International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

# Machine Learning-Based Phishing Detection: Improving Accuracy and Adaptability

**Akhil Mittal, Pandi Kirupa Gopalakrishna Pandian**

**Submitted**: 02/05/2024 **Revised**: 15/06/2024 **Accepted**: 22/06/2024

**Abstract:** In this modern era, phishing has become a great problem. Because of this, it can be observed that the personal information of people are leaked from emails & websites. Hence, it is needed that these instances of phishing are to be reduced. In doing one of the best tools that can be used is Machine Learning. This is a process of using historical data for making prediction of future scenarios. In this project, the details of the approached that can be used for the detection of phishing are analyzed. Moreover, the algorithms that are used by ML for this purpose are also envisaged here. The description of the process of collection of data is presented here. In addition to this, the results that shows the effectiveness of ML in the detection of phishing is also discussed here.

***Key Words**: Machine Learning, Neural Network, Decision Tree, Random Forest, Feature Engineering*

## Introduction

One of the most critical threats that are observed these days is the threat of the stealing of information through websites. In this way, the personal info of the users gets stolen. One of the traditional methods for the detection of this problem is the blacklisting of such sites. However, it has become irrelevant as the number of sites has increased a lot over the past few years. The best tool that can be used in this current scenario is “Machine Learning”.

This has the feature of checking a huge amount of data and analyzing them to find out the websites that are pure and have no risks of leaking information. The things that are analyzed in this process are the content present in the email, URLs, & the source codes. Here, the details of the process of “phishing detection” are discussed. For this, literature based on the detection of phishing was studied in order to identify the processes of this.

## Literature Review

### Machine Learning-based solutions for phishing website detection

According to Tang & Mahmoud, 2021, the use of ML shows good results in terms of detecting phishing on websites. It is considered to be one of the most critical threats or the users. There are different models of ML that can be used for this purpose. This contains both “supervised” & “unsupervised” models. The data that these models analyses are collected from the URLs, and content of the websites. “Supervised learning” has attained success in the detection of “phishing websites”. The approaches that it uses are “decision tree”, “neural network”, and SVM. This is based on the datasets with the help of which it is predicted that how often a website can be phishing (Tang & Mahmoud, 2021). These processes can easily be interpreted and their accuracy is fair enough.

In the current era, the use of CNNs & RNNs has become very popular. The results are good in terms of detection of phishing. The main benefit of these is that they are able to collect necessary information from the existing data. This lowers the need for “feature engineering”. The main characteristic of CNN is that it can differentiate a good and a phishing website visually. On the other hand, the use of RNN is mainly observed in the processing of “sequential data”.

### Phishing URL detection using lexical-based machine learning in a real-time environment

According to Gupta *et al*. 2021, the prime focus of this method is on finding out the composition & structure of URLs. In this way, phishing is detected. The process includes checking out the length of the URL, looking for the presence of keywords that are suspicious in nature and also the presence of “special characters”. The “lexical features” are such things that can easily be extracted and also analysed for the detection of phishing in real-time.

There are a lot of examples of how features based on lexical can be helpful in the detection of URLs that are phishing in nature. There are some indicators with the help of which websites that are phishing can be detected. These are unusual “port numbers”, more than one subdomain, and long lengths of URLs. ML is a good tool in the detection of this with processes such as SVM, regression, and “gradient boosting” (Gupta *et al*. 2021). This is very helpful in the increase of the detection process.

[1]*Independent Researcher, USA.*
[2]*Independent Researcher, USA.*

In the process of detection of phishing, the “lexical features” play an important role. It is because these are indicators of the emergency responses that are to be taken against the threats. If ML is combined with browsers & emails, then it can result in the analysis of URLs in an instant.

### Internet phishing detection based on log data

According to Obaid *et al*. 2021, there can be different forms of “log data”. These are logs of the activities of the user, logs of accessing the servers, and the log of traffic. These all are data that can provide enough information on the activities of phishing. ML is a crucial tool that can analyse these “log data” for the identification of the patterns of phishing (Obaid *et al*. 2021). The insights that “log data” provides are accessing attempts that are not usual attempts, failure in the process of login, and abnormality in the volume of data to be transferred.

