TIEVS - FRAUDULENT CAR ADVERTISEMENT DETECTION SYSTEM

2021-195

Project Proposal Report

Ravihari J.M.S – IT18089400

Supervisor – Ms. Manori Gamage

Co-Supervisor – Ms. Suriyaa Kumari

B.Sc. (Hons) Degree in Information Technology Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

March 2021

TIEVS - FRAUDULENT CAR ADVERTISEMENT DETECTION SYSTEM

2021-195

Project Proposal Report

Ravihari J.M.S – IT18089400

Supervisor – Ms. Manori Gamage

Co-Supervisor – Ms. Suriyaa Kumari

B.Sc. (Hons) Degree in Information Technology Specializing in Information Technology

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

March 2021

DECLARATION

We declare that this is our own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Ravihari J.M.S	IT18089400	Ravihari

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor Date

....

ABSTRACT

At present ,with the development of advanced technology, classified advertisements could reach a higher range of target audience through online classified advertisements. This proposal is especially keen on implementing a fraud detection system, specifically aimed at the items for sale on our 'Tievs' online classified ads platform, especially on Cars. Because when we need to purchase, most of us used to looking through classified advertisement websites without thinking twice moreover comparative affordability and the widespread adoption of the leasing culture, the population of used cars has been rising at an increasing rate. Though most people are genuine sellers, it is become very important to know how to spot classified scams, so that it reduced getting caught out. If any advertisements which fraudulent or consist of inappropriate content were to be posted, with having the intention of misleading the buyer or mistakenly, the system will identify and notify to prevent such events from occurring. With this regard, this research suggests a fraudulent website detection model based on the textual contents of an advertisement, supervised machine learning and Active learning techniques. The proposed model consists of some primary phases which the data acquisition phase, preprocessing phase, the feature extraction phase, clustering and the classification phase. Moreover, supervised machine learning techniques to construct the fraudulent content detection model and to resolve the highly imbalanced classification problem, use active learning methods to pick instances for judgement. The developed model will be integrated with the implemented rich user interface to facilitate visualization of the detection results to determine whether a car advertising is being spammed prior to submission of the advertisements that the advertisers do not want to be published. It will operate innovatively, to move forward in leaps and bounds to uplift online classified advertising to the next level

Key Words – fraudulent content detection ,supervised machine learning, Active learning, feature extraction

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
SVM	Support Vector Machine
AL	Active Learning
RF	Random Forest
PAM	Partitioning Around Medoids
UI	User Interface

1 INTRODUCTION

1.1 Background & Literature survey

At present, the public does not utilize tangible information sources such as newspapers, magazines, booklets, Leaflets, etc., especially when considering that a pandemic situation has occurred in the world, much more easily and securely even without needing to travel personally for long distances and exchanging money with other people, just to apply for or to receive the advertisement post(s), risking our health and safety. Since that online classified advertisement has become highly attracted market to buying and selling purpose. Additionally, comparative affordability and the widespread adoption of the rental culture, the population of used cars has been rising at an increasing rate. Currently, classified ads can reach more target groups using online classified advertising platforms, through advent of sophisticated technology.

Moreover , Presently the growth of the Internet, ease of communication, and recent technology covered the way criminals conduct fraudulent actions which consequence of the loss of dollars globally each year. Online classified advertisements have enhanced an essential element of the advertisement market. As below Figure shows the use of online classified ads sites doubled , according to the pew internet and American life project survey which conducted based on the internet users on 26th of March -29th of April in 2009, ensuing that economical online classified advertisement sites such as craigslist, eBay classifieds, and oodle have attracted plenty of posts and visits. Due to its high commercial potential, the online classified advertisement domain is a target for spammers, and this has become one of the biggest issues hindering further development of online advertisement [1].

According to market researcher classified intelligence, the U.S market for online classified advertisement was $14.1 billion in 2003, and it has increased quickly since then [1]. Craigslist, for instance, receives about 50 million new posts every month1 and is ranked the 7th most visited site in the U.S And the 35th most visited site in the world, according to alexa2[1].Since these sites attract a massive amount of advertisement on daily basis, online classified advertisement domain had become one of the biggest targets for spammers, and detection of fraudulent classified advertisement content has been an interesting topic among researchers.

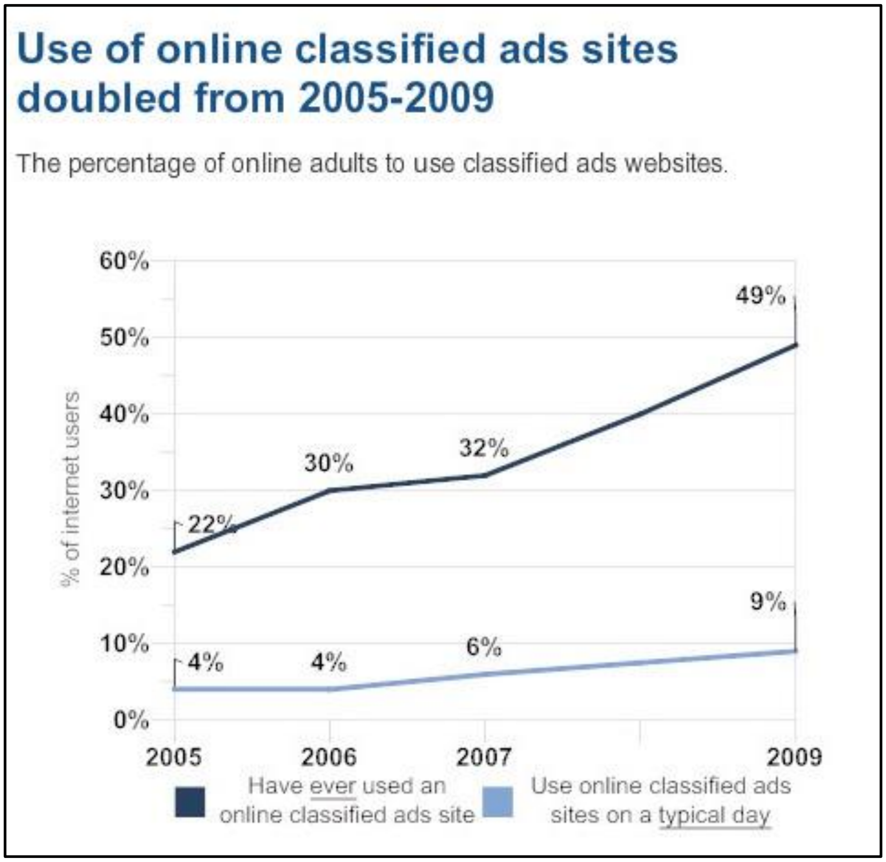


Figure 1-1 Use of online classified ads

To estimate the volume of fraud that happens in online classifieds advertisement sites we might consider the popularity of such site and the amount of revenue. Furthermore, Having Intention of cheating the buyers, normally spammers used to post fake advertisement and having an intention of violating the site, they post an advertisement with inappropriate/irrelevant content.

