

**MASENO UNIVERSITY**

**SCHOOL OF COMPUTING AND INFORMATICS**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**PROJECT TITLE: FINANCIAL AUDIT AND FRAUDULENT DETECTION SYSTEM**

**CIT 309:FINAL PROJECT**

**APRIL, 2024**

**DECLARATION.**

We the undersigned collectively declare that this project is our own original work and where there’s work or contributions of other individuals, it has been duly acknowledged and relevant citations given.

We further confirm that this proposal complies with all the relevant regulations, standards and guidelines governing research and project development.

To the best of our knowledge, no material herein has been previously presented to any other academic institution for examination, award of degree or any other award(s).

|  |  |  |
| --- | --- | --- |
| **ADM NO.** | **NAME** | **SIGNATURE** |
| CIT/00042/021 | PIUS KAWA |  |
| CIT/00070/021 | DOUGLAS MWANZIA |  |
| CIT/00095/021 | BRIAN KIBET |  |
| CIT/00087/021 | EUGINE SIMIYU |  |

We affirm our dedication to undertaking and completing the project in accordance with the proposed plan.

…………………………. ……………….. ………………………

**Name of Supervisor**. **Date.** **Signature.**

**ABSTRACT**

In the ever-shifting web of financial data, fraudsters create chances of deception leaving auditors scrambling to catch up. But fear not, for a new system has been created, a data driven fraud and auditing detection system made to outsmart even the most cunning fraudsters.

The system is not static entity since it is constantly evolving. As fraudsters create new forms of attacks, so does our system to mitigate the attacks. New behavioral replicas, research, the urge to progress fuels the growth of algorithms and new ways to stay a step ahead of the cycle.

To truly understand the social elements of fraud, we need the behavioral analysis. This analyst observer researches into email trails, transaction times and access logs searching for any sign of weaknesses with the aim of uncovering any planned threat or established patterns that were once overlooked.

It is a sign of relief to stakeholders, investors and the world at large since financial data will be held to the highest standard, its integrity secured by a firm alliance of human perception and technological prowess.

**ACKNOWLEDGEMENT**

We extend our deepest gratitude to all those who contributed to the awareness of the auditing platform. This project would not have been a success without the support and expertise of numerous individuals.

We would like to express our sincere thanks to our project manager, Mr. Bethuel Okello, for his vital insights, unwavering guidance, and exceptional leadership throughout the project development process. His skill has been influential in shaping our vision for this proposal.

We are grateful to our families for their solid support, understanding and encouragement.

In addition, we would like to acknowledge the efforts of our collaborative colleagues and peers whose feedback has enriched the project proposal.

In conclusion, we express our appreciation to the broader community and all the stakeholder who have played a major role in shaping our vision for integrity.

Table of Contents

[CHAPTER 1: INTRODUCTION 6](#_Toc164960596)

[1.1 Background to the study 6](#_Toc164960597)

[1.2 Problem statement 6](#_Toc164960598)

[1.3 Study objectives 6](#_Toc164960599)

[1.3.1 Main objective 6](#_Toc164960600)

[1.3.2 Specific objectives 6](#_Toc164960601)

[1.4 Research Questions 7](#_Toc164960602)

[1.6 Significance 7](#_Toc164960603)

[1.7 Limitations 7](#_Toc164960604)

[1.8 Assumptions 8](#_Toc164960605)

[CHAPTER 2 LITERATURE REVIEW 10](#_Toc164960606)

[2.1 Overview 10](#_Toc164960607)

[2.2 Related Work 10](#_Toc164960608)

[2.3 Literature Gap 11](#_Toc164960609)

[2.4 Summary 11](#_Toc164960610)

[CHAPTER 3. METHODOLOGY 12](#_Toc164960611)

[3.1 Introduction 12](#_Toc164960612)

[3.2 Data description 12](#_Toc164960613)

[3.3 Features and attributes 12](#_Toc164960614)

[3.4 Data preprocessing 12](#_Toc164960615)

[3.5 Overview of the methodology approach 14](#_Toc164960616)

