IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

Machine Learning for Fraud Detection In Financial Transactions

Precious Ngulube[1] and MS Fanny Chatola[2]
Student, DMI St John The Baptist University, Lilongwe, Malawi[1]
Lecture II, Department of Computer Science, DMI St John The Baptist University, Lilongwe, Malawi[2]
juhass1998@gmail.com and fannychatolatmsic2022@gmail.com

Abstract: *Fraud detection in credit card transactions is a critical application of machine learning, leveraging techniques like supervised learning (e.g., logistic regression, decision trees, and neural networks) and unsupervised methods (e.g., anomaly detection). Effective feature engineering enhances model performance, while metrics such as precision, recall, and ROC-AUC address the class imbalance challenge. This study highlights a robust pipeline for fraud detection, addressing evolving fraud tactics and balancing accuracy with user satisfaction.*

Keywords: Fraud detection

I. INTRODUCTION

The exponential growth of credit card usage has led to increased electronic transactions and fraud. Traditional rule-based and statistical methods for fraud detection, though initially effective, struggle to adapt to evolving and sophisticated fraud tactics. Machine learning (ML) techniques offer a scalable and adaptive alternative, leveraging algorithms like logistic regression, decision trees, and neural networks for supervised learning, as well as anomaly detection and clustering for unsupervised learning. Feature engineering, focusing on transaction characteristics such as amount, time, and location, enhances detection accuracy, making ML a cornerstone for modern fraud detection systems.

1.1 OBJECTIVES

To improve fraud detection accuracy, advanced machine learning techniques like ensemble learning and deep learning will be utilized, with optimization through hyperparameter tuning and feature selection to minimize false positives and negatives. A robust real-time system will leverage continuous model updates to ensure swift and reliable threat identification. Scalability will be achieved using cloud infrastructure and distributed computing, enabling efficient handling of large transaction volumes.

Data security and compliance with regulations such as GDPR and PCI-DSS will be ensured through encryption, secure storage, and regular audits. To address class imbalance, methods like SMOTE, weighted loss functions, and tailored ensemble approaches will enhance detection of fraudulent transactions, boosting system reliability and performance.

II. EXISTING SYSTEMS VS PROPOSED SYSTEM

The existing fraud detection system relies on rule-based approaches with predefined thresholds and basic statistical methods, offering limited adaptability to evolving fraud patterns and moderate accuracy with high false positives and negatives. It struggles to handle class imbalance, leading to under-detection of fraudulent transactions, and lacks real-time detection capabilities. Additionally, the existing system faces scalability issues, often failing to efficiently manage large transaction volumes during peak periods. Security and compliance measures are minimal, with basic encryption and insufficient adherence to privacy regulations, resulting in suboptimal user experience due to frequent false alarms. The rigidity of the system and manual update requirements make it ill-equipped to handle sophisticated and rapidly changing fraud tactics.

In contrast, the proposed system leverages advanced machine learning techniques such as ensemble learning, deep learning, and anomaly detection for high adaptability and improved accuracy. It incorporates methods like SMOTE, weighted loss functions, and tailored ensemble methods to address class imbalance effectively, enhancing fraud

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

detection rates while reducing false positives and false negatives. The system supports real-time detection with a robust and scalable design powered by cloud infrastructure and distributed computing, ensuring seamless performance during peak transaction periods. Security and compliance are prioritized through robust encryption, secure storage, and adherence to standards like GDPR and PCI-DSS, ensuring data integrity and user trust.

The proposed system also focuses on proactive fraud prevention by identifying potential vulnerabilities in transaction patterns and alerting stakeholders before fraud occurs. This predictive capability, driven by advanced analytics and machine learning, reduces financial losses and enhances the overall security framework. Additionally, the system incorporates explainable AI (XAI) to provide transparency in decision-making, enabling stakeholders to understand and trust the detection process.

Furthermore, the system's modular design ensures flexibility and easy integration with existing financial infrastructures, minimizing disruption during deployment. Regular updates and scalability features allow it to grow alongside increasing transaction volumes, ensuring long-term effectiveness. This comprehensive approach makes the proposed system a robust, reliable, and efficient solution to modern fraud detection challenges.

