HIGHER ACCURACY OF DETECTING PHISHING WEBSITES USING DECISION TREE ALGORITHM COMPARING WITH LOGISTIC REGRESSION ALGORITHM

Panga Satheesh[1], K. Malathi[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The main objective of the research study is to improve the accuracy for Detecting phishing websites using the Decision Tree Algorithm against Logistic Regression machine learning algorithm. **Materials and Methods**: The study used 20 samples with two groups of algorithms with the G-power value of 80% percent and the phishing attack data were collected from various web sources with recent study findings and threshold 0.05 and confidence interval 96.49% with mean and standard deviation. To predict the phishing attacks by improving the Logistic Regression Algorithm has found 92.65% of accuracy, therefore this study needs to find the better accuracy for Phishing Attack prediction with the Decision Tree Algorithm machine learning algorithm. **Result**: This research study found 96.59% of accuracy for Detecting phishing websites using the Decision Tree algorithm with a significant value of two tailed tests is 0.002 ($p<0.05$) with 96.49% confidence interval. **Conclusion**: This study concludes that the Decision Tree algorithm on Innovative phishing website Detection is significantly better than the Logistic Regression algorithm.

Keywords: Innovative Phishing Website Detection, Machine Learning, Decision Tree Algorithm, Logistic Regression Algorithm, Statistical Analysis, Supervised Learning

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode:602105.

[2*] Associate Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

1. Introduction

In daily life, all carry out most of my work on digital platforms. The research study is to improve the accuracy for Detecting phishing websites using the Decision Tree Algorithm against Logistic Regression machine learning algorithm (Gu 2021). Phishing websites are used as a technique to deceive users and trick them into submitting sensitive information such as their authentication details, which can include username, password as well as unique codes associated with multi factor authentication (Pascariu and Bacivarov 2021). It is a tool used by cyber criminals to steal personal information from the user. Cyber-attacks using malicious URLs have emerged as an addressing the issue of social problems (Jang, Song, and Kim 2022). The use of online virtual entertainment and E-trade sites among individuals increments step by step. It causes millions of transactions to occur online (Saravanan and Subramanian 2020). The criminals will create a fake website that looks the same as the real websites (Zaini et al. 2020) .With the widespread usage of the Internet for online banking and trade, phishing attacks and forms of identity theft-based scams are becoming extremely popular among the hacker communities (Razaque et al. 2020). The user will get fraud by purposefully entering their confidential information such as password, bank details and account credentials into the fake websites (Zaini et al. 2020). Phishing is a social engineering attack that aims at exploiting the weakness found in the system at the user's end (Patil et al. 2018) . There are a lot of ways in which attackers lure the victims into clicking the malicious links which may result in the victim losing their personal data and even money in some cases(Ghimire et al. 2021). Social engineering attack is a means of influencing the victim who has fewer awareness about these kinds of attack (Nadar et al. 2021) .
Phishers use multiple methods, including email, Uniform Resource Locators (URL), instant messages, forum postings, telephone calls, and text messages to steal user information (Dutta 2021).

Related work of machine learning and Supervised Learning have been applied with reference to Anti-phishing ways involved in educating web users and technical defense.Found that 4,950 papers on google scholar related to this title and most cited articles are Detection of phishing websites using an efficient feature-based machine learning framework (Rao and Pais 2019), (Jain and Gupta 2019), (Itoo, Meenakshi, and Singh 2021), (Lokesh and BoreGowda 2021). The most cited website is (Rao and Pais 2019). They have used a publicly available dataset which has 11055 values of the dataset of website phishing. This research related work was presented and published in more than 80 indexed journals. The best study of Detecting phishing websites (Lokesh and BoreGowda 2021).