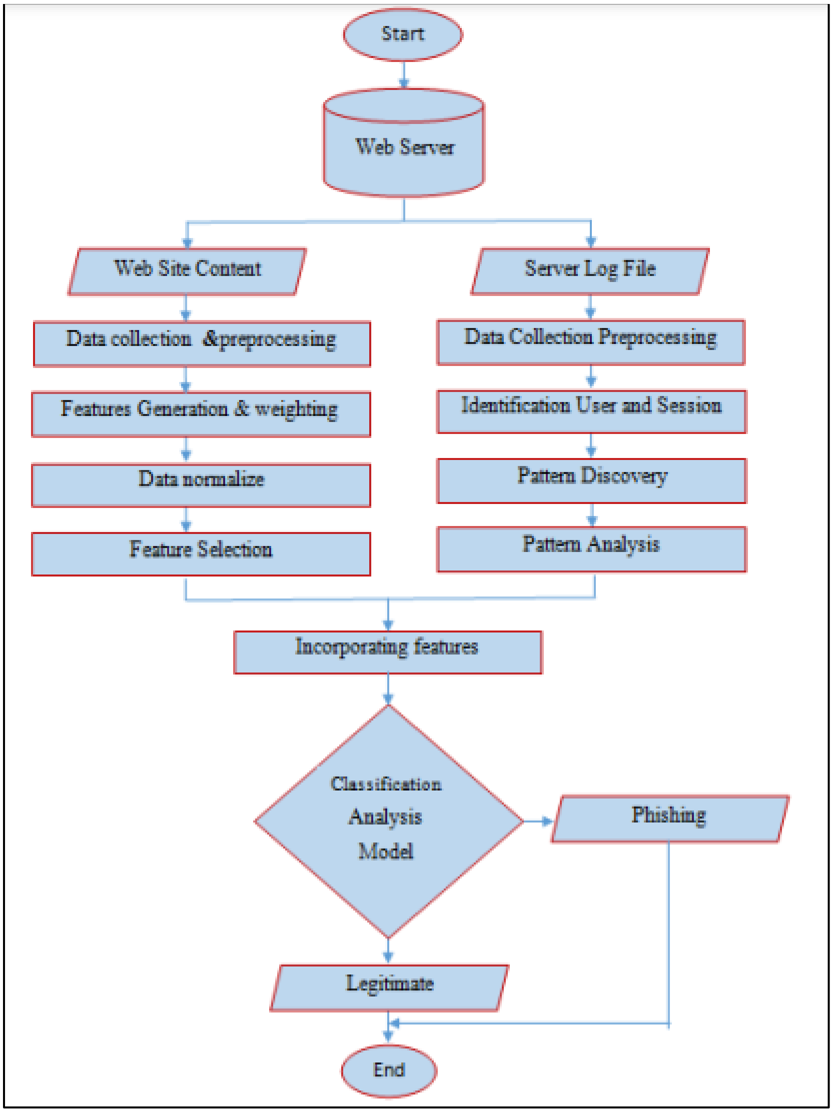

**Fig 1:** System, Flowchart

(Source: Obaid *et al*. 2021)

One of the best methods that are used for the detection of phishing is “clustering”, and “outliner detection”. These are helpful in distinguishing between the usual behaviour and unusual behaviour (phishing).

This form of detection of phishing is beneficial from the point of view of the use of the models of ML in order to use “historical data”. Also, when the “log data” is monitored continuously, it results in the adaptation of new methods of phishing and on the basis of that alerts are given.

## Methods

### Data collection & processing

Data is the resource on which the effectiveness of the ML models for the detection of phishing is dependent. The better the quality of input data the better results are obtained. This data in mainly important for the “training” & “evaluation”.

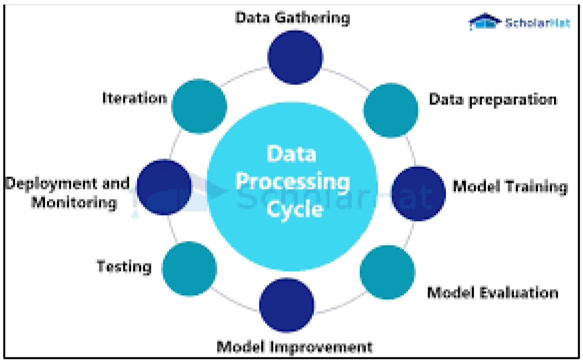

**Fig 2:** Data Processing

(Source: data:image/png;base64)

The data contains a set of URLs that contains both good and phishing URLs. The main sources for the collection of data are datasets of public, "phishing database", and " crawlers (web)". In order to achieve better results it is wise to collect data on phishing that contains the different techniques of phishing. Also, it should collect data from good websites for making comparisons. After the collection of data, it is processed before the analysis of it (Salahdine *et al*. 2021). This method of pre-processing data includes the below.

Cleaning of data-

This consists of the removal of data that are not relevant and provides decisions that can be biased.

Extraction-

This is the next step of the process of pre-processing. At this stage, the important contents are extracted from the URLs & websites. These contents are "lexical features", "behavioural features", and the features based on content.

Normalization-

At this stage, the data are prepared so that it can be fit to become a good input for ML.

**Design of Machine Learning models**

There are models that can be helpful for the detection of phishing activities. ML models are developed in different stages for this purpose. This stage includes the selection of essential algorithms, and changing them in such a way that can result in achieving good results. This starts with the selection of ML techniques (Deval *et al*. 2021). These techniques should possess the "classical algorithms" and method of "deep learning". The details of the stages are provided below.