Moreover, spammers used to mislead search engines using keywords and offering a price that is too good to be true. At present, fraudulent advertisements have become a major threat to the advertisement platform. Hence, there are more than a few types of research have been conducted to moderate the same.

Over the years, researchers have established several techniques to detect online spam. supervised learning is the most common existence of these studies. Moreover, In the past literature, two categories of features have been utilized to distinguish spam and non-spam studies so far, either using link-based features or content-based features as n-grams. However, the Link-based feature does not support it appropriately since most of the time it is rarely linked web pages together in advertisements.

To overcome implied issues, though our implementation will be designed and developed in the future to accommodate to various types of classified section, as well as more enhancements relevant to each requirement, we sought to decrease our research processes, implementations, and executions to ‘Car' classifieds in order to accomplish the project within twelve months.Ntoulas et al. [7] worked on a dataset of more than 105 million web pages downloaded by the MSN Search crawler.In a sample of 17,168 pages drawn from English pages, 86.2% are non-spam and 13.6% are spam. They extracted more than 20 features from pages' content that can provide information to distinguish a spam page from the normal ones [1].

A study conducted by Urvoy et al. [8] introduced a spam website detection model based on style association measures of textual features in HTML source code. Jaccard correlation index has been applied to measure the relationships. They also introduced a method to cluster a large group of reports according to this measure. Their suggested technique is particularly valuable to recognize pages crossed different sites which share an equal design. Furthermore, several other studies including [9], [10], [11], and [12] experimented to identify spam websites using supervised machine learning procedures.

1.2 Research Gap

Content-based fraudulent detection methods in online classified advertisements depend on the advertisement contents, components, and metadata such as domain registration details, body, URL and image features, links, etc. to detect fraudulent advertisement in online classified ads. Despite the fact that some researchers have sought to apply an excellent model designed to detect fraudulent advertisement content in used or reconditioned cars with reasonable accuracy levels, far more unrevealed novel techniques remain to be acknowledged. Various studies utilized these mechanisms to recognize the legality or fraudulency of classified advertising.

Many types of research have considered the following metabolic variables.

- Current market price
- Presence of URL or emails
- Time of the ad posting

According to the available research papers [1] ,[2] and resources there are several researches regarding spam detection and fraudulent website content detection. The most common way of filtering spam is predicting whether spam or not for a section of text by doing a classification task. Moreover, former researchers were dependent on the used data sets for the training and testing process. For example, the Own dataset was built by Delany et al [2] which consists of spam and non-spam messages. And they have been able to analyses the messages by using content-based clusters, whether they are spam or not and 9 to 10 clusters have been explicitly categorized.

Many researches are mainly concerned below attributes (table 1.1) Fetterly et al. were among the first groups to take content into account for detecting spam pages using statistical analysis of two datasets: DS1 represents 150 million URLs and DS2 includes 409 million HTML pages.[1]

Research A has built a spam detection classifier using attributes such as presence of URL or emails and the time of the ad posting in order to advance the traditional NLP based classifier. They have used the exploit external resources such as Kelley Blue Book (KBB)3 to get an estimated price for that car and compare it with the asking price.[1]
There is a significant restriction of research A, since they have used domain-specific features such as the market price of the vehicle which is being posted by using an outside source which is challenging to generalize,

Research B which is done by McCormick et al focusing on detecting classified fraud post is very similar to legitimate posts and the activities usually done by spammers over the phone or via email and also these kinds of fraudulent post acquire much detailed evidence or the information to perform the research development well. Moreover ,the research B is most focusing on the private data from the web server running the classified advertisement such as the user account details (age ,IP address of the publisher).
Gyongyi et al [3]. present a comprehensive taxonomy of web spam techniques including ones targeting content-based ranking algorithms. They also provide some suggestion for web spam countermeasures [1] .

The table shows a tabularized format of the explanation.

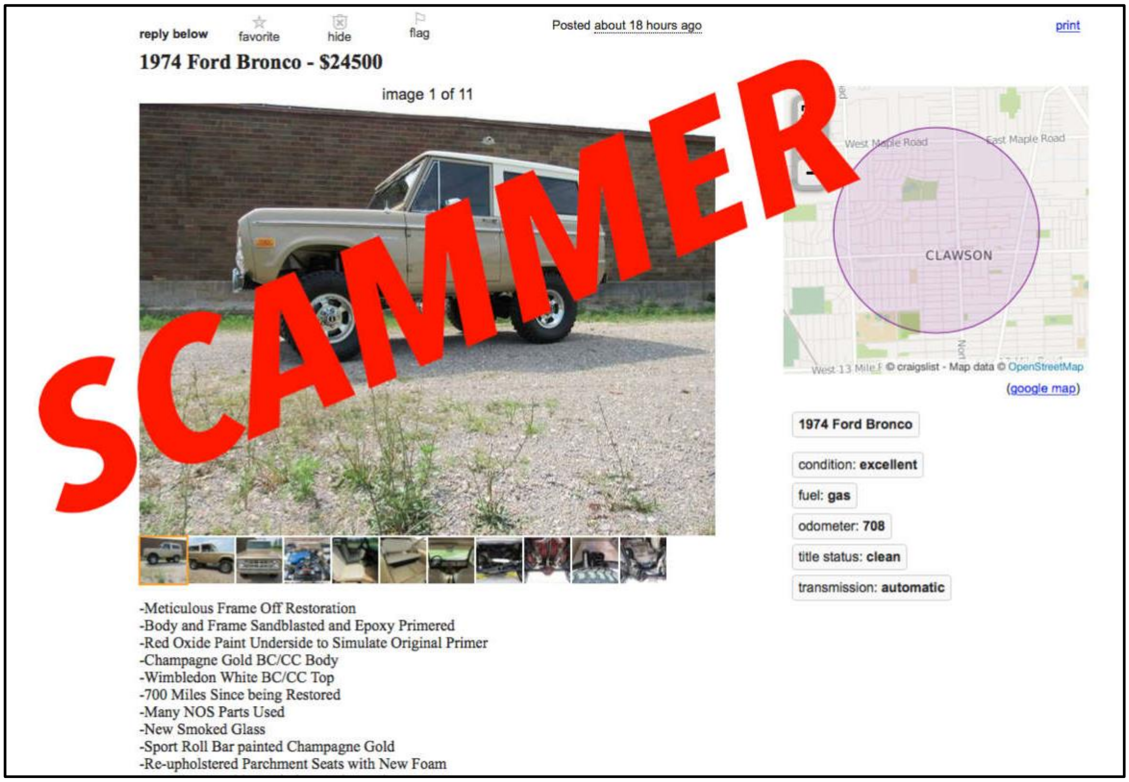
Product	Current market Price	Outgoing URL	Phone number	Time	Active learning
Research A [1]	✓	✓	✕	✓	✕
Research B [2]	✕	✓	✕	✓	✕
Research C[3]	✕	✓	✕	✕	✕
Proposed system	✓	✓	✓	✓	✓