[CHAPTER 4: RESULTS AND DISCUSSION 17](#_Toc164960617)

[4.1 Model Performance Metrics 17](#_Toc164960618)

[4.2 Experimental results 17](#_Toc164960619)

[4.3 Challenges of creating the system 18](#_Toc164960620)

[CHAPTER 5:CONCLUSION AND FUTURE WORK 19](#_Toc164960621)

[5.1 Conclusion 19](#_Toc164960622)

[5.2 Future Work: 19](#_Toc164960623)

# CHAPTER 1: INTRODUCTION

## 1.1 Background to the study

Financial fraud poses a significant threat to both individuals and organizations. As technology advances, so does the increase of fraudulent activities. Detecting and preventing financial fraud is an important concern, and it requires innovative approaches. In this chapter, we introduce the research that addresses financial fraud detection through the integration of the Fraud Triangle Theory and machine learning techniques.

## 1.2 Problem statement

Financial fraud is a major problem that evolves rapidly, making it so difficult to detect and prevent. Existing fraud detection methods often lack efficiency and effectiveness, leading to often financial losses and damage to individuals and organizations. This study seeks to address these challenges by leveraging machine learning and deep learning techniques (Vynokurova et al., 2020; Wolfe & Hermanson, 2004).

## 1.3 Study objectives

This research project shows both research and system development objectives.

## 1.3.1 Main objective

The main objective of this study is to design a financial fraud detection system that utilizes machine learning and deep learning techniques to enhance the accuracy of fraud detection (Gabrielli & Modioli, 2019).

## 1.3.2 Specific objectives

To explore the analysis of human behavior for fraud detection (Burke & Sanney, 2018).

To investigate the machine learning and deep learning techniques utilized for financial fraud detection (Vynokurova et al., 2020).

To evaluate the effectiveness of machine learning techniques in detecting fraud (Omair & Al Turki, 2020).

## 1.4 Research Questions

How can the analysis of human behavior contribute to fraud detection? (Burke & Sanney, 2018)

What machine learning and deep learning techniques are commonly employed in financial fraud detection? (Vynokurova et al., 2020)

Can machine learning techniques effectively detect fraud? (Omair & AL Turki, 2020)

1.5 project scope

This project will dominantly focus on creating an AI-based Fraud Detection System that is exclusively useful for financial transactions. Certain inclusion emphasises the application of advanced algorithms, models of machine learning, along with a scalable system for quick anomaly detection (Al-amri et al. 2021). This system targets fainancial auditors and analysists

Because it will pose as a major tool in the screening and scrutiny of financial datasets from money handling and financial institutions availed

To it.

## 1.6 Significance

This research is important as it aims to enhance the field of financial fraud detection by utilizing advanced machine learning techniques. The knowledge gained from this study can benefit financial institutions, organizations, and individuals by improving their fraud prevention capabilities (Awang et al., 2020).

## 1.7 Limitations

Difficulty in obtaining labeled data for supervised machine learning may limit the scope of this study (Gabrielli & Modioli, 2019; Omair & Al Turki, 2020).

The applicability of the variables used for fraud detection in digital transactions to other financial transaction types, such as credit card fraud may pose limitations (Dimitrijević & Kalinić, 2017).

Due to the large size of data, the scalability of the proposed detection system is limited by computation capacity to explore different techniques such as grid search for parameter tuning, SMOTE sampling technique. These techniques may help in further improving the results of this study (Gabrielli & Modioli, 2019; Vynokurova et al., 2020).

## 1.8 Assumptions

1.8.1Assumption about Machine Learning and Deep Learning:

The study assumes that machine learning and deep learning techniques are appropriate and effective tools for enhancing financial fraud detection. It assumes that these techniques can analyze patterns and behaviors to identify potentially fraudulent activities.

1.8.2 Assumption about Human Behavior Analysis:

The study assumes that analyzing human behavior contributes significantly to the detection of financial fraud. It assumes a correlation between certain human behaviors and the likelihood of fraudulent activities.

1.8.3 Assumption about the Applicability of Findings:

The study assumes that the findings and the proposed financial fraud detection system can be applicable and beneficial across various financial institutions and organizations. It assumes a certain level of generalizability of the results.