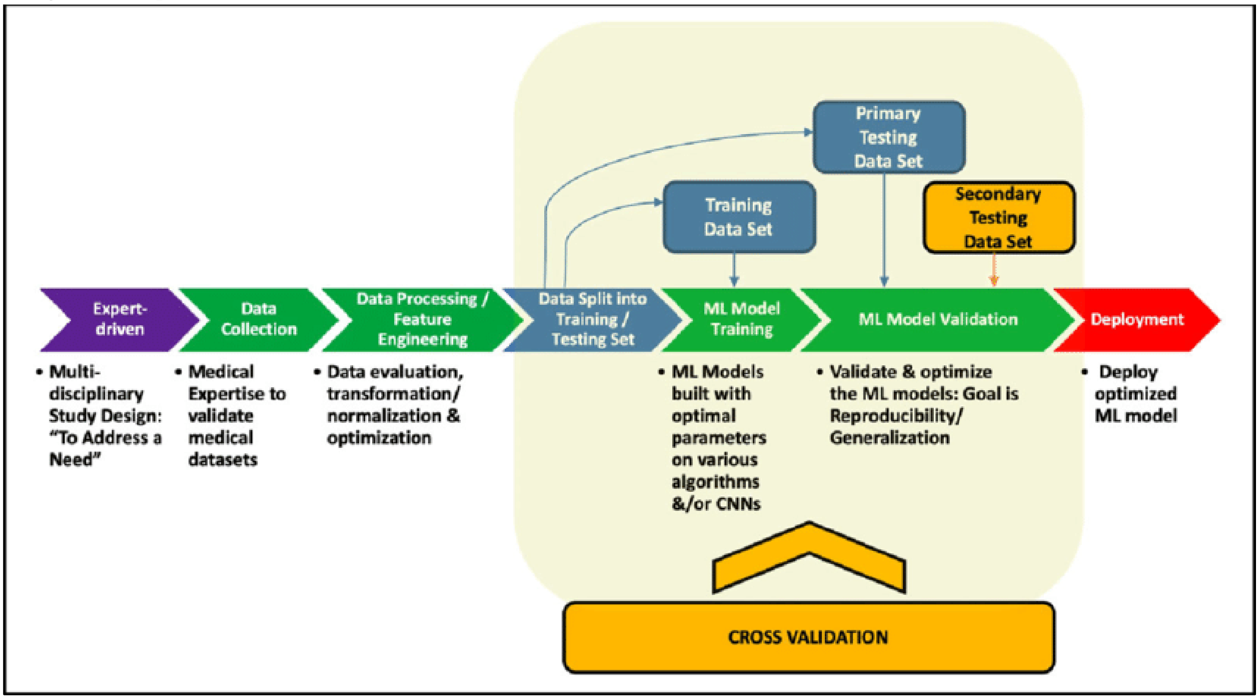
Primary Testing Data Set
Training Data Set
Secondary Testing Data Set
Expert-driven
Data Collection
Data Processing / Feature Engineering
Data Split into Training / Testing Set
ML Model Training
ML Model Validation
Deployment
• Multi-disciplinary Study Design: "To Address a Need"
• Medical Expertise to validate medical datasets
• Data evaluation, transformation/ normalization & optimization
• ML Models built with optimal parameters on various algorithms &/or CNNs
• Validate & optimize the ML models: Goal is Reproducibility/ Generalization
• Deploy optimized ML model
CROSS VALIDATION

**Fig 3:** Supervised ML model

(Source: https://www.researchgate.net)

Selection of algorithm-

There are some algorithms that are mostly used for the detection of phishing. These are SVM, “decision tree”, and “neural network”. Among these, the algorithms are decided on the basis of system requirements.

Training of models-

After the completion of the pre-processing, the data is divided into two halves. The first one is called “training set” The use of this is to build models. Another one is the “validation set” (Somesha *et al*. 2020). It is used for the evaluation of the performance of the prepared models.

Feature Engineering-

This is a process that is used for the betterment of the prediction of the models. This consists of the selection of features that provide good information and reduction of dimensionality for the increase of capability of computation.

**Deployment & Implementation**

This is the next process after the completion of training the models. The steps that are included in this are given below.

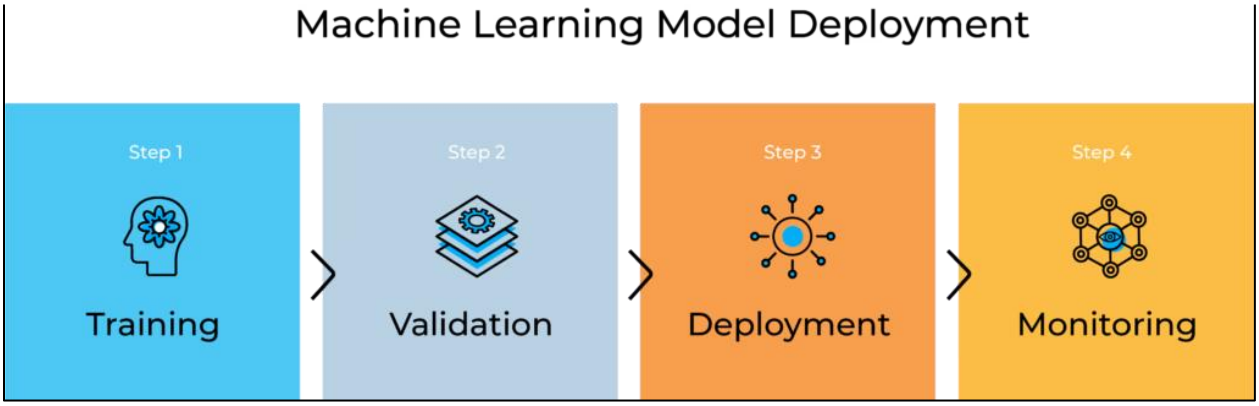

**Fig 4:** Deployment of ML model

(Source: https://framerusercontent.com)