Table1-1 Comparison of former researches

1.3 Research Problem

Even though it is distinctly apparent that online classified advertising platforms have already been implemented by numerous parties in recent years, there is yet to be a highly ameliorated application that incorporates complex machine learning technology to enhance the prominent functionalities to facilitate customer expectations and requirements efficiently.

Moreover, these online advertisements have become the most essential part of the advertising market since the development of the internet. And also, those online classified advertisement sites have a huge number of advertisements daily (Eg: eBay, Craigslist).



Posted about 18 hours ago

1974 Ford Bronco - $24500

image 1 of 11

1974 Ford Bronco

condition: **excellent**

fuel: **gas**

odometer: **708**

title status: **clean**

transmission: **automatic**

-Meticulous Frame Off Restoration
-Body and Frame Sandblasted and Epoxy Primered
-Red Oxide Paint Underside to Simulate Original Primer
-Champagne Gold BC/CC Body
-Wimbledon White BC/CC Top
-700 Miles Since being Restored
-Many NOS Parts Used
-New Smoked Glass
-Sport Roll Bar painted Champagne Gold
-Re-upholstered Parchment Seats with New Foam

Figure 1-2 Scammer Advertisement on Craigslist

Though the popularity of the online advertisement sector, it was highly targeted by spammers. Hence the proactively recognizing, detecting fraudulent advertisement content information, preventing it from being posted in the application, is a crucial problem at the present time. The widespread accessibility of the web has an unwanted effect of attracting online scammers who pose as genuine sellers by posting fake advertisements in an effort to defraud would-be buyers. Scammers can steal millions of dollars from unsuspecting users and threaten the reputation and utility of online ad services [4].

From the applications' developers and owner's point of view, they must implement strict program logic to identify or prevent fraudulent advertisements from being posted (fake spam/ irrelevant descriptions) accidentally or deliberately by users of the system, through verification and validation processes internally. This will require inspecting posts separately and meticulously, even after the ad being posted to the customers and if so, customers will view those fraud ads and lose confidence in using the platform for their needs, enabling loss of application reputation as well [2].

Most classified advertising products currently in use do not prioritize detecting and preventing fake advertisement or characteristic details from being submitted, let alone inspect them internally and inform the respected users who are in the process of submitting the advertisement. Nevertheless, many models of machine learning constructed either by regression or classification processes were inventively analyzed by the research, and the accuracy obtained in some cases were successful, while many others did not demonstrate commendable recognition.

Moreover, in fear of detection model complexity, the analysis of the characteristics which correspond to the description should be expanded from up to four characteristics while preserving the functionality of the model not carefully examined by most researchers. Even more ,many implemented online classified advertising systems, simply do not exhibit rich user interfaces for smooth functionality or promote quality user experience and user-friendliness, in consonance with the latest trending technology.

Recent machine learning approaches require making a user label instances of both spam and ham (not spam) descriptions so that an algorithm can learn to classify legitimate advertisement. It is exceedingly difficult for systems that process a huge number of advertisements to classify all descriptions handled over a given period of time. To minimize the impact, the messages to be labeled are often chosen at random (passive learning) or by a predictive algorithm (active learning) and also this is a highly imbalanced classification problem as shown below Figure , if we randomly pick a sample of instances for judgment, there are very few spam instances in the sample. To overcome implied issue, I plan to use active learning technique to pick instances that are likely to be spam .

