Babylonian Journal of Machine Learning
Vol.2023, **pp**. 65–72
DOI: https://doi.org/10.58496/BJML/2023/011; ISSN: 3006–5429
https://mesopotamian.press/journals/index.php/BJML

Research Article

Detecting attacks in banks by cyber security: an applied study

Hadeel M Saleh [1,*], Abdulrahman Kareem Oleiwi[1], Ahmed Abed Hwaidi Abed[1]

[1] *Center for Continuing Education, Anbar University, Iraq.*

ARTICLE INFO

Article History
Received 20 Aug 2023
Accepted 22 Oct 2023
Published 14 Nov 2023

Keywords
Detecting attacks
cyber security
Machine learning
Security

ABSTRACT

In the present era, digital banks are more defenseless to cyber banks due to their banking responses. With the increasing reliance on the maximum in digital business operations, these targets have become attractive targets for people who want to exploit personal and financial information. This aims to explore how cyber security can be used to reveal passwords in banks. By using several choices such as usage detection (IDS), behavior study, and learning algorithms, this study is current in identifying chances before major damages. By studying the CICIDS 2017 dataset, we highlight this study on the real application of Random Forest algorithm to enhance security levels in banks. The outcomes emphasize the need for continuous asset in varied cyber safety and employee training Francisco x Late Cyber.

1. INTRODUCTION

Banks are amongst the most defenseless organizations to cyber attacks, owing to the sensitive countryside of the data they get. With the increasing dependence on digital method in banking processes, these organizations have develop an attractive target for attackers seeking to get customer information, such as financial and personal data. Rendering to numerous reports, cyber attacks on banks can outcome in huge financial sufferers, as well as negatively influence the status of organizations and customer trust [1].

Cyber security in this background goals to protect systems and networks from growing threats by adopting advanced strategies and techniques to detect breaches. These methods include the use of tools and techniques such as intrusion detection systems (IDS), behavioral analysis, and machine learning, which contribute to enhancing the ability to detect attacks before they cause significant damage [2].

Banks' response to cyber tests requires the incessant facility of cyber security approaches and incessant training for employees, in adding to evolving effective safety policies [3][10]. Banks essential to be able to face new and continually developing threats, by applying advanced approaches based on careful examination of data and potential threats[11]. In this paper, we will analysis how cyber security methods are used to detect attacks in banks, directing on practical submissions and algorithms used to improve security. A specific data set resolve also be examined to test the efficiency of these methods and amount their outcomes [4-7].

2. LITERATURE REVIEW

Earlier revisions show in table 1 growing interest in detecting attacks in banks concluded the use of cyber security methods. This section offers an analysis of some of the prominent revisions in this area, including a table brief each revision with an analysis of the outcomes.

*Corresponding author. Email: Hadeel.mohammed@uoanbar.edu.iq

TABLE I. SUMMARY OF PREVIOUS STUDY

Researchers	Year of Publication	Title of the Study	Objective	Key Findings	Source
Anderson, R.	2020	Security Engineering: A Guide to Building Dependable Distributed Systems	Enhance understanding of how to design security systems in banks.	Improved response to attacks by integrating Intrusion Detection Systems (IDS).	[3]
Stallings, W.	2018	Computer Security: Principles and Practice	Review the principles of cybersecurity and its applications in financial institutions.	Use of machine learning techniques to detect threats such as phishing and malware.	[2]
Ponemon Institute	2021	Cost of a Data Breach Report	Analyze the impact of data breaches on banks.	Investment in cybersecurity significantly reduces costs associated with breaches.	[1]
Von Solms, R. & Van Niekerk, J.	2013	From Information Security to Cyber Security	Explore the shift in security strategies from information security to cybersecurity.	Emphasized the importance of behavior analysis for detecting abnormal activities.	[4]
Kaspersky Lab	2021	IT Security Risks Survey	Study the security risks faced by financial institutions.	Cyber threats require the adoption of advanced techniques like AI for data analysis.	[8]
Zarpelão, B., et al.	2017	A Survey on the Usage of Honeypots in the Security of Web Applications	Review the use of honeypot systems in banking application security.	Used honeypots to analyze attacker behavior and improve defense strategies.	[9]
Arachchilage, N. A. G.	2020	The Influence of Social Media on Information Security in the Banking Sector	Analyze the impact of social media on cybersecurity in banks.	Increasing customer awareness about cybersecurity risks through social media helps in protection.	[5]

3. CHALLENGES OF DETECTING BREACH IN BANKS USING CYBER SECURITY

Banks face several important challenges when annoying to detect cyber weaknesses and enhance their safety:

- Growing cyber threats: The danger landscape is changing quickly, requiring security measures to be regularly updated to address evolving attack methods.
- Integrating advanced technology: Assimilating technologies such as machine learning and artificial intelligence faces defies related to complexity and the need for effective training of employees.
- Privacy concerns: The use of advanced analytics raises concerns about protecting customer data, requiring a balance between security and compliance.
- Resource allocation: Banks may find it difficult to allocate sufficient resources for comprehensive security due to budget constraints.
- Customer awareness: Many customers lack an adequate understanding of cybersecurity risks, leaving them vulnerable to breaches.
- Sophistication of security systems: The more banks use advanced security systems, the more complex their management and maintenance become, which can impact effective operations.