III. SYSTEM DESIGN AND IMPLEMENTATION

The design and implementation of the fraud detection system for credit card transactions focus on leveraging advanced machine learning algorithms and real-time processing to identify fraudulent activities. The system is built to be scalable, secure, and efficient in detecting and mitigating fraud. The implementation involves various components, including data preprocessing, model training, real-time detection, and security measures to ensure effective fraud management.

Design Process

The fraud detection system follows a modular design, starting with Data Ingestion and Preprocessing, where real-time transaction data is cleaned and formatted. Feature Engineering extracts key features like transaction amount and location to create a training dataset. Model Training uses supervised and unsupervised learning algorithms to detect fraud patterns. In Real-Time Fraud Detection, the model classifies transactions and generates alerts for suspicious activity. Model Evaluation and Continuous Improvement ensure ongoing monitoring and updates to improve accuracy and adapt to evolving fraud tactics.

Hardware and Software Requirements

Hardware Requirements:

- High-Performance CPU: A multi-core processor capable of handling complex machine learning algorithms and real-time data processing for fraud detection.
- RAM: A minimum of 16GB is required for efficiently processing large volumes of transaction data and running machine learning models.
- Storage: At least 500GB of SSD storage to store large transactional datasets and model training data securely.
- GPU (Optional): NVIDIA GPUs to accelerate deep learning models for fraud detection, improving processing speed and model performance.
- Internet Connectivity: A high-speed internet connection or LAN for real-time data streaming and communication between system components.
- Cloud Infrastructure: If deploying on the cloud (AWS or Azure), Virtual Machine instances with appropriate CPU, RAM, and storage specifications for scalability

Software Requirements

- Data Ingestion and Preprocessing: Collecting and cleaning real-time transaction data, handling missing values, encoding categorical variables, and scaling numerical features.
- Feature Engineering: Extracting features such as transaction amount, time, location, merchant category, and user history to build datasets for model training.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

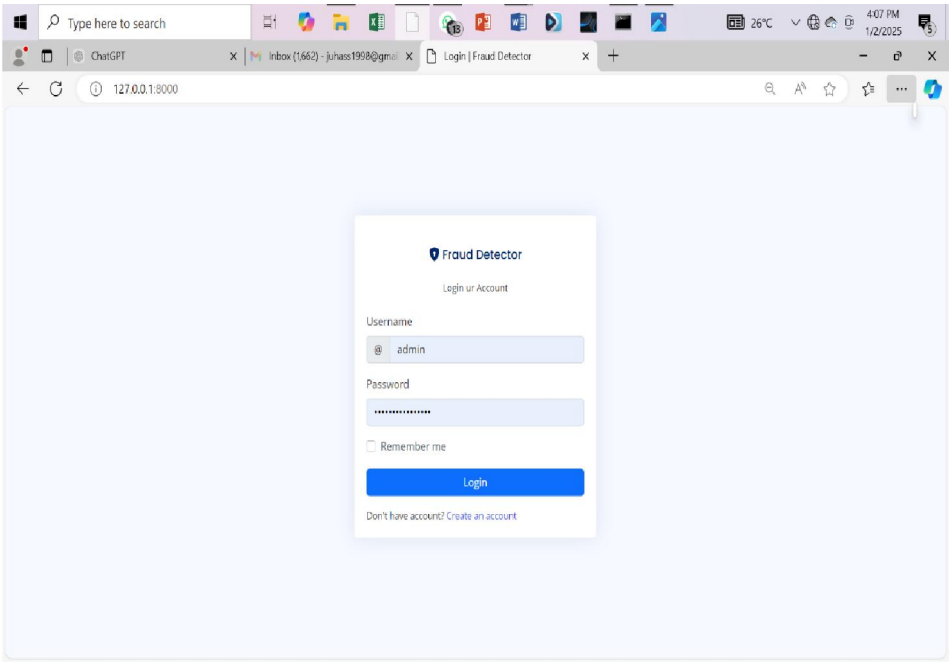
- Fraud Detection Models: Using machine learning algorithms like decision trees, random forests, gradient boosting, and neural networks, as well as unsupervised methods like anomaly detection for identifying fraud patterns.
- Real-Time Fraud Detection: Implementing models for classifying transactions as legitimate or fraudulent and generating alerts for suspicious activities in real-time.
- Model Evaluation and Continuous Improvement: Regularly evaluating model performance with metrics like precision, recall, and F1-score, and retraining models using new transaction data to adapt to evolving fraud tactics.