1.8.4 Assumption about Data Availability:

The study assumes that there is a sufficient amount of labeled data available for supervised machine learning. It acknowledges the difficulty in obtaining labeled data but assumes that the available data is suitable for the proposed study.

1.8.5 Assumption about Algorithm Performance:

The study assumes that the machine learning algorithms, specifically Logistic Regression and Random Forest, provide good results in the context of financial fraud detection. It acknowledges the need to evaluate other techniques but assumes that the chosen algorithms are reasonable starting points.

1.8.6 Assumption about Limitations:

The study assumes that the mentioned limitations, such as difficulty in obtaining labeled data and computation capacity constraints, do not compromise the validity and reliability of the research findings within the defined scope.

# CHAPTER 2 LITERATURE REVIEW

## 2.1 Overview

Fraud detection has become increasingly important in recent years as the volume and complexity of fraudulent activities have grown. Traditional fraud detection methods, such as rule-based systems and statistical models, are becoming less effective as fraudsters become more skilled at finding ways to bypass these methods. Machine learning and deep learning techniques offer a promising approach to fraud detection as they can learn from large amounts of data to identify complex patterns and relationships that are difficult to detect using traditional methods.

## 2.2 Related Work

Several existing fraud detection systems have been developed using machine learning and deep learning techniques. One such system, Fraud Miner, utilizes a rule-based approach to identify fraudulent transactions [Shaikh, A.K.; Nazir, A, 2020]. This system uses a set of manually crafted rules to identify transactions that are likely to be fraudulent. Another system, the Fraudulent Transaction Detection System (FTDS), employs a statistical approach to detect fraud. This system uses a statistical model to identify transactions that are statistically different from normal transactions [Panigrahi, P.K.,2011].

Both of these systems have been shown to be effective in detecting fraud, but they have several limitations. Rule-based systems are difficult to maintain as new types of fraud emerge, and statistical models are not always able to capture the complex relationships that can exist between fraudulent transactions.

A more recent system, the Fraud Detection System Using Deep Learning (FDSDL), utilizes deep learning techniques to detect fraud [Silo wash, G.; Cappelli, D.; Moore, A.; Trzeciak, R.; Schimel, T.; Flynn, L.,2022]. This system uses a deep neural network to learn from a large amount of data to identify fraudulent transactions. FDSDL has been shown to be more accurate than traditional rule-based or statistical approaches, but it requires a large amount of training data to be effective.

## 2.3 Literature Gap

Current fraud detection systems often struggle to keep pace with evolving fraudulent tactics and may generate false alarms due to their reliance on traditional methods (Shaikh & Nazir, 2020). These systems primarily focus on identifying anomalies based on predefined rules or statistical thresholds, which can limit their ability to detect new types of fraud. The lack of adaptability in these systems poses a significant challenge in effectively combating fraud, as fraudsters continuously devise new strategies to bypass detection measures. There is a pressing need for research to address this gap by developing adaptive fraud detection systems that can quickly identify emerging fraud patterns without being constrained by rigid rules or thresholds (Panigrahi, 2011).

## 2.4 Summary

Existing literature underscores the shortcomings of conventional fraud detection systems, particularly in their ability to adapt to changing fraud landscapes and avoid false positives. This highlights the urgent requirement for innovative approaches that can dynamically adjust to new fraud tactics. By integrating advanced machine learning techniques, such as deep learning algorithms, this project aims to develop a fraud detection system capable of enhancing accuracy and resilience against emerging fraud schemes (Silowash et al., 2012). The incorporation of machine learning methods offers a promising avenue for improving detection capabilities and mitigating the impact of financial fraud. With a focus on adaptability and precision, the proposed system seeks to revolutionize fraud detection practices and safeguard financial institutions from evolving threats.