Our institution is passionate about high quality evidence based research and has excelled in various domains (Vickram et al. 2022; Bharathiraja et al. 2022; Kale et al. 2022; Sumathy et al. 2022; Thanigaivel et al. 2022; Ram et al. 2022; Jothi et al. 2022; Anupong et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Palanisamy et al. 2022). The drawbacks of this detecting phishing website is If the Internet connection fails, this system won't work and Loss of Customers. Loss of Data and all websites related data will be stored in one place. There are more relative articles with an accuracy score from the DATA classifier for innovation of Anti-Phishing to predict website phishing attacks. Therefore the aim of this study is to increase the accuracy of Innovative phishing website Detection vulnerability and improve the prediction model using the DTA.

2. Materials and Methods

This research study was carried out at the DBMS Laboratory, Department of Computer Science and Engineering, Saveetha School of Engineering, Chennai. The two Supervised Learning groups of classification algorithms used for the study. Group 1 and Group 2 are the Decision Tree algorithms and Logistic Regression algorithms respectively als their ranges are shown in the Fig.1. Each sample size was predicted using the G-power tool with version 3.1.10 and resulting in 20 sample sizes with 96% of G-power station values and the threshold two tailed significant values is set to 0.05 and the confidence interval as 96% (Kankrale and Kankrale 2021) .

The phishing, anti-phishing dataset which is to be imputed for the proposed work is collected from (Noor n.d.). one of the more popular online communities for data scientists and machine learning practitioners. It also provides a customizable personal Google co-laboratory with a free online GPU. The dataset used here consists of 42 attributes and contains 5 features that can be used to predict the website phishing attacks. The dataset has 11055 rows which consists of data for the symptoms that are related to Phishing Attack and also includes many sites in the dataset (Website, Noor n.d.)(Noor n.d.)(Website, Noor n.d.). Nearly 4.6 billion active internet users are there in 2020, a record for that year throughout the world.

Decision Tree Algorithm:
Decision tree classifiers are utilized as a notable order procedure. A decision tree is a flowchart-like tree structure where an inside node addresses a component or characteristic , the branch addresses a decision rule and each leaf node addresses the result . The highest node in a decision tree is known as the root node. It figures out how to parcel in light of the quality worth. It segments the tree in a recursive way called recursive parceling. This specific element gives the tree classifier a higher goal to manage an assortment of informational indexes, whether mathematical or downright information. Additionally, decision trees are great for managing nonlinear connections among traits and classes. Routinely, a pollutant is not set in stone to evaluate the nature of the division for every node and the Gini Variety Index is utilized as a known standard for the complete presentation. The decision tree is adaptable as it can be without much of a stretch model of nonlinear or flighty connections. It can decipher the communication between indicators. It can likewise be deciphered very well in view of its binary structure. Notwithstanding, the choice tree has different downsides that will more often than not abuse information. Plus, refreshing a decision tree with new examples is troublesome.
The following pseudocode comes under the Decision Tree Algorithm formula to use on the center pictures dataset and additionally works with the tree model. The pseudocode can take the datasets as input and therefore the final output of the pseudocode is sent through the parameters Accuracy and the classification.The entropy of the decision tree algorithm must be calculated by using below equation (1).

Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no) (1)

Pseudocode of the DTA Algorithm:

Input: Training Dataset
Output: Accuracy
1. Read the training dataset as input
2. Preprocess the dataset and split to train and test
3. Define class
   Logistic Regression(test attribute)
   if(condition satisfy)
   return accuracy
   else
   return previous step
   end
4. Classifiers predicted accuracy.

Logistic Regression Algorithm:

Logistic Regression is a supervised learning algorithm. It provides accurate results when new data is given to the trained model. It is a predictive analysis algorithm in view of the idea of probability. The sigmoid capacity is a numerical capacity used to plan the anticipated worth of probabilities. The worth of Logistic Regression should be somewhere in the range of 0 and 1 which can be determined utilizing the underneath condition (2).