IV. METHODOLOGY

The Agile methodology is a project management approach that involves breaking the project into phases and emphasizes continuous collaboration and improvement. Teams follow a cycle of *Meetings, Planning, Design, Develop, Test and Evaluate*. The stages in this methodology are iterative, allowing for flexibility and the ability to revisit any stage to implement updates. This iterative nature ensures that the development process remains adaptable, enabling continuous refinement and improvement. After completing all the stages, the system can be revisited and modified as needed, fostering a dynamic and responsive development approach.

Input Design Screenshots

These Screenshots depict the system' input channels including user authentication such as a Login page, Account creation page, Personal details input areas.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

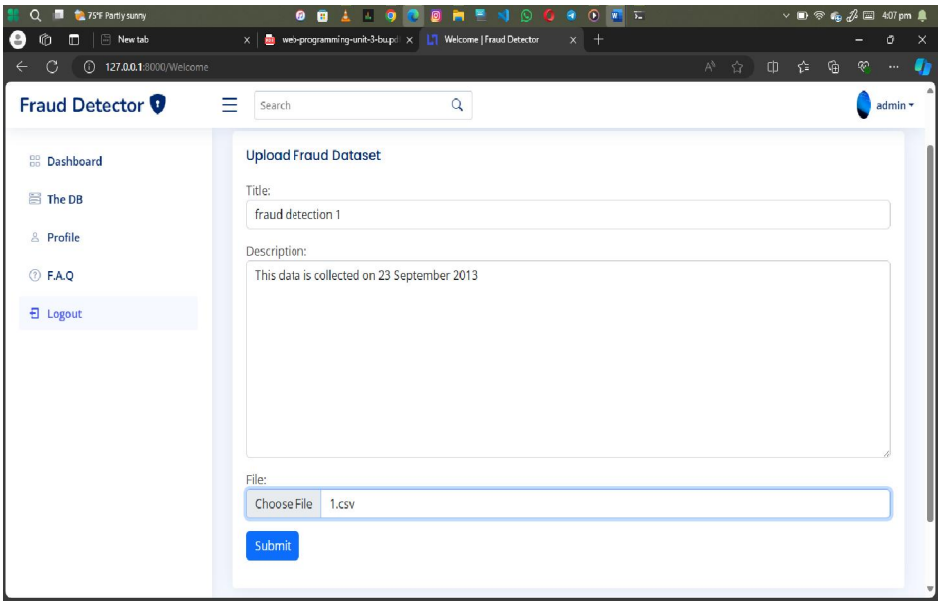
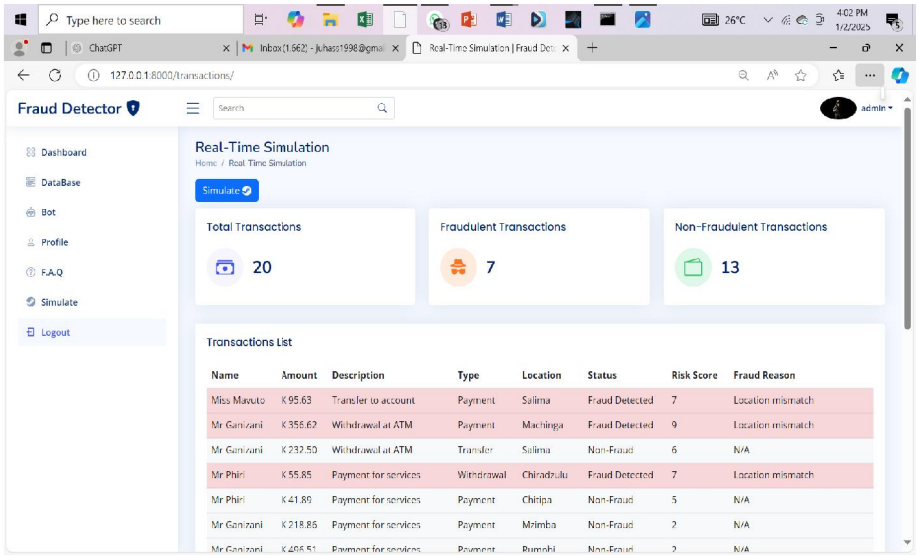


