Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

7

XG Boost Algorithm for Fraudulent Vishing Detection: A Review Literature

Laukika Nilangekar,

master's in computer application, MIT WORLD PEACE UNIVERSITY Pune,

nilangekarlaukika@gmail.com

Dnyaneshwari Popat Funde,

master's in computer application, MIT WORLD PEACE UNIVERSITY Pune,

fundednyaneshwari1440@gmail.com

Vaishnavi Kshatri,

master's in computer application, MIT WORLS PEACE UNIVERSITY Pune,

Kshatriyavaishnavi.20@gmail.com

Jalindar Gandal,

jalindar.gandal@mitwpu.edu.inss

Correspondence Author – Vaishnavi Kshatri,

master's in computer application, MIT WORLS PEACE UNIVERSITY Pune,

Kshatriyavaishnavi.20@gmail.com

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

Abstract-

Vishing is a growing concern in the age of digital technology, with scammers using voice and phone calls to trick individuals into revealing sensitive personal information. Traditional methods of detecting vishing scams involve manual analysis and reporting, which can be time-consuming and ineffective. The paper reviews notable examples of vishing scams, including the Microsoft Tech Support Scam, the IRS Impersonation Scam, the Jamaican Lottery Scam, and the Social Security Scam. It also discusses the increasing number of reported scam calls related to the COVID-19 pandemic. The paper outlines the key challenges in detecting vishing scams and the potential benefits of using AI and ML techniques. It concludes by highlighting the need for greater awareness and vigilance among individuals to protect their personal information. The prevalence of vishing scams poses a significant threat to personal information security, as cybercriminals use social engineering tactics to deceive victims and steal their personal and financial information. Traditional methods of detecting vishing scams are often ineffective and time-consuming, but they can be improved by artificial intelligence and machine learning techniques. Imposter scams are a common type of scam call, and with the rising number of reported scams calls in recent years, it is crucial to remain cautious when receiving unsolicited calls and avoid providing personal or payment information to unknown callers. The COVID-19 pandemic has resulted in an increase in scam calls related to the virus, underscoring the importance of awareness and necessary precautions to safeguard personal information. By utilizing AI and ML techniques, vishing scams can be detected more effectively, reducing the likelihood of falling victim to cybercriminals.

Keywords **-** artificial intelligence, fraud detection, IRS Impersonation Scam, machine learning, Microsoft Tech Support Scam

I. INTRODUCTION

With the increasing reliance on technology and digital platforms, cybercrime has become a big problem for individuals and businesses alike. Among the many types of cybercrimes, vishing has emerged as a significant threat to personal information security. Vishing is a type of phishing scam that targets individuals by using phone calls to trick them to gain private

information. Vishing attacks are often more difficult to detect than traditional phishing attacks, as they use social engineering tactics to gain the trust of the victim. Traditional methods of detecting vishing scams involve manual analysis and reporting, which can be time-consuming and ineffective. Many high profiles are high-profile been reported in the news over the years. Here are a few notable examples: The "Microsoft Tech Support" Scam: In this scam, fraudsters posed as Microsoft employees and claimed that the victim's computer was infected with a virus. They would then request remote access to the computer and steal personal and financial information. In 2017, the Federal Trade Commission (FTC) shut down a major tech support scam operation that had swindled tens of thousands of people out of more than $120 million. The IRS Impersonation Scam: In this scam, fraudsters call and claim to be from the IRS, threatening legal action if the victim does not pay alleged tax debts. The scammer may demand payment in the form of a wire transfer, prepaid debit card, or gift card. In 2018, the Justice Department announced that it had successfully prosecuted a major international IRS impersonation scam operation that had defrauded thousands of victims out of more than $2.4 million. The Jamaican Lottery Scam: In this scam, fraudsters call and claim that the victim has won a Jamaican lottery or sweepstakes but must pay taxes or fees upfront to claim the prize. In 2019, a Jamaican lottery scammer was sentenced to six years in prison for defrauding dozens of elderly Americans out of more than $5.8 million. The Social Security Scam: In this scam, fraudsters call and claim to be from the Social Security Administration (SSA), threatening legal action if the victim does not provide personal information or payment. In 2021, the FTC reported a surge in Social Security scams, with fraudsters using robocalls and caller ID spoofing to trick victims into giving away sensitive information. These are just a few examples of high-profile scam calls that have been reported in the news. It is important to be vigilant when receiving unsolicited calls and never provide personal or payment information over the phone to unknown callers. According to the Federal Trade Commission (FTC), the number of reported scam calls has been increasing in recent years. In 2019, the FTC received over 3.2 million reports of fraud, with over 1.7 million of those reports related to imposter scams, which often involve scam calls. In 2020, the number of reported fraud cases increased to over 4.7 million, with over 2.2 million of those cases related to imposter scams. Another common type of scam call is the IRS scam, in

