EasyChair Preprint

№ 8412

Feature Selection & ML Based Prediction of Phishing Websites

Anjaneya Awasthi and Noopur Goel

July 10, 2022

Feature Selection & ML Based prediction of Phishing Websites

*Anjaneya Awasthi[1[0000-0003-4033-936X]] and Noopur Goel[2[0000-0003-3351-3761]]

[1] Department of Computer Applications

VBS Purvanchal University, Jaunpur, UP, India

[1] anjaneyaawasthi@gmail.com

[2] Department of Computer Applications

VBS Purvanchal University, Jaunpur, UP, India

[2] noopurt11@gmail.com

Abstract

The Internet and cloud based technologies increase online purchases and transactions, in recent years. Phishing be a well-known assault that deceives consumers into seeing harmful material in exchange on the behalf of their personal information. However, due to ineffective security systems, the number of victims will grow exponentially. The Internet's anonymous and unregulated foundation makes it more vulnerable to phishing attempts. In this paper, we use a point wise mutual information method to offer a phishing detection system that uses features from the website URLs. We built a supervised machine learning system for phishing website detection. We used "Embedding, Sentiment, and Lexicon characteristics, as well as PMI-semantic orientation", in the study. The methods "SVM, Naïve Bayes, KNN, Decision Tree and algorithm Random Forest," were used to apply extracted features. Experiments using our suggested framework in a multi-class scenario, as well as in a binary setting, show promise in terms of "values of Kappa, enhanced accuracy, and calculated f-values". These findings suggest that the framework we've provided be a viable option on the behalf of detecting malicious phishing behavior with severity of online links of social networks and other fake web URLs. Finally, we used different machine learning techniques to compare the outcomes of suggested and baseline characteristics. 10 fold cross validation calculated 90.363 highest accuracy and all four experiment evaluated always high accuracy on the behalf of Random Forest on the behalf of training dataset on 80%. The test result also calculated enhanced accuracy on the behalf of Random Forest on 20% or test dataset.

Keywords: Algorithm Naïve Bayes, KNN , Decision Tree (J48), Random Forest algorithm and SVM.

Introduction:

Phishing is said to be a deceptive tactic that uses social and technological trickery to steal a person's identity and financial information. Users can utilize bogus websites to provide financial data such as usernames and passwords by using faked e-mails from real firms and agencies. Hackers frequently utilize systems to intercept usernames and passwords on the behalf of online accounts of customers. Phishers utilize a variety of tactics to collect user information, including email, URLs, instant chats, forum comments, phone calls, and text messages. Phishing material has a structure that is similar to legitimate content in order to fool people into accessing it in order to get sensitive information. The main goal of Phishing is to get specific personal information on the behalf of financial gain or to commit identity theft. Phishing assaults are wreaking havoc on businesses all around the world [1], the

majority of phishing attempts target financial/payment institutions in addition to webmail. Criminals create illegal copies of legitimate web sites and email, in order to get private data [2–4]. In addition to slogans, this e-mail should be displayed with the logos of a respectable firm. In addition to the structure of HTML, the design allows on the behalf of the copying of pictures or a full website [5]. It's also one of the reasons on the behalf of the Internet's rapid expansion as a communication medium, since it allows on the behalf of the misuse of brands and trademarks [6–8]. The "spooled" emails send by phisher to as many individuals as possible in order to catch users. When users read these e-mails, they are often directed away from the actual company and toward a faked website. There's a good probability that user information will be exploited. As a result, phishing has become extremely urgent, difficult to detect and predict, and unduly crucial in modern culture [9 - 11]. However, there may be a dearth of efficient anti-phishing approaches to identify dangerous URLs within a business in order to safeguard its users. In the case that harmful code is placed on the website, hackers may steal precious and crucial user information as well as install malware, posing a major threat to both cyber security and user privacy. Malicious URLs on the Internet may be quickly discovered using Machine Learning (ML) techniques [12 – 18]. As a result, conventional methods are unable to detect new dangerous URLs. Researchers proposed approaches based on machine learning to identify harmful URLs in order to overcome the limitations of the blacklist-based system [19–21]. Malicious URL detection may be thought of as a binary classification problem with two possible outcomes: malicious and benign [22–24]. When compared to the blacklist technique, this strategy offers higher generalization ability on the behalf of detecting unknown harmful URLs. One of the ML approaches that gives a solution on the behalf of difficult real-time challenges be the (RNN) and (LSTM) . RNN can store inputs on the behalf of a longer length of time using LSTM. It be comparable to the idea of computer storage. Furthermore, each feature will be handled in accordance with the uniform distribution [25].

Research Background:

S. Nisha and A. N. Madheswari discussed about Phishing assaults. Phishing assaults are now available in a variety of forms. Messages requiring users to verify account information, requesting that users re-enter their information, bogus account charges, unwanted account changes, new free services requiring immediate action, and many other malicious sites are sent to a large number of recipients in the hopes that the unsuspecting person will react by clicking on a link to or signing on a fake site [26].

H. B. Kazemian and S. Ahmed discussed about phishing websites. The disadvantage of this strategy be that blacklists cannot generally include all phishing websites since it takes a long time on the behalf of a newly built fraudulent website to be added [27].

K. Thomas et al., analyzed about phishing assaults. Malware be usually sent in the form of an email attachment that may be opened and downloaded. Malware be often installed. A blacklist-based strategy, a content-based approach, and a heuristic-based approach are all employed to combat phishing assaults. A blacklist be a collection of harmful URLs [28].

A. Firdaus et al., & M. F. A. Razak et al., considered about Malware-based phishing. Malware-based phishing refers to assaults that cause malicious software to be installed and executed on consumers' systems [29, 30].

J. A. Chaudhry et al., discussed about Key loggers and screen grabbers, spyware. Malware be usually sent in the form of an email attachment that may be opened and downloaded. Key loggers and screen grabbers, spyware that collects and logs input keyboards or displays the screen and provides information to the phisher, are two types of malware widely used in phishing attempts. In some circumstances, the attacker's goal be to take control of the victim's computer [31].