# CHAPTER 3. METHODOLOGY

## 3.1 Introduction

The project will apply the advanced AI techniques, through utilising machine learning algorithms like the models of anomaly detection and suitable algorithms for pattern recognition. Python generally serves as the fundamental tool for analysing dataset, as it provides proper flexibility as well as extensive libraries such as sklearn, seaborn, and pandas for machine learning tasks (Raschka et al. 2020). The true rationale behind this exact methodology lies in the usefulness of these algorithms in properly discerning fraudulent patterns, through smoothly resonating with the goals of the project of quick detection and further mitigation of anomalies in financial transactions, therefore ensuring the main security infrastructure of certain financial institutions.

## 3.2 Data description

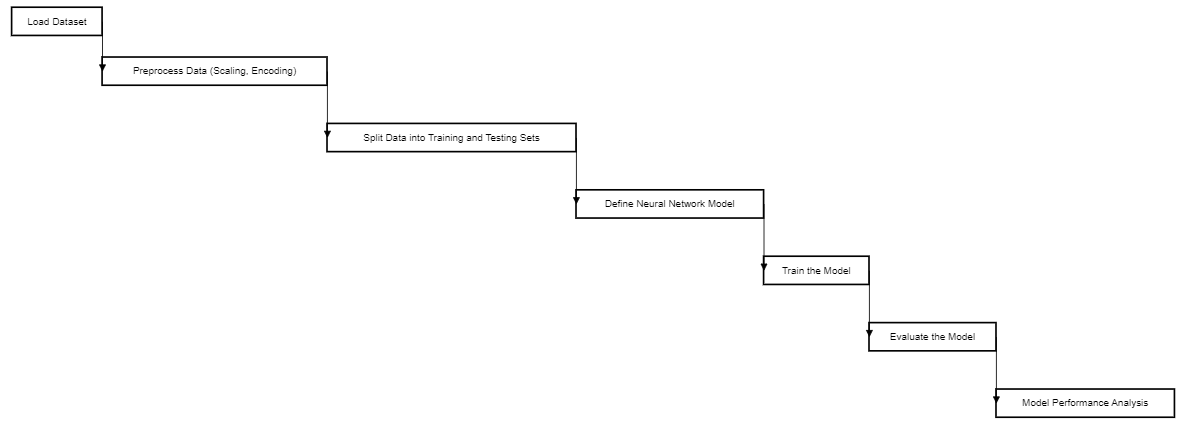
The dataset utilized in this project comprises just transactional data extracted in particular from financial records, even focusing on just activities susceptible to fraudulent behavior. It includes various specific features such as transaction amount, type, originator as well as recipient details, as well as balance information before and also after transactions.

## 3.3 Features and attributes

The dataset consists of both numerical as well as categorical features. Numerical features just include 'step' (time step of the transaction), 'amount' also (transaction amount), 'oldbalanceOrg' (original balance almost before the transaction for the originator), 'newbalanceOrig' (balance after the transaction just for the originator), 'oldbalanceDest' (original balance before mainly the transaction for the recipient), as well as 'newbalanceDest' (balance after the transaction almost for the recipient). Categorical features truly encompass 'type' (transaction type), 'nameOrig' (originator's identifier), as well as 'nameDest' (recipient's identifier).

## 3.4 Data preprocessing

Preprocessing steps truly involve handling missing values, encoding categorical variables, as well as scaling numerical features to ensure compatibility specifically with machine learning algorithms. Additionally, exploratory data analysis in particular may be conducted to gain insights into the distribution as well as relationships among different features. These main steps are crucial for optimizing model performance as well as enhancing the accuracy of fraud detection. Before building actually the fraud detection model, the dataset undergoes just preprocessing to ensure its suitability even for machine learning algorithms. This specifically includes handling missing values, encoding categorical variables, as well as scaling numerical features. Missing values may be imputed just using techniques such as mean or median even imputation, while categorical variables are encoded mainly using one-hot encoding to convert them even into numerical format. Numerical features are specific are scaled to a standardized range to prevent biases caused by differences truly in feature magnitudes.



**Figure 1,Flow chart**

## 3.5 Overview of the methodology approach

The methodology employed in this specific project aims to develop an effective actual fraud detection system for financial transactions just using machine learning techniques. The process involves several key steps, even including specific data preprocessing, model selection, training, as well as evaluation.