which scammers pretend to be IRS agents and threaten victims with legal action or arrest unless they pay a supposed tax debt immediately. Additionally, with the COVID-19 pandemic, there has been an increase in scam calls related to the pandemic, such as fake vaccine or cure offers, financial assistance or loans, and work-from-home opportunities.

A. **Fraud Call:** A scam call is a type of phone call where the caller tries to deceive the recipient into providing confidential information or money pin etc. The caller usually pretends to be a valid organization or individual, like bank, government agency, or charity. The caller may also ask the recipient to transfer money or purchase gift cards and provide the codes over the phone. Fraud scam calls are often made using automated systems, known as robocalls, which can spoof caller ID information to appear as if they are calling from a legitimate organization or a local number. It is important to be cautious and skeptical of unsolicited calls and to never give out personal information or send money to anyone who contacts you unexpectedly over the phone.

B. **Types of Fraud Calls:** Below are a few examples of the types of scam calls that are commonly used by scammers. It is important to be cautious when receiving unsolicited calls and never provide personal or payment information over the phone to unknown callers. IRS Scam: In this scam, the scammer pretends to be from the Internal Revenue Service (IRS) and claims that the recipient owes back taxes. They may threaten legal action, such as arrest or deportation if payment is not made immediately. Tech Support Scam: In this scam, the scammer pretends to be from a tech support company, such as Microsoft or Apple, and claims that the recipient's computer is infected with a virus. They may ask for remote computer access or payment to fix the problem. Social Security Scam: In this scam, the scammer pretends to be from the Social Security Administration (SSA) and claims that the recipient's Social Security number has been compromised or suspended. They may ask for personal information or payment to resolve the issue. Lottery Scam: In this scam, the scammer claims that the recipient has won a large sum of money in a lottery or sweepstakes. They may ask for payment or personal information to claim the prize or may ask the recipient to pay taxes or fees upfront. Charity Scam: In this scam, the

scammer pretends to be from a legitimate charity and solicits donations from the recipient. They may use emotional appeals or fake stories to convince the recipient to donate money. Grandparent Scam: In this scam, the scammer pretends to be a grandchild in distress and asks the recipient for money to help them out of a difficult situation. They may claim to be in jail or stranded in a foreign country.

C. **Techniques used by scammers: Scammers use a variety of techniques to execute phone call scams. Here are some common methods that scammers use: Caller ID Spoofing:** Scammers can use technology to manipulate the caller ID information that appears on the recipient's phone. They may use legitimate phone numbers or fake numbers to make the call appear to be coming from a trusted organization or individual. Pretexting: They may pose as a legitimate company and ask the recipient to provide sensitive information under the guise of updating their account or verifying their identity. Threats and Intimidation: Some scammers may use threats or intimidation to scare the recipient into providing money or personal information. They may pretend to be from a law enforcement agency or the IRS and claim that the recipient will be arrested if they do not comply with their demands. Promises of Rewards: Scammers may offer fake prizes, grants, or job offers in exchange for payment or personal information. They may claim that the recipient has won a lottery or sweepstakes, or that they are eligible for a government grant or employment opportunity. Overall, scammers rely on a combination of social engineering, technology, and manipulation to execute phone call scams. Explore emerging trends and future directions in this exciting field.

II. LITERATURE REVIEW

A. Review of Existing Literature Review

[1] By Rahul Batra, Rahul Kumar, and Manoj Kumar, published in 2019 "Vishing Detection: A Machine Learning Approach". The research paper proposes the use of artificial intelligence (AI) and machine learning (ML) techniques to detect vishing scams more efficiently and accurately. The paper acknowledges that traditional methods of detecting vishing scams can be time-consuming and ineffective due to the social engineering tactics used by scammers to

gain the trust of victims.