R. Gowtham and I. Krishnamurthi analyzed about some phishing method. Injection of content be a phishing method in which a phisher modifies a portion of the information on a trusted internet page. This be done to redirect the visitor away from the genuine website to a page where personal information must be submitted [32].

G. Xiang et al., analyzed about Heuristic-based systems. Using term-frequency-inverse document frequency (TF–IDF) measures, a content-based approach on the behalf of detecting phishing websites. Heuristic-based systems collect characteristics from websites in order to determine if they are phishing or not [33].

Methodology:

Explaining Collecting data, defining phishing traits, creating a model, testing, and ultimately comparing the results are the five components of the phishing detection system. The component of the phishing detection system be shown in algorithms description section, which are addressed in the following sections.

Data Description:

The collection of data is the initial step in the implementation. The dataset phase be critical on the behalf of ensuring the correctness of the results. The dataset will help to clarify and explain phishing and legitimate actions. The dataset be then evaluated on the behalf of additional investigation and the results are utilized to forecast or anticipate future phishing occurrences.

All of the characteristics were gathered from UCI dataset. There shape of dataset is (11055, 16)

phishing website characteristics in all that have been collected. This information was mostly gathered from a well-known phishing database. The categorical_val of datset are:

```
['having_IPhaving_IP_Address',
'URLURL_Length',
'Shortining_Service',
'having_At_Symbol',
'double_slash_redirecting',
'Prefix_Suffix',
'having_Sub_Domain',
'port',
'HTTPS_token',
'Request_URL',
'URL_of_Anchor',
'Links_in_tags',
'SFH',
'Submitting_to_email',
'Abnormal_URL',
'Result']
```

All the values are categories by histogram blue and red color as:

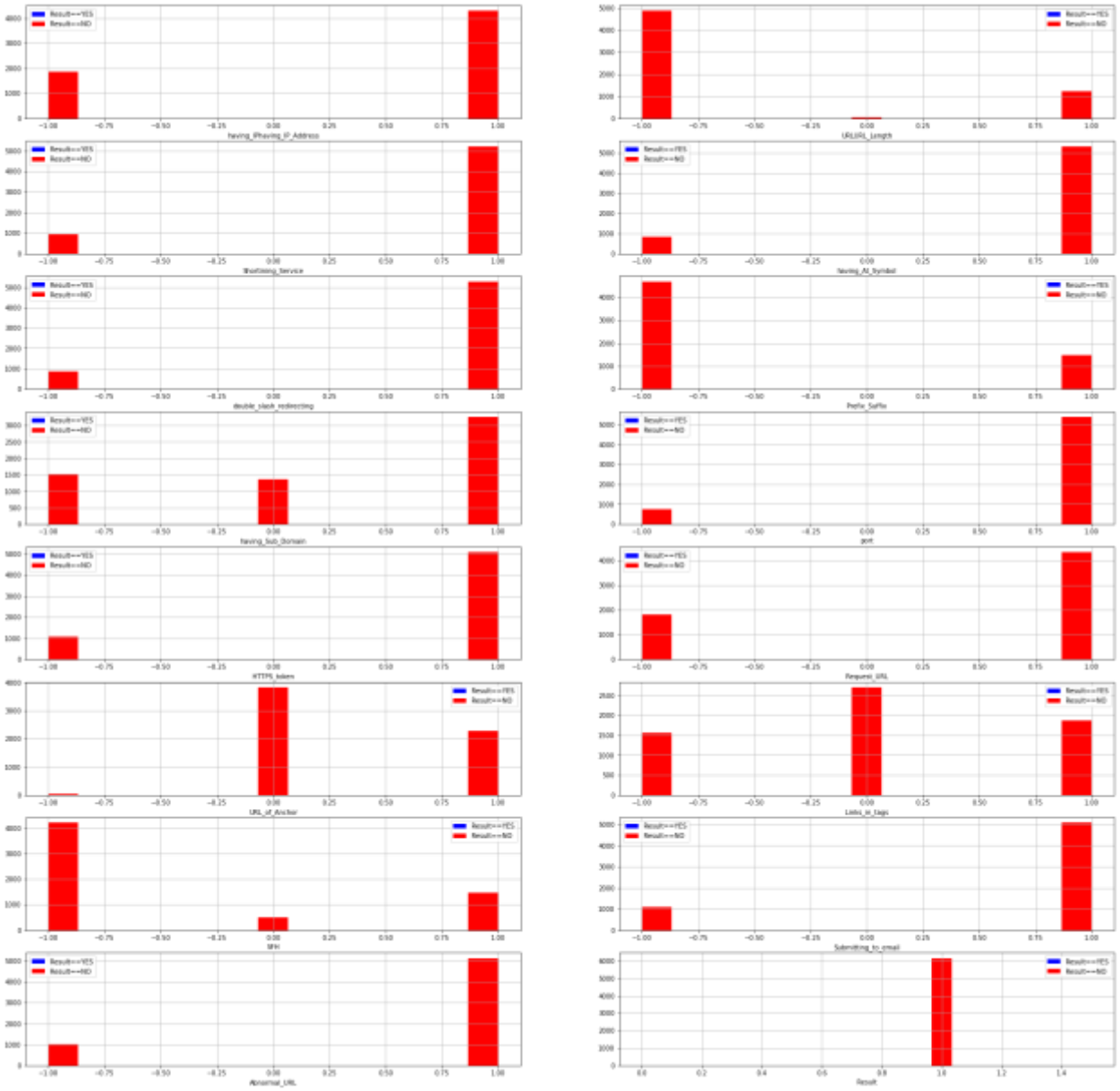

Figure 1. Representation of Histogram Plotting of Phishing dataset

The categorical values represents in figure 1., the gathered dataset have been converted to numerical values by substituting the values "1," "0," and "-1" .

