*International Journal of Engineering Applied Sciences and Technology, 2022*
**Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 160-164**
Published Online November 2022 in IJEAST (http://www.ijeast.com)

# ADVERTISEMENT CLONE DETECTIONUSING SENTIMENT ANALYSIS

N.M.K. Ramalingam Sakthivelan
Associate Professor,
Department of Computer Science and Engineering
Paavai Engineering College, Pachal, Namakkal.

Silambarasan V, Thavasi S, Vijaya Shankar P
UG Student,
Department of Computer Science and Engineering
Paavai Engineering College, Pachal, Namakkal.

*Abstract***: As most of human activities are being moved to cyberspace, phishers and other cybercriminals are making the cyberspace unsafe by causing serious risks to users and businesses as well as threatening global security and economy. Nowadays, phishers are constantly evolving new methods for luring user to reveal their sensitive information. To avoid falling victim to cybercriminals, a phishing detection algorithm is very necessary to be developed. Machine learning or data mining algorithms are used for phishing detection such as classification that categorized cyber users in to either malicious or safe users or regression that predicts the chance of being attacked by some cybercriminals in each period. Many techniques have been proposed in the past for phishing detection but due to dynamic nature of some of the many phishing strategies employed by the cybercriminals, the quest for better solution is still on. In this project, we can implement the framework for classify the phishing advertisements using multiple machine learning algorithms such as decision tree, random forest, and Gradient boosting algorithm. Experimental results shows that the Gradient boosting algorithm provides improved accuracy rate than the other two algorithm and applied on benchmark datasets that are collected from KAGGLE website.**

*Index Terms* – **Sentiment Analysis, Decision Tree, Randon Forest, Recommendation System.**

## I. INTRODUCTION

The word phishing was first coined in 1996 as a form of online identity theft after an attack by hackers on America Online account and the first phishing lawsuit was filed in 2004 against a California teenager who created an imitation of the advertisement "America Online" to gain access to user sensitive information including credit card details causing them huge financial lost. Phishing is a cyber-crime which involves the fraudulent act of illegally capturing private information like credit card details, usernames, password, account information by pretending to be authentic and esteemed in instant messaging, email, and various other communication channels. The traditional approaches used by majority of the email filters for identifying these emails are static which make it weak to deal with latest developing patterns of phishing since the defrauders are dynamic in actions and keep on modifying their activities to dodge any kind of detection. Phishers are sending fake emails to their victims pretending to be from legitimate and well-known organizations such as banks, university, communication network etc., where they will require updating some personal information including their passwords and usernames to avoid losing access right to some of the services provided by that organization. Phishers use this avenue to obtain users sensitive information which they in turn use it to access their important accounts resulting in identity theft and financial loss

Primarily, phishing is a felonious activity carried out by an intruder or a hacker through a well-versed technical deceptions and social engineering. The sole purpose is to steal the sensitive PII (Personal Identity Information) as well as the pecuniary credentials of the users or customers. The social engineering strategy utilizes baiting as well as spoofed emails to trick the beneficiaries to disclose their financial data. The received mail claims to be from a legitimate user or business houses which in turn misleads the beneficiary to be deceived. Another interesting fact is the technical stratagem strategy which steals the sensitive credentials from the user's computer system by installing malicious software in it. The main intent of this scheme is to intercept the users' online credentials such as username as well as passwords.

Data breaches such as Privacy leaks, property thefts as well as identity thefts happens due to the well-known phishing method that will be used as computer network attacks. Throughout the year 2017, 29.4% of networked computers were attacked by at least one malware-based web attack in

*International Journal of Engineering Applied Sciences and Technology, 2022*
**Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 160-164**
Published Online November 2022 in IJEAST (http://www.ijeast.com)

accordance with the Kaspersky Laboratory statistics. And, a number of 99,455,606 unique URLs were acknowledged as malevolent URLs by the antivirus components. Additionally in the same year, out of all phishing detection, the financial phishing has grown to a larger extent from 47.5% to 54% approximately. Consequently, this kind of data breach has become one of the major security threats over the internet.

## II. EXISTING SYSTEM

Two machine learning classification model Decision Tree and Random Forest has been selected to detect phishing advertisements.

### 2.1 DECISION TREE ALGORITHM

One of the most widely used algorithm in machine learning technology. Decision tree algorithm is easy to understand and easy to implement. Decision tree begins its work by choosing best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to build tree until it finds the leaf node. Decision tree creates training model which is used to predict target value or class in tree representation each internal node of the tree belongs to attribute and each leaf node of the tree belongs to class label. In decision tree algorithm, gini index and information gain methods are used to calculate these nodes.

