ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

An Effective monitoring of cyberattacks and acknowledging to user using ML

Mr. B. Kishore Kumar,[1]T Amulya[2], Ch V S Vyshnavi[3], M. Manogna[4], L Meghana[5]

[1]Asst. Professor, Department of Computer Science and engineering

[2,3,4,5]Student, Department of Computer Science and engineering

[1,2,3,4,5]QIS College of Engineering and Technology

Abstract-In order to steal critical information from internet consumers and businesses, Cyber-Attackers entrap them. Attackers get access to sensitive data such as usernames, passwords, credit card numbers, and bank account numbers on corporate systems. Cyber-attacks include Phishing Attacks, when attackers trick internet users into believing their websites are real and stealing their personal information. In a malware attack, an attacker sneaks a malicious programme into a corporate server or an internet user's computer without the victim's knowledge and proceeds to steal all the data on the server or computer in question. Malware attacks are becoming more common. Intrusion refers to an attack on a network in which the attacker will steal all of the network's resources. Blacklist or whitelist, heuristic and visual similarity-based techniques are just a few of the cyber attack solutions that have been presented, but they all have their drawbacks. Intrusion Detection Systems (IDS) have been proposed as well as signature and anomaly-based methods. Because to a lack of cyber knowledge and unprotected HTTP sites. The number of cyberattacks is steadily rising. Anti-phishing approaches have a number of drawbacks, and a new categorization model is developed to address these shortcomings. In the proposed model, several Machine Learning techniques are combined to create an ensemble approach.

Keywords— Cyberattacks, Intrusion Detection System, HTTP Sites,Attackers.

I. INTRODUCTION

Most individuals in this digital environment interact with each other using a computer or a digital gadget that is linked to the Internet.The ease, comfort, and help offered by online banking, shopping, and other services has led to an increase in the number of individuals who use them.Online service websites may be compromised by an attacker who sees it as a chance to make money or achieve notoriety.Some of the techniques to take sensitive information from consumers include phishing, malware, and intrusion. An internet user is tricked into divulging personal information by a website that seems to be from a trustworthy source.Five machine learning methods may be used to identify cyber assaults.Feature selection, Ensemble technique, and Machine learning algorithms are all handled by the proposed system. Our study's goal was to discover a fast algorithm that also had the best level of accuracy possible.

II. RELATEDWORKS

Some existing techniques have been applied to detect cyber attacks.

- Heuristic-based techniques
- Visual Similarity-based Approach
- Machine learning-based techniques
- Signature-based-detection
- Anomaly-based-method

Disadvantages of Existing System

- It has less accuracy compared to list-based techniques as there is no guarantee of existence of these features in all phishing websites.
- An attacker can bypass the heuristic features once he knows the algorithm or features used in detecting phishing sites thereby reaches his goal of stealing sensitive information.
- mage comparison of suspicious website with entire legitimate database store takes more time complexity.
- More space to store legitimate image database.
- Web page with animated website compared with phishing website leads to the low percentage of similarity that leads to high false negative rate. This technique fails, when the background of web page is slightly changed without deviating f rom visual appearance of legitimate site.
- These techniques won’t work efficiently on the large sets of data.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

III. PROPOSED SYSTEM ARCHITECTURE

Add new features with machine learning algorithms to reduce the false positives in detecting Cyber attacks.Made an attempt to identify the best machine learning algorithm to detect Cyber attacks with high accuracy than the existing techniques. Used five machine learning algorithms (Logistic regression (LR), KNN, Random Forest (RF), support vector machine (SVM) and Decision Tree) to classify the websites as legitimate and phishing as shown in fig.1.

Benefits of Proposed System

The proposed system consists of dataset which contains detailed information about the different cyber attacks which would be helpful for detection and prediction of the attacks.As the proposed system contains different Machine Learning algorithms the result is taken based on the highest accuracy.For each algorithm the metric values are generated which can be considered as predicted results.Ensemble Approach Called Voting Classifier is used to compare the models to produce high accuracy and less errors.