Features Selection Method:

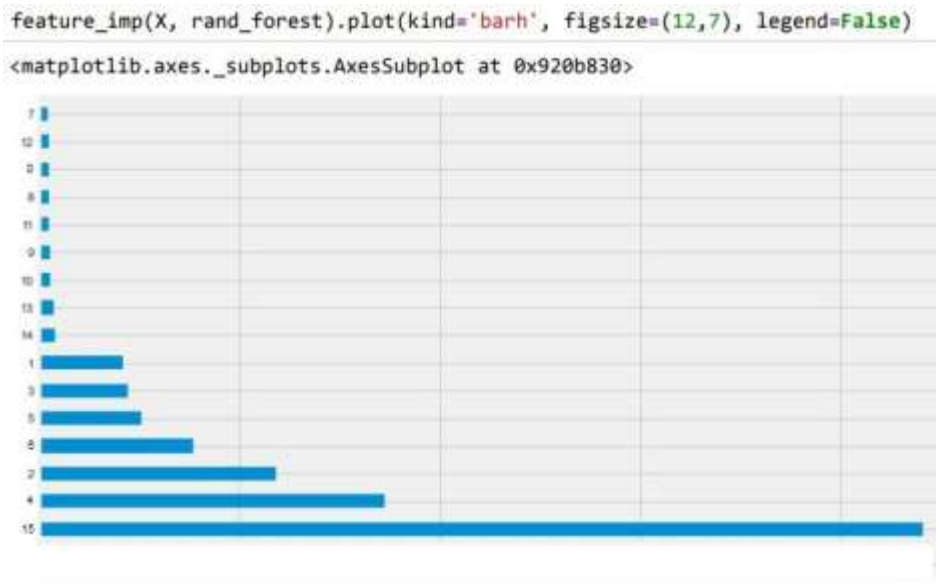

Figure 2. Representation of Phishing dataset features selection by Random Forest Algorithm

The data science provide an environment for play with features as like Random Forest used as feature selection technique or tree based algorithms. This tree based method ranked how increase node purity. The “nodes with the largest drop in impurity are found at the beginning of the trees, while the nodes with the least drop decrease in impurity are found at the conclusion”. We may produce a subset of the most essential characteristics by trimming trees below a certain node. Fig 2., shows the target variable “Results” have highly important position in this phishing datset.

Algorithms Description:

Naïve Bayes Algorithm:

Abhilash, P. M., and D. Chakradhar. Introduced about effective inductive learning algorithms. Naïve Bayes is one of the “most efficient and effective inductive learning algorithms in the world of machine learning”, and it has been employed as an excellent classifier in various social media research.

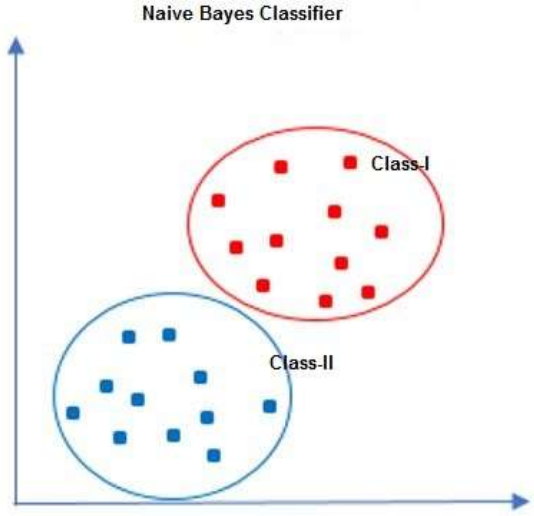

Figure 3. Representation of Naïve Bayes Classification on Phishing dataset

Nave Bayes text classification has been widely utilised in document categorization assignments since the 1950s, and it can categorise any sort of data, including text, network characteristics, phrases, and so on. This method be referred to as a generative model, and it describes how a dataset be created using a probabilistic model in figure 3. It can produce fresh data that be comparable to the data on which the model be being trained by sampling from this model. On the behalf of textual characteristics and word embeddings, we employed the most basic version of the Naïve Bayes classifier in our research [34].

KNN Algorithm:

Khorshid, Shler Farhad, and Adnan Mohsin Abdulazeez introduce about KNN supervised algorithm. The K-Nearest Neighbors (KNN) technique be a supervised learning algorithm and one of the most straightforward instance-based learning algorithms on the behalf of multi-class problems.

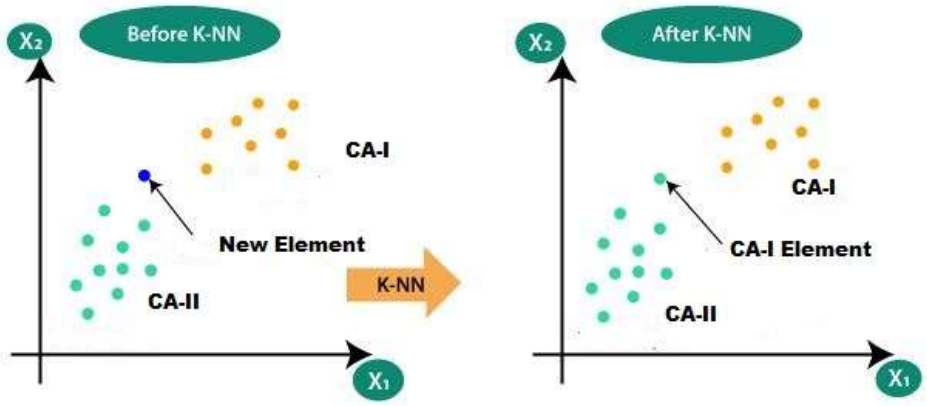

Figure 4. Representation of K-NN Classification on Phishing dataset

The distance between a fresh sample and its neighbor be employed in this approach to categorize it. As a result, the training set's K-nearest neighbours are found, and an item be assigned to the class with the most members among its k nearest neighbours in figure 4. KNN be a non-parametric lazy learning method that makes no assumptions about the distribution of the underlying data [35].

Decision Tree Algorithm:

Charbuty, Bahzad, and Adnan Abdulazeez introduce about decision tree. In decision tree be a well-known classification algorithm and one of the most extensively used inductive learning methods in machine learning in figure 5.

