DOI: 10.xx/ijdsml.2023.xx

DETECTION OF MALIGN PAGES USING MACHINE LEARNING TECHNIQUES: DESCRIPTION AND ANALYSIS

Shubham Pandey and A.K. Malviya

Department of Computer Science and Engineering, Kamla Nehru Institute of Technology, India

Abstract

Malicious websites hosting drive-by download exploits have become more common as the internet has increased in popularity. To facilitate malware, attackers are increasingly redirecting users away from ordinary websites to attack pages. In the interim, the page's CSS settings are being tweaked to prevent modifications to the aesthetic impacts of specific pages. According to the researchers, positive examples include malicious drive-by-download web pages, while negative instances include benign web pages. Malign pages are discussed in the unsecured work and spam. To detect malign pages from pages of any website, first, manually crawl a list of safe URLs to collect the data. Furthermore, ensemble classifiers are used in work to train and test the model, i.e., Random Forest, Logistic Regression, Naive Bayes, and lastly, an ensemble of all the models is used. At last, a comparison is shown based on the accuracy score of different classifiers.

Keywords:

Alexa, Spam Scatter, Logistic Regression, Random Forest, Naive Bayes

1. INTRODUCTION

People's choices in everyday lives have been made immeasurably easier thanks to the Internet. Convenient web applications have attracted cybercriminals who unlawfully earn using phishing websites, spam adverts, and malware propagation [1] [2]. All of these criminal operations need ignorant individuals to browse the websites given by the adversary with the intention of attack, no matter how diverse their goals and methods may be. As a result, these pages are referred to as malicious sites [3] [4]. Static and dynamic characteristics are the most often utilized features for identifying various forms of malicious website pages, with each category having its features [5]. Web page static information is the primary source of static features. There are many, but extracting them is straightforward [6] [7].

Bypassing the firewall's detection and effectively implanting malicious malware on a user's computer without their awareness is possible with this client attack. It is dubbed a Drive-by-Download assault [8][9]. Research on extracting features for classification and identification from CSS files is relatively uncommon [10]. Overall relevant CSS properties have been the subject of theoretical investigation and experimental verification of their efficacy in categorizing malicious web pages [11] [12].

2. RELATED WORKS

There is a variety of work provided by many writers, which is listed below:

Bhargavi et al. [13] Used artificial intelligence classifiers such as support vector machine, arbitrary timberland, etc., to examine and detect harmful websites from a different order system. Innocent Bayes, calculated relapse, and several unusual URL classifiers are ready to identify harmful websites in light of deleted highlights. Ninety-five percent accuracy was achieved by the random forest classifier on the sample data it provided., compared to other AI classifiers. Pernicious pages, artificial intelligence, recognition, URL, and pernicious websites are all catchphrases.

Sachdeva and Machine [14] proposed a technique for identifying harmful web addresses. Before entering the search engine database, the created focused crawler verifies the URL. The URL feature set is created by identifying the parts of URLs that spammers can exploit to detect malicious URLs. As a result, a database that is not malignant is constructed. The search procedure triggers a query through this search engine database. The proposed crawler is faster at its search operations than the baseline one. Average search times for ten queries show that the focused crawler recommendation is 13% quicker than the baseline crawler. Quality parameters, like recall and precision, are calculated to check the performance of the suggested focused crawler, and they are determined to be 92.3 percent and 94.73 percent, respectively.

Rakesh et al. [15] described that implementing new security measures has addressed Concerns about online security. User information is collected by leading a user to an unauthorized URL, a sort of phishing. The attacker uses cookies or the user's current session to finish the attack by diverting the user's browser to the target site. Cross-Site Request Forgery (CSRF) vulnerabilities may be detected using a modified algorithm. The user will be warned if a User Interface code that utilizes the information gained in such an exploit is identified.

Gerža et al. [16] explained that the module, the fundamental building block of the Isis remote experiments available to users via websites, is examined in depth for various malicious threats. E-laboratories are sometimes used to refer to these types of labs. A state-of-the-art is described, along with some of its most fundamental characteristics and concepts. The following section analyzes software, hardware, and unique dangers for remote labs. The next step is identifying malicious attacks that may target the Measure server function.