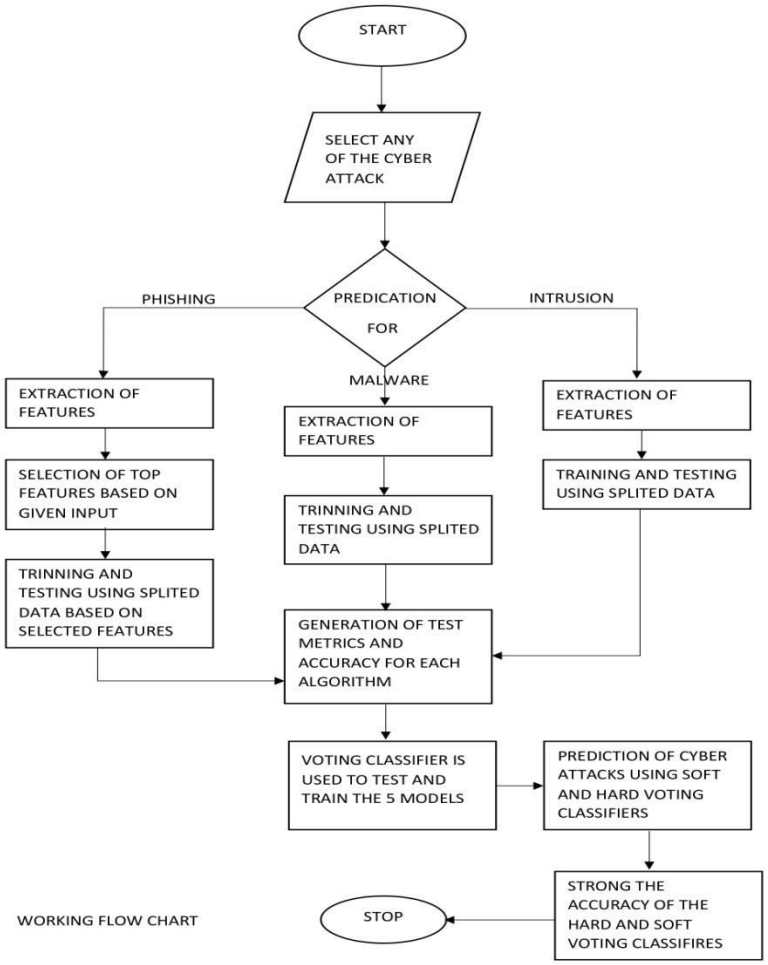

Fig.1 Flow chart

The System follows the following steps for implementation:

Step-1: Select any of the cyber-attacks and the system starts the process.

Step-2: if the selected cyber-attack is phishing then user have to specify how many features to be selected.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Step-3: The system will train and test different types of machine learning algorithms using the dataset for the selected cyber-attack and extracts accuracy.
Step-4: Now voting classifier is applied using the five models and voting classifier will compare all the models and produces the accuracy.

DATASET PRE-PROCESSING
In any Machine Learning process, Data Pre-processing is that step in which the data gets transformed, or Encoded, to bring it to such a state that now the machine can easily parse it. In other words, the features of the data can now be easily interpreted by the algorithm. A dataset can be viewed as a collection of data objects, which are often also called as a records, points, vectors, patterns, events, cases, samples, observations, or entities. Data objects are described by a number of features, that capture the basic characteristics of an object, such as the mass of a physical object or the time at which an event occurred, etc.. Features are often called as variables, characteristics, fields, attributes, or dimensions. There are many features available in the given dataset and we are going to select only the top most features in the given dataset by using Feature Selection for Phishing dataset. Malware detection and Intrusion detection are huge datasets, hence data processing will be slow. So feature scaling is applied here to covert the values in dataset in such a way that the values are in between 0 to 1 as shown in Fig. 2.