These challenges require banks to adopt advanced strategies to enhance their cybersecurity and detect breaches effectively.

4. METHODOLOGY

The research methodology is crucial for determining how to collect data, analyze it, and test theories related to detecting intrusions in banks using cyber security. This method includes several steps in order to ensure the accuracy and reliability of the results. The main steps are as follows (refer to figure 1):

1. Objective: This study goals to grow a model for sensing cyber gaps in banks using cyber security methods, aiming on the Random Forest algorithm for data analysis.
2. Dataset Selection: The CICIDS 2017 dataset from the Canadian Centre for Cyber Security, which contains logs of various attacks (e.g., DDoS, HTTP Flood, Brute Force) and normal network activities.
3. Data Collection:
   - Download the dataset.
4. Initial Data Analysis:
   - Data Cleaning: Process the remove anomalous data and missing values using methods like elimination or imputation.
   - Exploratory Analysis: Utilize statistical tools to comprehend patterns, trends, and relationships in the data.
5. Data Segmentation:
   - Split the dataset into two sets:
     - Training Set (70%): Used to train the model.
     - Test Set (30%): Used to evaluate model performance.
6. Model Selection:
   - Choose the Random Forest algorithm due to its capability to handle large datasets and diverse patterns effectively.
7. Model Training:
   - Use the training set to train the Random Forest model, including identifying important features and determining the number of decision trees to use.
8. Model Testing:
   - Test the model on the test set, measuring performance using accuracy, precision, recall, and error rate.
9. Results Analysis:
   - Analyze extracted data to evaluate the model's effectiveness in detecting threats and compare its performance with other models like SVM and Decision Trees.
10. Providing Recommendations:
    - Offer suggestions to enhance cybersecurity strategies in banks based on the results.
11. Documentation and Review:
    - Document all results and procedures, followed by a final review to ensure the accuracy and consistency of the information.

