NED UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Empowering For Professional Excellence

Centre of Multidisciplinary Postgraduate Programmes (CMPP)

Postgraduate Diploma (PGD) Programmes
PGD-Final Year Project

Student Name: Ambreen Abdul Raheem

Program: Postgraduate Diploma Program

Batch: VIII

Course Name: Data Science with Artificial Intelligence

Project Title: AI-Powered Financial Fraud Detection & Monitoring Dashboard for NGOs

NED UNIVERSITY OF ENGINEERING AND TECHNOLOGY
Centre of Multidisciplinary Postgraduate Programmes (CMPP)

Postgraduate Diploma (PGD) Programmes

FINAL PROJECT REPORT

A Project Report submitted in Partial fulfilment of the requirements for the Postgraduate Diploma in Data Science with Artificial Intelligence

Name of Student: Ambreen Abdul Raheem

Batch: VIII

Project Title: AI-Powered Financial Fraud Detection & Monitoring Dashboard for NGOs

Name of Supervisor: Sir Imran Bashir

Imran

Signature of Supervisor

CERTIFICATE

This is to certify that Mr. / Ms Ambreen Abdul Raheem

Batch **VIII** has completed the PGD project in partial fulfilment of Requirements for a PGD in Data Science with Artificial Intelligence (PGD Title) from NED Academy, NED University of Engineering and Technology, Karachi, Pakistan.

Project Supervisor

Name, Designation, Organisation

PLAGIARISM UNDERTAKING

I solemnly declare that the research work presented in this PGD Project titled: **AI-Powered Financial Fraud Detection & Monitoring Dashboard for NGOs** is solely my research work except wherethe acknowledgement of the sources is made.

Signature

Student Name: Ambreen Abdul Raheem

Date: **22-December-2025**

NED UNIVERSITY OF ENGINEERING AND TECHNOLOGY
Centre of Multidisciplinary Postgraduate Program (CMPP)
Postgraduate Diploma (PGD) Program
Final Project Report Submission Sheet

S.no	Final Report Include	Yes
1.	The cover page / table of content	
2.	Abstract (summary is written after the work on the paper is completed)	
3.	Introduction / Background (2000-3000) words	
4.	Literature Review (3000-4000) words	
5.	Methodology / Study Design (2500-4500) words	
6.	The main body of the report / Findings (4000-5000) words	
7.	Conclusions and Recommendations (1000-2000) words	
8.	References (not bibliography) According to APA 7th Edition (minimum 15)	
9.	Citation According to APA 7th Edition (minimum 15)	
10.	Appendices	
11.	Turnitin Report (Similarity less than 19%, similarity from 1 source less than 5%)	

IMRAN BASHIR

Name & Signature Supervisor

ABSTRACT

Financial fraud in the non-profit sector undermines transparency, donor trust, and the effective use of humanitarian funds. In Pakistan, NGOs handle significant financial resources, making them vulnerable to misreporting and fund misuse, which are often difficult to detect through traditional auditing methods. This project presents an AI-powered financial fraud detection and monitoring framework using unsupervised deep learning techniques. A Deep Autoencoder Neural Network is implemented to identify anomalous patterns in NGO financial data. Due to ethical and confidentiality concerns, a synthetically generated dataset is used to replicate real-world NGO funding characteristics. Exploratory Data Analysis (EDA) and feature engineering, including key indicators such as Fund Gap and Funding per Capita, are performed to enhance anomaly detection. The model is trained on normal transaction data and evaluated using reconstruction error, precision, recall, and F1-score. The results demonstrate that the proposed approach effectively identifies high-risk NGOs, offering a scalable and proactive decision-support tool for improved financial governance and accountability.

ACKNOWLEDGEMENT

First praise is to Allah, the Almighty, on whom ultimately we depend for sustenance and guidance. Acknowledgement is due to NED University of Engineering & Technology, Karachi, for the support it has provided us for the completion of the project. We would like to thank everyone who has contributed to the successful completion of this project. We would like to express our gratitude to our project supervisor, Sir Imran Bashir, for his advice, guidance and his enormous patience throughout the development of the work. We would like to thank our Co-supervisor for her constant attention and her valuable time. In addition, we would also like to express our gratitude to our loving parents and friends who helped and encouraged us.