3. TECHNIQUES USED

3.1 CASCADING STYLE SHEETS (CSS)

Cascade style sheets (CSS) are frequently used in malicious covert redirect attacks to obscure the attack page and reduce the likelihood of detection by the victim's browser. To avoid detection, a drive-by download assault is disguised by hiding most of its components. The investigation into the recognition of potentially hazardous web pages is still in its infancy [17].

- The total number of items with height and width values is 0 (zero). An attacker can alter an element's footprint by manipulating the width and height characteristics. Meaning the element's aesthetic impact will be changed. The element's value can be altered if it's set to zero.

- The number of show attribute instances with the value none. An element's display attribute specifies the type of box it will generate. No display of the element will occur if the value is set to none.
- The number of occasions where a value for the visibility attribute is hidden. The element's visibility property determines whether it is visible. The element is obscured if its visibility is set to hidden.
- It is possible to have an attribute with a value not displayed on the overflow page. An element's content may overflow the borders of its container, defined by the overflow attribute, which determines what happens.
- The material is reduced, and the rest is hidden when the setting is set to hidden.
- The value of the z-index attribute's value. It is determined by the z-index attribute how elements are stacked in the document.

3.2 ALEXA

The global website rankings of Alexa are made public. Daily, it collects more than 1000 gigabytes of data, generating millions of URL hyperlinks and categorizing and ranking each according to their significance. Currently, Alexa is the website with the most URLs and delivers the most comprehensive ranking information [15].

3.3 SPAM SCATTER

Researchers use Spam scatter, a spam-collection infrastructure, to extract URLs from spam messages. Spam scatter offers URLs for spamming malware that is false. All of the redirect URLs are retrieved by researchers [18].

3.4 ENSEMBLE CLASSIFIER

Superior performance can be attained through ensemble machine-learning techniques, which combine the insights of several learners. As a general rule, ensemble classifiers outperform individual classifiers. The ensemble classifiers combine the opinions of several students by using a subset of features chosen at random. Powerful classifiers are widely used because they are useful in a variety of contexts [19].

- **Logistic Regression**

Logistic regression is utilized to compute the possibility of binary predictor variables being forecasted when a dataset contains an independent variable that is predicted based on a binary dependent variable.

Although linear regression generates a curve, logistic regression produces a curve because it generates a straight line. This makes logistic regression equivalent to linear regression in this regard. When utilizing one or more predictor or independent variables, logistic regression generates logistic curves that illustrate the range of data from 0 to 1 [20]. Logistic curves can be created.

- **Random Forest**

A machine learning algorithm for artificial intelligence uses this technique as part of an ensemble methodology to increase its prosperity and precision in its machine learning algorithms. Furthermore, empirical research has found that it is also the most effective choice regarding forecast accuracy [21], and countless data already demonstrate its significance in selecting multiple options for each shrub.

The abundance of high-quality trees in this woodland inspired its name. When the data from these trees is pooled, we get the most precise predictions imaginable. Compared to a single decision tree, which only offers one conclusion and a handful of groups, a forest of decision trees guarantees a more accurate outcome by containing a bigger number of groups and possibilities. Choosing the proper feature between a random selection of traits also has the added benefit of injecting uncertainty into the model [22].

- **Naive Bayes**

Indeed, this is a Predictive theory-driven statistical procedure that selects the most probable rulings. Bayesian likelihood has been used to evaluate the possibility of unknown outcomes arising from well-established value systems. As a result, existing knowledge and logic can be implemented in unpredictable assertions and logic in inconsistent statements. According to the first method, there is a legally binding independence presumption surrounding attributes throughout the data set. A major assumption of the Naive Bayes technique is that the predictor variables have a high degree of (naive) statistical independence from one another. In addition to being a powerful classification method that is easily understandable and interpreted makes, it is a good choice even though the complexity of inputs is high. The mathematical formula for Bayes' theorem is as follows:

$$P(A\backslash B) =(P(B\backslash A)P(A))/(P(B)) \quad (1)$$

Assuming that both A and B are events and $P(B)$ is negative. A posteriori probabilities, or the likelihood of occurrence occurring in light of available data, are denoted by $P(A|B)$ [23].