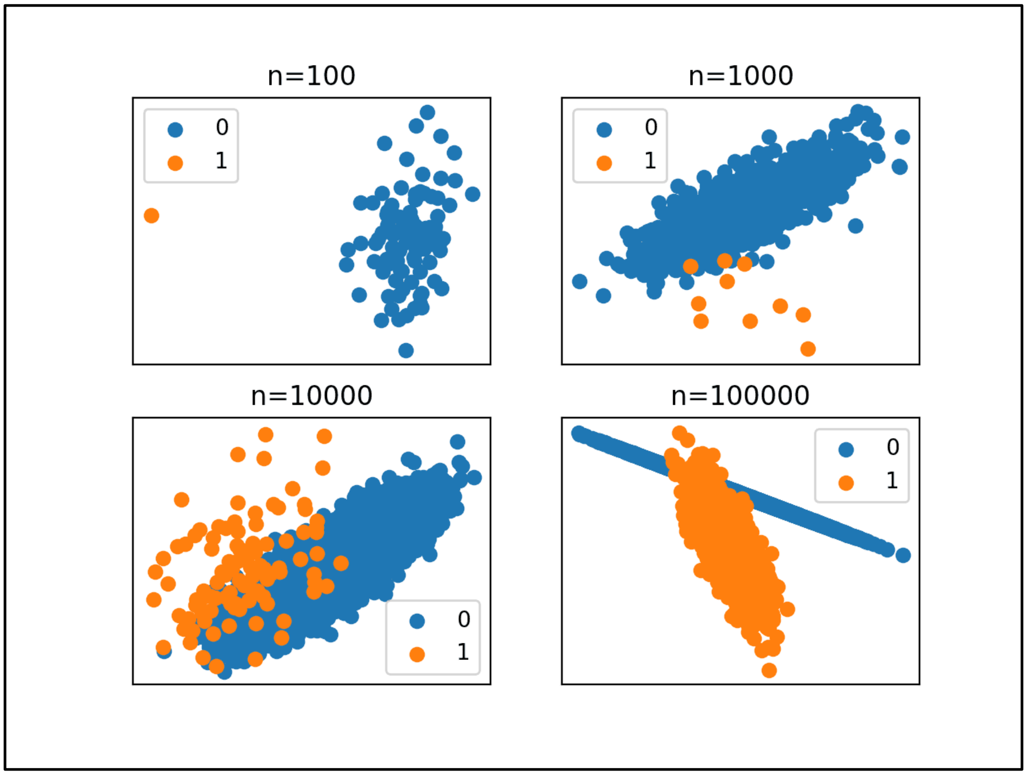


Figure 1-3 Imbalanced Classification

2 OBJECTIVES

2.1 Main Objectives

A SMART, systematic, and rich user-friendly solution to enable interested parties to post their advertisements efficiently and accurately without any inconveniences, in addition to providing an optimized, interactive search facility within the system, to search and find items to buy, according to specific customer requirements, seamlessly and productively moreover, design and develop an internal process, using machine learning appliances, to proactively identify fraudulent advertisement information context having inappropriate/irrelevant content being uploaded into the application. (Will be targeting one type of classified such as vehicles)

The objective of this research is to use a machine learning algorithm to introduce a fraudulent advertising detecting system to keep the advertising site from being spammed by spammers or people with misleading intentions who are in the prior to submitting an advertisement for an item of a particular type (such as Cars) during the year. As a result, the main objective is to use machine learning appliances to recognize and prevent fraudulent advertising content information from being published in the app.

When a customer is about to submit their advertisement(s) into the system, proactive inspections will be done through the system to find out whether the current advertisement is being inspected consisting of spam or inappropriate content information. If so, the customer will be prompted with a warning message, indicating such matters present in the advertisement, which is yet to be submitted, so that the customer could proceed with the necessary measures to be taken to avoid such schemes. Nevertheless, because of it's scope and time limitations imposed on the system, these functional requirements are implemented and aim one kind of classified cars. The emphasis will therefore be on identifying and presenting, according to requirements described by consumers on the advertised form they fill before submissions, fraudulent advertisements of the used or second-hand vehicle.

2.2 Specific Objectives

To reach the concept of proactively identifying and preventing fraudulent advertisement content information, from being posted in the application, using machine learning appliances the specific objectives that needs to be accomplished is as listed below,

- Data collection

  Congregate the appropriate data set using external sources (e.g.car_ad.csv/ vehicles.csv) or survey using craigslist region to train the model and test the model built.

- Explore the appropriate algorithm

  Explore the most suitable machine learning algorithm by considering the Accurateness and / Interpretability of the output Speed or Training time, Linearity etc.,

- Develop and train the model

  Develop and Train the model to detect, whether the advertisement being inspected by inappropriate /irrelevant content.

- Integration of the developed model with the application without inconveniences. select an initial set of email messages to be labeled as training examples using clustering and build , train an optimal model using active learning

- Build the client and server components

  Build the appropriate client and server-side components and extract the data from advertisement form, prior to posting. If the advertisement is consisting of the inappropriate / irrelevant content, the client will be prompted with a warning message, indicating such issues present in the advertisement, which is yet to be submitted.

3 METHODOLOGY

This sector will emphasis about how the development will be handled and implemented. Also, consists of a brief discussion about the feasibility study. As this is a research project implementation Agile development method will be used to implement the system module by module [5]. The below Figure 4-1 indicates the high diagram of the proposed components system for fraudulent vehicle advertising detection. Figure 3.1 shows. The Python machine learning API can first obtain the dataset from its MongoDB database through a content management system called 'Strapi.' Once obtaining a dataset, the data set will be prepared before shifting in the models by different attribute analysis methods, features extraction techniques, and other techniques used for data preprocessing.

3.1 Feasibility Study

3.1.1 Technical feasibility

While reviewing earlier research papers we Observed over the existing technology stack and discussed the possibilities. It confirmed the technical feasibility of existing technology stack along with the proposed framework. Though we need to enhance our knowledge in machine learning appliances since it is highly required throughout the development process.

3.1.2 Scheduling Feasibility

Checked the Scheduling Feasibility by creating tasks according to the deadlines since the development process should be end before the deadline and final product needs to be presented on the scheduled date.

