of the tag path depicts the path from root element of the Web page to the actual element, including the actual element. The second part of the tag path contains the tag paths of all child elements which are simply connected together to a single string. The asterisk in square brackets is used to separate the element path to the actual element and the following part of the path which includes the tag names of all child elements as well as for being able to distinguish different elements which otherwise would build the same path, e.g. and elements of the same table. Light Extraction assumes the majority of the product records of the analysed Web page in the cluster which contains the maximum number of elements. Light Extraction expects that not all product records within a Web page include exactly the same child nodes, since there are products including some additional information like an old product price or user reviews. Following these assumptions, Light Extraction considers all elements having the same tag path until the mark in the form of the asterisk in square brackets as elements of the same data region. Thus, the algorithm returns all elements having the same tag path from the start to the marking as the identified cluster which includes the maximum number of elements as the product records of the Web page.

C.Product Attribute Identification and Extraction

For processing the collected data in further steps it is essential to assign the extracted attributes to pre-defined product features to be aware of the meaning of the extracted attributes. Thus, information about the format of the product attributes or common knowledge of the structure of product records and especially about the product attributes inside the records can be exploited for extracting the product attributes. The proposed approach identifies and extracts the product attributes by exploiting common knowledge about the format and content of the attributes which was collected by analysing the 50 e-shop websites.

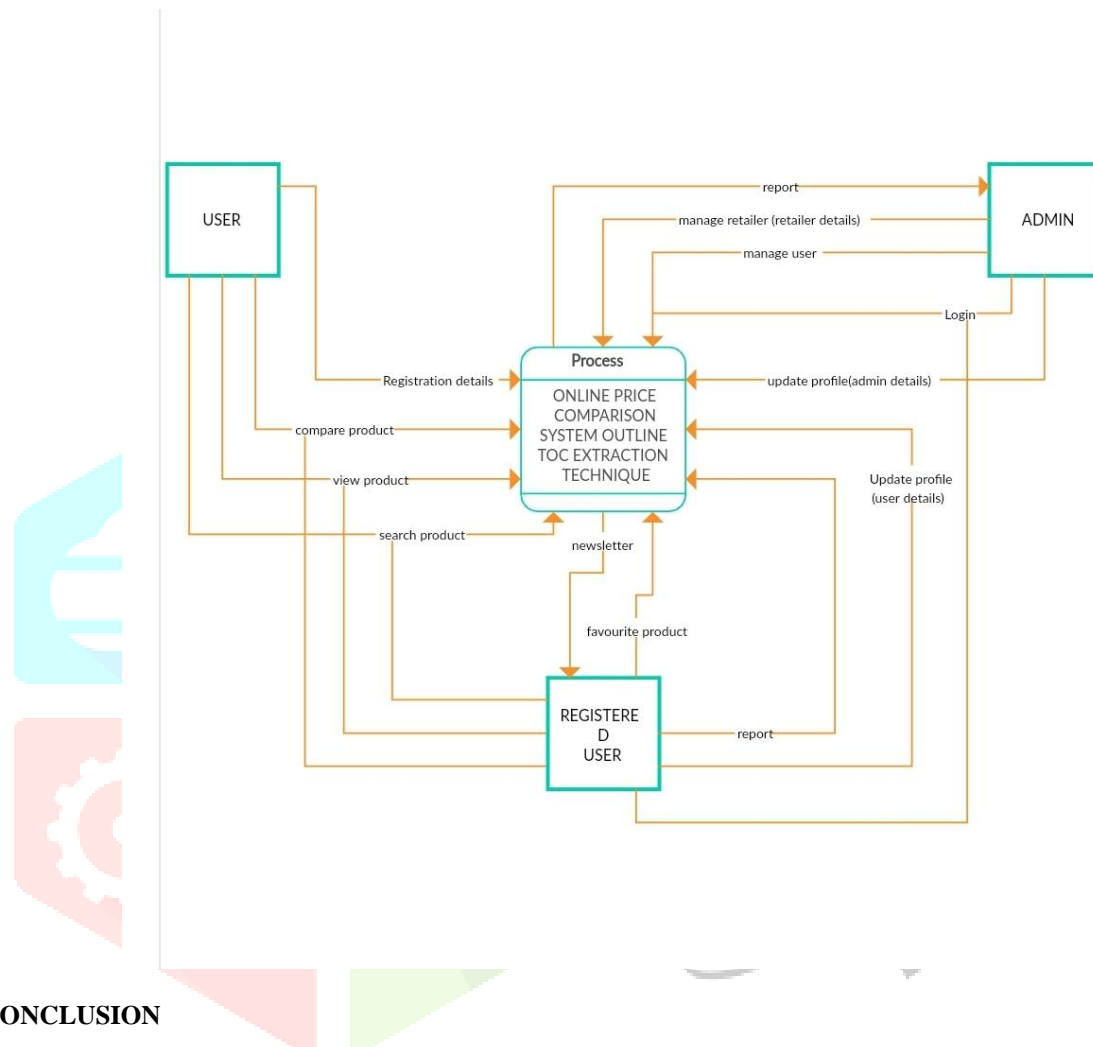
1) **Product Image:** In 100% of the 50 analysed e-shops the product image was found in the scr attribute of the Error! Filename not specified. Element within the product records which contains the image with the biggest file size. Thus, as shown in Algorithm 2 the approach analyses all Error! Filename not specified. elements within the product record, it loads the images from the URL given in the scr attributes of each Error! Filename not specified element and collects the image having the biggest file size.