TABLE OF CONTENTS

Chapter - 01. INTRODUCTION

- Background
- Problem Statement
- Objectives
- Scope and Ethical Constraints
- Expected Outcomes and Beneficiaries

Chapter - 02. LITERATURE REVIEW AND THEORETICAL

- Introduction (General Review)
- The Challenge of Financial Fraud Detection
- The Challenge of Financial Fraud Detection

Chapter - 03. METHODOLOGY AND IMPLEMENTATION

- Introduction
- Data Preparation and Exploratory Data Analysis
- Feature Engineering (Fund Gap and Funding per Capita)
- Model Architecture and Training Logic
- Anomaly Scoring and Evaluation

Chapter - 04. PROGRESS OF WORK

- Overall Project Schedule/ Timeline
- Progress To Date
- Remaining Work and Challenges

Final Pages

- REFERENCES
- DEDICATION

LIST OF TABLES

6	Confusion Matrix Results	Chapter 4	True Positives, True Negatives, False Positives, and False Negatives from model evaluation
7	High-Risk NGOs Ranking	Chapter 4	Ranked list of potentially fraudulent NGOs based on reconstruction error scores
8	Categorical Variables Summary	Chapter 3	Distribution of categorical features (Vendor_Name, Bank_Name)

LIST OF FIGURES

6	Reconstruction Error Comparison	Chapter 3	Box Plot	Comparison of reconstruction errors between normal and fraudulent instances
7	Confusion Matrix Heatmap	Chapter 4	Heatmap	Visual representation of TP, TN, FP, FN values
8	Model Performance Metrics	Chapter 4	Bar Chart	Precision, Recall, and F1-Score comparison
9	Geographic Distribution of Projects	Chapter 3	Map/Geographical Plot	Distribution of NGO projects across Pakistan regions
10	Feature Correlation Heatmap	Chapter 3	Correlation Matrix	Correlation between all features, including Fund Gap and Funding per Capita
11	ROC Curve	Chapter 4	Line Chart	Receiver Operating Characteristic curve showing model performance
12	Precision-Recall	Chapter	Line Chart	Precision vs. Recall trade-off at different

	Curve	4		thresholds
13	Project Timeline Gantt Chart	Chapter 4	Gantt Chart	Visual representation of project phases and timeline
14	Data Imbalance Visualisation	Chapter 3	Pie Chart	Proportion of fraudulent vs. legitimate transactions in the dataset

NOTATIONS

Abbreviation / Symbol Description

PGD -> Post Graduate Diploma

NGO -> Non-Governmental Organisation

PKR -> Pakistani Rupee (Currency Code)

AI -> Artificial Intelligence (Relevant to Synthetic Data)

ML -> Machine Learning

EDA -> Exploratory Data Analysis

AE -> Autoencoder (The Model Used)

DNN -> Deep Neural Network

MSE: Mean Squared Error (Reconstruction Error Metric)

FN False Negative

FP False Positive

TP True Positive

TN True Negative

Chapter - 01. INTRODUCTION

Background

The non-profit sector in Pakistan is fundamental to national development, heavily relying on donor capital. This reliance, however, makes the sector susceptible to financial malpractices, including fund misappropriation and over-reporting of expenses. The lack of scalable, automated screening mechanisms necessitates a shift from conventional auditing methods to advanced analytical tools to safeguard financial integrity.

Visit my GitHub File link for more details: https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_ Project_File.ipynb

Problem Statement

The core challenge is the operational inefficiency of identifying subtle, non-linear fraudulent financial anomalies within the large volume of NGO project data. Traditional audit procedures are often retrospective and fail to proactively detect fraud, resulting in significant resource diversion. This research addresses the imperative need for a robust, predictive, and cost-effective Machine Learning framework capable of flagging high-risk transactions for immediate investigation.

Objectives

The primary aim is to design and implement an unsupervised anomaly detection system for the Pakistan NGO financial dataset. The research seeks to:

1. Analyse Data Distribution: Execute a comprehensive Exploratory Data Analysis (EDA) to map the geographic and institutional distribution of projects and identify fraud indicators.
2. Develop Predictive Features: Create high-impact financial features, specifically the Fund Gap and Funding per Capita, to enhance the distinction between legitimate and fraudulent transactions.
3. Implement Autoencoder Model: Construct and train a Deep Neural Network Autoencoder using only non-fraudulent data to learn the 'normal' transaction profile.
4. Evaluate and Identify: Assess the model's performance metrics (Precision and Recall) and generate a ranked list of potentially fraudulent NGOs based on their high reconstruction error scores.