PHISHING DETECTION
After data processing using feature selection process the selected top most features are now tested and trained using 5 different types of algorithms in such a way that 75% of dataset is used for training and 25% is used for testing. Accuracy and the metric values for the algorithms are generated and stored.

INTRUSION & MALWARE DETECTION Intrusion and Malware detection datasets are huge so the dataset is converted in such a way that the values will come in the range of 0-1. This is done by using Feature Scaling process. Once the values are converted then testing and training process will start using 75% of dataset is for training and 25% is for testing. Accuracy and metric values f or all the five algorithms are generated and stored as shown in fig.3.

VOTING CLASSIFIER In this module we test and train the dataset using five different models and compare them using the hard and soft voting in such a way that hard voting works based on the majority and soft voting works based on the probability as shown in fig.4.

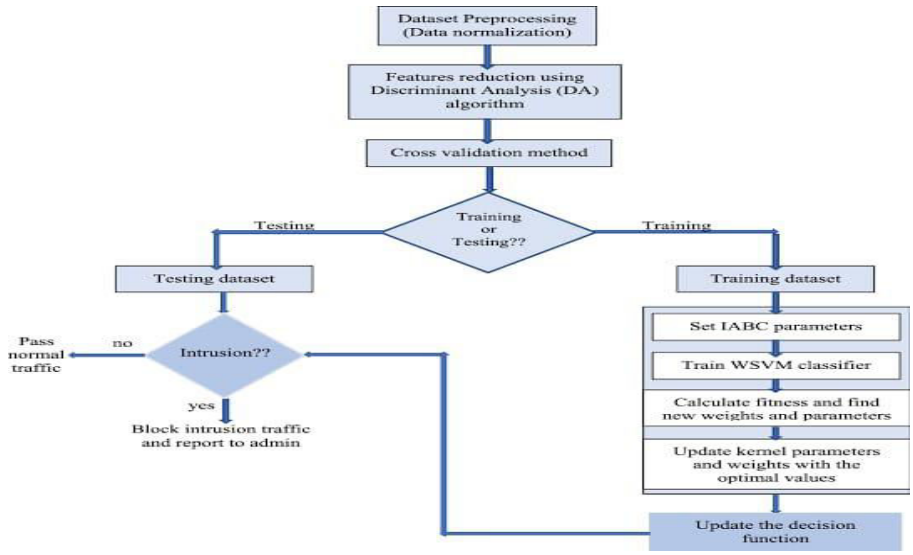

Fig.2 Intrusion Methodology

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

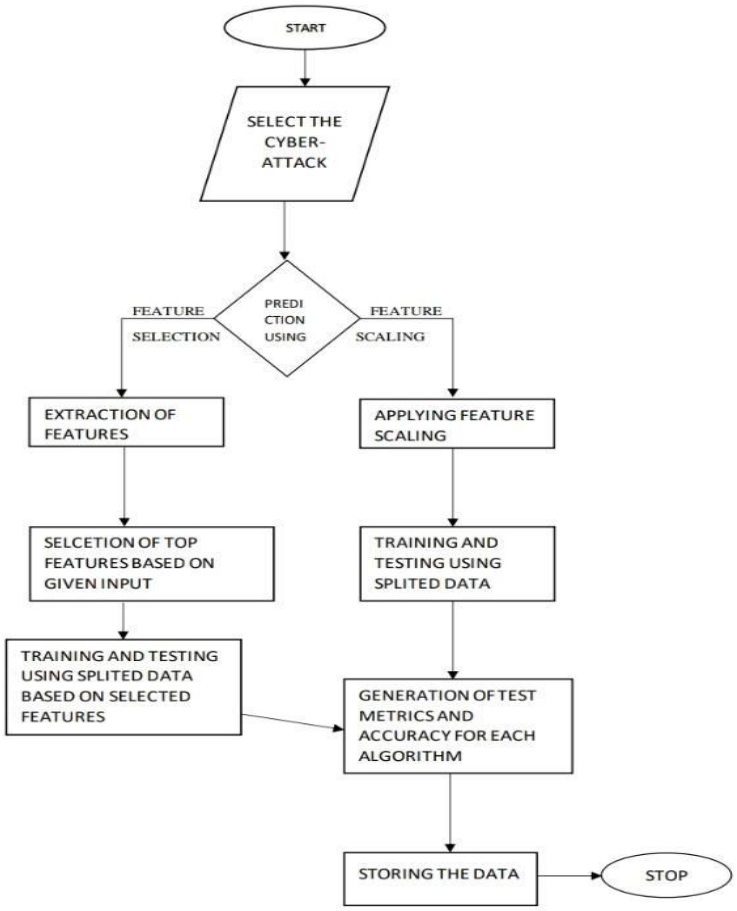

Fig. 3 Intrusion and Malware flow chart

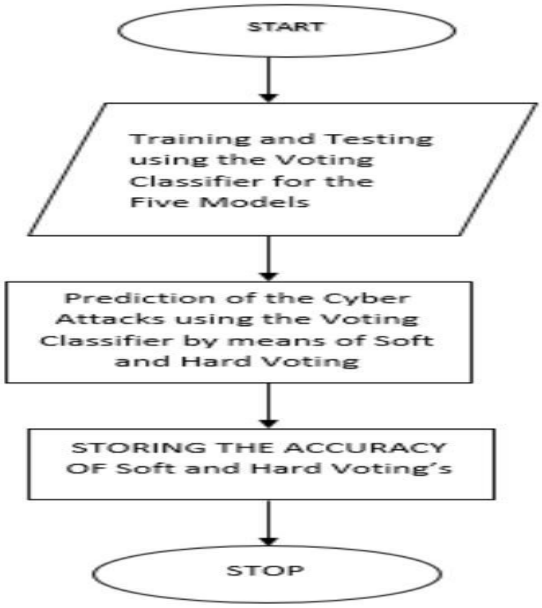

Fig.4 Voting Classifier

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

.

IV. RESULTS AND DISCUSSION

The results obtained after executing the implementation code is shown from Fig.5 to Fig.11.