Integration-

These modes can be used with the integration of emails & browsers. With this use of this, it will be possible to analyses the URLs and the contents present in the emails (Gandotra & Gupta, 2021). As a result of this, it protects the users against phishing.

Scalability-

In this stage, it is ensured that the system is capable of attaining a large “traffic volume”. Hence, features like the efficiency of the system and speed are analyzed.

Monitoring-

This involves checking the system on a continuous basis in order to increase the performance of the models.

**Result**

**Accuracy Improvement**

The models based on ML have some significantly good results in terms of the detection of phishing cases. This is possible because of the use of algorithms such as “random forest”, “decision tree”, and SVM.

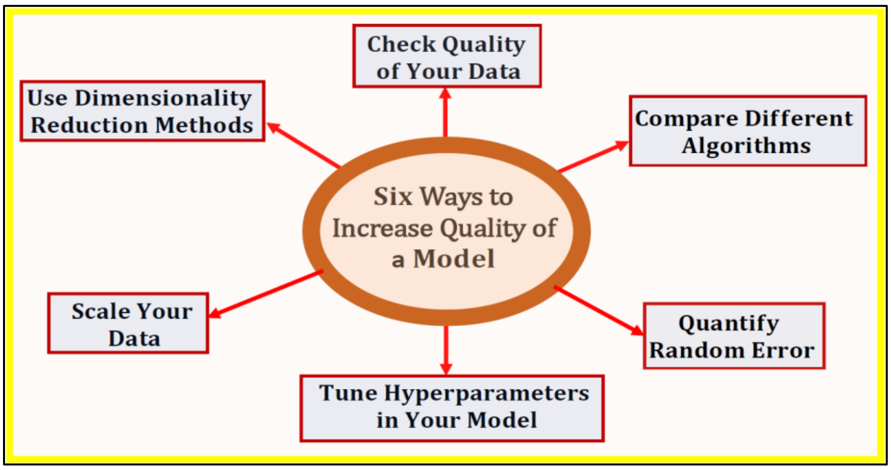
Six Ways to Increase Quality of a Model
Check Quality of Your Data
Use Dimensionality Reduction Methods
Compare Different Algorithms
Scale Your Data
Quantify Random Error
Tune Hyperparameters in Your Model

**Fig 5:** Accuracy of ML models

(Source: https://encrypted-tbn0.gstatic.com)

In many cases, the level of accuracy has crossed 95%. There are different forms of features that are integrated together such as lexical, on the basis of content, and on the basis of behavior (Odeh *et al*. 2021). All of these results in finding out the phishing websites from the good websites.

### Enhancement of adaptability

Adaptability is a feature that is very much essential for the improvement of the effectiveness of the ML models. This involves making improvements on a continuous basis. The system of learning in real-time makes it easy to know about the new patterns of phishing. Also, it helps in maintaining a good accuracy of detection of phishing (Aljofey *et al*. 2020). It can be noticed that when hybrid methods are used different models of ML are integrated together. In this way, the system becomes more adaptable.

### Detection in real-time

There are many good results that can be observed when these n\models are combined with the emails & browsers. Also, another good result of this is that the gateway of security gets enhanced for the detection of phishing.