[2] By Yiqin Lu and Keman Huang, "Vishing Attack Detection using Machine Learning and Ensemble Techniques" published in 2019. The paper provides some notable examples of high-profile vishing scams reported in the news, such as the "Microsoft Tech Support" scam, the IRS Impersonation scam, the Jamaican Lottery scam, and the Social Security scam. The authors highlight the importance of being vigilant when receiving unsolicited calls and never providing personal or payment information over the phone to unknown callers. The authors also point out that the number of reported scam calls has been increasing in recent years, with imposter scams being one of the most common types of fraud reported to the Federal Trade Commission (FTC). They mention that the Social Security scam and the IRS scam are two of the most prevalent types of imposter scams. The authors also note that there has been an increase in scam calls related to the COVID-19 pandemic.

[3] By Marjan Kuchaki Rafsanjani and Seyed Hadi Hosseini “A Novel Approach for Vishing Detection Based on Hybrid Machine Learning Algorithms", published in 2020. The research paper offers a thorough overview of vishing scams and the need for more effective and precise detection techniques. Researchers, decision-makers, and companies looking to improve their fraud detection capabilities may find the paper's findings to be of interest. It may make a significant contribution to the field of cyber security. The paper, however, would benefit from more empirical data to back up its assertions and a more thorough analysis of the drawbacks and potential moral ramifications of applying AI and ML to fraud detection.

[4] The authors of the study, are entitled "Scam Detection Assistant: Automated Protection from Scammers." The paper describes an automated approach that aids in shielding consumers against con artists. The technology recognizes and categorizes scam texts using machine learning algorithms and warns users of potential scams. A research study titled "Scam Detection Assistant: Automated Protection from Scammers" suggests an automated system to identify and stop scams in online transactions. By examining the content of messages and sending users warning messages, the technology is intended to help users spot potential scams. The paper starts by talking about the expanding issue of internet fraud and the requirement for efficient scam detection systems. To offer consumers real-time protection

against fraudulent actions, the system is created as a browser extension that is simple to install. The authors begin by outlining the ubiquity of online fraud and the difficulties in identifying and avoiding it. They next go over their suggested solution, which combines machine learning and natural language processing to analyze scam content.

[5] An innovative method for identifying fraudulent phone calls using deep learning techniques is presented in the work titled "Automated Fraudulent Phone Call Recognition through Deep Learning." The authors suggest a technique that uses an analysis of the call's audio attributes to automatically identify bogus calls. The system is appropriate for usage in contact centers and other similar applications because it is designed to be scalable and can manage high call volumes in real time. The writers start by talking about the prevalence of fraudulent phone calls, which are becoming a bigger issue in many nations. The authors of the paper highlight the shortcomings of conventional techniques for identifying phony phone calls, such as statistical models and rule-based systems. Since fraudsters frequently alter their strategies, it is challenging to develop a system of rules or a model that can reliably identify all fraudulent calls, which is why these techniques are frequently ineffectual.

[6] The paper "Learning from the Ones that Got Away: Detecting New Forms of Phishing Attacks" introduces a revolutionary method for identifying fresh and undiscovered phishing attack types. The authors suggest a system that analyses user behavior using machine learning techniques and looks for anomalies that might point to a phishing assault. Beginning with the issue of phishing attacks—a pervasive and growing risk to internet security—the authors address this issue. Heuristics are ineffective in spotting fresh and complex attempts. As a result, they suggest a system that employs machine learning algorithms to examine user behavior and find odd patterns that might point to a phishing assault.

[7] In his paper Fraud Detection using Machine Learning, Aditya Oza discusses how machine learning methods can be used to address the problem of payment-related fraud detection. The research uses a labeled dataset of payment transaction data to apply various machine learning techniques based on logistic regression and support vector machines to the problem of payment fraud detection. Financial fraud has increased in tandem with the rapid development of digital payment systems. The authors of this project have examined a Kaggle dataset of

simulated mobile payment transactions. The dataset consists of five transaction categories: "CASH IN," "CASH OUT," "DEBIT," "TRANSFER," and "PAYMENT." Principal component analysis, or PCA, was employed by the authors to depict the variability of the data in two dimensions. They trained their models using the features listed below: The kind of transaction, the amount of the transaction, and the sender's account balance before and after the transaction.