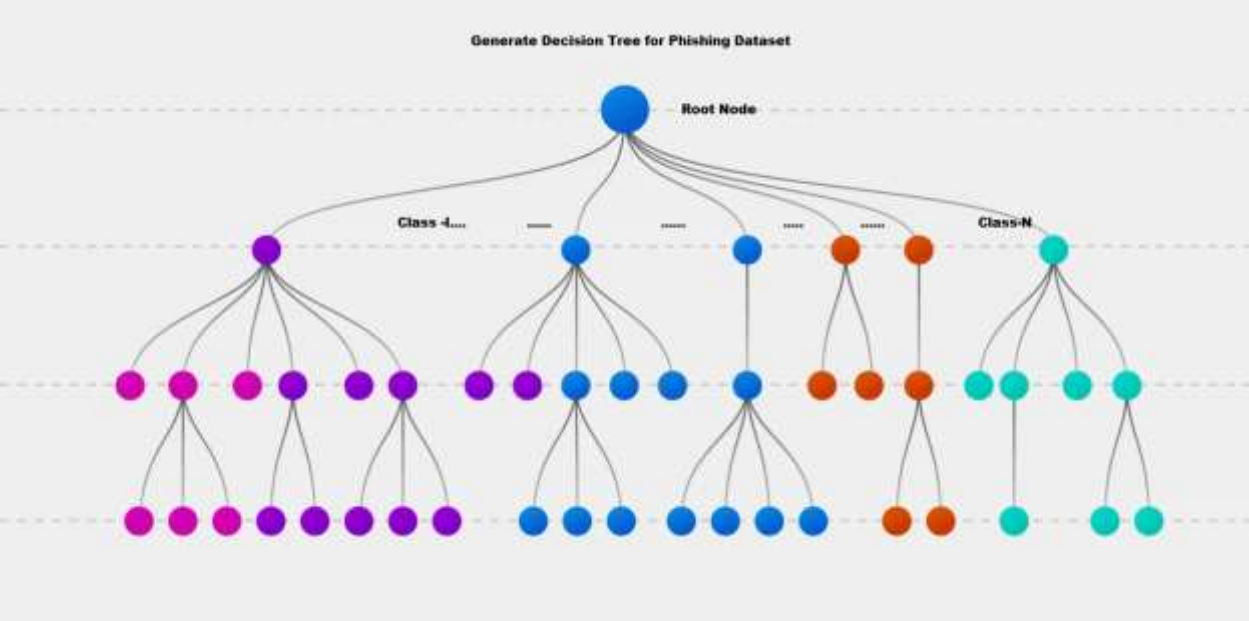

Figure 5. Representation of Decision Tree Classification on Phishing dataset

It can handle both continuous and discrete characteristics, as well as training data with missing values. The idea of information entropy be used to construct decision trees using labeled training data. Their capacity to understand disjunctive statements and tolerance to noisy data appear to make them excellent on the behalf of text categorization [36].

Random Forest Algorithm:

Zhang, W., et al., introduce about Random forest (RF) classification and regression technique that uses an ensemble of algorithms. On a random subset of data samples and characteristics, RF constructs numerous decision tree classifiers.

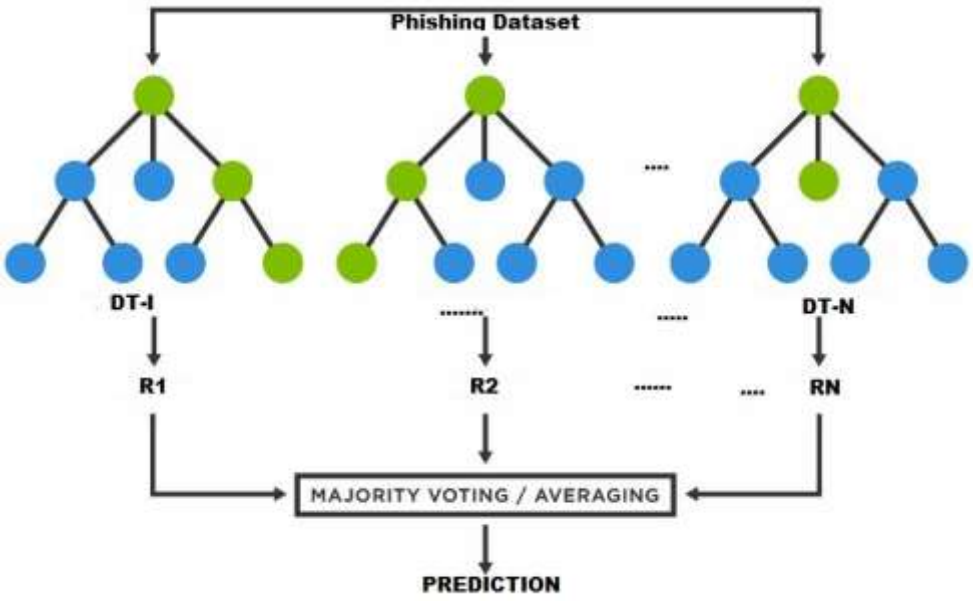

Figure 6. Representation of Random Forest Classification on Phishing dataset

The majority voting of decision trees be used to classify a fresh sample in figure 6. The fundamental benefit of RF is that it scales well to big datasets, be a solid approach on the behalf of predicting missing data, and provides excellent accuracy even when a significant amount of the data be missing [37].

Support Vector Machine Algorithm:

Chandra, Mayank Arya, and S. S. Bedi. Introduce about SVM supervised learning pattern recognition technique that can categorise both linear and non-linear data. SVM's main idea be to find separators that can best identify the different classes in a search space in figure 7.