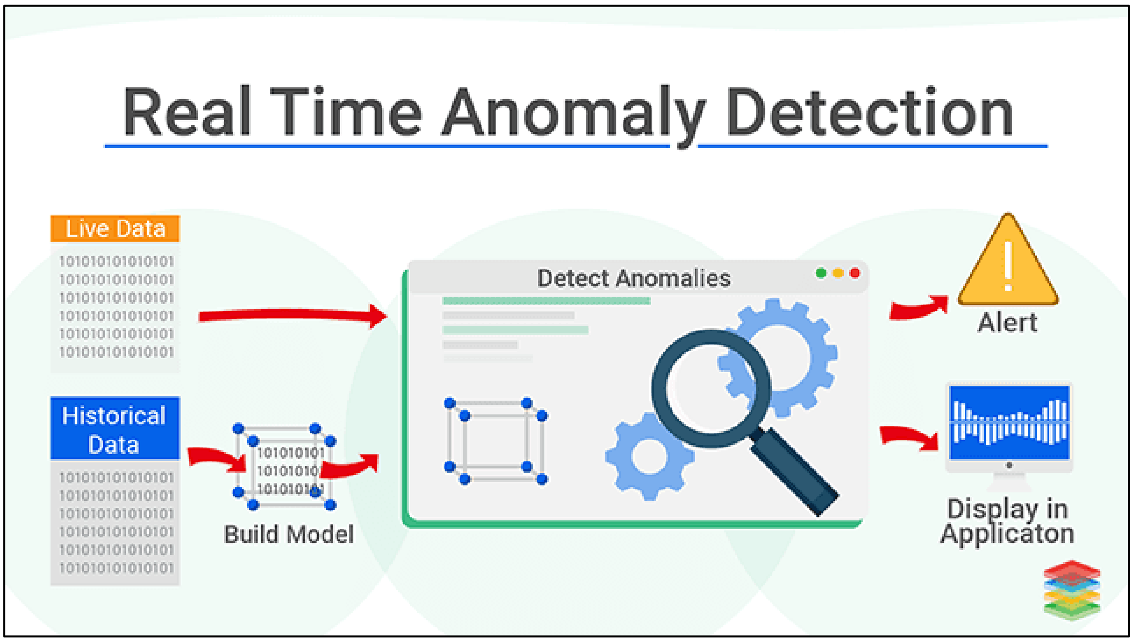

**Fig 6:** Real-Time threat detection

(Source: https://miro.medium.com)

In this way, the URLs are analyzed fast and give the users fast alerts. In the process where lexical data is involved, it results in a fast filtration of data. Hence, it is possible to analyse the content in depth. Hence, more security can be provided.

### Scalability

The importance of the detection of phishing is increasing at a rapid rate. In the systems where there can be seen the use of clouds, a good amount of data is to be handled there. Hence, continuous protection of the users of the system is required (Rashid *et al*. 2020). This can be done through monitoring the ML models on a continuous basis. With the help of scalability, it can be made sure that the activity of the internet increases. Also, it ensures that it is possible to detect phishing activities and protect the users.

### Discussion

Here, the advancements in the detection of phishing were checked. In particular detection of phishing with the use of models of ML was checked here. Also, how the ML can accurately determine the phishing possibilities is analysed here. With the use of "advanced algorithms" it is possible to make a difference between the good and phishing websites (Do *et al*. 2021). Also, the continuous monitoring of data helps keep on checking the possible threats to the systems. With the help of scalability is possible to make sure that a good amount of data is handled by the system. The two main characteristics of a good system are "adaptability" & "performance" which increases the strength of the system.

### Future Direction

The main focus of this in the future should be on the ease of interpretation of the process of the detection of phishing. This will make sure that the trust of the users is well-established in the models. Moreover, in this, the use of AI can also be useful. It can make further improvements in the process of detection of threats (Casimiro *et al*. 2021). In addition to this, a collaboration can be done for sharing details like threats among the different platforms. As a result, the system will benefit from the identification of new patterns.

## Conclusion

When ML is used in the system for the detection of phishing, good results can be observed. The main algorithms that are used by ML for this purpose are SVM, “decision trees”, & “nural networks”. This system has gained an accuracy of 95%. When the models are updated in a continuous manner then it results in having new data that highlights new methods of phishing. As a result of this more secure system can be made. In the present day, the scenario of phishing is increasing. Hence, it has become essential to use ML for the detection of this. Moreover, the use of AI is likely to increase the performance of these models.

Reference List

## Journals

[1] Aljofey, A., Jiang, Q., Qu, Q., Huang, M. and Niyigena, J.P., 2020. An effective phishing detection model based on character level convolutional neural network from URL. *Electronics*, *9*(9), p.1514.

[2] Casimiro, M., Romano, P., Garlan, D., Moreno, G.A., Kang, E. and Klein, M., 2021, September. Self-Adaptation for Machine Learning Based Systems. In *ECSA (Companion)*.

[3] Deval, S.K., Tripathi, M., Bezawada, B. and Ray, I., 2021, October. “X-Phish: Days of Future Past”: Adaptive & Privacy Preserving Phishing Detection. In *2021 IEEE Conference on Communications and Network Security (CNS)* (pp. 227-235). IEEE.

[4] Do, N.Q., Selamat, A., Krejcar, O., Yokoi, T. and Fujita, H., 2021. Phishing webpage classification via deep learning-based algorithms: an empirical study. *Applied Sciences*, *11*(19), p.9210.

[5] Gandotra, E. and Gupta, D., 2021. An efficient approach for phishing detection using machine learning. *Multimedia security: algorithm development, analysis and applications*, pp.239-253.

[6] Gupta, B.B., Yadav, K., Razzak, I., Psannis, K., Castiglione, A. and Chang, X., 2021. A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment. *Computer Communications*, *175*, pp.47-57.

[7] Obaid, A.J., Ibrahim, K.K., Abdulbaqi, A.S. and Nejrs, S.M., 2021. An adaptive approach for internet phishing detection based on log data. *Periodicals of Engineering and Natural Sciences*, *9*(4), pp.622-631.