3.2 System Overview

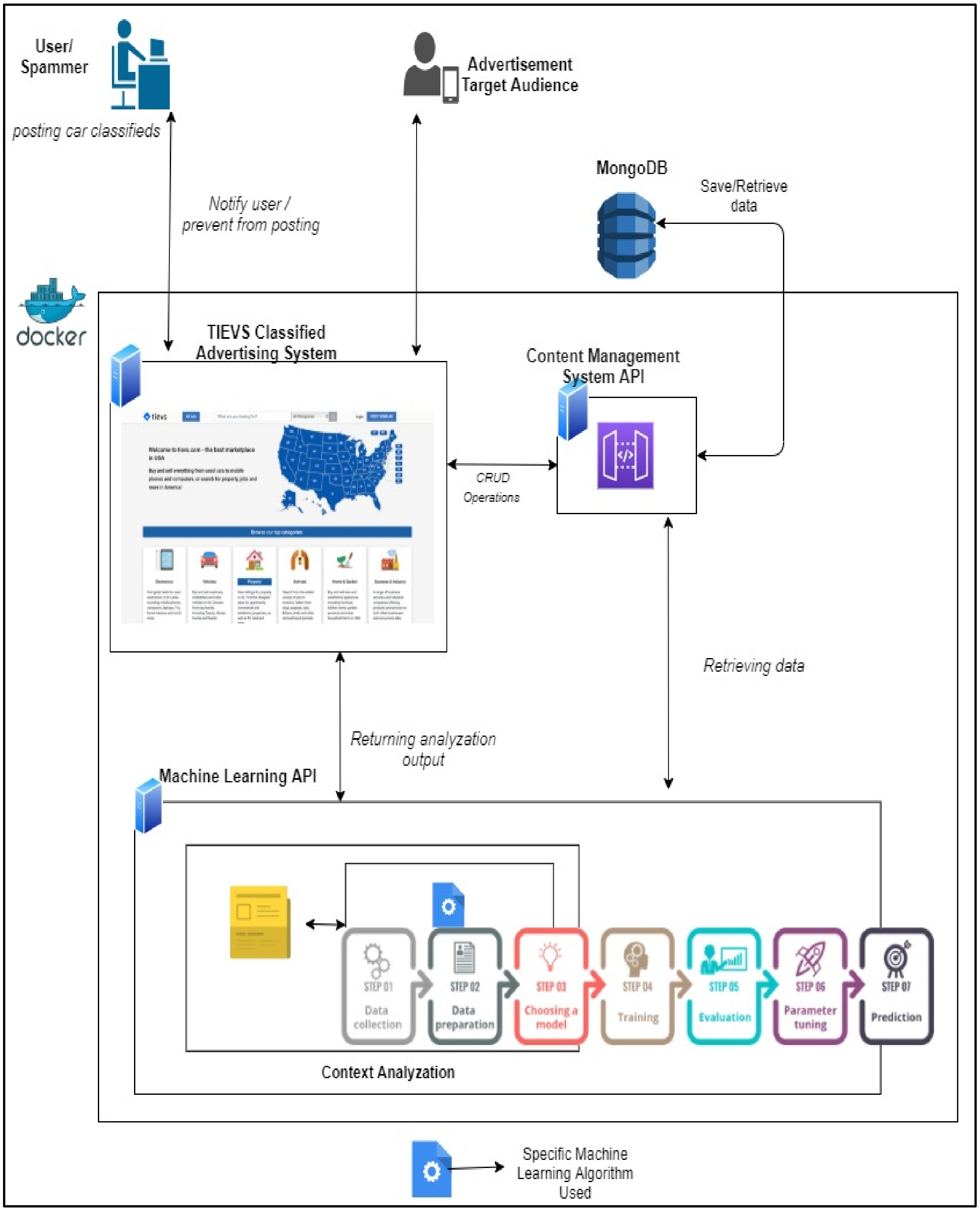


Figure 3-1System Overview Diagram for the solution proposed

3.3 Implementation Stage

The implementation phase complies with the development of the functionalities below,

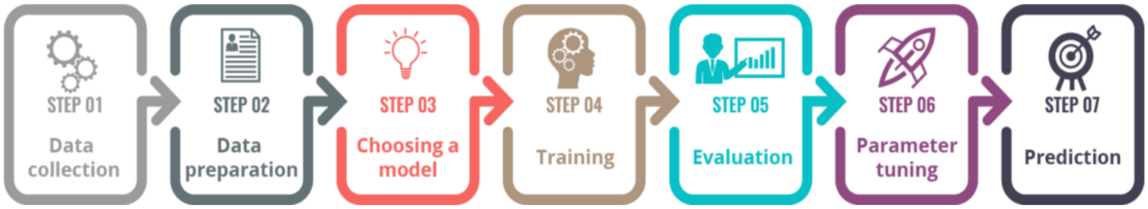


Figure 3-2 Implementation phase

- Data Acquisition Phase

  Data set collection to validate the discussed approach to detecting fraud in online classified advertisements, the relevant data set using external sources (car_ad.csv/vehicles.csv) or survey using craigslist region to train the model and test the model built. Whereas manual data gathering on cars would be exhausting and time-consuming, this proposed system would make use of an online dataset that is already available. Austin Reese first released and revised the dataset, which was scraped from one of the most popular and widely listed online classified advertising platforms, Craigslist.com, which is recognized and is used by many users in the United States.

- Data preparation phase

  This phase contains the method of converting raw data into features. It contains with data which is represent a well charactered to learning the algorithm. Hence, data can be decomposed into several chunks to capture the specific data associations. The final step is to split your data into two sets: one for training your algorithm, and another for evaluation purposes. Be sure to select non-overlapping subsets of your data for the training and evaluation sets in order to ensure proper testing [6] then can determine the consequence of the predictions back to the input data to enhance and optimize the models over the period.

- Develop and Train the model

Learning from past experiences is an attribute of humans while computers do not have this ability. In supervised or Inductive machine learning, our main goal is to learn a target function that can be used to predict the values of a data set. The process of applying supervised ML to a real-world problem is described in the below figure.