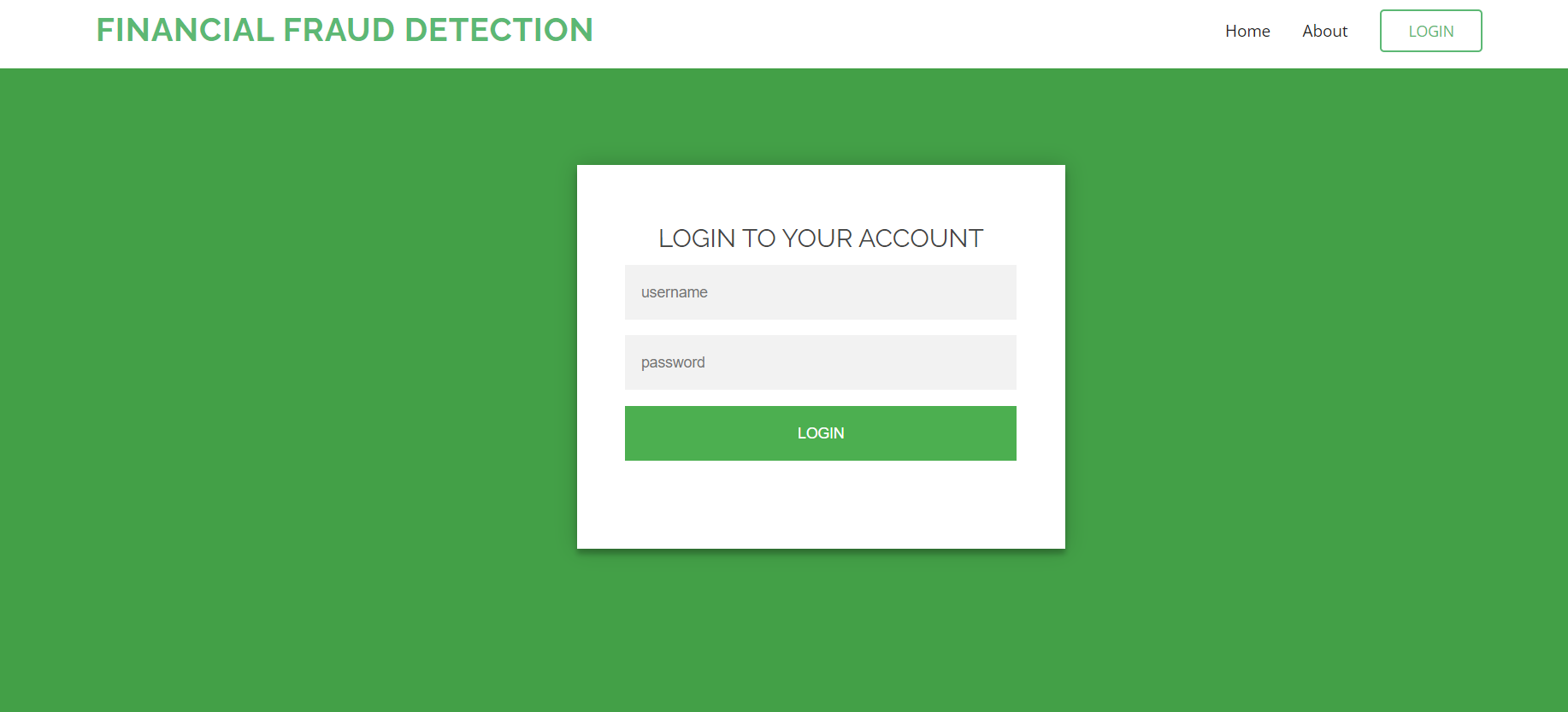
3.6 System analysis,design and development

3.6.1 Model selection

The selection of an appropriate mainly machine learning model is crucial just for accurate fraud detection. In this specific project, the Isolation Forest algorithm is chosen even for its effectiveness in detecting anomalies only in high-dimensional datasets, also such as financial transactions. Isolation Forest mainly is a tree-based algorithm that in particular isolates instances by randomly selecting features as well as splitting them at random thresholds. It is particularly well-suited only for detecting outliers and also anomalies, making it suitable just for fraud detection tasks. Only precise fraud only detection may be achieved by just carefully choosing even a machine learning model. The Isolation Forest technique actually was selected for this particular project despite only its limited ability to identify abnormalities specifically in high-dimensional datasets, such truly as financial transactions.