Value(V)=1/(1+e^-value) (2)

Where, e is base of the natural algorithms

Pseudocode: Logistic Regression Algorithm
Input: Training dataset
Output: Classifier predicted accuracy
1. Peruse the training dataset into the classifier
2. Calculate cost function, gradient descent
3. Repeat
4. Calculate sigmoid function for each iteration
5. While the condition satisfy
6. Define class
   define Logistic Regression(test attribute)
   if(condition satisfy)
   return accuracy
   else
   return previous step
   end
7. Classifiers predicted accuracy.

Experiment Setup

The stage used to assess the machine learning algorithms and Supervised Learning was the jupyter lab. The equipment designs were Intel center Ryzen processors with a RAM size of 8GB. The system type used was 64-bit, OS, X64 based processor with SSD or 512 GB. The working framework utilized was Windows and the device utilized was jupyter lab with python programming language. The dataset is fake and real news is collected. Data preprocessing has to be done. Data cleaning like removing the unnecessary attributes from the dataset and concatenating and shuffling also need to be done. Information investigation shows the items present in the dataset. Convert the dataset that contains just the information required for the classifier. Part the dataset into a training set and testing set. Presently carry out the machine learning classifier and utilize the training dataset to prepare the classifier. Subsequent to training the

classifier utilizes a testing dataset to test the trained classifier to get the anticipated exactness from the classifier.

The SPSS apparatus is utilized to play out the measurable computations for the outcomes that are acquired from classifiers for different test sizes. The text part in the training dataset is an independent variable though the text part in the testing dataset is subject to the training dataset. The comparison of the performances of Decision Tree algorithm andLogistic Regression algorithm is done.

Statistical Analysis

The IBM SPSS device is utilized to play out the statistical analysis of the outcomes that have been produced after the Innovative Phishing Website Detection Using a dataset involving machine learning classifiers for the different test sizes that are significant (p=0.002) .

3. Results

The accuracy of the Decision Tree algorithm is approximately 97% and Logistic Regression algorithm is approximately 92%. The accuracy varies for different test sizes in decimals. The accuracy varies due to random change in the test size of the algorithm from Table 1.
Group Statistics, mean precision and standard deviation for Decision Tree calculations is 96.4950 and 0.75188. Logistic Regression algorithm is 92.4470 and 0.20902. In performing statistical analysis of 20 examples, the Decision Tree calculation got 0.75188 standard deviation with 0.23776 standard blunder while the Logistic Regression calculation acquired 0.20902 standard deviation with 0.06610 standard error from Table 2. The significance value showed that hypothesis holds good.

Have done the Independent sample test while performing the statistical analysis where have compared the accuracy of the Decision Tree algorithm and the Logistic Regression algorithm with the significance value less than p<0.05.Independent Samples Test, the comparison of accuracy for Innovative phishing website Detection using Decision Tree algorithm and Logistic Regression algorithm with significance rate 0.002 and standard error difference 0.24678. When compared with the other algorithms, performance of the proposed Decision Tree classifier achieved better performance than the Logistic Regression classifier from Table 3 .

It is known as the Innovative phishing website Detection architecture. The architecture defines the steps which are performed to develop phishing websites. It consists of the steps as Data Pre-processing, Database, Data Extraction, Modeling Classifier, Implementation and Predicted Accuracy from Fig.1.

Simple Bar Mean of Accuracy by DTA, LRA, the bar diagram addressing the examination of mean accuracy of Decision Tree calculation is 96.4950% and Logistic Regression calculation is 92.4470%. Decision Tree algorithm with error rate of 0.23776 and Logistic Regression algorithm have error rate about 0.06610. Independent t-test was used to compare the accuracy of two algorithms and a statistically significant difference was noticed P < 0.05. The Decision Tree model obtained 96.4950% accuracy from Fig.2. When compared with the other algorithms performance of the proposed Decision Tree classifier achieved better performance than Logistic Regression Algorithm

4. Discussion

Decision Tree Algorithms have better accuracy rates than Logistic Regression Algorithms. The results are collected by performing multiple iterations of the experiment for identifying different scales of accuracy rate. Further, performed the statistical analysis calculations using the SPSS tool with the results which are obtained from the experiment. Independent samples t-test is performed. In this study of Innovative phishing website Detection , the Decision Tree Algorithm has higher accuracy approximately 96% in comparison to Logistic Regression algorithm approximately 92%. Decision Tree algorithm has better significance 0.002 while using the independent samples t-test.