2) **Product Prices and Currency:** As the e-shop analysis has shown a product record includes up to 5 different prices of the following price types: (1) product price including tax, (2) product price excluding tax, (3) regular price, (4) voucher amount, (5) unit price, (6) price range, (7) savings amount, (8) tax, (9) shipping costs, (10) minimum amount of purchase for free shipping, (11) deposit for product container and (12) start price for bids. Elements including prices can be detected by using regular expressions for possible price formats and currencies. According to the analysis results the actual product price is usually the price element having the biggest font size given by its font-size or font attribute. In the most cases the regular product price has the text-decoration attribute value line-through. The currency of a price element can be found within the text of its parent element (including the text of all child elements) as the currency could be found in 86% of the analysed e-shops within the price element itself, in 4% in the parent element of the price element, in other 4% in a child element and in 6% in a sibling. Hence, as shown in Algorithm 3 the approach runs through all elements within the product record and searches for elements including text matching a regular expression for prices and which parent element contains text including a currency. The approach considers such elements as price elements. Further the approach analyses the fontsize, font and text-decoration attributes of these elements to find the product price with the biggest font size which is expected to be the product price as well as a crossed out price which is considered as regular price.

3) **Product Name and Link to Product Detail Page:** Following the analysis of the 50 e-shops the approach identifies and extracts the product name and the link to the product's detail page as shown in Algorithm 4. In 96% of the analysed e-shops the element including the product name has been an anchor tag or has been embedded in an anchor tag. In 98% of the cases where the product name has been included in an anchor element the product name occurred as text, in 24% it has additionally been included as value of the anchor's title attribute whereof in 2 cases the text has been only a short version of the product name and the complete name could only be found in the title attribute. As in 86% of the analysed e-shops the product name has been the value of the alt or title attribute of the product image whereof in 82% cases this has been an additional occurrence of the product name to its occurrence embedded in an anchor element the algorithm runs through all anchors and compares their text and title attribute with the alt and title attributes of the product image which has been detected as described in Section IV-C1. Does the text or title value of an anchor match the alt or title value of the product image it is expected to be the product name. Does no text or title attribute of the anchor match the alt or title value of the product image the algorithm takes the title value or text of the anchor with the biggest font-size and font-weight, excepted from this check are anchors including a price element since the actual product price usually has the biggest font-size within the product record. This approach was chosen since in 86% of the cases that the product name was embedded in an anchor element the product name element had the biggest font-size expected the font-size of the actual product price. The font weight is additionally considered as in 4% of the analysed cases the product name element had the same font-size as other anchor elements and could only be determined by the font-weight. Since in 96% of the analysed cases the link to the product detail page could be found within the anchor including the product name element or it was embedded into

the product name element the link to the product detail page is most likely the URL inside the href attribute of the detected anchor element. In 4% of the considered cases the product name only occurred in the alt or title value of the product image and the link to the product detail page could only be found in the href attribute of an ancestor of the product image. Thus, if no product name and link to the detail page was found inside an anchor element the algorithm checks the alt and title attributes of the product image as well as the href attributes of its ancestors.

Architectural Design



V. CONCLUSION

As the demand for the mobile looking is increasing the necessity for a lot of secure, safe and reliable dealing is of utmost demand. Smartphones, that became Associate in Nursing important a part of today's life, have reduced all the efforts that square measure needed for looking. There are 2 blessings of it: 1st no got to substitute the queue for a protracted time in malls only for scanning the item, second there'll be no scope for the frauds that happen in mobile shopping. The transactions which will ensue oftentimes with the retailer's cloud is going to be created secure. the trend of shopping online has come back to remain. Online shops are open twenty-four hours of the day and might be accessed from anyplace wherever there's a web affiliation. the benefit and convenience of shopping online can invariably lure a lot of customers thereto. However, consumers must be alert and responsive to the risks concerned and take additional care once looking on-line. Due to the openness and fight of the net market, most business invariably attempt to maintain the very best customary of security yet as a user centred web site to spice up their business.

VI. REFERENCES

1. Andrea Horch, 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 1st International Conference on Advances in Information Technology (ICAIT)
4. Jinhua Sun and Yanqi Xie," A Web Data Mining Framework for E-commerce Recommender Systems" 2009 International Conference on Computational Intelligence and Software Engineering
5. 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)