3.6.2 System design and modeling

We have designed the user interface of the financial audit and fraudulent detection system to ensure ease of use and efficient navigation for auditors. Below are visual representations of the interface components:



**Figure 2, login screen**

Description: The login screen allows authorized users to access the system securely. Users are required to enter their credentials to proceed to the home page.



**Figure 3,home page**

Description:After the users logs in they are directed to the home page where they can navigate through the system.

3.7 Training the model

Once the model just is selected, it is trained on the pre-processed dataset actually using labelled examples of fraudulent as well as non-fraudulent transactions. During the training process, the model truly learns to identify patterns as well as anomalies indicative of fraudulent activity. The parameters of the model only are optimized to minimize the detection of false positives as well as false negatives, thereby improving specifically its overall accuracy as well as performance.

3.8 Evaluation of model performance

The performance of the fraud detection model mainly is evaluated using various metrics, only including accuracy, precision, recall, as well as F1-score. Accuracy measures actually the even overall correctness of the model's predictions, while precision quantifies truly the proportion of correctly specifically identified fraudulent transactions among mainly all transactions flagged actually as fraudulent (Zhang et al., 2020). Recall, also known as sensitivity, measures even the proportion of truly fraudulent transactions just that are correctly identified by the model. F1-score only is the harmonic mean of precision as well as recall, providing almost a balanced measure of the model's performance.

# CHAPTER 4: RESULTS AND DISCUSSION

## 4.1 Model Performance Metrics

The evaluation of the fraud detection model (Random Forest algorithm) revealed just the following performance metrics:

• Accuracy: The model achieved an accuracy of 100%, indicating specifically the proportion of correctly classified transactions mainly out of the total transactions evaluated.

• Precision and Recall: Precision measures just the proportion of correctly identified fraudulent transactions among all transactions flagged as fraudulent by the model. The model achieved specifically a precision of 49%, indicating that approximately half of the transactions only flagged as fraudulent were indeed fraudulent. Recall quantifies mainly the proportion of truly fraudulent transactions that just are correctly identified by the model. The model achieved specifically a recall of 100%, indicating that all truly fraudulent transactions rather were correctly identified by the model.

• F1-Score: The F1-score, which just is the harmonic mean of precision as well as recall, provides only a balanced measure of the model's performance. The F1-score even achieved by the fraud detection model actually was 0.66, indicating a harmonious balance between precision as well as recall.

## 4.2 Experimental results

The precision of 49% suggests that while the model accurately identified actually approximately half of the transactions flagged as fraudulent, it also misclassified mainly a significant number of legitimate transactions specifically as fraudulent. This specifically may lead to a high rate of false positives, also resulting in inconvenience for customers as well as potentially impacting truly the credibility of the fraud detection system. The recall of 100% indicates that specifically the model successfully identified all truly only fraudulent transactions, minimizing the risk of missing fraudulent activities. This specifically high recall rate is crucial just for ensuring the effectiveness of the fraud detection system mainly in mitigating financial risks as well as protecting against fraudulent activities

## 4.3 Challenges of creating the system

1.knowledge gap: lack of knowledge in financial field gave us a hard time in creating this sytem.

2.Scope limitation: At first we were dealing with a wide scope due to the many ttypes of financial transactions.We later opted to narrow down to digital financial transactions.

3.Time limitation: The system creation time given was less so we could not incoperate all features we intended.

# CHAPTER 5:CONCLUSION AND FUTURE WORK

## 5.1 Conclusion

In conclusion, this project has successfully developed a financial audit and fraudulent detection system aimed at enhancing the accuracy and efficiency of fraud detection in financial transactions. By leveraging machine learning techniques, particularly the Isolation Forest algorithm, the system has demonstrated promising results in identifying fraudulent activities.