The mean accuracy and standard deviation for the Decision Tree algorithm is 96.4950 and 0.75188.For Logistic Regression, the algorithm is 92.4470 and 0.20902. Decision Tree Algorithms appear to create the most predictable outcomes with negligible standard deviation. In paper(Kudarvalli and Fiaidhi, n.d.), they implemented a Random Forest which provides accuracy of 97%. In paper (Ahmad et al. 2020), Random forest and KNN machine learning algorithms are carried out to distinguish the fake news with an exactness of 97% and 95%. These two papers saw that the Random Forest calculation proposed has better exactness. But In paper(Lyu and Lo 2020), observing that the machine learning classifier Random Forest has an accuracy rate of 97.26% more than Decision Tree. Based on the literature survey it is demonstrated that the

Decision Tree calculation which is carried out has better accuracy contrasted with previous works.

There is a statistically significant distinction in precision for two calculations is $p<0.05$, by performing independent sample tests in the SPSS factual device. Mean and standard deviation are additionally determined utilizing the SPSS factual instrument. Standard blunder distinction defines the error the Decision Tree algorithm with error rate of 0.23776 and Logistic Regression algorithm have error rate about 0.06610. There are only slight differences while assuming with and without variances.

After performing the statistical analysis and independent sample test in the IBM SPSS tool the significance is $p<0.05$. The limitations that the research attributes that the dataset contains are not very many to predict accuracy(%) for Innovative phishing site Detection. The more the independent and dependent variables the more accuracy will be improved. The future work will consider the dataset with many attributes. Hence, the classifier can work effectively and can further develop the forecast precision. Ascribes like profile, source, verifications can bring about superior accuracy and exact precision values.

5. Conclusion

The approach of classifying the phishing website Detecting manually requires more knowledge of the domain. In this research, discussed the problem of classifying phishing website Detecting articles using machine learning models. The accuracy of innovative phishing website Detection using Decision Tree algorithm has better accuracy in comparison with Logistic Regression algorithms. The significance rate is 0.002 which indicates that hypothesis holds true.

Declarations

Conflict of interests

No conflict of interest in this manuscript.

Author Contribution

Author PS was involved in data collection, data analysis, manuscript writing. Author KM was involved in conceptualization, data validation and critical review of manuscript.

Acknowledgement

The creators might want to thank the board, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for giving the potential open doors and offices to explore study.

Funding

The authors thank the following organizations for providing financial support that enabled us to complete the study.

1. Saveetha University
2. Saveetha Institute of Medical And Technical Sciences
3. Saveetha School of Engineering
4. Sterling Software Pvt. Ltd.

6. References

Anupong, Wongchai, Lin Yi-Chia, Mukta Jagdish, Ravi Kumar, P. D. Selvam, R. Saravanakumar, and Dharmesh Dhabliya. 2022. "Hybrid Distributed Energy Sources Providing Climate Security to the Agriculture Environment and Enhancing the Yield." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102142.

Bharathiraja, B., J. Jayamuthunagai, R. Sreejith, J. Iyyappan, and R. Praveenkumar. 2022. "Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil." Bioresource Technology 351 (May): 126956.

Dutta, Ashit Kumar. 2021. "Detecting Phishing Websites Using Machine Learning Technique." PLOS ONE. https://doi.org/10.1371/journal.pone.0258361.

Ghimire, Awishkar, Avinash Kumar Jha, Surendrabikram Thapa, Sushruti Mishra, and Aryan Mani Jha. 2021. "Machine Learning Approach Based on Hybrid Features for Detection of Phishing URLs." 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). https://doi.org/10.1109/confluence51648.2021.9377113.