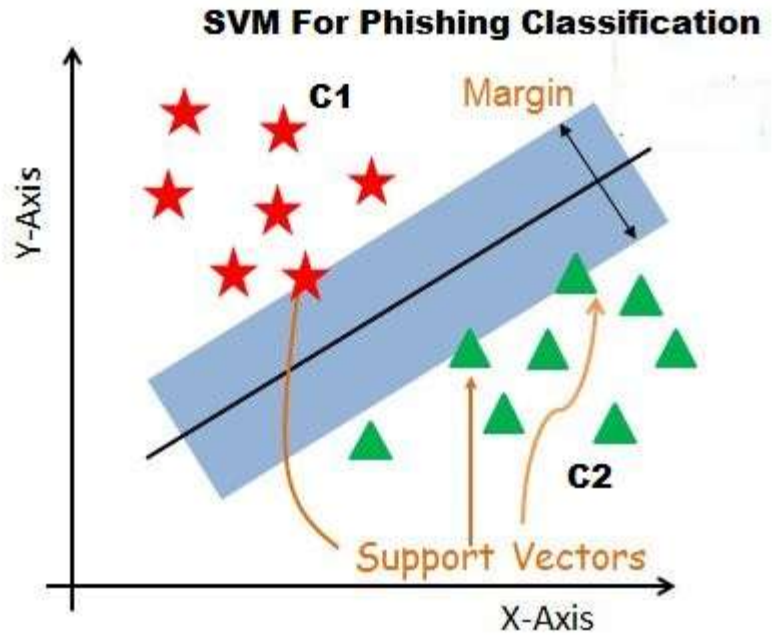

Figure 7. Representation of SVM Classification on Phishing dataset

Support vectors are the data points that separate one or more hyperplanes utilising crucial training tuples [38].

Proposed Model:

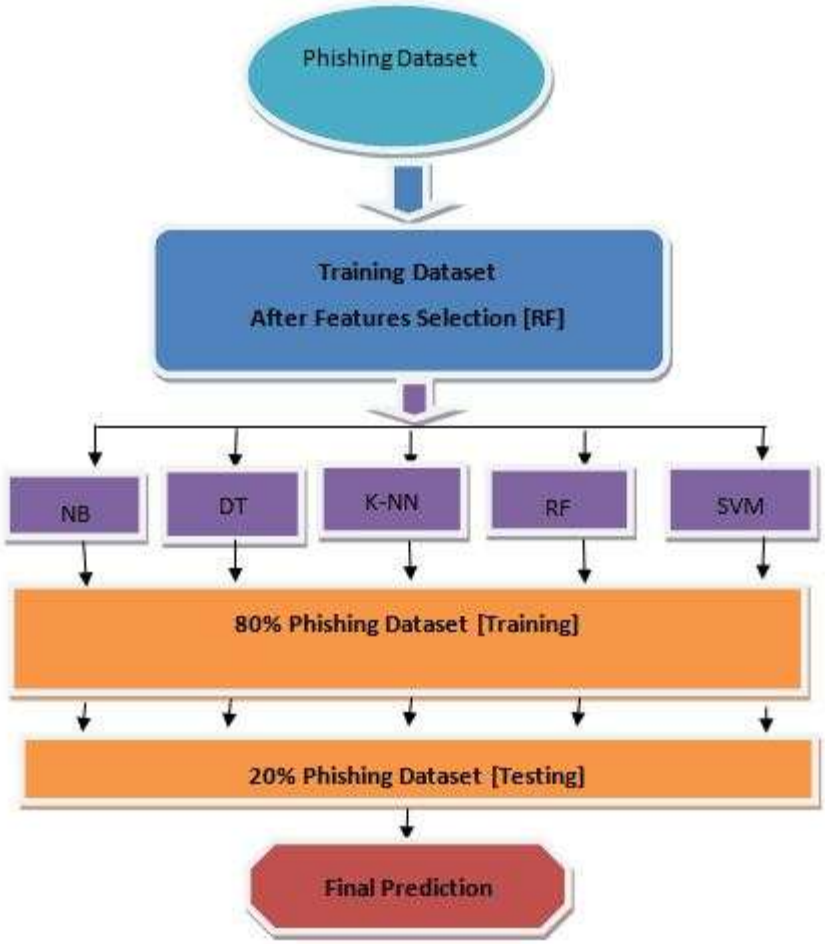

Figure 8. Representation of Proposed Model on Phishing dataset

The tagged synthetic data be then used to train a model that may be used to actual election data to assess if there be evidence that a voting precinct be at danger. The model's outcome variable be divided into two categories: Training and Testing in figure 8. We train this model with Naïve Bayes, decision tree, K-NN, Random Forest and SVM. Random Forest be an ensemble supervised machine learning technique and features selection method also that has previously been demonstrated to be select and in identifying importance probability of dataset. To evaluate our model's performance, we divide the synthetic data into 10 folds, train it on 80% and then test it on 20%, the one fold of data that was kept on the behalf of testing. The results enhanced the prediction accuracy on the behalf of the ten tests.

Performance evaluation:

Performance measurements assess certain aspects of categorization task performance and do not always give the same information. Any classification method requires an understanding of how a

model works. Different evaluation measures may have different underlying mechanics, thus understanding what each of these metrics reflects and what sort of information they are trying to transmit be critical on the behalf of comparison. A classifier's performance may be measured in a variety of ways, including accuracy, F-measure and kappa values [39].

True Positive (TP)	The number of instances correctly labeled as belonging to the positive class
True Negative (TN)	Negative instances correctly classified as negative
False Positive (FP)	Instances incorrectly labeled as belonging to the class
False Negative (FN)	Instances not labeled as belonging to the positive class but should have been
Accuracy	Calculating the accuracy of a classifier is the easiest evaluation approach among the metrics listed above, however it does not work on the behalf of imbalanced datasets. In multi-class classification with imbalanced data, accuracy might be deceiving, thus we choose on the behalf of precision and recall, or a combined precision and recall measure known as the f-measure. However, other from being the harmonic mean of accuracy and recall, f-measure does not have a particularly strong intuitive explanation.
Kappa Statistic	The Kappa statistic to psychology as a measure of agreement between two judges, and it has since been utilized in the literature as a performance measure in categorization. It essentially indicates how much better a classifier performs when compared to a classifier that guesses at random based on the frequency of each class. The Kappa statistic is used to assess the agreement between a dataset's expected and observed categorizations while accounting on the behalf of chance agreement. It's simply a normalized version of the proportion of right classifications (classification accuracy), where normalization be done in relation to a random classifier's performance. It illustrates how much a classifier improves over a random one at a look. Kappa be always more than or equal to one. Closer values near 1 suggest that the classifier be successful, while values closer to 0 indicate that it be ineffective. Kappa was created to account on the behalf of the risk of guessing, but the assumptions it makes about rater independence and other characteristics are not well supported, and as a result, the agreement estimate might be unduly reduced.