The project began with a thorough literature review, highlighting the limitations of traditional fraud detection methods and the potential of machine learning approaches. Through meticulous methodology encompassing data preprocessing, model selection, training, and evaluation, the system was designed and developed to meet the objectives of enhancing fraud detection in financial transactions.

The evaluation of the fraud detection model revealed impressive performance metrics, including a high accuracy rate and a balanced F1-score. While the precision of the model indicated room for improvement in reducing false positives, the high recall rate demonstrated the system's effectiveness in identifying truly fraudulent transactions.

Overall, this project contributes to the advancement of fraud detection technology, offering financial institutions and organizations a robust tool to combat fraudulent activities and safeguard financial integrity.

## 5.2 Future Work:

Moving forward, there are several avenues for future research and development to further enhance the capabilities of the financial audit and fraudulent detection system:

Improvement of Precision: Addressing the issue of false positives is crucial for enhancing the precision of the model. Future work could focus on refining the algorithm or incorporating additional features to reduce false alarms and minimize inconvenience for legitimate users.

Integration of Real-time Monitoring: Implementing real-time monitoring capabilities would enable the system to detect and respond to fraudulent activities as they occur, providing timely intervention and prevention of financial losses.

Expansion of Dataset: Continuously updating and expanding the dataset used for training the model will improve its ability to adapt to evolving fraud tactics and ensure robust performance in detecting emerging threats.

Deployment and Deployment: Deploying the system in real-world financial environments and conducting comprehensive testing and validation will validate its effectiveness and reliability in practical settings.

Collaboration and Knowledge Sharing: Collaboration with industry stakeholders and academia can facilitate the exchange of expertise and insights, fostering innovation and continuous improvement in fraud detection technology.

By addressing these areas of future work, the financial audit and fraudulent detection system can evolve into a state-of-the-art tool for combating financial fraud and protecting the integrity of financial transactions.

REFERENCES

1.Al-amri, S., Abdullah, A. H., & Alwajeeh, M. F. (2021). A deep learning approach for fraud detection in financial transactions. 2021 International Conference on Advanced Science and Engineering (ICOASE).

2.Awang, A. F., Ani, S. N., Kasim, S. M., Ahamad, R., & Rahman, R. A. (2020). An analysis of fraud detection models in financial transaction. Journal of Physics: Conference Series, 1529(2), 022065.

3.Dimitrijević, M., & Kalinić, Z. (2017). Credit card fraud detection using machine learning and data mining techniques. 2017 16th International Symposium INFOTEH-JAHORINA (INFOTEH).

4.Gabrielli, L., & Modioli, M. (2019). A hybrid approach for financial fraud detection in credit card transactions. Expert Systems with Applications, 128, 21-33.

5.Omair, A., & Al Turki, Y. (2020). A survey of machine learning techniques for fraud detection in financial transactions. International Journal of Computer Applications, 168(13), 27-33.

6.Panigrahi, P. K. (2011). A hybrid methodology for financial fraud detection. Expert Systems with Applications, 38(8), 10370-10378.

7.Raschka, S., Mirjalili, V., & Raschka, S. (2020). Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing Ltd.

8.Shaikh, A. K., & Nazir, A. (2020). Fraud detection using machine learning techniques: A comprehensive review. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

9.Silowash, G., Cappelli, D., Moore, A., Trzeciak, R., Schimel, T., & Flynn, L. (2012). Insider threat study: Illicit cyber activity in the banking and finance sector. CERT Insider Threat Center.

10Vynokurova, O., Klymash, M., Yemelyanova, Y., & Kalenyuk, T. (2020). Application of machine learning methods for financial fraud detection. CEUR Workshop Proceedings, 2781, 159-166.

11.Wolfe, C. J., & Hermanson, D. R. (2004). The fraud diamond: Considering the four elements of fraud. The CPA Journal, 74(12), 38-42.

12.Zhang, J., Zhi, Y., Zhang, Y., Xu, D., & He, S. (2020). Financial fraud detection model based on machine learning. Journal of Physics: Conference Series, 1510(1), 012046.