Gu, Chenyu. 2021. "A Lightweight Phishing Website Detection Algorithm by Machine Learning." 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML). https://doi.org/10.1109/conf-spml54095.2021.00054.

Itoo, Fayaz, Meenakshi, and Satwinder Singh. 2021. "Comparison and Analysis of Logistic Regression, Naïve Bayes and KNN Machine Learning Algorithms for Credit Card Fraud Detection." International Journal of Information Technology. https://doi.org/10.1007/s41870-020-00430-y.

Jain, Ankit Kumar, and B. B. Gupta. 2019. "A Machine Learning Based Approach for Phishing Detection Using Hyperlinks Information." Journal of Ambient Intelligence and Humanized Computing. https://doi.org/10.1007/s12652-018-0798-z.

Jang, Minhae, Jaeju Song, and Myongsoo Kim. 2022. "A Study on the Detection Method for Malicious URLs Based on a Number of Search Results Matching the Internet Search Engines Combining the Machine Learning." Journal of Electrical Engineering & Technology. https://doi.org/10.1007/s42835-021-00888-1.

Jothi, K. Jeeva, K. Jeeva Jothi, S. Balachandran, K. Mohanraj, N. Prakash, A. Subhasri, P. Santhana Gopala Krishnan, and K. Palanivelu. 2022. "Fabrications of Hybrid Polyurethane-Pd Doped ZrO2 Smart Carriers for Self-Healing High Corrosion Protective Coatings." Environmental Research. https://doi.org/10.1016/j.envres.2022.113095.

Kale, Vaibhav Namdev, J. Rajesh, T. Maiyalagan, Chang Woo Lee, and R. M. Gnanamuthu. 2022. "Fabrication of Ni–Mg–Ag Alloy Electrodeposited Material on the Aluminium Surface Using Anodizing Technique and Their Enhanced Corrosion Resistance for Engineering Application." Materials Chemistry and Physics. https://doi.org/10.1016/j.matchemphys.2022.125900.

Kankrale, Prof Rajendra, and Rajendra Kankrale. 2021. "Phishing Website Detection Using Machine Learning." International Journal for Research in Applied Science and Engineering Technology. https://doi.org/10.22214/ijraset.2021.35671.

Kudarvalli, Harika, and Jinan Fiaidhi. n.d. "Detecting Fake News Using Machine Learning Algorithms." https://doi.org/10.36227/techrxiv.12089133.v1.

Lokesh, Gururaj Harinahalli, and Goutham BoreGowda. 2021. "Phishing Website Detection Based on Effective Machine Learning Approach." Journal of Cyber Security Technology. https://doi.org/10.1080/23742917.2020.1813396.

Lyu, Shikun, and Dan Chia-Tien Lo. 2020. "Fake News Detection by Decision Tree." 2020 SoutheastCon. https://doi.org/10.1109/southeastcon44009.2020.9249688.

Nadar, Vinitha Kumaresan, Bhavesh Patel, Vidyullata Devmane, and Uday Bhave. 2021. "Detection of Phishing Websites Using Machine Learning Approach." 2021 2nd Global Conference for Advancement in Technology (GCAT). https://doi.org/10.1109/gcat52182.2021.9587682.

Noor, Ahmad. n.d. "Website Phishing Dataset." Accessed April 25, 2022. https://www.kaggle.com/ahmednour/website-phishing-data-set.

Palanisamy, Rajkumar, Diwakar Karuppiah, Subadevi Rengapillai, Mozaffar Abdollahifar, Gnanamuthu Ramasamy, Fu-Ming Wang, Wei-Ren Liu, Kumar Ponnuchamy, Joongpyo Shim, and Sivakumar Marimuthu. 2022. "A Reign of Bio-Mass Derived Carbon with the Synergy of Energy Storage and Biomedical Applications." Journal of Energy Storage. https://doi.org/10.1016/j.est.2022.104422.

Pascariu, Cristian, and Ioan C. Bacivarov. 2021. "Detecting Phishing Websites Through Domain and Content Analysis." 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). https://doi.org/10.1109/ecai52376.2021.9515165.