- **K-Nearest Neighbors (KNN):**

The KNN classifier appears to be a simple, easy-to-implement supervised machine-learning technique that could tackle classification and regression issues. The term "K-nearest neighbor" can be shortened to "K-nn." This method is used to uncover fraudulent auto insurance claims and track down cardholders who have fallen behind on their payments. Following is a schematic representation of the K-NN network topology [24].

4. PROPOSED METHODOLOGY

This section outlines the phases of the approach employed in the intended work. The intended model works on the classification of benign and malicious web pages. The classifier (KNN, Random Forest, Logistic Regression, and Naive Bayes) is utilized individually, and then their ensembled are created to improve the accuracy of the intended work. Fig.1 depicts the block diagram of the suggested technique.

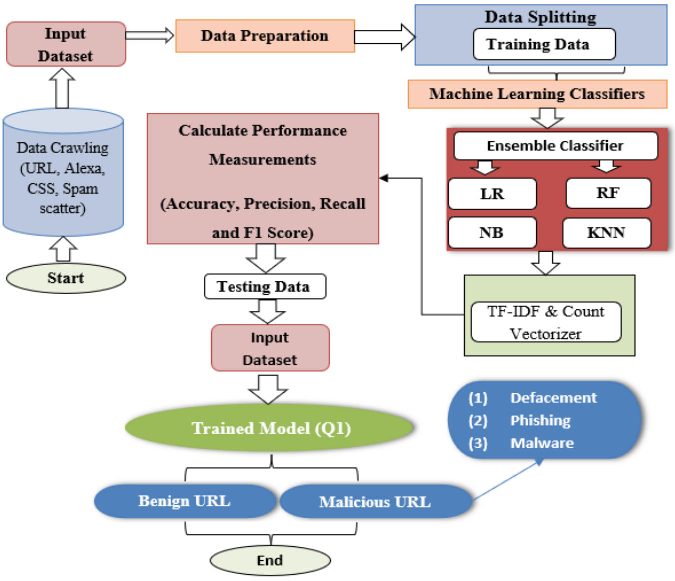


Fig.1. Block Diagram of Proposed Methodology (Training Model)

The stepwise explanation of the proposed methodology is defined in the below steps:

Step 1: First, gather webpages from the internet with the web crawler's help and then store all the pages' information in a file.

Step 2: In this step, classify the stored data into two URLs, i.e., benign URL and Malicious URL. Secured websites are shown by benign URLs (for example, Alexa- https://www.alexa.com), and malicious URLs show unsecured websites (for example, Spam scatter. A filter is applied to the Malicious URL to redirect the current URL.

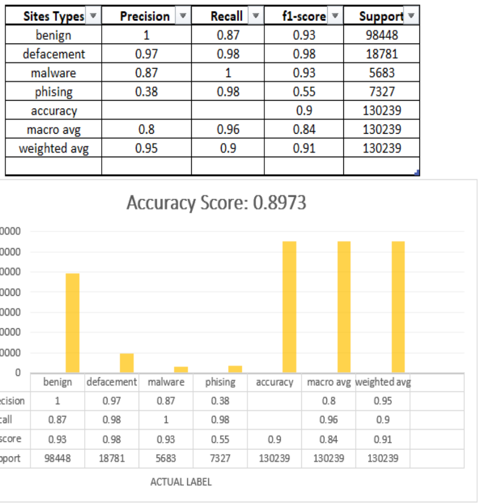
Step 3: After crawling and analyzing the original URL set of data, extract CSS links of all web pages and then extract new features recommended for further processing. Data collection (Total Data samples - 651192), (Training data samples - 520953), (Testing data samples - 130239).

Step 4: In this step, classification algorithms are used on features of both URL and CSS to train the model. Classification algorithms are used in work, i.e., Random Forest, Logistic Regression, and Naive Bayes.

5. IMPLEMENTATION RESULTS

The experiments in this part evaluate the efficacy of the proposed methodology. The proposed strategy analyzed the performance of the different classifiers based on the accuracy score. The fraction of all subjects correctly classified is referred to as accuracy.