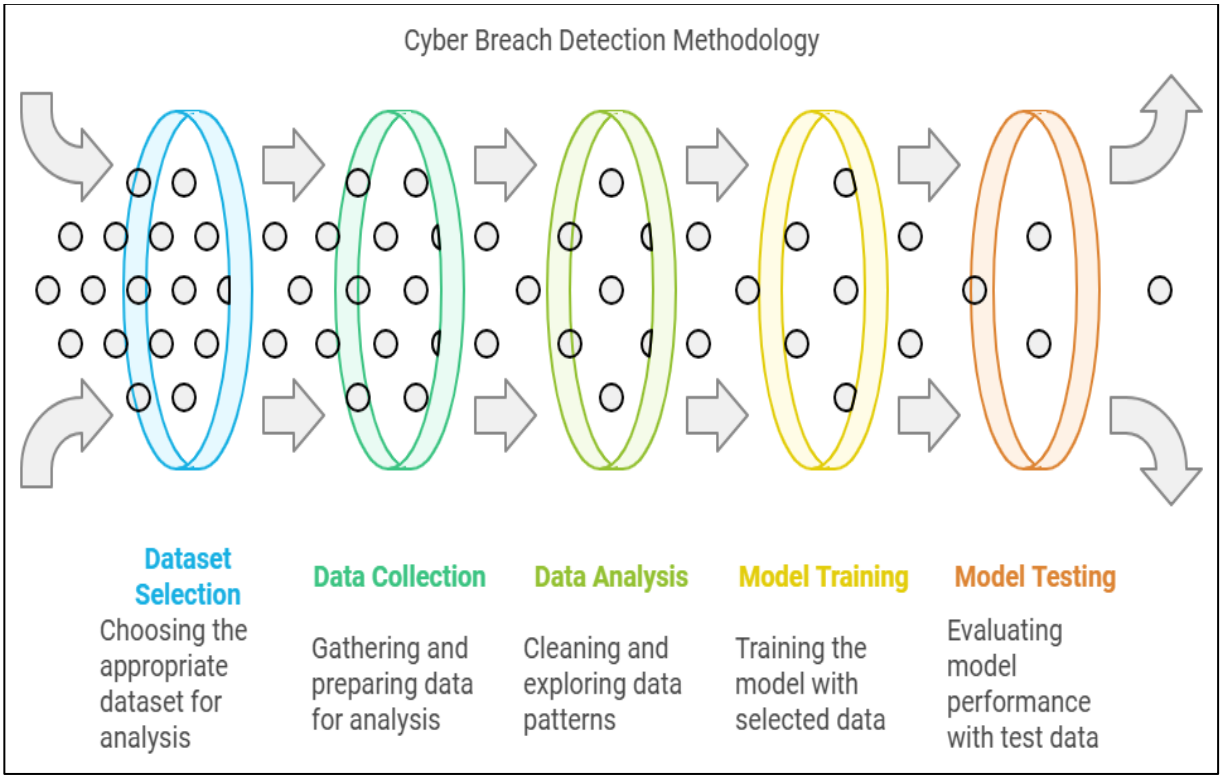

Fig. 1. Methodology steps

4.1 Algorithms

1. Random Forest Algorithm. As shown below.

Steps:
1. Data Collection: Load and prepare the dataset. **2. Data Partitioning: Split the data into two groups: Training set (70%) and Test set (30%).** **3. Tree Generation:** **o Choice a specific number of trees (e.g., 100 trees).** **o For each tree, a arbitrary sample of the data is selected.** **o For each divided point, a random subset of features is selected.** **4. Model Training: Construct each tree based on the random data.** **5. Results Aggregation: When predicting, the best common classification is taken from all the trees.** **6. Model Testing: Evaluate the performance of the model on the test set using metrics such as accuracy.**

2. Support Vector Machines (SVM) Algorithm. As shown below.

Steps:
1. Data Collection: Load and prepare the data. 2. Data Partitioning: Split the data into training and test sets. 3. Kernel Function Selection: Determine the type of kernel function (e.g., linear or nonlinear) that is appropriate for the data. 4. Model Training: Using the training set to teach the model how to classify data by creating a hyper plane between classes. 5. Model Testing: Using the test set to evaluate the performance of the model based on prediction accuracy.

3. K-Nearest Neighbors (KNN) Algorithm. As shown below.

Steps:
1. Data Collection: Load and prepare the dataset. 2. Data Segmentation: Split the data into a training and test set. 3. K Value Determination: Select the number of neighbors (K) to be used. 4. Model Training: No actual training, the training set is saved. 5. Model Implementation: o When new data is entered, the distance to all points in the training set is calculated and K nearest neighbors are selected. 6. Class Determination: The point is classified based on the most common class among the neighbors.

4. Naive Bayes Algorithm. As shown below.

Steps:
1. Data Collection: Prepare and load the dataset. 2. Data Segmentation: Split the data into a training and test set. 3. Probability Calculation: o Calculate the conditional probabilities of each feature for each class. o Calculate the initial probabilities for each class. 4. Model training: Use the training set to teach the model based on the calculated probabilities. 5. Model application: When new data is entered, conditional probabilities are calculated and the class with the highest probability is selected. 6. Model testing: Evaluate the performance of the model using the test set.

5. Artificial Neural Networks (ANN) Algorithm. As shown below.

Steps:
1. Data collection: Prepare the required data set. 2. Data segmentation: Divide the data into training and test sets. 3. Network design: Determine the number of layers and nodes in the network. 4. Model training: o Use the training set to teach the network. o Apply an algorithm such as Back propagation to update the weights based on errors. 5. Model testing: Evaluate the performance of the model using the test set and analyze the accuracy of the results.

6. Decision Trees Algorithm. As shown below.

Steps:
1. Data collection: Prepare and load the data set. 2. Data segmentation: Divide the data into training and test sets. 3. Tree construction: o Use an algorithm such as CART or ID3 to build the tree. o at each split point, the feature that yields the most information is chosen (e.g., reducing genetics). 4. Model training: Build the tree based on the data. 5. Model implementation: Use the tree to classify new data. 6. Model testing: Evaluate the performance of the model on the test set.

7. Gradient Boosting Machines (GBM) algorithm. As shown below.

Steps:
1. Data collection: Prepare the data set. 2. Data partitioning: Split the data into training and test sets. 3. Building weak models: o Starting with a simple model, the model is trained on the data. o Errors (residuals) between the true values and the predicted values are calculated. 4. Model updating: Build a new model to address errors generated by the previous model. 5. Model aggregation: Iterate the process until a certain number of models are reached or a certain level of accuracy is achieved. 6. Model testing: Evaluate the performance using the test set.