Patil, Vaibhav, Pritesh Thakkar, Chirag Shah, Tushar Bhat, and S. P. Godse. 2018. "Detection and Prevention of Phishing Websites Using Machine Learning Approach." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). https://doi.org/10.1109/iccubea.2018.8697412.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Rao, Routhu Srinivasa, and Alwyn Roshan Pais. 2019. "Detection of Phishing Websites Using an Efficient Feature-Based Machine Learning Framework." Neural Computing and Applications. https://doi.org/10.1007/s00521-017-3305-0.

Razaque, Abdul, Mohamed Ben Haj Frej, Dauren Sabyrov, Aidana Shaikhyn, Fathi Amsaad, and Ahmed Oun. 2020. "Detection of Phishing Websites Using Machine Learning." 2020 IEEE Cloud Summit. https://doi.org/10.1109/ieeecloudsummit48914.2020.00022.

Saravanan, Priya, and Selvakumar Subramanian. 2020. "A Framework for Detecting Phishing Websites Using GA Based Feature Selection and ARTMAP Based Website Classification." Procedia Computer Science. https://doi.org/10.1016/j.procs.2020.04.116.

Sumathy, B., Anand Kumar, D. Sungeetha, Arshad Hashmi, Ankur Saxena, Piyush Kumar Shukla, and Stephen Jeswinde Nuagah. 2022. "Machine Learning Technique to Detect and Classify Mental Illness on Social Media Using Lexicon-Based Recommender System." Computational Intelligence and Neuroscience 2022 (February): 5906797.

Thanigaivel, Sundaram, Sundaram Vickram,

Nibedita Dey, Govindarajan Gulothungan, Ramasamy Subbaiya, Muthusamy Govarthanan, Natchimuthu Karmegam, and Woong Kim. 2022. "The Urge of Algal Biomass-Based Fuels for Environmental Sustainability against a Steady Tide of Biofuel Conflict Analysis: Is Third-Generation Algal Biorefinery a Boon?" Fuel. https://doi.org/10.1016/j.fuel.2022.123494.

Vickram, Sundaram, Karunakaran Rohini, Krishnan Anbarasu, Nibedita Dey, Palanivelu Jeyanthi, Sundaram Thanigaivel, Praveen Kumar Issac, and Jesu Arockiaraj. 2022. "Semenogelin, a Coagulum Macromolecule Monitoring Factor Involved in the First Step of Fertilization: A Prospective Review." International Journal of Biological Macromolecules 209 (Pt A): 951–62.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." Fuel. https://doi.org/10.1016/j.fuel.2022.123814.

Zaini, Nur Sholihah, Deris Stiawan, Mohd Faizal Ab Razak, Ahmad Firdaus, Wan Isni Sofiah Wan Din, Shahreen Kasim, and Tole Sutikno. 2020. "Phishing Detection System Using Nachine Learning Classifiers." Indonesian Journal of Electrical Engineering and Computer Science 17 (3): 1165.

Narayanasamy, S., Sundaram, V., Sundaram, T., & Vo, D. V. N. (2022). Biosorptive ascendency of plant based biosorbents in removing hexavalent chromium from aqueous solutions–Insights into isotherm and kinetic studies. Environmental Research, 210, 112902.

Tables and Figures

Table 1: Accuracy Table DTA and LRA the accuracy of the Decision Tree algorithm is approximately 96% and Logistic Regression algorithm is approximately 92%

Test Size	0.2	0.21	0.22	0.23
Decision Tree Algorithm	96.59	97.28	96.9	96.62
Logistic Regression Algorithm	92.69	92.56	92.43	92.5

Table 2: Group Statistics, that the mean accuracy and standard deviation for Decision Tree algorithms is 96.4950 and 0.75188. Logistic Regression algorithm is 92.4470 and 0.20902

	DTA,LRA	**N**	**Mean**	**Std. Deviation**	**Std. Mean Error**
Accuracy	DTA	10	96.4950	0.75188	0.23776
	LRA	10	92.4470	0.20902	0.6610

Table 3: Independent Samples Test, the comparison of accuracy for Innovative phishing website Detection classification using Decision Tree algorithm and Logistic Regression algorithm with significance rate 0.002 and standard error difference 0..