All the comprehensive experiments to evaluate the performance of each of the five classifiers, namely Algorithm NB, KNN, technique Decision Tree, machine learning Random Forest, and predictor Support Vector Machine.

All five classifiers were put to the test in a variety of scenarios. We also tested our suggested framework in a binary environment to evaluate if our multi-class strategy on the behalf of detecting phishing attempt behavior in tweets works better in a binary classification challenge. On the behalf of the goal of standardization of best findings to cross compare across each layer of characteristics that provide experimental results, we omitted trials with low performance from the list. All experiments were conducted using a 10-fold cross validation scheme.

Results:

This section compares the effectiveness of several classifiers when it comes to categorizing tweets into various levels.

Table 1. Classifiers performance under various settings in multi-class classification.

Features selection Method	Cross Fold Validation	Classifier	Accuracy	Kappa Statistics	F-Measure
Random Forest	5	.NB	067.214	000.397	000.744
		.KNN	086.692	000.416	000.864
		.DT	089.714	000.475	000.886
		.RF	**089.759**	**000.474**	**000.886**
		.SVM	086.576	000.417	000.864
Random Forest	10	.NB	076.910	000.276	000.794
		.KNN	086.679	000.415	000.864
		.DT	089.731	000.479	000.887
		.RF	**090.363**	**000.471**	**000.889**
		.SVM	089.747	000.475	000.886
Random Forest	15	.NB	076.710	000.256	000.774
		.KNN	085.679	000.401	000.873
		.DT	087.731	000.461	000.874
		.RF	**090.253**	**000.311**	**000.779**
		.SVM	088.747	000.451	000.756
Random Forest	20	.NB	075.680	000.232	000.761
		.KNN	084.655	000.381	000.753
		.DT	086.691	000.371	000.694
		.RF	**089.321**	**000.279**	**000.671**
		.SVM	086.747	000.621	000.696

Table 1 displays the training results on 80% dataset of multi-class classification on the behalf of each classifier under various circumstances. With the features selection method "Random Forest" all the classifiers: Naïve Bayes, KNN, Decision Tree (J48), Random Forest and SVM, calculate classification accuracy, F- Measure and Kappa Statistics in environment of cross fold 5, 10 , 15 and 20 validations.

Table 2. Classifiers performance under various settings in binary classification.

Random Forest	10	.NB	076.810	000.281	000.786
		.KNN	087.679	000.315	000.764
		.DT	088.731	000.479	000.887
		.RF	**091.363**	**000.281**	**000.789**

		.SVM	89.891	0.515	0.716

Table 2 displays the test results on remaining 20% datset of multi-class classification on the behalf of each classifier under various circumstances. With the features selection method “Random Forest” all the classifiers: Naïve Bayes, KNN, Decision Tree (J48), Random Forest and SVM, calculate classification accuracy, F- Measure and Kappa Statistics in environment of cross fold 10 validation.

Discussion:

The current study takes a step forward by highlighting the shortcomings of the current phishing attempt detection method. We developed a holistic framework on the behalf of determining the results of phishing attempt on Twitter in this study, which be based on past research from other fields. Random Forest calculated highest values of accuracy in each iteration with different cross fold 5, 10, 15 and 20 validations

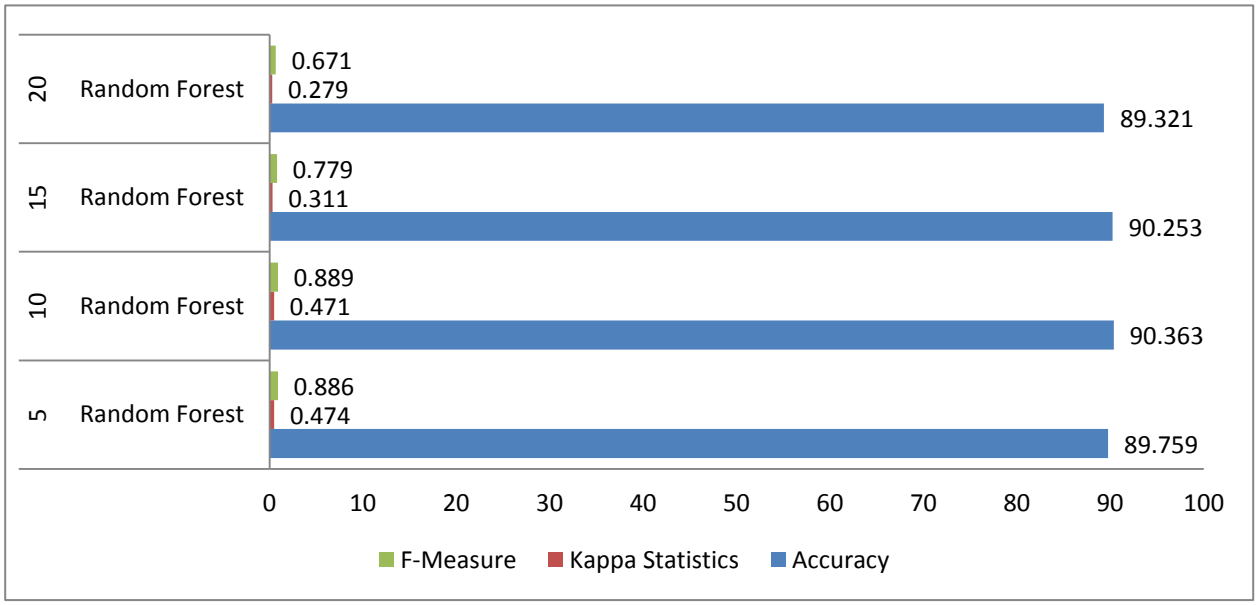

Figure 9. Representation of Random Forest on the behalf of training dataset on 80% with various cross validation fold.