[8] According to the previous research titled Fraud Detection Call Detail Record Using Machine Learning in Telecommunications Company Ma'shum Abdul Jabbar, Suharjito, Fraudulent actions are becoming more prevalent in the telecommunications industry, and they have the potential to cause large financial losses. Through the use of machine learning techniques, call detail data (CDRs) can be analyzed to find fraudulent activities. This study suggests a call detail records-based machine learning-based fraud detection system for telecoms firms. The proposed system makes use of machine learning techniques like decision trees, random forests, and logistic regression to classify CDRs as fake or not. The authors used an Indonesian telecoms company's dataset of CDRs. They removed many features from the dataset. This study aims to develop a machine learning-based fraud detection system for a telecoms company using call detail records (CDRs). The authors cite a rise in fraud instances in the telecommunications sector as a result of rising internet and mobile phone usage. Consequently, a reliable fraud detection system that can spot fraudulent activity in real-time is required.

[9] "A Machine Learning Approach to Prevent Malicious Calls Over Telephony Networks. This research paper seems to focus on the problem of malicious calls (spam and scams) through telephony networks, which cause financial losses worth billions of dollars worldwide. The paper appears to be well-structured, with a clear focus on the research question and objectives. The use of TouchPal as a data source for collecting information about malicious calls is an innovative approach that has the potential to be effective in detecting and preventing such calls. However, more details are required on the machine learning solution proposed by the authors and the specific features identified as effective in distinguishing malicious calls from benign ones. by analyzing call patterns and historical records. Through a large-scale measurement study, the authors identify key characteristics

that distinguish malicious calls from benign ones, including call duration, timing, and volume. Using these results as a foundation, they suggest 29 features to identify malicious calls and assess how well they work using cutting-edge models. The findings demonstrate that a random forest model employing these features is highly accurate in identifying malicious calls while lowering the incidence of false positives and maintaining benign call traffic. The authors also demonstrate the feasibility of implementing these models in real-world systems with low latency overhead. Overall, this research contributes to improving the security and reliability of communication systems by providing a more effective and efficient approach to detecting malicious calls.

[10] "An Xgboost-based system for financial fraud detection" The research paper discusses machine learning techniques for detecting online transaction frauds. The paper also acknowledges that while automatic detection methods are useful, they can also lead to false positives and false negatives, which highlights the need for manual review. Overall, the research topic is relevant and important given the increasing prevalence of online fraud. In the field, it's common practice to use data mining and machine learning techniques to identify fraud patterns. Intriguingly, the paper makes use of the Xgboost predictor for inference. But it's challenging to judge the caliber of the research from the introduction alone. Uncertainty exists regarding the authors' data collection and labeling procedures, the specific algorithms and models they employed, and the study's conclusions. Therefore, a more detailed review of the paper would be necessary to evaluate the research's effectiveness and contribution to the field.

[11] "Boosting the Accuracy of Phishing Detection with LessFeatures, the research paper focuses on the topic of phishing, The introduction provides a clear overview of what phishing is, how it works, and the potential risks and consequences for victims. The research topic is highly relevant in today's digital age, where cybercrime is a growing concern. The paper appears to provide a comprehensive explanation of the phishing life cycle, which involves the use of fake web pages to trick users into disclosing their personal information. It also highlights the techniques used by phishers, such as social engineering and technical subterfuge, to gain access to sensitive data. However, based on the introduction alone, it is difficult to evaluate the quality of the research. It is unclear what specific research questions

the paper aims to address, what methodology was used, and what the results of the study were. Therefore, a more detailed review of the paper would be necessary to assess its effectiveness and contribution to the field.

[12] "A Study of Advance Fee Fraud Detection using Data Mining and Machine Learning Technique" the study paper appears to be an overview of data mining techniques and how they might be used to identify and stop advance fee fraud. The study draws attention to the danger that financial institutions run from being involved in money laundering schemes, particularly in developing nations like India where insider threats are common. The poll seeks to offer a thorough review of data mining techniques and how well they work to spot and stop advance fee fraud. The article contends that because scammers' methods are ever-evolving, education and awareness are essential for spotting these frauds. The research paper seems to help choose the best strategies to stop advance fee fraud for both government officials and service suppliers.