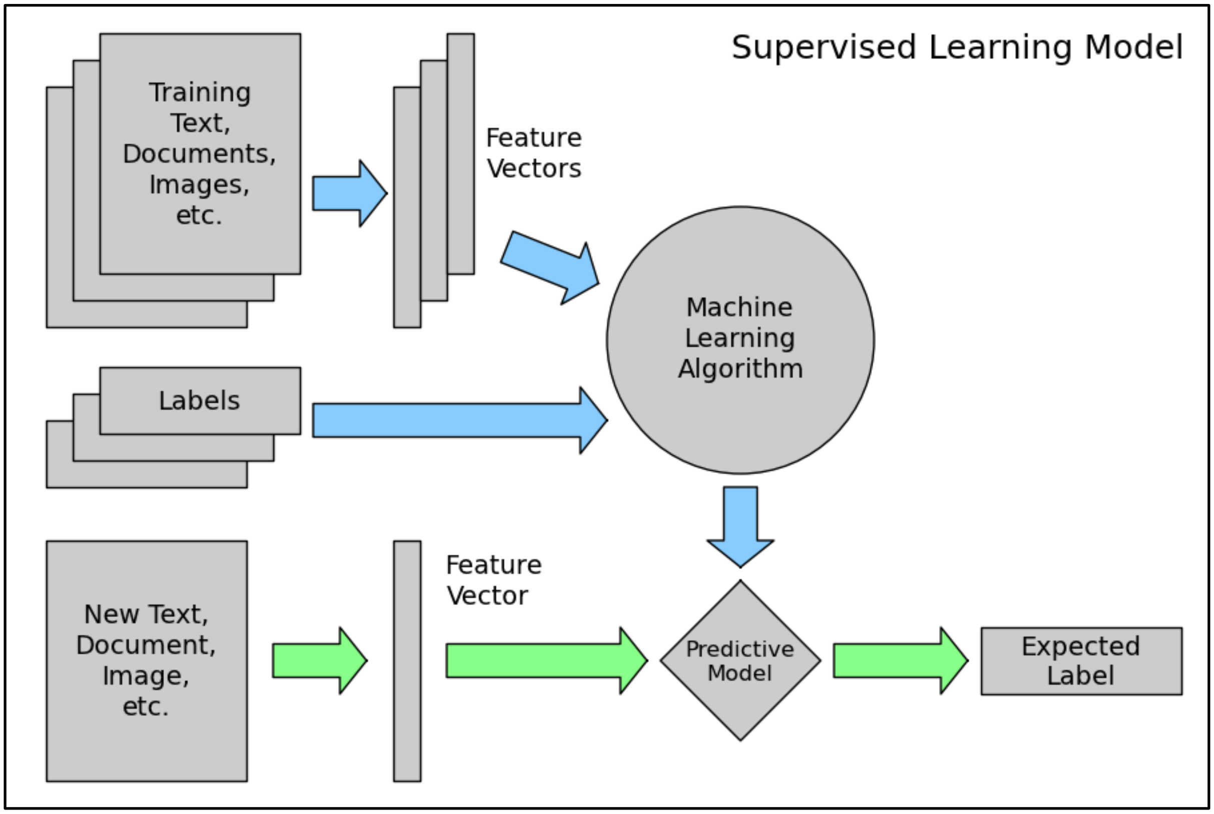

Supervised Learning Model
Training Text, Documents, Images, etc.
Feature Vectors
Machine Learning Algorithm
Labels
New Text, Document, Image, etc.
Feature Vector
Predictive Model
Expected Label

Figure 3-3 The process of applying supervised ML

- Evaluation and detection using ML and AL.

  Discover the most appropriate machine learning algorithm by considering the Interpretability of the output Speed or Training time, Accuracy, Linearity, etc., to identify fraudulent advertisement information context having irrelevant content and after the prediction label as spam or non-spam. In other application areas, clustering has been combined with active learning [13]. Clustering is used in this work to pick an initial set of email messages to be labeled as training examples by using or PAM.Partitioning Around Medoids (PAM) [14] is an algorithm for partitioning around medoids has been used to cluster a standardized random sampling of 25% of the training pool's contents. There are many contexts in which active learners might propose queries, and also are numerous query processing strategies that have been used to determine which instances are most important. Using Active learning an initial machine learning model is built from a labeled sampling of instances, and the algorithm then requests for labels for unlabeled instances where the classification algorithm is most ambiguous. Figure 3-4 shows a learner may start with a small number of instances from the labeled training set L, request labels for one or more carefully chosen instances, learn from the query results, and then use its new knowledge to decide which instances to query next [17] .

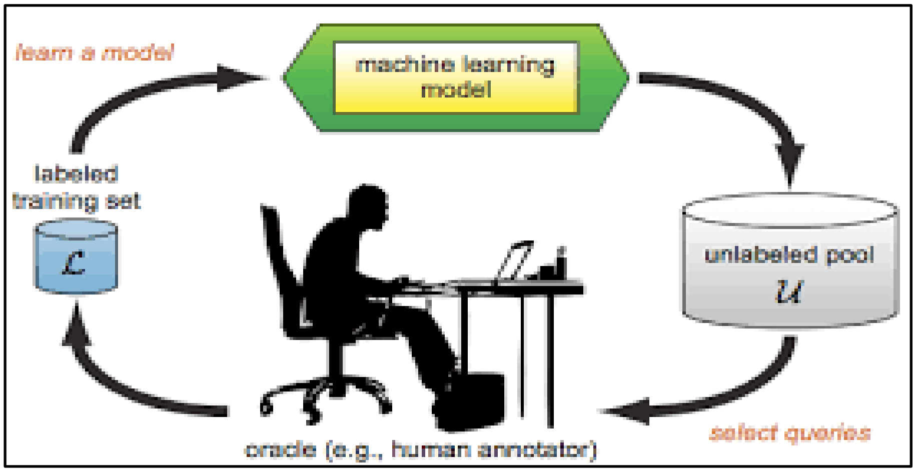


Figure 3-4 Active Learning Cycle

- Build the appropriate client and server-side components

  Extract the data from advertisement form, prior to posting. When a customer is about to submit their advertisement(s) into the system , if the advertisement is being spam the customer will be prompted with a warning message, indicating such issues present in the advertisement, which is yet to be submitted. These functions will be developed with the web and mobile User Interfaces