Fig 1. Input Screenshots

Output design Screenshots

These images depict the output design for the system which include the fraud that is detected, dataset of the system.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

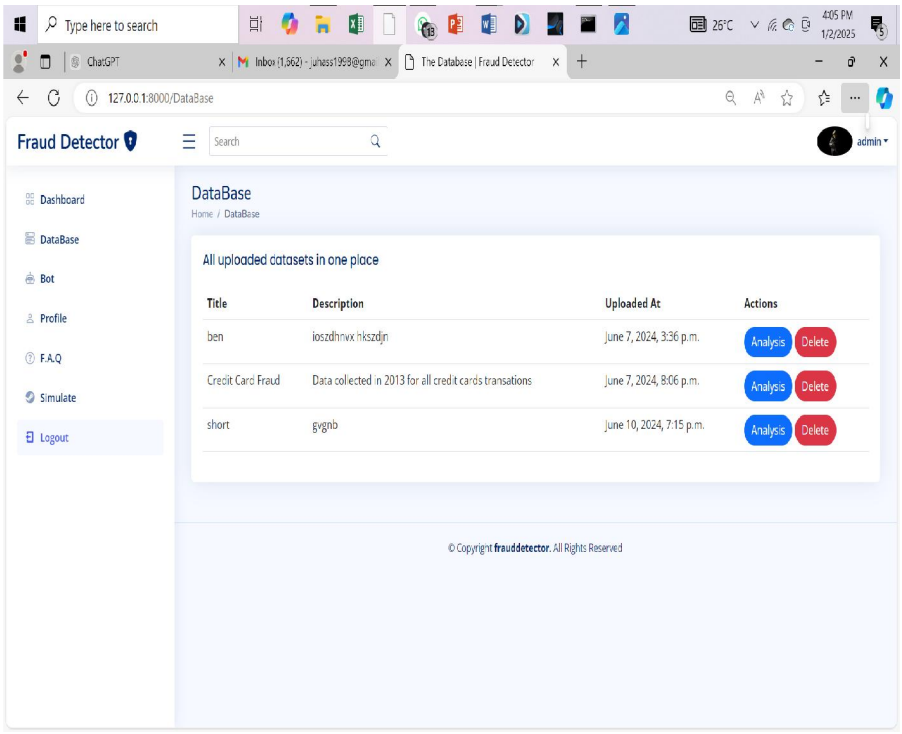


Fig 2. Application Output Screenshots

System Architecture

The fraud detection system processes transaction data in real-time using Kafka and Spark, with models deployed via microservices. Data is stored securely in databases, monitored by Prometheus and Grafana, and continuously improved through feedback. A web interface provides real-time insights.