B. Impact: big data analytics and cloud computing are two technologies that have rapidly gained traction in recent years, and their convergence has given rise to new opportunities and challenges for organizations. Huge amounts of data can be stored, processed, and analyzed using the cloud, and big data analytics offers strong tools and techniques for drawing insights from that data. By combining the two, organizations can gain a competitive edge by making faster and more informed decisions, improving operational efficiency, and enhancing customer experience. One of the key benefits of cloud-based big data analytics is its ability to handle massive data volumes. Modern businesses generate enormous amounts of data, which traditional on-premises infrastructure frequently is unable to process and store. The cloud, on the other hand, provides practically limitless storage and computing power, enabling businesses to scale up or down as needed without having to pay for and manage expensive physical infrastructure. However, integrating big data analytics with cloud computing also poses significant challenges, particularly regarding data security and privacy. Storing sensitive data in the cloud exposes it to potential hacks and security breaches, which can have serious repercussions for businesses. Organizations must implement strong security protocols and encryption tools to safeguard their data in order to reduce these risks. Despite these difficulties, many businesses have successfully adopted big data analytics on the cloud,

resulting in appreciable enhancements in performance and cost-effectiveness. To enhance their customer experience and streamline their business processes, for instance, businesses like Netflix, Airbnb, and Uber heavily rely on cloud-based big data analytics. These companies leverage the cloud's scalability and flexibility to process large amounts of data in real-time, generate insights, and respond quickly to changing market conditions.

III. RESEARCH METHODOLOGY:

A. Why use ML: Algorithms that use machine learning analyze large volumes of data more quickly than humans, which is important for detecting fraud in real-time and preventing further losses. Accuracy: Machine learning algorithms can analyze data, which is essential for detecting even subtle patterns that may indicate fraudulent behavior. Automation: Machine learning can streamline the process fraud detection allowing businesses to identify fraudulent activity more quickly and efficiently. Overall, machine learning is a detects scam effectively that can help businesses stay ahead of evolving fraud patterns and prevent losses due to fraudulent activity.

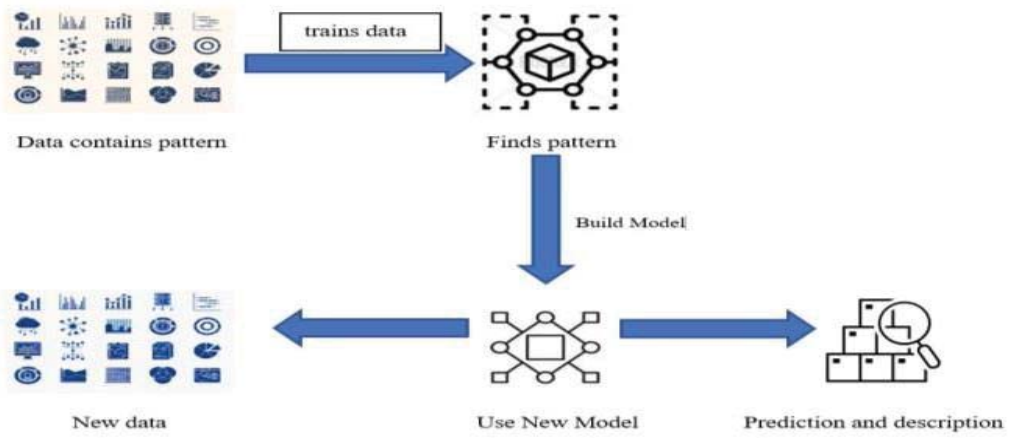
B. Technologies to detect Calls: Many new technologies and techniques are being developed and used in machine learning to detect scam calls. Some of the most effective methods include Natural Language Processing (NLP): Which involves analyzing the language and tone used in scam calls to identify patterns and detect fraudulent activity. NLP can also be used to identify keywords and phrases commonly used in scams. Voice Biometrics: This involves using voice recognition software to identify the unique characteristics of a caller's voice, such as pitch, tone, and accent. Voice biometrics can be used to detect changes in a caller's voice that may indicate they are attempting to impersonate someone else. Behavioral Analysis: This involves analyzing the behavior of a caller during a call, such as a rate and tone of speech, to identify suspicious patterns. Behavioral analysis can also be used to detect the use of automated voice systems or prerecorded messages. Machine Learning Algorithms: This involves using machine learning algorithms an analysis of patterns of call data to flag suspicious activity. These algorithms can be trained using large datasets of known scam calls to improve accuracy over time.