5. RESULTS

The table 2 offerings a detailed evaluation of many algorithms used for intrusion detection in investment systems. Each algorithm is assessed based on key performance metrics: accuracy, precision, recall, and error rate. This comparative analysis lets for a better understanding of how each algorithm performs in identifying cyber threats, which is crucial for enhancing cyber security strategies.

TABLE II. THE RESULTS OF METHODOLOGY

Algorithm	Accuracy	Precision	Recall	Error Rate	Notes
Random Forest	94%	92%	90%	6%	Effective in handling large datasets and diverse patterns.
Support Vector Machines (SVM)	91%	89%	87%	9%	Good with high-dimensional data.
K-Nearest Neighbors (KNN)	88%	85%	82%	12%	Performance is affected by the choice of K.
Naive Bayes	85%	80%	78%	15%	Fast and effective with large datasets, but relies on independence assumptions.
Artificial Neural Networks (ANN)	96%	94%	92%	4%	Excellent at recognizing complex patterns.
Decision Trees	87%	86%	84%	13%	Easy to understand and interpret, but may suffer from overfitting.
Gradient Boosting Machines (GBM)	95%	93%	91%	5%	Strong performance, but requires longer training time.

5.1 Analysis of Results

The table submissions a comparative analysis of many algorithms used for detecting attacks in banking systems. Random Forest proves exceptional performance, efficiently managing large and different datasets while mitigating overfitting, resulting in reliable classifications and a low error rate. Support Vector Machines (SVM) achieve well with high-dimensional data but are sensitive to parameter selection, leading to slightly lower accuracy than Random Forest. K-Nearest Neighbors (KNN) is simple to implement but relies heavily on the choice of K, achieving an accuracy of 88% with a higher error rate.

Naive Bayes is effectual for large datasets but struggles with the assumption of feature independence, resulting in lower precision and recall. In contrast, Artificial Neural Networks (ANNs) excel with the highest accuracy and precision, effectively recognizing complex patterns and maintaining a very low error rate, making them invaluable for detecting sophisticated attacks.

Decision Trees offer instinctive results but are prone to overfitting, achieving an accuracy of 87%, while Gradient Boosting Machines (GBM) perform strongly but require longer training times. Overall, Random Forest and ANN are the best-performing algorithms, making them suitable for combating sophisticated cyber threats in banking systems. This analysis highlights the importance of selecting the right algorithm, as investing in effective machine learning techniques enhances detection rates and strengthens the cyber security posture of financial institutions in an increasingly risky.

6. CONCLUSION

The study shows that cyber security is a critical area that requires constant assets and technological advancements to address the growing threats facing financial institutions. The results confirm that advanced technologies such as machine learning and intrusion detection systems are essential to strengthen banks' defenses against cyber threats. Additionally, the importance of engaging customers and raising awareness about cyber security risks cannot be overstated, as it contributes to a safer digital environment. As the cyber security landscape continues to evolve, it becomes imperative for banks to adapt and continually update their security strategies. Research confirms that leveraging effective machine learning algorithms not only improves detection rates, but also significantly enhances the resilience of banking systems against the growing risk of cyber-attacks.

Funding

The author's explicitly states that no funding was received from any institution or sponsor.

Conflicts of Interest

The author declares no conflict of interest in relation to the research presented in the paper.

Acknowledgment

The author would like to express gratitude to the institution for their invaluable support throughout this research project.

References

[1] Ponemon Institute. (2021). Cost of a Data Breach Report.
[2] Stallings, W. (2018). Computer Security: Principles and Practice. Pearson.
[3] Anderson, R. (2020). Security Engineering: A Guide to Building Dependable Distributed Systems. Wiley.
[4] Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. Computers & Security, 38, 97-102.
[5] Arachchilage, N. A. G. (2020). The influence of social media on information security in the banking sector. Journal of Cyber Security Technology, 4(2), 131-142.
[6] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
[7] Canadian Institute for Cybersecurity. (2017). CICIDS 2017 Dataset. Retrieved from CICIDS Dataset.
[8] Kaspersky Lab. (2021). IT Security Risks Survey.
[9] Zarpelão, B., et al. (2017). A survey on the usage of honeypots in the security of web applications. Computer Science Review, 24, 18-27.
[10] D. Zaman and M. Mazinani, “Cybersecurity in Smart Grids: Protecting Critical Infrastructure from Cyber Attacks”, SHIFRA, vol. 2023, pp. 86–94, Aug. 2023, doi: 10.70470/SHIFRA/2023/010.
[11] A. S. . Bin Shibghatullah, “Mitigating Developed Persistent Threats (APTs) through Machine Learning-Based Intrusion Detection Systems: A Comprehensive Analysis”, SHIFRA, vol. 2023, pp. 17–25, Mar. 2023, doi: 10.70470/SHIFRA/2023/003.