I-Random Forest, 89.759, 0.474, 0.886, II-Random Forest, 90.363, 0.471, 0.889 III- Random Forest, 90.253, 0.311, 0.779 and Random IV-Forest, 89.321, 0.279, 0.671. Figure 9. Shows 10 fold cross validation calculated 90.363 highest accuracy and all four experiment evaluated always high accuracy on the behalf of Random Forest on the behalf of training dataset on 80% . The test result also calculated enhanced accuracy on the behalf of Random Forest on 20% or test dataset.

To identify results in tweets, a large number of trials were conducted using a binary scheme (either Phishing behaviour present in the tweet or not) and a multi-classification strategy. The current study's main goal and contribution was to establish a systematic technique to apply target variable levels to phishing attempt behavioural text using multi-class classification.

In binary classification, our suggested method on the behalf of detecting phishing attempt behaviour outperforms numerous feature engineered techniques and methodologies reported. Random Forest had the greatest overall classifier performance.

Feature selection contributes to enhancing prediction accuracy by lowering dimensionality of the dataset and utilized to provide improved results in text mining domain. The capacity to limit the number of selected characteristics while keeping as much overall prediction information as feasible be a fundamental requirement on the behalf of successful feature selection. The majority of the published literature focuses on structured data approaches. Previously developed feature selection methods were created without considering the impact of class distribution on the learning problem. As a result, many of them only produce a marginal improvement in performance. Multi-minority classes and creating new discriminatory characteristics of data that increase classifier accuracy.

Conclusion:

Although the internet and social media offer demonstrable benefits on the behalf of society, its widespread usage may have severe negative implications. In Twitter, we built a model on the behalf of identifying Phishing attempt and its severity result. In binary and multi-class classification, be the most efficient strategy on the behalf of dealing with class imbalance, where misclassification on the behalf of minority class (es) has a larger cost in terms of its influence on detection model reliability. The proposed model is a feature-based model "Random Forest" that leverages characteristics from message content to construct a machine learning classifier on the behalf of categorizing messages or URLs as non-phishing and determining severity of results as 1 & -1. The training and test results also calculated enhanced accuracy on the behalf of Random Forest on 80% & 20% dataset. Other social media platforms (such as Facebook, YouTube, and others) should be looked at to determine whether there be a similar trend of phishing attempt intensity.

Future Work:

Future research could improve automated machine learning model that can detect phishing attempt behaviour and severity, which could be a step toward automated systems on the behalf of analysing contemporary social online behaviours from written text and visual content that can harm mental health. The detection programme might analyse the phisher's messages and then align them to a predetermined level of severity, allowing on the behalf of early identification of Phishing attacks.

References:

[1] Jain A.K., Gupta B.B. "PHISH-SAFE: URL Features-Based Phishing Detection System Using Machine Learning", *Cyber Security*. *Advances inside Intelligent Systems in addition toward Computing*, vol. 729, 2018, https://doi. org/10.1007/978-981-10-8536-9_44

[2] Purbay M., Kumar D, "Split Behavior of Supervised Machine Learning Algorithms on the behalf of Phishing URL Detection", Lecture Notes inside Electrical Engineering, vol. 683, 2021, https://doi.org/10.1007/978-981- 15-6840-4_40

[3] Gandotra E., Gupta D, "An Efficient Approach on the behalf of Phishing Detection using Machine Learning", Algorithms on the behalf of Intelligent Systems, Springer, Singapore, 2021, https://doi.org/10.1007/978-981-15-8711-5_ 12.