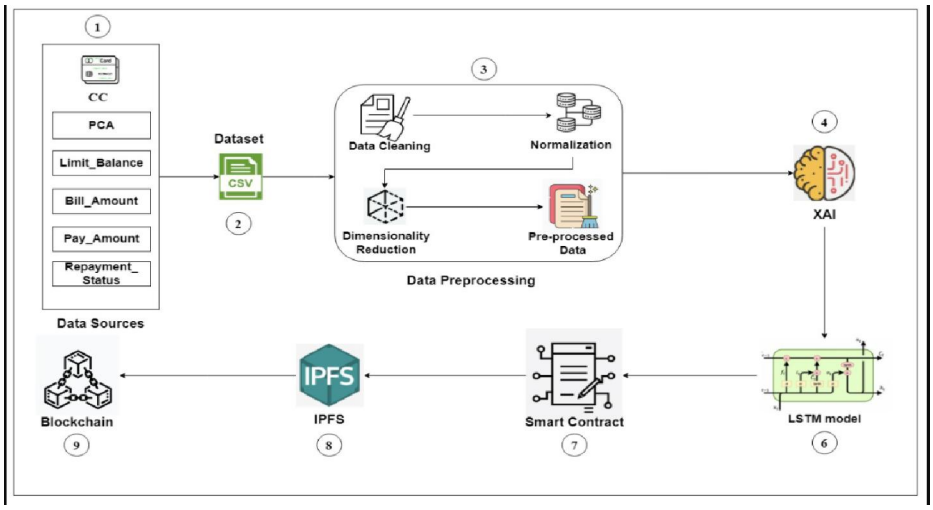


Figure 3. System Architecture

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

Data Flow Diagram

The fraud detection system collects transaction data from issuers and payment gateways, storing it in a raw data store. After preprocessing (cleaning, normalization, feature extraction), the data is stored in the preprocessed data store. Machine learning models analyze the data for fraud, with results saved in the model output store. Suspicious transactions are flagged and stored for fraud analysts' review, whose feedback is used to improve models. System administrators monitor system performance using metrics stored in system logs to maintain efficiency and reliability.

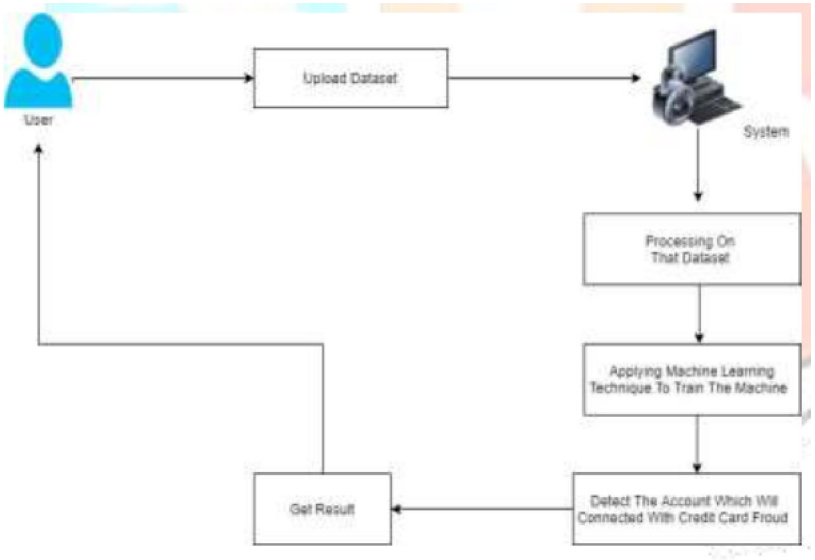


Figure 4 Data Flow Diagram

Class Diagram

The class diagram for the fraud detection system encompasses several key classes and their relationships. The Transaction class includes attributes like transactionID, cardNumber, and transactionAmount, with methods to validate transactions and calculate risk scores. The User class, with attributes such as userID and role, has subclasses Cardholder, FraudAnalyst, and SystemAdmin, each with specific attributes and methods: Cardholder initiates transactions, FraudAnalyst reviews and provides feedback on flagged transactions, and SystemAdmin monitors the system and updates the machine learning models. The MachineLearningModel class includes attributes like modelID and modelType, with methods for training, predicting, and updating models using the Dataset class, which handles data loading, preprocessing, and balancing. The SystemStatus class tracks system health, active users, and transaction rates. Relationships include cardholders initiating multiple transactions, fraud analysts reviewing transactions, system admins updating models, and machine learning models utilizing datasets for training and updates

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

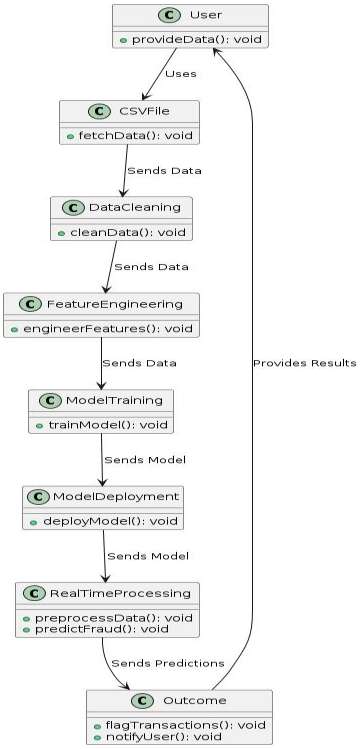


Figure 5. Class Diagram

V. RESULTS

Upon completion, the fraud detection system will provide processed transaction data that is cleaned and normalized for analysis, alongside fraud detection results from machine learning models identifying potentially fraudulent transactions. Suspicious transactions will be flagged for further review by fraud analysts, with their feedback used to refine and improve the detection models. Additionally, system performance metrics will be collected, providing insights into the system's operational efficiency, reliability, and performance trends, ensuring the system effectively detects and prevents fraud while continuously improving.