### 2.2 RANDOM FOREST ALGORITHM

Random forest algorithm is one of the most powerful algorithms in machine learning technology and it is based on concept of decision tree algorithm. Random forest algorithm creates the forest with number of decision trees. High number of trees gives high detection accuracy. Creation of trees are based on bootstrap method. In bootstrap method features and samples of dataset are randomly selected with replacement to construct single tree. Among randomly selected features, random forest algorithm will choose best splitter for the classification and like decision tree algorithm; Random Forest algorithm also uses gini index and information gain methods to find the best splitter. This process will get continue until random forest creates n number of trees.

## III. PROPOSED SYSTEM

Phishing has been for a long time a difficult threat in every society as it changes form with time and it has taken billions of dollars from governments, companies, and individuals alike. It is an identity theft which employs a kind of social engineering attack to get vital information from individuals or group of individuals. In this paper we focus on studying various features employed in different phishing attacks. It is mostly employed in classification task where it is used as a classifier for mapping input pattern into a specific class. It is a recent supervised learning algorithm that implements a process known as boosting to improve the performance of gradient boosted trees. Among its strengths are better regularization ability which helps to reduce over fitting, high speed and performance owing to the parallel nature in which trees are built, flexibility due to it costume optimization objectives and evaluation criteria, and inbuilt routines for handling missing values. These and many other advantages of XGBOOST have made it an excellent tool of choice for manyresearchers in data science and machine learning. Some of the researchers employed these techniques.

## IV. ARCHITECTURE

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g., the behavior) between them. It can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture, collectively these are called architecture description languages (ADLs).

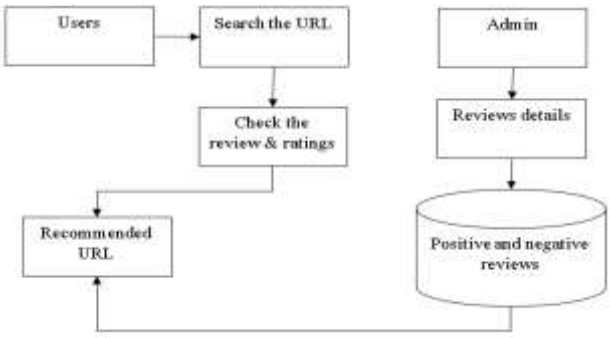


FIG 4.1 SYSTEM ARCHITECTURE DIAGRAM

### 4.1 ADMIN ENROLMENT

As there are more than millions of apps on the App store, there is many competitions between apps to be on top of the leader board based on popularity. The higher rank on the leader board leads to huge number of downloads & million doll or of profit. Apps give advertisement to promote their apps on the leader board. Many apps use fraudulent means to boost their ranking on the leader board of the App store. There are various means to increase downloads & ranking of the app which is done by "bot farms" or "human water armies". In this module, we can design the framework to analyze the applications. Admin can be responsible to

*International Journal of Engineering Applied Sciences and Technology, 2022*
**Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 160-164**
Published Online November 2022 in IJEAST (http://www.ijeast.com)

handle the all feedbacks and User can search the application and buy the applications

### 4.2 ADD ADVERTISEMENTS DETAILS

In an E –Commerce Web application (Web app) is an application program that is stored on a remote server and delivered over the Internet through a browser interface. Admin can add the advertisements in web site.

### 4.3 WORD TRAINING

Admin collect reviews and have various types of reviews. Reviews may be rating reviews, text reviews and smileys reviews. All reviews are stored in database for future evaluation. Ratings, reviews, and emoticons are stored in database. Rating, Reviews and Emoticons are the evaluation or assessment of something, in terms of quality quantity or some combination of both.

### 4.4 SENTIMENT ANALYSIS

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and ratings for app that range from marketing to customer service to buy the application efficiently. Admin can analyse whether the app is positive or negative. In star rating, we can calculate star count values. In text reviews, extract keywords and matched with database. Then smileys reviews are calculated based positive and negative symbols.

### 4.5 RECOMMENDATION SYSTEM

Recommender systems are a subclass ofinformation filtering system that seek to predict the "Reviews" or "preference" that a user would give to an URL. User can search the URL in search bar. And view the list of URL based on ratings and review details. Implement the machine learning algorithm to classify the URL such as positive or negative. Positive URL is display in recommendation panel based on ratings and reviews. If the URL has negative review means, automatically the positive URL in recommendation panel.