The Fig.2 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use the accuracy, recall, and F1 score for Naive Bayesian with a TF-IDF accuracy value of 0.8973.

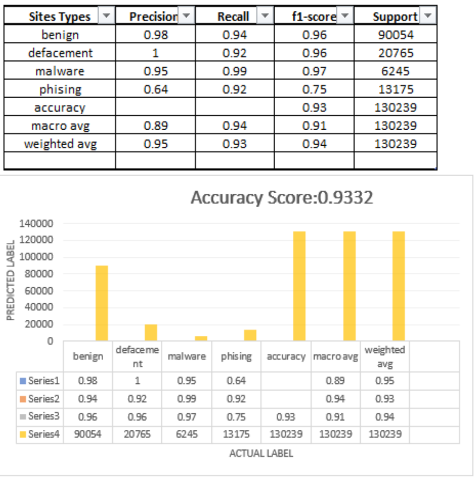


Sites Types	Precision	Recall	f1-score	Support
benign	1	0.87	0.93	98448
defacement	0.97	0.98	0.98	18781
malware	0.87	1	0.93	5683
phising	0.38	0.98	0.55	7327
accuracy			0.9	130239
macro avg	0.8	0.96	0.84	130239
weighted avg	0.95	0.9	0.91	130239

	benign	defacement	malware	phising	accuracy	macro avg	weighted avg
■ Precision	1	0.97	0.87	0.38		0.8	0.95
■ Recall	0.87	0.98	1	0.98		0.96	0.9
■ f1-score	0.93	0.98	0.93	0.55	0.9	0.84	0.91
■ Support	98448	18781	5683	7327	130239	130239	130239

Fig.2. Naive Bayesian with TF-IDF

The Fig.3 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and F1 scores for Logistic Regression with a Count Vectorizer accuracy value of 0.9198.

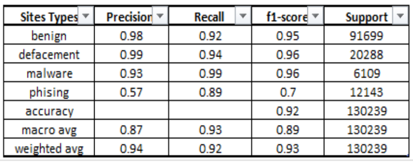


Sites Types	Precision	Recall	f1-score	Support
benign	0.98	0.94	0.96	90054
defacement	1	0.92	0.96	20765
malware	0.95	0.99	0.97	6245
phising	0.64	0.92	0.75	13175
accuracy			0.93	130239
macro avg	0.89	0.94	0.91	130239
weighted avg	0.95	0.93	0.94	130239

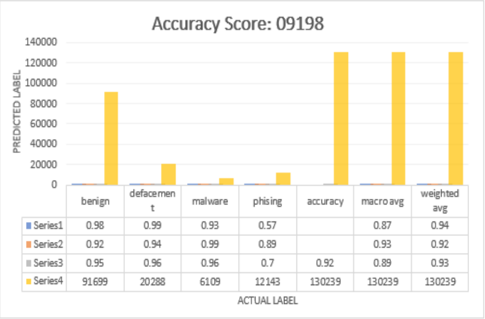
	benign	defacement	malware	phising	accuracy	macro avg	weighted avg
■ Series1	0.98	1	0.95	0.64		0.89	0.95
■ Series2	0.94	0.92	0.99	0.92		0.94	0.93
■ Series3	0.96	0.96	0.97	0.75	0.93	0.91	0.94
■ Series4	90054	20765	6245	13175	130239	130239	130239

Fig.3. Logistic Regression with Count Vectorizer

The Fig.4 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and F1 scores for the Logistic Regression with a TF-IDF accuracy value of 0.9332.