[8] Odeh, A., Keshta, I. and Abdelfattah, E., 2021. PHIBOOST-a novel phishing detection model using Adaptive boosting approach. *Jordanian Journal of Computers and Information Technology (JJCIT)*, *7*(01).

[9] Rashid, J., Mahmood, T., Nisar, M.W. and Nazir, T., 2020, November. Phishing detection using machine learning technique. In *2020 first international conference of smart systems and emerging technologies (SMARTTECH)* (pp. 43-46). IEEE.

[10] Salahdine, F., El Mrabet, Z. and Kaabouch, N., 2021, December. Phishing attacks detection a machine learning-based approach. In *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0250-0255). IEEE.

[11] Somesha, M., Pais, A.R., Rao, R.S. and Rathour, V.S., 2020. Efficient deep learning techniques for the detection of phishing websites. *Sādhanā*, *45*, pp.1-18.

[12] Tang, L. and Mahmoud, Q.H., 2021. A survey of machine learning-based solutions for phishing website detection. *Machine Learning and Knowledge Extraction*, *3*(3), pp.672-694.

[13] Kaur, Jagbir. "Streaming Data Analytics: Challenges and Opportunities." International Journal of Applied Engineering & Technology, vol. 5, no. S4, July-August 2023, pp. 10-16.https://romanpub.com/resources/ijaetv5-s4-july-aug-2023-2.pdf

[14] Pandi Kirupa Kumari Gopalakrishna Pandian, Satyanarayan kanungo, J. K. A. C. P. K. C. (2022). Ethical Considerations in Ai and Ml: Bias Detection and Mitigation Strategies. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 248–253. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/10511

[15] Ashok : "Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.")

[16] Kaur, J. (2021). Big Data Visualization Techniques for Decision Support Systems. Jishu/Journal of Propulsion Technology, 42(4). https://propulsiontechjournal.com/index.php/journal/article/view/5701

[17] Ashok : "Choppadandi, A., Kaur, J.,Chenchala, P. K., Nakra, V., & Pandian, P. K. K. G. (2020). Automating ERP Applications for Taxation Compliance using Machine Learning at SAP Labs. International Journal of Computer Science and

Mobile Computing, 9(12), 103-112. https://doi.org/10.47760/ijcsmc.2020.v09i12.014

[18] Chenchala, P. K., Choppadandi, A., Kaur, J., Nakra, V., & Pandian, P. K. G. (2020). Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. International Journal of Open Publication and Exploration, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127

[19] Kaur, J., Choppadandi, A., Chenchala, P. K., Nakra, V., & Pandian, P. K. G. (2019). AI Applications in Smart Cities: Experiences from Deploying ML Algorithms for Urban Planning and Resource Optimization. Tuijin Jishu/Journal of Propulsion Technology, 40(4), 50-56.

[20] Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service . (2019). International Journal of Transcontinental Discoveries, ISSN: 3006-628X, 6(1), 29-34. https://internatioHappy Guru Purnima sir charan sparsh aljournals.org/index.php/ijtd/article/view/98

[21] Kaur, J., Choppadandi, A., Chenchala, P. K., Nakra, V., & Pandian, P. K. G. (2019). Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service. International Journal

[22] of Transcontinental Discoveries, 6(1), 29-34. https://internationaljournals.org/index.php/ijtd/article/view/98

[23] Choppadandi, A., Kaur, J., Chenchala, P. K., Kanungo, S., & Pandian, P. K. K. G. (2019). AI-Driven Customer Relationship Management in PK Salon Management System. International Journal of Open Publication and Exploration, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128

[24] Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.

[25] Ashok Choppadandi et al, International Journal of Computer Science and Mobile Computing, Vol.9 Issue.12, December- 2020, pg. 103-112. ( Google scholar indexed)

[26] Choppadandi, A., Kaur, J., Chenchala, P. K., Nakra, V., & Pandian, P. K. K. G. (2020). Automating ERP Applications for Taxation Compliance using Machine Learning at SAP Labs. International Journal of Computer Science and Mobile Computing, 9(12), 103-112. https://doi.org/10.47760/ijcsmc.2020.v09i12.014

[27] [Chenchala, P. K., Choppadandi, A., Kaur, J., Nakra, V., & Pandian, P. K. G. (2020). Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. International Journal of Open Publication and Exploration, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127]

[28] AI-Driven Customer Relationship Management in PK Salon Management System. (2019). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128

[29] Pradeep Kumar Chenchala. (2023). Social Media Sentiment Analysis for Enhancing Demand Forecasting Models Using Machine Learning Models. International Journal on Recent and Innovation Trends in Computing and Communication, 11(6), 595–601. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10762

[30] Tilala, Mitul, Saigurudatta Pamulaparthyvenkata, Abhip Dilip Chawda, and Abhishek Pandurang Benke. "Explore the Technologies and Architectures Enabling Real-Time Data Processing within Healthcare Data Lakes, and How They Facilitate Immediate Clinical Decision-Making and Patient Care Interventions." European Chemical Bulletin 11, no. 12 (2022): 4537-4542. https://doi.org/10.53555/ecb/2022.11.12.425.