## V. RESULT

**Fig 5.1 Admin Login**

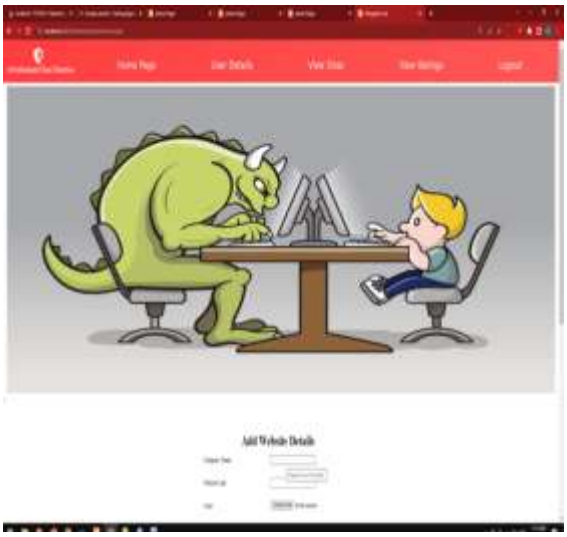

**Fig 5.2 User Login**

**Fig 5.3 User Details**

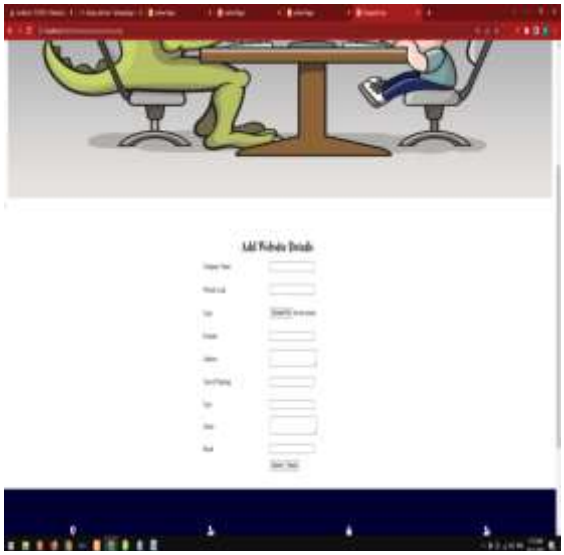

**Fig 5.4 Add Advertisement Details**

*International Journal of Engineering Applied Sciences and Technology, 2022*
**Vol. 7, Issue 7, ISSN No. 2455-2143, Pages 160-164**
Published Online November 2022 in IJEAST (http://www.ijeast.com)