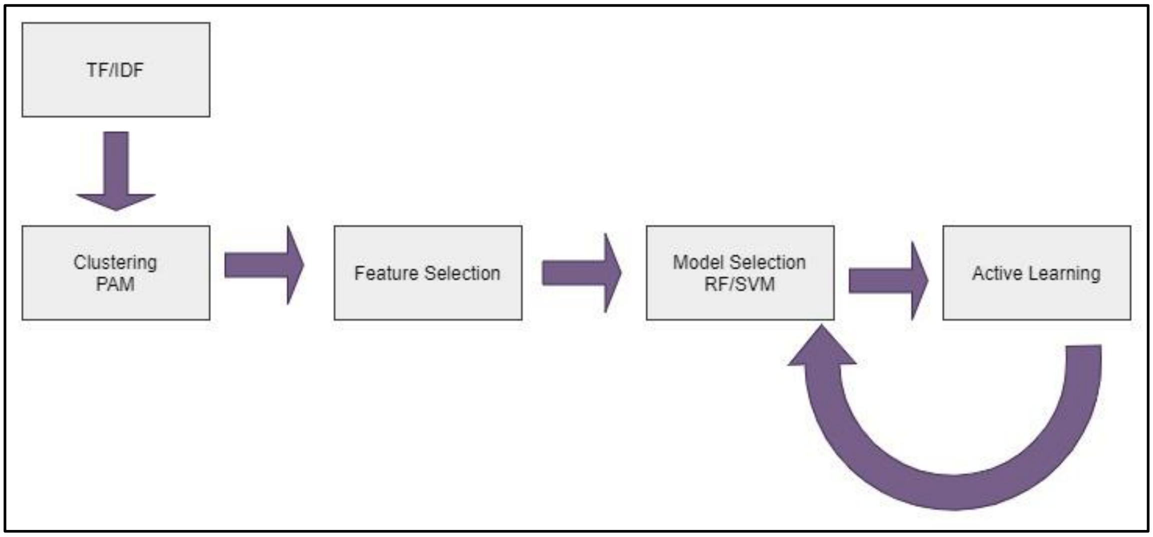


Figure 3-5 Process Diagram

3.4 Software Development Life Cycle

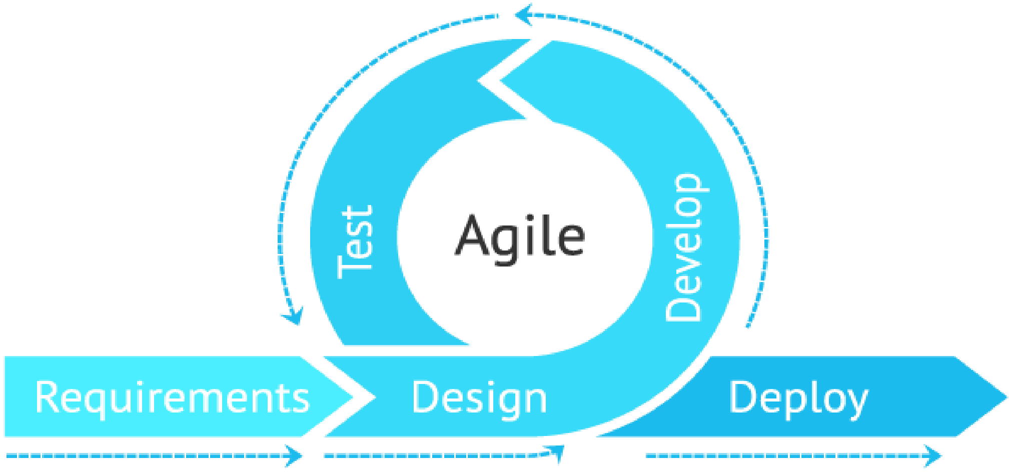


Figure 3-6 Agile Methodology

Agility is the ability to respond to unpredictable changes with quick response and profitability [15].Many companies benefit from enhanced job organization, cooperation with industry, and predictability of performance metrics. This effort will be led during the year by the notorious agile development and constant improvement aspirated, change inviting, Agile methodology strategy, as shown in Figure 3-5.Moreover, Agile principles that help increase the growth of almost any implementation since these processes are versatile and the rules do not have to be thoroughly understood. Although a set plan is constructed to aid for the time management and delivery schedule alignments with the help of a Gantt chart, according to the Agile values, if impediments were to occur, the necessary adaptations must be taken into place in a timely manner, without interrupting the continuous development processes [16]. At these kinds of moments, reacting to change efficiently and effectively, as well as ensuring the consistency of the project timeline, are critical. As Agile approach is used throughout the software development cycle, that would be much simpler to consider the challenges by turning them into opportunities for success. Ultimately, Scrum, a lightweight edition of the massive Agile framework, can be used to accommodate high-quality incremental goals and objectives in short durations, as well as team performance and application development.

3.5 Project Requirements

3.5.1 Functional requirements

- Successful customer navigation to application
- The proper visual interface of advertising process elements should be implemented for a convenient representation of the different features of a car.
- The 'Submit' button should preserve its precise functionality.
- Extraction of accurate data in order to establish the fraud detection model.
- Integration between the Machine learning API and the fraud detection system.
- Proactively identify and prompt a warning message prior to submit the advertisement

3.5.2 Non-functional requirements

- Accuracy
- Security
- Ease of use
- Adaptability
- Scalability
- Efficiency
- Ease of use
- Timesaving
- Accessibility
- Maintainability

4 GANTT CHART

The Gantt chart of the development process is as depicted in Figure 6.1.