		Levene's Test for Equality of Variances (1)	**Levene's Test for Equality of Variances (2)**	**T-test for Equality of Means (3)**	**T-test for Equality of Means (4)**	**T-test for Equality of Means (5)**
		F	**Sig.**	**Std.Error Difference**	**95% Confidence lower**	**95% Confidence upper**
Accuracy	**Equal Variances assumed**	13.401	0.002	0.24678	3.52953	4.56647
	Equal Variances			0.24678	3.50087	4.59513

	not assumed					

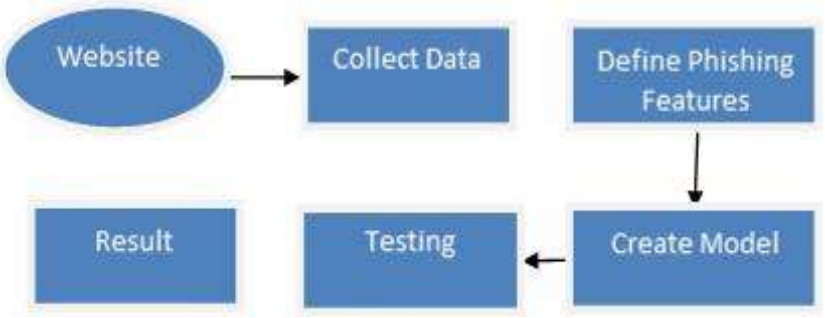

Fig.1.Machine Learning Classifier Architecture

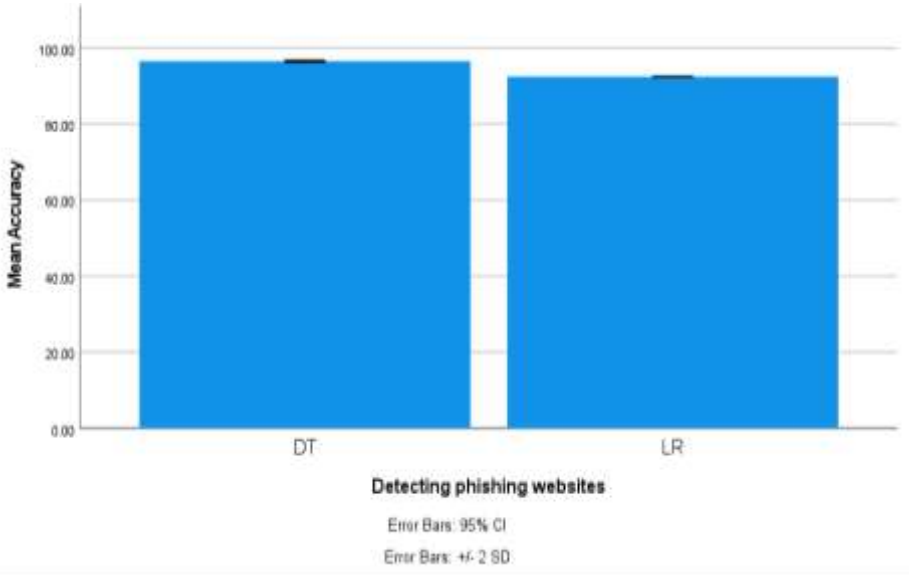

Fig. 2: Simple Bar Mean of Accuracy by DTA, LRA, the bar chart representing the comparison of mean accuracy of the Decision Tree algorithm is 96.4950 and Logistic Regression algorithm is 92.4470. X-Axis: Decision Tree algorithm vs Logistic Regression algorithm. Y-Axis: Mean accuracy of detection ± SD.