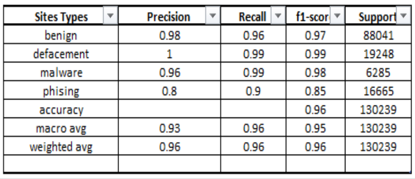
Sites Types	Precision	Recall	f1-score	Support
benign	0.98	0.92	0.95	91699
defacement	0.99	0.94	0.96	20288
malware	0.93	0.99	0.96	6109
phising	0.57	0.89	0.7	12143
accuracy			0.92	130239
macro avg	0.87	0.93	0.89	130239
weighted avg	0.94	0.92	0.93	130239



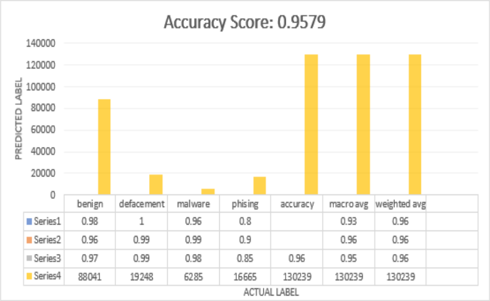
	benign	defacement	malware	phising	accuracy	macro avg	weighted avg
Series1	0.98	0.99	0.93	0.57		0.87	0.94
Series2	0.92	0.94	0.99	0.89		0.93	0.92
Series3	0.95	0.96	0.96	0.7	0.92	0.89	0.93
Series4	91699	20288	6109	12143	130239	130239	130239

Fig.4. Logistic Regression with TF-IDF

The Fig.5 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and F1 scores for Naive Bayesian with a Count Vectorizer accuracy value of 0.9579.



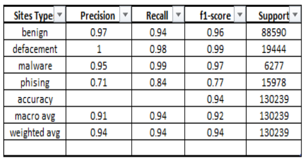
Sites Types	Precision	Recall	f1-scor	Support
benign	0.98	0.96	0.97	88041
defacement	1	0.99	0.99	19248
malware	0.96	0.99	0.98	6285
phising	0.8	0.9	0.85	16665
accuracy			0.96	130239
macro avg	0.93	0.96	0.95	130239
weighted avg	0.96	0.96	0.96	130239



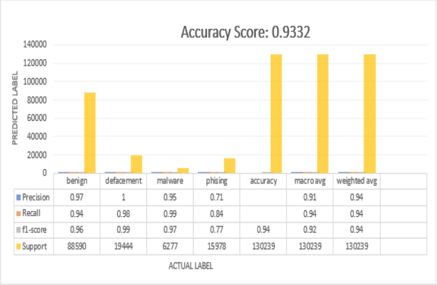
	benign	defacement	malware	phising	accuracy	macro avg	weighted avg
Series1	0.98	1	0.96	0.8		0.93	0.96
Series2	0.96	0.99	0.99	0.9		0.96	0.96
Series3	0.97	0.99	0.98	0.85	0.96	0.95	0.96
Series4	88041	19248	6285	16665	130239	130239	130239

Fig.5. Naive Bayesian with Count Vectorizer

The Fig.6 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and F1 scores for KNN with a Count Vectorizer accuracy value of 0.9579.



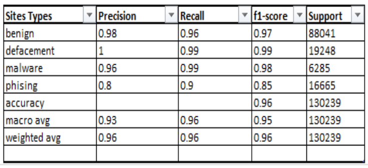
Sites Type	Precision	Recall	f1-score	Support
benign	0.97	0.94	0.96	88590
defacement	1	0.98	0.99	19444
malware	0.95	0.99	0.97	6277
phising	0.71	0.84	0.77	15978
accuracy			0.94	130239
macro avg	0.91	0.94	0.92	130239
weighted avg	0.94	0.94	0.94	130239



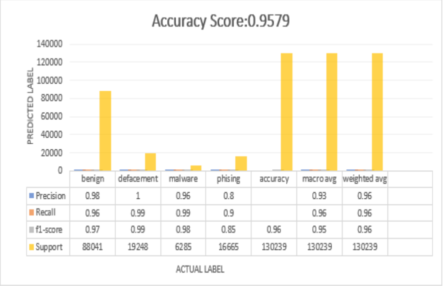
	benign	defacement	malware	phising	accuracy	macro avg	weighted avg
Precision	0.97	1	0.95	0.71		0.91	0.94
Recall	0.94	0.98	0.99	0.84		0.94	0.94
f1-score	0.96	0.99	0.97	0.77	0.94	0.92	0.94
Support	88590	19444	6277	15978	130239	130239	130239

Fig.6. KNN with Count Vectorizer

The Fig.7 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and the F1 score for the Random Forest- TFIDF accuracy value is 0.9332.