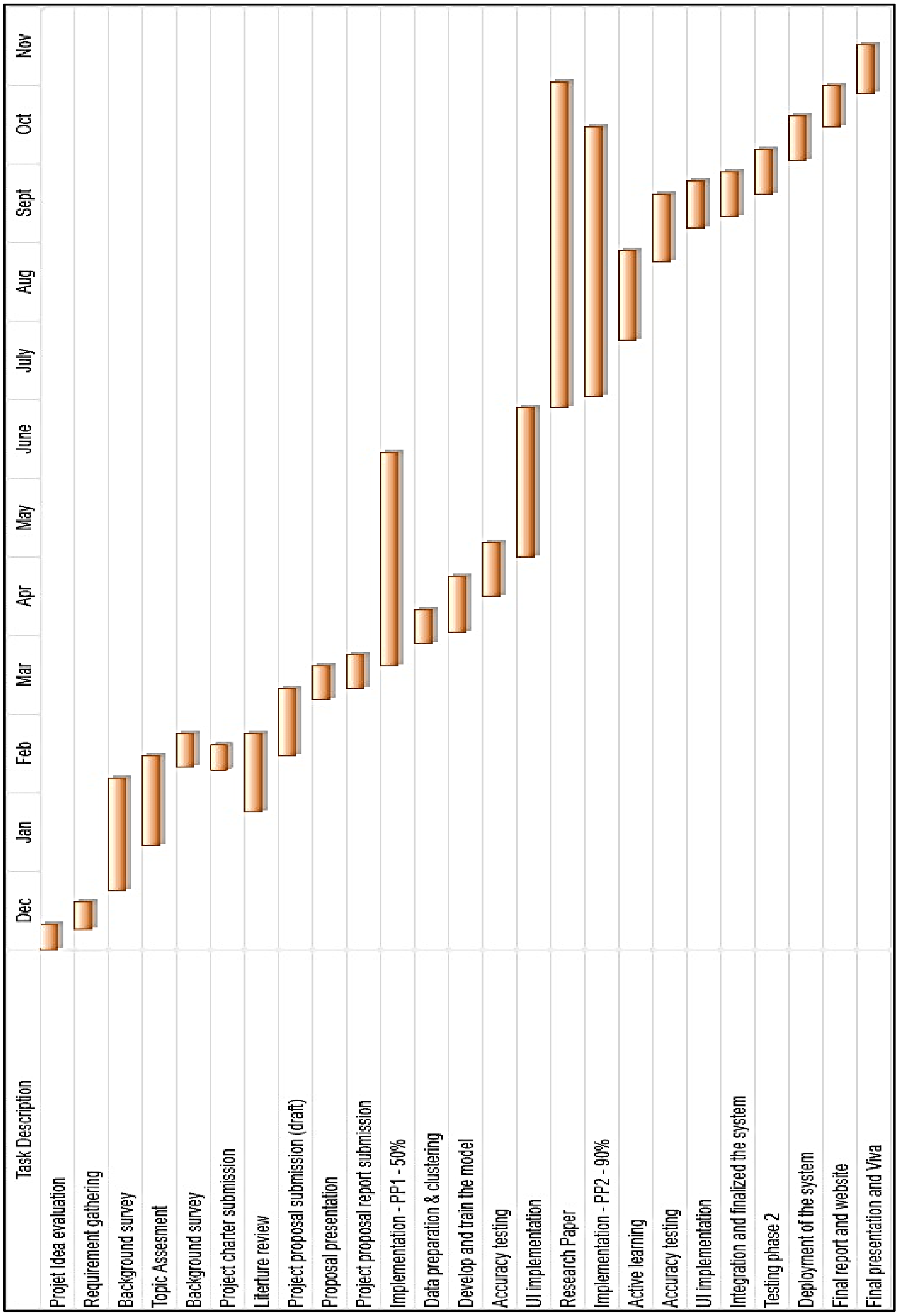


Figure 4-1 Gantt Chart

5 DESCRIPTION OF PERSONAL AND FACILITIES

5.1 Fraudulent content detection system

- .Utilization of the ‘Used cars dataset’ in Kaggle (vehicles.csv).

- Train and develop the model RF or SVM, and for clustering PAM, with variable refinements, resolving the imbalanced classification issue utilizing active learning for labeling, using the training set portion of the ‘Used Cars Dataset' obtained from Kaggle, and achieving the most competent model

- Selected data set will be classified into training and testing data sets.

- Acquired the data from the post to evaluate the machine learning algorithm in order to detect fraudulent advertisement.

- Selects the most appropriate ML algorithm in order to archive the high accuracy and efficiency .

- Using Active learning primary machine learning model is constructed from a labeled sampling of instances.

- Create the client-side and server-side components required for data extraction of an item from an advertising form prior to posting.

- Prevention of submitting the fraudulent advertisement by notify the user prompting an warning message while proceed with external requirements if possible.

6 BUDGET AND BUDGET JUSTIFICATION

Component	Amount (USD)	Amount (LKR)
Internet Connection(for 10 months)	7	14000
Domain Name Registration	12	2400
Total	19	16000

Table 6-1 Budget

REFERENCE

[1] H. Tran, T. Hornbeck, V. Ha-Thuc, J. Cremer, and P. Srinivasan, “Spam detection in online classified advertisements,” *ACM Int. Conf. Proceeding Ser.*, vol. 11, Apr. 2011, doi: 10.1145/1964114.1964122.

[2] M. Maktabdar Oghaz, A. Zainal, M. Maarof, and M. Kassim, *Content based Fraudulent Website Detection Using Supervised Machine Learning Techniques*. 2017.

[3] Z. Gyongyi and H. Garcia-Molina, “Web Spam Taxonomy,” presented at the First International Workshop on Adversarial Information Retrieval on the Web (AIRWeb 2005), Chiba, Japan, Apr. 2005, Accessed: Feb. 26, 2021. [Online]. Available: http://ilpubs.stanford.edu:8090/771/.

[4] A. McCormick and W. Eberle, “Discovering Fraud in Online Classified Ads,” p. 6.

[5] “Agile Methodology: What is Agile Software Development Model?” https://www.guru99.com/agile-scrum-extreme-testing.html (accessed Feb. 26, 2021).

[6] “Six Steps to Master Machine Learning with Data Preparation,” *KDnuggets*. https://www.kdnuggets.com/six-steps-to-master-machine-learning-with-data-preparation.html/ (accessed Feb. 26, 2021).

[7] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. “Detecting spam web pages through content analysis. In WWW '06: Proceedings of the 15th international conference on World Wide Web, pages 83{92, New York, NY, USA, 2006. ACM.

[8] Ntoulas, A., Hall., B., Najork, M., Manasse, M. and Fetterly, D.: Detecting Spam Web Pages through Content Analysis. In Proceedings of *15th Int. Conf. World Wide Web*, pp. 83–92 (2006).

[9] Shen, G., Gao, B. Liu, T. Y., Feng, G., Song, S. and Li, H.: Detecting link spam using temporal information. In Proceedings IEEE Int. Conf. Data Mining, ICDM, no. 49, pp. 1049–1053 (2006).

[10] Becchetti, L., Donato, D., Baeza-yates, R. and Leonardi, S.: Link Analysis for Web Spam Detection. ACM Transactions on the Web.vol. 2 no. 1, pp. 1-42 (2007).

[11] Drost, I. and Scheffer, T.: Thwarting the nigritude ultramarine: Learning to identify link spam. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3720 LNAI, pp. 96–107 (2005).

[12] Abbasi, A.: Detecting Fake Medical Web Sites Using Recursive Trust Labeling. In Proceedings 14th Int. Worldw. Web Conf., vol. 30, no. 4, pp. 1-22 (2012).

[13] Nguyen, H.T. and Smeulders, A. "Active Learning Using Pre-clustering", Proceedings of the 21st International Conference on Machine Learning, 2004, pp. 623-630,http://www.aicml.cs.ualberta.ca/_banff04/icml/pages/papers/94.pdf.

[14] Kaufman, L. and Rousseeuw, P.J., "Partitioning Around Medoids", Finding Groups in Data, Wiley-Interscience, 2005, pp. 68-125.

[15] Erande, Ameya S., and Alok K. Verma. "Measuring agility of organizations-a comprehensive agility measurement tool (CAMT)." International Journal of Applied Management and Technology 6.3 (2008).

[16] ADITI. (2005). “Agile Methodology based Services,” [Online] Available at: http://www.aditicorp.com/services/agile-methodology-based-services/ [Accessed Feb. 22, 2021].

[17] B. Settles, “Active Learning Literature Survey,” p. 47.

7 Appendix

7.1 Plagiarism Report