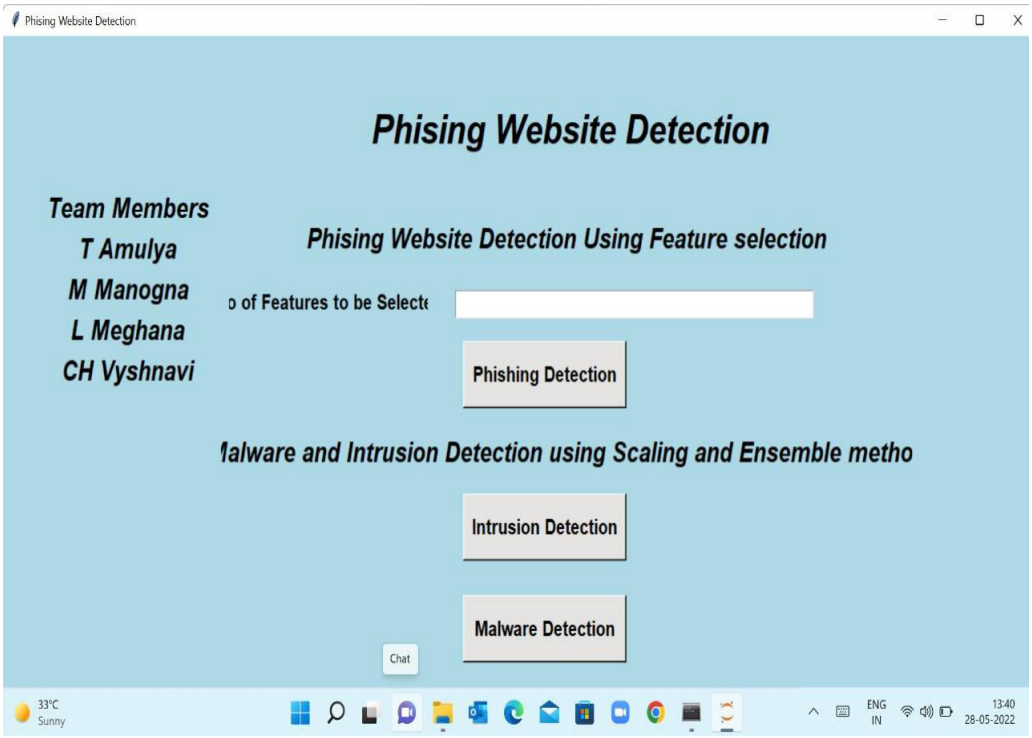

Fig.5 Home Page

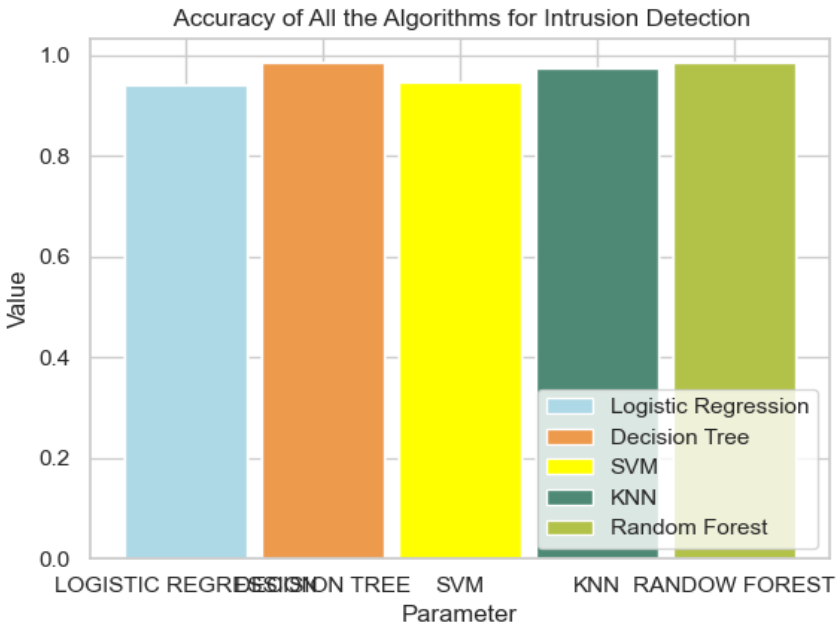

Fig.6 Accuracy average for Intrusion detection

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

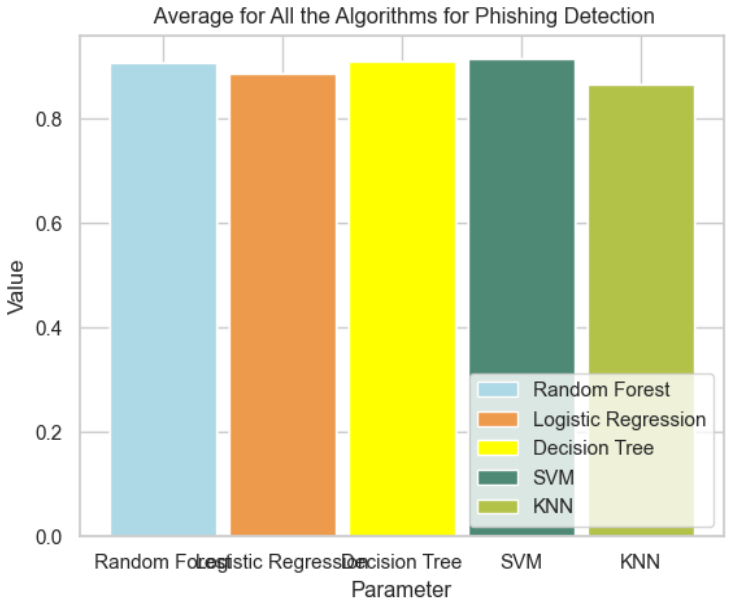

Fig.7 Accuracy average for Phishing detection

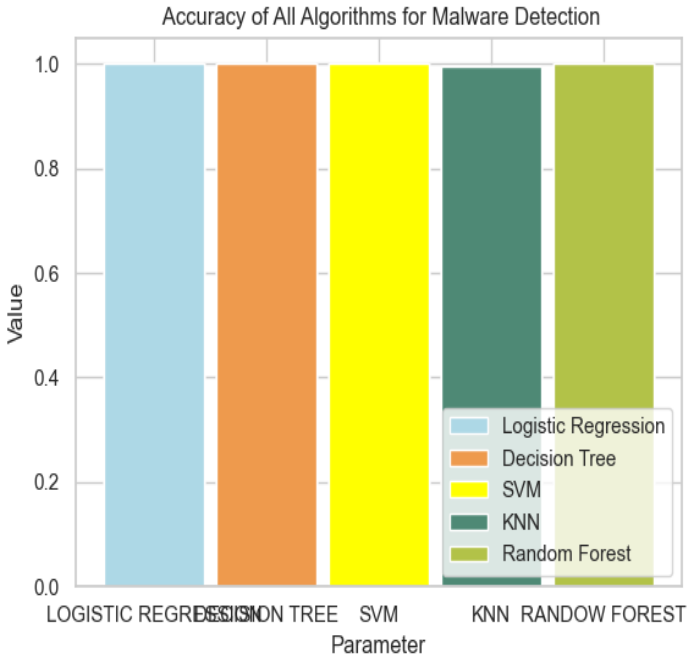

Fig.8 Accuracy average for Malware detection

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

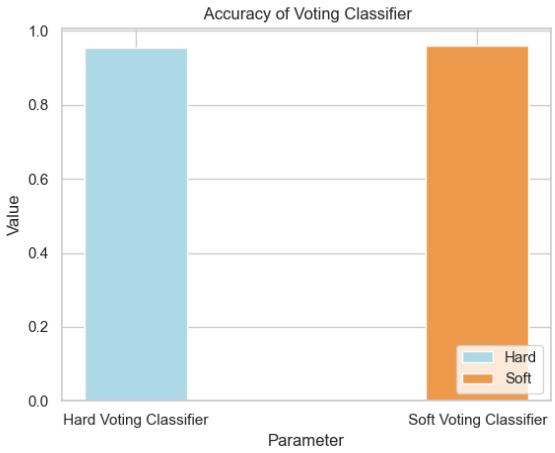

Fig.9 Accuracy of voting classifier for Malware detection

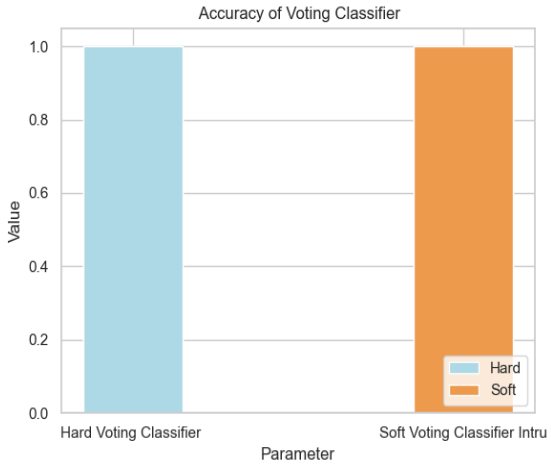

Fig.10 Accuracy of voting classifier for Phishing detection

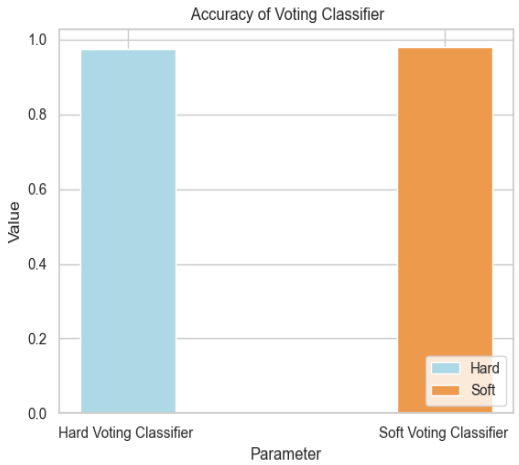

Fig.11 Accuracy of voting classifier for Intrusion detection

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

V. FUTURE SCOPE AND CONCLUSION

Phishing,Malware and Intrusion are some cyber-crime procedure utilizing online services to grab individual sensitive data. The System is done using website phishing, Malware detection, KDD dataset utilizing different classification algorithms to provide the algorithm with best accuracy. This system is designed in such a way that it overcomes the disadvantages of existing systems by using website Phishing, Malware detection and KDD dataset utilizing different Machine Learning algorithms. In future we would like to design and host a website using Unsupervised MachineLearning algorithms as it can recognise any type of Cyber attacks efficiently than SupervisedMachine Learning algorithms.