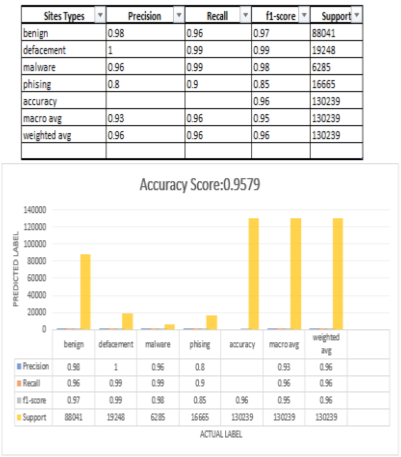
Sites Types	Precision	Recall	f1-score	Support
benign	0.98	0.96	0.97	88041
defacement	1	0.99	0.99	19248
malware	0.96	0.99	0.98	6285
phising	0.8	0.9	0.85	16665
accuracy			0.96	130239
macro avg	0.93	0.96	0.95	130239
weighted avg	0.96	0.96	0.96	130239



	benign	defacement	malware	phising	accuracy	macro avg	weighted avg
Precision	0.98	1	0.96	0.8		0.93	0.96
Recall	0.96	0.99	0.99	0.9		0.96	0.96
f1-score	0.97	0.99	0.98	0.85	0.96	0.95	0.96
Support	88041	19248	6285	16665	130239	130239	130239

Fig.7. Accuracy of Random Forest -TFIDF

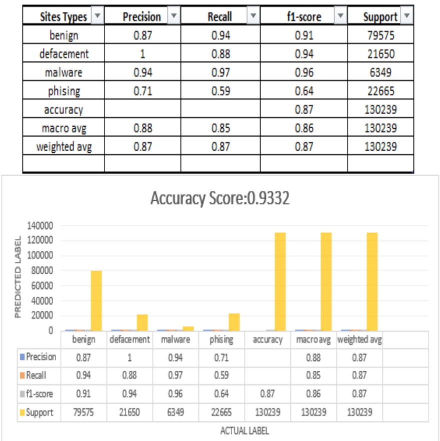
The Fig.8 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and F1 score for Random Forest with a Count Vectorizer accuracy value is 0.9579.



Sites Types	Precision	Recall	f1-score	Support
benign	0.98	0.96	0.97	88041
defacement	1	0.99	0.99	19248
malware	0.96	0.99	0.98	6285
phising	0.8	0.9	0.85	16665
accuracy			0.96	130239
macro avg	0.93	0.96	0.95	130239
weighted avg	0.96	0.96	0.96	130239

Fig.8. Accuracy of Random Forest with Count Vectorizer

The Fig.9 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and the F1 score for the KNN-TFIDF accuracy value is 0.9332.



Sites Types	Precision	Recall	f1-score	Support
benign	0.87	0.94	0.91	79575
defacement	1	0.88	0.94	21650
malware	0.94	0.97	0.96	6349
phising	0.71	0.59	0.64	22665
accuracy			0.87	130239
macro avg	0.88	0.85	0.86	130239
weighted avg	0.87	0.87	0.87	130239

Fig.9. Accuracy of KNN-TFIDF

The Fig.10 below shows the accuracy performance indicators. The precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and the F1 score for the accuracy of ensemble classifier value is 0.96.

	Precision	**Recall**	**F1-score**	**Support**
Benign	0.98	0.96	0.97	88041
Defacement	0.93	0.98	0.95	19248
Malware	0.96	0.91	0.98	6285
Phishing	0.92	0.90	0.85	16665
Accuracy	-	-	0.96	130239
Macro Avg	0.93	0.96	0.95	130239
Weighted Avg	0.96	0.96	0.96	130239

Fig.10. Accuracy of the ensemble classifier

5.1 COMPARATIVE ANALYSIS

This comparative study's primary premise is to compare the accuracy of research on the statistical and data science challenges and opportunities using machine learning classifiers, which offers a prediction model based on machine learning to determine the data analytics. Comparative analysis of studies is performed to identify the performance of machine learning classifiers by assessing their accuracy comparisons among classifiers such as NB-TFIDF, LR with CV, LR- TFIDF, NB with CV, KNN-CV, RF-TFIDF, RF with CV, and