[31] Mitul Tilala, Abhip Dilip Chawda, Abhishek Pandurang Benke, Akshay Agarwal. (2022). Regulatory Intelligence: Leveraging Data Analytics for Regulatory Decision-Making. International Journal of Multidisciplinary Innovation and Research Methodology, ISSN: 2960-2068, 1(1), 78–83. Retrieved from https://ijmirm.com/index.php/ijmirm/article/view/77

[32] Mitul Tilala. (2023). Real-Time Data Processing in Healthcare: Architectures and Applications for Immediate Clinical Insights. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1119–1125. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10629

[33] Tilala, Mitul, and Abhip Dilip Chawda. "Evaluation of Compliance Requirements for Annual Reports in Pharmaceutical Industries." NeuroQuantology 18,

no. 11 (November 2020): 138-145. https://doi.org/10.48047/nq.2020.18.11.NQ20244.

[34] Dodda, Suresh, Navin Kamuni, Venkata Sai Mahesh Vuppalapati, Jyothi Swaroop Arlagadda Narasimharaju, and Preetham Vemasani. "AI-driven Personalized Recommendations: Algorithms and Evaluation." Propulsion Tech Journal 44, no. 6 (December 1, 2023). https://propulsiontechjournal.com/index.php/journal/article/view/5587

[35] Kamuni, Navin, Suresh Dodda, Venkata Sai Mahesh Vuppalapati, Jyothi Swaroop Arlagadda, and Preetham Vemasani. "Advancements in Reinforcement Learning Techniques for Robotics." Journal of Basic Science and Engineering 19, no. 1 (2022): 101-111. ISSN: 1005-0930.

[36] Dodda, Suresh, Navin Kamuni, Jyothi Swaroop Arlagadda, Venkata Sai Mahesh Vuppalapati, and Preetham Vemasani. "A Survey of Deep Learning Approaches for Natural Language Processing Tasks." International Journal on Recent and Innovation Trends in Computing and Communication 9, no. 12 (December 2021): 27-36. ISSN: 2321-8169. http://www.ijritcc.org

[37] Jigar Shah , Joel lopes , Nitin Prasad , Narendra Narukulla , Venudhar Rao Hajari , Lohith Paripati. (2023). Optimizing Resource Allocation And Scalability In Cloud-Based Machine Learning Models. Migration Letters, 20(S12), 1823–1832. Retrieved from https://migrationletters.com/index.php/ml/article/view/10652

[38] Joel lopes, Arth Dave, Hemanth Swamy, Varun Nakra, & Akshay Agarwal. (2023). Machine Learning Techniques And Predictive Modeling For Retail Inventory Management Systems. Educational Administration: Theory and Practice, 29(4), 698–706. https://doi.org/10.53555/kuey.v29i4.5645

[39] Narukulla, Narendra, Joel Lopes, Venudhar Rao Hajari, Nitin Prasad, and Hemanth Swamy. "Real-Time Data Processing and Predictive Analytics Using Cloud-Based Machine Learning." Tuijin Jishu/Journal of Propulsion Technology 42, no. 4 (2021): 91-102.

[40] Nitin Prasad. (2022). Security Challenges and Solutions in Cloud-Based Artificial Intelligence and Machine Learning Systems. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 286–292. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10750

[41] Varun Nakra, Arth Dave, Savitha Nuguri, Pradeep Kumar Chenchala, Akshay Agarwal. (2023). Robo-Advisors in Wealth Management: Exploring the Role of AI and ML in Financial Planning. European Economic Letters (EEL), 13(5), 2028–2039. Retrieved from https://www.eelet.org.uk/index.php/journal/article/view/1514

[42] Varun Nakra. (2023). Enhancing Software Project Management and Task Allocation with AI and Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1171–1178. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10684

[43] Joel lopes, Arth Dave, Hemanth Swamy, Varun Nakra, & Akshay Agarwal. (2023). Machine Learning Techniques And Predictive Modeling For Retail Inventory Management Systems. Educational Administration: Theory and Practice, 29(4), 698–706. https://doi.org/10.53555/kuey.v29i4.5645

[44] Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[45] Shah, J., Prasad, N., Narukulla, N., Hajari, V. R., & Paripati, L. (2019). Big Data Analytics using Machine Learning Techniques on Cloud Platforms. International Journal of Business Management and Visuals, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[46] Cygan, Kamil J., Ehdieh Khaledian, Lili Blumenberg, Robert R. Salzler, Darshit Shah, William Olson, Lynn E. Macdonald, Andrew J. Murphy, and Ankur Dhanik. "Rigorous Estimation of Post-Translational Proteasomal Splicing in the Immunopeptidome." bioRxiv (2021): 1-24. https://doi.org/10.1101/2021.05.26.445792

[47] Shah, Darshit, Ankur Dhanik, Kamil Cygan, Olav Olsen, William Olson, and Robert Salzler. "Proteogenomics and de novo Sequencing Based Approach for Neoantigen Discovery from the Immunopeptidomes of Patient CRC Liver Metastases Using Mass Spectrometry." The Journal of Immunology 204, no. 1_Supplement (2020): 217.16-217.16. American Association of Immunologists.