[4] Hung Le, Quang Pham, Doyen Sahoo, in addition toward Steven C.H. Hoi, “URLNet: Learning a URL Representation with Deep Learning on the behalf of Malicious URL Detection”, Conference’17, Washington, DC, USA, arXiv:1802.03162, July 2017.
[5] Hong J., Kim T., Liu J., Park N., Kim SW, “Phishing URL Detection with Lexical Features in addition toward Blacklisted Domains”, Autonomous Secure Cyber Systems. Springer, https://doi.org/10.1007/978-3-030-33432- 1_12.
[6] J. Kumar, A. Santhanavijayan, B. Janet, B. Rajendran in addition toward B. S. Bindhumadhava, “Phishing Website Classification in addition toward Detection Using Machine Learning,” 2020 International Conference on Computer Communication in addition toward Informatics (ICCCI), Coimbatore, India, 2020, pp. 1–6, 10.1109/ ICCCI48352.2020.9104161.
[7] Hassan Y.A. in addition toward Abdelfettah B, “Using case- based reasoning on the behalf of phishing detection", Procedia Com- puter Science, vol. 109, 2017, pp. 281–288.
[8] Rao RS, Pais AR. Jail-Phish: An improved search engine based phishing detection system. Computers & Security. 2019 Jun 1; 83:246–67.
[9] Aljofey A, Jiang Q, Qu Q, Huang M, Niyigena JP. An effective phishing detection model based on char- acter level convolutional neural network from URL. Electronics. 2020 Sep; 9(9):1514.
[10] AlEroud A, Karabatis G. Bypassing detection of URL-based phishing attacks using generative adversar- ial deep neural networks. In: Proceedings of the Sixth International Workshop on Security in addition towardPrivacy Analytics 2020 Mar 16 (pp. 53–60).
[11] Gupta D, Rani R, “Improving malware detection using big data in addition towardensemble learning”, Computer Elec- tronic Engineering, vol. 86, no.106729, 2020.
[12] A. Awasthi and N. Goel, "Generating Rules to Detect Phishing Websites Using URL Features," 2021 1st Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology(ODICON), 2021, pp. 1-9, doi: 10.1109/ODICON50556.2021.9429003.
[13] Awasthi, A., Goel, N. (2021). Phishing Website Prediction: A Comparison of Machine Learning Techniques. In: Jeena Jacob, I., Kolandapalayam Shanmugam, S., Piramuthu, S., Falkowski-Gilski, P. (eds) Data Intelligence and Cognitive Informatics. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-15-8530-2_50
[14] Awasthi, A., Goel, N. (2021). Phishing Website Prediction: A Machine Learning Approach. In: Panigrahi, C.R., Pati, B., Pattanayak, B.K., Amic, S., Li, KC. (eds) Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, vol 1299. Springer, Singapore. https://doi.org/10.1007/978-981-33-4299-6_12
[15] J. Anirudha in addition towardP. Tanuja,”Phishing Attack Detection using Feature Selection Techniques “, Proceed- ings of International Conference on Communication in addition towardInformation Processing (ICCIP), 2019, http:// dx.doi.org/10.2139/ssrn.3418542
[16] Wu CY, Kuo CC, Yang CS,” A phishing detection system based on machine learning” In: 2019 Interna- tional Conference on Intelligent Computing in addition towardits Emerging Applications (ICEA), pp 28–32, 2019.
[17] Chiew KL, Chang EH, Tiong WK,”Utilisation of website logo on the behalf of phishing detection”, Computer Security, pp.16–26, 2015.
[18] Srinivasa Rao R, Pais AR, “Detecting phishing websites using automation of human behavior”, In: Pro- ceedings of the 3rd ACM workshop on cyber-physical system security, ACM, pp 33–42, 2017.
[19] Sahingoz OK, Buber E, Demir O, Diri B, “Machine learning based phishing detection from URLs”, Expert System Application, vol. 117, pp. 345–357, 2019.
[20] Zamir A, Khan HU, Iqbal T, Yousaf N, Aslam F et al., “Phishing web site detection using diverse machine learning algorithms”, The Electronic Library, vol.38, no.1, pp. 65–80, 2019
[21] Almseidin M, Zuraiq AA, Al-kasassbeh M, Alnidami N, “Phishing detection based on machine learning in addition towardfeature selection methods”, International journal of interactive mobile technology, vol. 13, no.
[22] Tan CL, Chiew KL, Wong K, “PhishWHO: phishing webpage detection via identity keywords extraction in addition towardtarget domain name finder”, Decision Support Systems, vol. 88, pp 18–27, 2016.
[23] Gull S in addition towardSA Parah, “Color image authentication using dual watermarks”, In: 2019 fifth international conference on image information processing (ICIIP), pp 240–245, 2019.

[24] Giri KJ, Bashir R, Bhat JI, “A discrete wavelet based watermarking scheme on the behalf of authentication of medical images”, International journal of E-Health medical Communication, pp 30–38, 2019.
[25] Gandotra E, Bansal D, Sofat S, “Malware threat assessment using fuzzy logic paradigm”, Cybernetics in addition towardsystems, pp. 29–48, 2016.
[26] S. Nisha and A. N. Madheswari, “Secured authentication on the behalf of internet voting in corporate companies to prevent phishing attacks,” vol. 22, no. 1, pp. 45–49, 2016.
[27] H. B. Kazemian and S. Ahmed, “Comparisons of machine learning techniques on the behalf of detecting malicious webpages,” Expert Syst. Appl., vol. 42, no. 3, pp. 1166–1177, 2015.
[28] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, “Design and Evaluation of a Real-Time URL Spam Filtering Service,” 2011 IEEE Symp. Secur. Priv., pp. 447–462, 2011.
[29] A.Firdaus, N. B. Anuar, M. F. A. Razak, I. A. T. Hashem, S. Bachok, and A. K. Sangaiah, “Root Exploit Detection and Features Optimization: Mobile Device and Blockchain Based Medical Data Management,” J. Med. Syst., vol. 42, no. 6, 2018.
[30] M. F. A. Razak, N. B. Anuar, F. Othman, A. Firdaus, F. Afifi, and R. Salleh, “Bio-inspired on the behalf of Features Optimization and Malware Detection,” Arab. J. Sci. Eng., 2018.
[31] J. A. Chaudhry, S. A. Chaudhry, and R. G. Rittenhouse, “Phishing attacks and defenses,” Int. J. Secur. its Appl., vol. 10, no. 1, pp. 247–256, 2016.
[32] R. Gowtham and I. Krishnamurthi, “A comprehensive and efficacious architecture on the behalf of detecting phishing webpages,” Comput. Secur., vol. 40, pp. 23–37, 2014.
[33] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, “Cantina+,” ACM Trans. Inf. Syst. Secur., vol. 14, no. 2, pp. 1–28, 2011.
[34] Abhilash, P. M., and D. Chakradhar. "Sustainability improvement of WEDM process by analysing and classifying wire rupture using kernel-based naive Bayes classifier." *Journal of the Brazilian Society of Mechanical Sciences and Engineering* 43.2 (2021): 1-9.
[35] Khorshid, Shler Farhad, and Adnan Mohsin Abdulazeez. "breast cancer diagnosis based on k-nearest neighbors: A review." *PalArch's Journal of Archaeology of Egypt/Egyptology* 18.4 (2021): 1927-1951.
[36] Charbuty, Bahzad, and Adnan Abdulazeez. "Classification based on decision tree algorithm on the behalf of machine learning." *Journal of Applied Science and Technology Trends* 2.01 (2021): 20-28.
[37] Zhang, W., Wu, C., Zhong, H., Li, Y., & Wang, L. "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization." *Geoscience Frontiers* 12.1 (2021): 469-477.
[38] Chandra, Mayank Arya, and S. S. Bedi. "Survey on SVM and their application in image classification." International Journal of Information Technology 13.5 (2021): 1-11.
[39] Mahmood Moghimi , Ali Yazdian Varjani. “New rule-based phishing detection method”, Expert Systems with Applications, Volume 53, 2016, Pages 231-242, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2016.01.028.