KNN-TFIDF are performed to gain accuracy. NB-TFIDF and NB with CV and RF with CV classifier show the highest and lowest accuracy among all classifiers. NB-TFIDF has a minimum accuracy rate of 0.8973, and RF with CV has a maximum accuracy rate of 0.9579. The accuracy performance indicators, the precision, confusion matrix, recall, and F1 score offer greater visibility into the forecast. Information retrieval, word segmentation, named object identification, and many other applications use accuracy, recall, and the F1 score for the accuracy of ensemble classifier value is 0.96.

The Fig.11 shows the comparative analysis of the accuracy of different classifiers below.

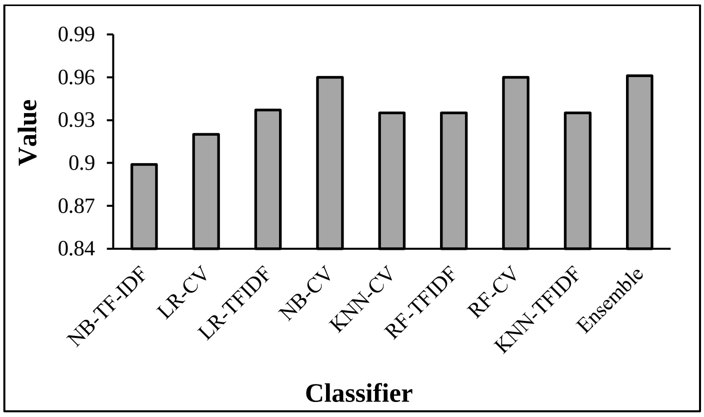


Fig.11. Graphical representation of comparative analysis

6. CONCLUSIONS

Static and dynamic characteristics are the most often utilized features for identifying various forms of malicious website pages, with each category having its own set of features. Various measurement criteria, such as performance, scalability, throughput, and others, are studied in the literature to compare current static and dynamic load-balancing methods. The accuracy score of various classifiers are compared and observed the results of the ensembled classifiers outperform better in comparison to the individual classifier.

REFERENCES

[1] A. Rasool, C. Bunterngchit, L. Tiejian, M.R. Islam, Q. Qu and Q. Jiang, "Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis", *International Journal of Environmental Research and Public Health*, Vol. 19, No. 6, pp. 3211-3221, 2022.

[2] N. Spirin and J. Han, “Survey on Web Spam Detection: Principles and Algorithms”, *ACM SIGKDD Explorations Newsletter*, Vol. 13, No. 2, pp. 1-12, 2012.

[3] Z. Li, S. Alrwais, Y. Xie, F. Yu and X.F. Wang, “Finding the Linchpins of the Dark Web: A Study on Topologically Dedicated Hosts on Malicious Web Infrastructures”, *Proceedings of ACM Symposium on Security and Privacy*, pp. 12-18, 2013.

[4] J.W. Zhuge, Y. Tang, X.H. Han and H.X. Duan, “Honeypot Technology Research and Application”, *Ruanjian Xuebao/Journal of Software*, Vol. 24, No. 4, pp. 825-842, 2013.

[5] H.L. Zhang, W. Zou and X.H. Han, “Drive-by-Download Mechanisms and Defenses”, Ruanjian Xuebao/Journal of Software, Vol. 25, No, 2, pp. 768-779, 2013.

[6] P. Zhao and S.C.H. Hoi, “Cost-Sensitive Online Active Learning with Application to Malicious URL Detection”, *Proceedings of ACM International Conference on Knowledge Discovery and Data Mining*, pp. 1-10, 2013.

[7] H.Z. Sha, Q.Y. Liu, T.W. Liu, Z. Zhou, L. Guo and B.X. Fang, “Survey on Malicious Webpage Detection Research”, *International Journal of u- and e- Service, Science and Technology*, Vol. 8, No. 5, pp. 195-206, 2015.