**Fig 5. 5 Advertisement Details**

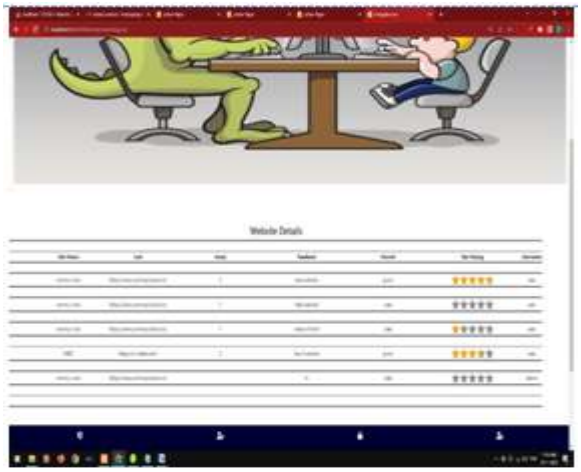

**Fig 5. 6 Ratings and Feedback**

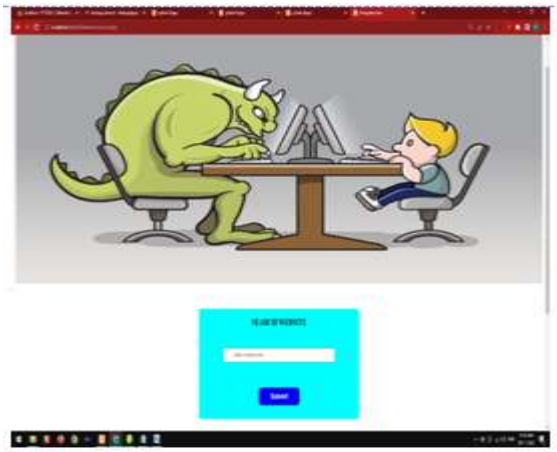

**Fig 5.7 Search URL**

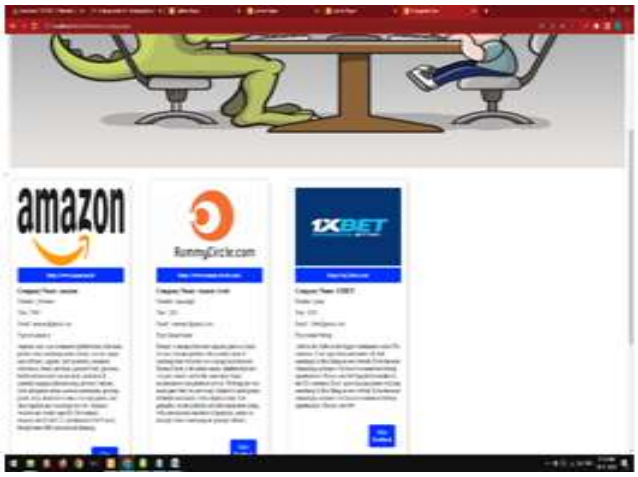


**Fig 5.8 Advertisement URL Recommendation**


**Fig 5.9 Give Ratings (star and Emoticon Ratings)**

## VI. CONCLUSION AND FUTURE ENHANCEMENTS

### 6.1 CONCLUSION

In this project, proposed a product recommendation system based on hybrid recommendation algorithm. The main advantages of method are a visual organization of the data based on the underlying structure, and a significant reduction in the size of the search space per result output. And user can easily search the websites anywhere and anytime. Ratings, reviews, and emoticons are analyzed and categorized as positive and negative sentiments. Search the websites based on filtering and reviews-based filtering. Medium Access Control (MAC) based filtering approach can be used to avoid fake websites. The current results are notably better than random approach. However, feel that with a better dataset and a few improvements to method, may achieve better results. Recommendations is one of the main modules of the system which helps overcome the drawbacks of the traditional

Collaborative and Content Based Recommendations. And have obtained promising results using current model.

**6. 2 FUTURE ENHANCEMENT**

We can extend the work with number of directions work can potentially take in the future. It is possible to modify algorithm to use an approach that lies between collaborative filtering and content-based filtering. There are multiple ways to do this. One way would be to include user-specific data such as information about products liked by a user's friends, and information from reviews written by the user and their similarity to other interests and so on, as inputs to the hybrid systems.

## VII. REFERENCES

[1]. L. Jin, H. Takabi, and J. B. Joshi.(Feburary 2011), "Towards active detection of identity clone attacks on online social networks," in Proceedings of the first ACM conference, San Antonio, TX, USA,

[2]. (pp. 27).

[3]. M. Conti, R. Poovendran, and M. Secchiero. (August 2012), "FakeBook: Detecting fake profiles in on-line social networks," in Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012, Turkey, (pp. 1071–1078).

[4]. Z. Shan, H. Cao, J. Lv, C. Yan, and A. Liu. (January2013), "Enhancing and identifying cloning attacks in online social networks," in Proceedings of the 7th International Conference, Kota Kinabalu, Malaysia, (pp. 1–6).

[5]. S. Gilda. (2017)," Notice of Violation of IEEE PublicationPrinciples: Evaluating machine learning algorithms for fake news detection", IEEE 15th Student Conference on Research and Development (SCOReD), (pp.110-115).

[6]. **DOI:**10.1109/SCORED.2017.8305411.

[7]. M. Granik and V. Mesyura.(2017), "Fake news detection using naive Bayes classifier," 2017 IEEE 1st Ukr. Conf. Electr. Comput. Eng. UKRCON 2017 - Proc., (pp. 900–903).

[8]. J. Kapusta, P. Hájek, M. Munk, and Ľ. Benko. (2020), "Comparison of fake and real news based on morphological analysis," Procedia Comput. Sci., vol. 171, (pp. 2285–2293).

[9]. Hajek, P., Barushka, A., and Munk, M. (2020). "Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining.", Neural Comput. Appl. 32, (pp. 17259–17274). **DOI:**10.1007/s00521-020-04757-2.

[10]. He, D., Pan, M., Hong, K., Cheng, Y., Chan, S., Liu, X., et al. (2020). "Fake review detection based on pu learning and behavior density.", IEEE Network (pp. 34, 298–303)**.DOI:**10.1109/MNET.001.1900542.

[11]. Hovy, D. (2016). "The enemy in your own camp: how well can we detect statistically-generated fake reviews -an adversarial study," in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, (pp. 351–356).**DOI:**10.18653/v1/P16-2057.

[12]. Jerripothula, K. R., Rai, A., Garg, K., and Rautela, Y. S. (2020)." Feature-level rating system using customer reviews and review votes.", IEEE Trans. Comput. Soc. Syst. 7, (pp.1210–1219).**DOI:**10.1109/TCSS.2020.3010807.

[13]. David Ndumiyana, Munyaradzi Magomelo, and Lucy Sakala. (April 2013), "Spam Detection using a Neural Network Classifier," Online Journal of Physical and Environmental Science Research, vol. 2, issue 2, (pp. 28-37).

[14]. Abbasi A., and Chen H. (2007). "Detecting Fake Escrow WebsitesUsing Rich Fraud Cues and Kernel Based Methods," in Proceedings of the 17th Workshop on Information Technologiesand Systems, Montreal, Canada, (pp. 55-60).

[15]. Xia Hu, Jiliang Tang, and Huan Liu.(2014),"Online socialspammer detection". In AAAI'14, (pg. 59–65).

[16]. Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C., and Subrahmanian, V. (2018). "REV2: fraudulent user prediction in rating platforms," in Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, WSDM '18, (pp. 333–341).**DOI:**10.1145/3159652.3159729.