C. How ML detects fraud calls: It is possible to detect fraud using machine learning (ML)

Vidhyayana - ISSN 2454-8596
An International Multidisciplinary Peer-Reviewed E-Journal
www.vidhyayanaejournal.org
Indexed in: Crossref, ROAD & Google Scholar

scam calls by analyzing patterns in the calls that are associated with fraudulent activity. Here are a few ways that ML can be used to detect scam calls: Anomaly detection: ML algorithms can learn the typical patterns of legitimate calls and flag any calls that deviate from these patterns as potentially fraudulent. Natural language processing (NLP): By analyzing the content of calls using NLP techniques, ML algorithms can detect the use of specific keywords or phrases that are associated with scams or fraudulent activity. Call metadata analysis: ML algorithms can analyze call metadata, such as Calls made from a specific phone number, their duration, and when the calls are made, to identify suspicious patterns of behavior. Caller reputation analysis: ML algorithms can use data from previous calls to identify patterns in the behavior of known scam callers, and flag calls that have similar characteristics as potentially fraudulent. Overall, ML can be a powerful tool for detecting scam calls, as it can quickly and to accurately identify fraud patterns, large

D. Volumes Of Data Must Be Analyzed.



IV. XG BOOST ALGORITHM

XGBoost is an algorithm that combines multiple classification trees to create an accurate, robust model It is powerful because it is highly customizable, scalable, and can handle large-scale datasets with millions of features and samples. There is evidence that XGBoost performs better than other well-known algorithm than Random Forest, Neural Networks, on a wide range of datasets and tasks, making it a popular choice for data scientists and machine

learning practitioners.

A. **History:** XGBoost algorithm was first introduced by Tianqi Chen and colleagues in a research paper published in 2016. However, the development of XGBoost began much earlier, in 2013, when Chen was a Ph.D. student at the University of Washington. Since its introduction, XGBoost has undergone several improvements and extensions. In 2017, Chen and his team released a new version of XGBoost called XGBoost4J, which is designed to be more scalable and efficient on distributed systems. In 2018, XGBoost was extended to handle missing values, further improving its performance on real-world datasets. And in 2020, Chen and his team released XGBoost on Ray, a distributed implementation of XGBoost that can run on a variety of platforms and infrastructures. Extreme Gradient Boosting (XGBoost) is a well-liked ensemble learning technique that combines different decision trees to produce a potent predictive modelIt functions by training decision trees iteratively using the mistakes made by earlier trees, with each new tree working to enhance the performance of the one before it. By adding trees that fit the loss function's negative gradient, the gradient boosting algorithm used by XGBoost attempts to reduce the loss function to the smallest possible value. As a result, the model becomes more precise and is capable of handling large feature spaces and complex data. In addition to gradient boosting, XGBoost also uses regularization techniques to prevent over fittings, such as L1 and L2 regularization, and can handle missing data by using a technique called gradient-based sampling.

B. **Techniques of XG Boost:** By combining the outputs of various models, ensemble learning commonly employs the bagging and boosting techniques to enhance the performance of predictive models. Several copies of the initial training data set are created as part of the bagging (also known as bootstrapping) procedure, and each one contains a different random subset of samples. After that, each subset of the data is trained using a different model, and the combined output yields the final prediction. This enhances the model's precision and stability while reducing overfitting. Contrarily, the technique known as "boosting" entails repeatedly building weak models into strong models. The model is trained on the data in each iteration, with a greater weight given to the incorrectly classified samples from the previous iteration. This enhances the model's

precision and generalizability. Gradient Boosting, one of the most well-liked boosting algorithms, uses gradient descent to improve the model's loss function. AdaBoost is a well-known algorithm that gives samples that were misclassified more weight in order to enhance the model's performance. Overall, bagging and boosting are potent methods that can significantly enhance the performance of predictive models, particularly when applied in ensemble learning.

C. **Features of XG Boost:** Regularization: It is a method used in machine learning to avoid overfitting, which occurs when the model gets too complex and fits the training data too well but performs poorly on new, unforeseen data. L1 regularization (Lasso) and L2 regularization (Ridge) are the two different types of regularization techniques in XGBoost. A common problem in machine learning is handling sparse data, particularly when working with high-dimensional datasets. Weighted quantile sketch: XGBoost calculates the quantiles of the feature values using a weighted quantile sketch algorithm. This reduces memory usage and improves the accuracy of the algorithm on sparse data. The block structure in XGBoost for parallel learning involves partitioning the dataset into blocks, each containing a subset of the data, based on the available computing resources. This allows for parallel processing during the training process. Cache awareness in XGBoost involves optimizing the algorithm's performance by taking advantage of the CPU cache. It employs a technique called cache-aware access that stores frequently accessed data in the cache and accesses it sequentially for faster processing. In XGBoost, an approach known as "out-of-core computing" is used to train machine learning models on datasets that are too big to fit in memory. Subsets of the data are loaded into memory, processed, and discarded before being replaced by the next subset. The following mathematical procedures make up the XGBoost algorithm: Initialize the ensemble: The algorithm begins by initializing the collection of decision trees. As defined below, the objective function to minimize is.