REFERENCES

[1] Somaiya Vidyaviharet.all,Phishing Website Detection using Machine Learning,IJCA, Volume 181-No.23,October 2018.

[2] Sadeh N, Tomasic A, Fette I. Learning to detect phishing emails. Proceedings of the16th international conference on World Wide Web. 2007: p. 649-656

[3] AndrBergholz, Gerhard Paa, Frank Reichartz, Siehyun Strobel, and SchloBirlinghoven. Improved phishing detection usingmodel-based features. In FifthConference on Email and Anti-Spam, CEAS, 2008

[4] UCI Machine Learning Repository." http://archive.ics.uci.edu/ml/, 2012.

[5] H. A. Chipman, E. I. George, and R. E. McCulloch.BART: Bayesian AdditiveRegression Trees. Journal of the Royal Statistical Society, 2006. Ser.B,Revised.

[6] S. Nawafleh, W. Hadi (2012). Multi-class associative classification to predictingphishing websites. International Journal of Academic Research Part A; 2012;4(6),302-306J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2.Oxford: Clarendon, 1892, pp.68–73.

[7] P. Tiwari, R. Singh International Journal of Engineering Research & Technology(IJERT) ISSN: 2278-0181Vol. 4 Issue 12, December-2015.

[8] J. P. Marques de Sa. Pattern Recognition: Concepts,Methods and Applications.Springer, 2001.

[9] D. Michie, D. J. Spiegelhalter, and C. C. Taylor. Machine Learning, Neural andStatisticalClassi_cation. Ellis Horwood, 1994.

[10] L. Breiman. Random forests. Machine Learning,45(1):5{32, October2001

[11] Mrs. Sayantani Ghosh, Mr. Sudipta Roy, Prof. Samir K.Bandyopadhyay, "A tutorial review on Text Mining Algorithms".

Authors

Mr. B. KishoreKumar currently working as Assistant professor (M Tech) of Computer Science & Engineering in Qis college of Engineering and Technology (Autonomous& NAAC 'A 'Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

T. Amulya pursuing B.Tech in the department of Computer Science & Engineering from Qis college of Engineering and Technology (Autonomous &NAAC'A'Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022 respectively.

CH. Venkata Siva Vyshnavi pursuing B.Tech in the department of Computer Science & Engineering from Qis college of Engineering and Technology (Autonomous &NAAC'A'Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022respectively.

M.Manogna pursuing B.Tech in the department of Computer Science & Engineering from Qis college of Engineering and Technology (Autonomous &NAAC'A'Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022respectively.

L. Meghana pursuing B.Tech in the department of Computer Science & Engineering from Qis college of Engineering and Technology (Autonomous &NAAC'A'Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist. Affiliated to Jawaharlal Nehru Technological University, Kakinada in 2018-2022respectively.