[8] S.M. Nair, “Detecting Malicious URL using Machine Learning: A Survey”, *International Journal for Research in Applied Science and Engineering Technology*, Vol. 25, No. 2, pp. 1-12, 2020.

[9] M. Krbecek, F. Schauer and I. Zelinka, “Possible Utilization of the Artificial Intelligence Elements in the Creation of Remote Experiments”, *International Journal of Online and Biomedical Engineering*, Vol. 10, No. 1, pp. 1-15, 2014.

[10] G. Stringhini, C. Kruegel and G. Vigna, “Shady Paths: Leveraging Surfing Crowds to Detect Malicious Web Pages”, *Proceedings of ACM International Conference on Computer and Networking*, pp. 133-144, 2013.

[11] T. Shibahara, T. Yagi, M. Akiyama, Y. Takata and T. Yada, “Detecting Malicious Web Pages based on Structural Similarity of Redirection Chains”, *Proceedings of ACM International Conference on Computer and Communication Security*, pp. 1-7, 2015.

[12] S. Kumar, P. Eugster and S. Santini, “Software-Based Remote Network Attestation”, *IEEE Transactions on Dependable and Secure Computing*, Vol. 19, No. 5, pp. 2920-2933, 2022.

[13] S.R.M. Bhargavi and Mahammad Shabana, “Designing A New Classification System to Analyze and Detect the Malicious Web Pages using Machine Learning Classifiers”, *Nveo-Natural Volatiles and Essential Oils Journal*, Vol. 46, No. 2, pp. 1295-1301, 2021.

[14] R. Sachdeva and S. Gupta, “A Novel Focused Crawler with Anti-spamming Approach and Fast Query Retrieval”, *Proceedings of International Conference on Inventive Computation and Information Technologies*, pp. 315-331, 2021.

[15] A.K. Rakesh, R.S. Muthuraijkumar, L. Sairamesh and M. Vijayalakmi, “Detection of URL based Attacks using Reduced Feature Set and Modified C4. 5 Algorithms”, *Advances in Natural and Applied Sciences*, Vol. 9, No. 5, pp. 304-311, 2015.

[16] M. Gerza, F. Schauer and R. Jasek, “Security of ISES measure Server Module for Remote Experiments against Malign Attacks”, *International Journal of Online and Biomedical Engineering*, Vol. 10, No. 2, pp. 41-56, 2014.

[17] B. Chen and Y. Shi, “Malicious Hidden Redirect Attack Web Page Detection based on CSS Features”, *Proceedings of IEEE International Conference on Computer and Communications*, pp. 1-6, 2018.

[18] Y. Sonmez, T. Tuncer, H. Gokal and E. Avci, “Phishing Web Sites features Classification based on Extreme Learning Machine”, *Proceedings of IEEE International Conference on Digital Forensic and* Security, pp. 22-25, 2018.

[19] A. Subasi, “Sensor based Human Activity Recognition using Adaboost Ensemble Classifier”, *Procedia Computer Science*, Vol. 140, pp. 104-111, 2018.

[20] D.W. Hosmer, S. Lemeshow and R.X. Sturdivant, “*Assessing the Fit of Regression Model*”, Wiley, 2013.

[21] N. Malini and M. Pushpa, “Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection”, Proceedings of IEEE International Conference on Advances in Electrical, Electronics, Information, Communication and Bio Informatics, pp. 1-7, 2017.

[22] M.S. Hossain and M. Shamsul Arefin, “Development of an Intelligent Job Recommender System for Freelancers using Client’s Feedback Classification and Association Rule Mining Techniques”, *Journal of Software*, Vol. 14, No. 7, pp. 312-339, 2019.

[23] S. Patil, V. Nemade and P.K. Soni, “Predictive Modelling for Credit Card Fraud Detection using Data Analytics”, *Procedia Computer Science*, Vol. 132, pp. 385-395, 2018.

[24] F. Schauer, M. Krbecek and M. Ozvoldova, “Controlling Programs for Remote Experiments by Easy Remote ISES (ER-ISES)”, *Proceedings of IEEE International Conference on Remote Engineering and Virtual Instrumentation*, pp. 1-5, 2013.