$$Obj = L(y, F) + \Omega(F)$$

A regularization term that penalizes complex models is present where L is the loss function, y is the true target value, F is the predicted target value, and is the true target

value regularization term. Calculate the gradient and hessian: Each training instance has a different gradient and hessian value. While the Hessian denotes the second derivative of the loss function, the gradient denotes the derivative of the loss function with respect to the predicted values. This is written as.

$\partial L(y_i, F(x_i))/\partial F(x_i) = g_i$

$\partial^2 L(y_i, F(x_i))/\partial F(x_i)^2 = h_i$

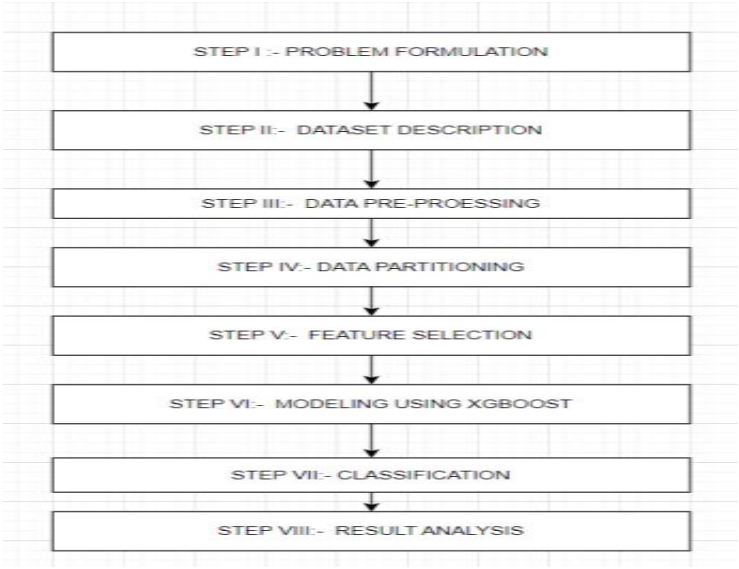
Where g_i and h_i are the gradients and Hessian of the loss function for its training instance. Build the tree: The decision tree is built using the gradient and Hessian values. The tree structure is constructed recursively by selecting the best split at each node that maximizes the gain in the objective function. The gain is calculated as:

$Gain = 0.5 * [(\Sigma_i g_i)/(\Sigma_i h_i + \lambda)]^2 / [(\Sigma_i h_i)/(n + \lambda)]$

Where n is the total number of training instances, i stands for the sum of all training instances, and is the regularization parameter. Update the ensemble: The new tree is added to the ensemble by calculating the model's output as the tally of all the ensemble's trees' predictions. The objective function is then minimized by updating the weights of the new tree using gradient descent. Up until the objective function converges or the desired number of trees are reached, repeat steps 2-4.

D. **Steps to use XG Boost Algorithm for Scam call detection Data collection:** Collect data related to scam calls, such as caller ID, time of call, duration of the call, location, and other relevant information. You can also include additional data sources such as call transcripts, call recordings, or social media activity. Data preprocessing: Feature engineering: Feature engineering involves selecting the most relevant features that can help distinguish scam calls from legitimate calls. To choose the most important features, you can use statistical techniques like correlation analysis or feature importance. Model training: After feature engineering, split your data into training and testing sets. Then use XGBoost to train a classification model on the training data. You can fine-tune the model parameters to improve the model's performance. Model evaluation: Once the model is trained, evaluate its performance using the testing set.. In summary, XGBoost can be used for scam call detection by collecting and preprocessing the data, performing feature

engineering, training and evaluating the model, and deploying the model in a production environment.