[48] Mahesula, Swetha, Itay Raphael, Rekha Raghunathan, Karan Kalsaria, Venkat Kotagiri, Anjali B. Purkar, Manjushree Anjanappa, Darshit

Shah, Vidya Pericherla, Yeshwant Lal Avinash Jadhav, Jonathan A.L. Gelfond, Thomas G. Forsthuber, and William E. Haskins. "Immunoenrichment Microwave & Magnetic (IM2) Proteomics for Quantifying CD47 in the EAE Model of Multiple Sclerosis." Electrophoresis 33, no. 24 (2012): 3820-3829. https://doi.org/10.1002/elps.201200515.

[49] Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[50] Cygan, K. J., Khaledian, E., Blumenberg, L., Salzler, R. R., Shah, D., Olson, W., & ... (2021). Rigorous estimation of post-translational proteasomal splicing in the immunopeptidome. bioRxiv, 2021.05.26.445792.

[51] Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment microwave and magnetic proteomics for quantifying CD 47 in the experimental autoimmune encephalomyelitis model of multiple sclerosis. Electrophoresis, 33(24), 3820-3829.

[52] Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment Microwave & Magnetic (IM2) Proteomics for Quantifying CD47 in the EAE Model of Multiple Sclerosis. Electrophoresis, 33(24), 3820.

[53] Raphael, I., Mahesula, S., Kalsaria, K., Kotagiri, V., Purkar, A. B., Anjanappa, M., & ... (2012). Microwave and magnetic (M2) proteomics of the experimental autoimmune encephalomyelitis animal model of multiple sclerosis. Electrophoresis, 33(24), 3810-3819.

[54] Salzler, R. R., Shah, D., Doré, A., Bauerlein, R., Miloscio, L., Latres, E., & ... (2016). Myostatin deficiency but not anti-myostatin blockade induces marked proteomic changes in mouse skeletal muscle. Proteomics, 16(14), 2019-2027.

[55] Shah, D., Anjanappa, M., Kumara, B. S., & Indiresh, K. M. (2012). Effect of post-harvest treatments and packaging on shelf life of cherry tomato cv. Marilee Cherry Red. Mysore Journal of Agricultural Sciences.

[56] Shah, D., Dhanik, A., Cygan, K., Olsen, O., Olson, W., & Salzler, R. (2020). Proteogenomics and de novo sequencing based approach for neoantigen discovery from the immunopeptidomes of patient CRC liver metastases using Mass Spectrometry. The Journal of Immunology, 204(1_Supplement), 217.16-217.16.

[57] Shah, D., Salzler, R., Chen, L., Olsen, O., & Olson, W. (2019). High-Throughput Discovery of Tumor-Specific HLA-Presented Peptides with Post-Translational Modifications. MSACL 2019 US.

[58] Srivastava, M., Copin, R., Choy, A., Zhou, A., Olsen, O., Wolf, S., Shah, D., & ... (2022). Proteogenomic identification of Hepatitis B virus (HBV) genotype-specific HLA-I restricted peptides from HBV-positive patient liver tissues. Frontiers in Immunology, 13, 1032716.

[59] Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[60] Pavan Ogeti, Narendra Sharad Fadnavis, Gireesh Bhaulal Patil, Uday Krishna Padyana, Hitesh Premshankar Rai. (2022). Blockchain Technology for Secure and Transparent Financial Transactions. European Economic Letters (EEL), 12(2), 180–188. Retrieved from https://www.eelet.org.uk/index.php/journal/article/view/1283

[61] Challa, S. S. S., Chawda, A. D., Benke, A. P., & Tilala, M. (2023). Regulatory intelligence: Leveraging data analytics for regulatory decision-making. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1426-1434. Retrieved from http://www.ijritcc.org

[62] Fadnavis, N. S., Patil, G. B., Padyana, U. K., Rai, H. P., & Ogeti, P. (2021). Optimizing scalability and performance in cloud services: Strategies and solutions. International Journal on Recent and Innovation Trends in Computing and Communication, 9(2), 14-23. Retrieved from http://www.ijritcc.org

[63] Challa, S. S. S., Tilala, M., Chawda, A. D., & Benke, A. P. (2021). Navigating regulatory requirements for complex dosage forms: Insights from topical, parenteral, and ophthalmic products. NeuroQuantology, 19(12), 971-994. https://doi.org/10.48047/nq.2021.19.12.NQ21307

**[64]** Fadnavis, N. S., Patil, G. B., Padyana, U. K., Rai, H. P., & Ogeti, P. (2020). Machine learning applications in climate modeling and weather forecasting. NeuroQuantology, 18(6), 135-145. https://doi.org/10.48047/nq.2020.18.6.NQ20194