VI. DISCUSSION

The fraud detection system leverages machine learning models to analyze transaction data and identify potentially fraudulent activity. By preprocessing the raw transaction data (cleaning, normalization, and feature extraction), the system ensures that the data is in an optimal form for accurate analysis. The machine learning models then flag suspicious transactions, allowing fraud analysts to manually review and provide feedback. This feedback loop is crucial for continuously improving the models, ensuring that the system adapts to new fraud patterns over time.The inclusion of system performance metrics and logs allows administrators to monitor and maintain the system's health, ensuring its reliability and efficiency. This data-driven approach not only enhances the system's fraud detection capabilities but also ensures that the system remains adaptable and responsive to evolving fraud tactics. The feedback-driven model also improves the accuracy of the fraud detection process, making the system more effective in real-world scenarios where fraud patterns are constantly changing.

Overall, the system's design supports both automated fraud detection and human oversight, creating a balanced approach to handling financial fraud. Continuous monitoring and model refinement make the system a robust solution for mitigating fraud risks in credit card transactions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 5, Issue 1, January 2025

VII. CONCLUSION

The fraud detection system provides an effective, adaptive solution for identifying and preventing fraudulent transactions. By utilizing machine learning models for analysis and incorporating feedback from fraud analysts, the system continuously improves its detection accuracy. The integration of performance monitoring ensures the system operates reliably and efficiently. Through a combination of automated detection and human oversight, the system enhances fraud prevention in credit card transactions, offering a dynamic and scalable approach to combat evolving fraud tactics. This system not only improves security but also ensures that the detection process remains accurate and responsive over time.

VIII. ACKNOWLEDGMENT

First and foremost, I am grateful to the almighty God for the strength, health, and above all life. I am thankful for the support from family for their unwavering support. I would like to thank the DMI – St. John the Baptist University Malawi, for providing me the opportunity to do the journal work as part of our curriculum. I also sincerely thank MS Fanny Chatola for his support in making this paper possible as well guiding throughout my curriculum project as well as Mr. Mtende Mkandawire , Head of Computer Science and Information Technology for their kind help during my project work by providing an opportunity to enhance my career.

REFERENCES

[1]. Chandran, S., & Sankaranarayanan, P. (2020). Fraud Detection in Financial Transactions Using Machine Learning Algorithms. Springer.
[2]. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The Application of Data Mining Techniques in Financial Fraud Detection: A Review and Future Research Directions. Expert Systems with Applications, 38(10), 11106-11115.
[3]. Khan, M. A., & Dey, L. (2016). Financial Fraud Detection Using Machine Learning Algorithms: A Comparative Study. Procedia Computer Science, 83, 379-385.
[4]. Buczak, A. L., & Guven, E. (2016). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials, 18(2), 1153-1176.
[5]. Sculley, D., Holt, G., Golovin, D., et al. (2015). Hidden Technical Debt in Machine Learning Systems. Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS).
[6]. Yang, C., & Li, H. (2019). A Machine Learning Framework for Financial Fraud Detection. Proceedings of the IEEE International Conference on Computational Intelligence and Security (CIS).
[7]. Miller, L., & Zhang, H. (2018). Building Robust Fraud Detection Systems Using Artificial Intelligence. Journal of Financial Technology, 11(2), 45-59.
[8]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).
[9]. Kumar, S., & Kaur, P. (2017). Fraud Detection in Financial Transactions Using Supervised Learning Algorithms. International Journal of Computer Applications, 169(9), 1-6.
[10]. Zhou, Y., & Wang, F. (2018). An Adaptive Fraud Detection System Using Hybrid Machine Learning Models. Journal of Computer Science and Technology, 33(4), 762-774.