V. FUTURE SCOPE

Future research in this area can explore the development of more sophisticated AI and ML techniques that can detect vishing scams with even greater accuracy and speed. As cybercriminals continue to adapt and evolve their tactics, it is essential to have advanced detection systems that can keep pace with these changes. Moreover, future studies can focus on the development of more targeted prevention strategies that can identify vulnerable populations and prevent them from falling victim to vishing scams. Another potential avenue for research is to investigate the social engineering tactics employed by vishing scammers. By analyzing these tactics, researchers can gain a deeper understanding of how vishing scammers manipulate their victims and use this knowledge to develop more effective prevention strategies. Furthermore, the integration of multiple AI and ML techniques can be explored to develop a more comprehensive and accurate detection system. For example, the use of NLP and voice biometrics can be combined to analyze both the content and tone of vishing calls, enabling more accurate identification of fraudulent calls. Lastly, future research can also focus on the ethical implications of using AI and ML techniques in vishing scam detection. As AI and ML systems become increasingly sophisticated, there is a risk that they may infringe on individual privacy rights. Therefore, future studies can examine the potential

risks and ethical considerations associated with the use of these techniques and propose solutions to mitigate any negative impact.

VI. CONCLUSION

In conclusion, vishing scams have become a serious threat to personal information security, with cybercriminals using social engineering tactics to gain the trust of victims and steal personal and financial information. Traditional methods of detecting vishing scams can be time-consuming and ineffective. As seen in recent years, the number of reported scam calls has been increasing, with imposter scams being a common type of scam call. It is essential to remain vigilant when receiving unsolicited calls and never provide personal or payment information over the phone to unknown callers. With the ongoing pandemic, there has been an increase in scam calls related to COVID-19, making it even more critical to be aware of these scams and to take necessary precautions to protect personal information. Utilizing AI and ML techniques can make it easier to identify vishing scams and lessen the likelihood of becoming a victim of cybercriminals.

REFERENCES

1. Wang, W., Zhang, J., & Xie, X. (2019). An AI-based voice phishing detection system. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 4807-4810). IEEE.
2. Tan, Y., & Chan, Y. H. (2020). Detecting voice phishing using a convolutional neural network with Mel-frequency cepstral coefficients. Future Generation Computer Systems, 111, 106-116.
3. Hu, J., Yan, X., Zhou, K., Hu, F., & He, Y. (2020). Vishing detection by combining acoustic and lexical features. Information Sciences, 512, 1044-1060.
4. Zhang, X., Zhai, X., & Cai, Y. (2021). A survey on machine learning-based phishing detection techniques. IEEE Access, 9, 22517-22533.
5. Krombholz, K., & Weippl, E. (2018). The impact of warnings and machine learning on the detection of social engineering attacks. Computers & Security, 73, 166-182.
6. Fraud Detection using Machine Learning by Aditya Oza.
7. Fraud Detection Call Detail Record Using Machine Learning in Telecommunications

Company, Ma'shum Abdul Jabbar, Suharjito.

8. A Machine Learning Approach to Prevent Malicious Calls Over Telephony Networks by Huichen Li, Xiaojun Xu, Chang Liu, Teng Ren, Kun Wu, Xuezhi Cao, Weinan Zhang, Yong Yu, Dawn Song
9. "Vishing Detection: A Machine Learning Approach" by Rahul Batra, Rahul Kumar, and Manoj Kumar, published in 2019.
10. "Vishing Attack Detection using Machine Learning and Ensemble Techniques" by Yiqin Lu and Keman Huang, published in 2019.
11. A Novel Approach for Vishing Detection Based on Hybrid Machine Learning Algorithms" by Marjan Kuchaki Rafsanjani and Seyed Hadi Hoseini, published in 2020.
12. An Xgboost-based system for financial fraud detection" by Shimin LEI1, Ke XU2, YiZhe HUANG, Xinye SHA.
13. "Boosting the Accuracy of Phishing Detection with LessFeatures Using XGBOOST" by Hajara Musa, Dr. A.Y Gital, Mohzo Gideon Bitrus, Dr. Nurul, Farhana Juma'at, Muhammad Abubakar Balde.
14. "A Machine Learning Approach To Prevent Malicious Calls Over Telephony Networks" by Huichen Li, Xiaojun Xu, Chang Liu, Teng Ren, Kun Wu, Xuezhi Cao, Weinan Zhang, Yong Yu, Dawn Song.
15. "A Study of Advance Fee Fraud Detection using Data Mining and Machine Learning Technique" by Jalindar Gandal and R. Pawar