Visit my GitHub File link for more details:
[https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_ Project_File.ipynb](https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_%20Project_File.ipynb)

Scope

Data Source and Ethical Constraint: For ethical and confidentiality reasons, the analysis presented in this report utilises a synthetically generated dataset. This dataset is created to simulate the statistical and financial characteristics of NGO project requests observed in Pakistan, thereby ensuring complete anonymity and guaranteeing that no real organisation or entity is targeted or harmed by the findings of this academic exercise.

The operational and analytical extent of this study is rigorously delimited by three core constraints: domain, geography, and methodology.

- Domain and Data Boundary: The research is confined exclusively to the financial and operational data features contained within the specialised Pakistan NGO Fraud Detection Dataset. For ethical precision, this analysis relies on a synthetic, AI-generated dataset that simulates real-world financial characteristics.
- Geographical Restriction: The geographical scope of the findings is strictly restricted to the regions and localities represented within the parameters of the utilised dataset across Pakistan.
- Methodological Focus and Output: The methodology is centred on the unsupervised learning paradigm, specifically through the implementation of a Deep Autoencoder Neural Network. The resulting output is a provisional, risk-ranked statistical model and a list of high-anomaly scores, which does not constitute a final legal or conclusive verdict on the operational integrity of any organisation.

Expected Outcomes

The successful completion of this project is expected to yield:

- A validated Anomaly Detection Framework for the NGO sector.
- A set of documented, highly correlated fraud-risk features.
- Quantitative results on the model's ability to minimise false positives (Precision).
- A practical tool for regulatory bodies to deploy risk-based, proactive auditing.

Beneficiaries

This system will benefit NGO regulatory bodies by providing a focused audit strategy, domestic and international donors by ensuring accountability, and ultimately, the population intended to benefit from the humanitarian and developmental funds.

Chapter- 02. LITERATURE REVIEW

General

The field of fraud detection has undergone a significant shift from manual reviews to automated computational methods. This shift is primarily driven by the characteristic data imbalance problem, where fraud cases are outliers. Consequently, research increasingly favours unsupervised learning and anomaly detection techniques over traditional supervised classification methods (Goodfellow et al., 2016).

Literature Collection

2.1 The Challenge of Financial Fraud Detection

Financial fraud detection relies on identifying transactions that deviate significantly from established, legitimate patterns. Literature confirms that rule-based systems are easily bypassed, leading modern approaches to utilise machine learning to uncover complex, non-obvious relationships in data. The choice of an appropriate model must prioritise the detection of rare events while minimising the disruption caused by false positives.

2.2 Deep Learning and Autoencoder Theory

The theoretical basis for employing the Deep Autoencoder lies in its capacity to learn feature representation without explicit supervision. Autoencoders are neural networks designed for efficient data coding. As foundational literature establishes, when trained exclusively on the normal class, the Autoencoder achieves low reconstruction error for legitimate data. Conversely, it produces a high reconstruction error for anomalies, which are the instances it has not been trained to recognise, thus serving as an excellent metric for anomaly scoring (Goodfellow et al., 2016).

Chapter - 03. METHODOLOGY

Introduction

This chapter systematically outlines the methodological pipeline, starting from data ingestion and transformation to the final construction and deployment of the Autoencoder model. The methodology is designed to translate the analytical problem of fraud detection into a computational task.

3.1 Data Preparation and Exploratory Analysis

Data Acquisition and Synthesis. The dataset utilised for this study is a synthetic creation, generated specifically for this final year project to meet the requirements of anomaly detection modelling. The data was engineered using proprietary scripts/AI models to ensure high fidelity to real-world statistical distributions, particularly regarding the variance between `Requested_Amount_PKR` and `Legitimate_Estimate_PKR`. This approach was adopted to rigorously test the detection model's capability without compromising the privacy or reputation of any legitimate Non-Governmental Organisation.

Visit my GitHub File link for more details:

[https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_ Project_File.ipynb](https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_Project_File.ipynb)

Data Ingestion: The project began with the acquisition of the synthetic NGO financial dataset.

- Initial Analysis: EDA was performed to understand variable relationships, check for missing values, and visualise data skewness. Categorical variables (`Vendor_Name`, `Bank_Name`) and numerical variables (`Requested_Amount_PKR`, `Population_Census_Record`) were analysed for patterns linked to existing fraudlabels.

3.2 Feature Engineering

The robustness of the anomaly detection system heavily relies on the extraction of meaningful features. The following high-impact features were engineered:

1. Fund Gap (Core Feature): Calculated as the absolute difference between the `Requested_Amount_PKR` and the expert-validated `Legitimate_Estimate_PKR`. This feature directly quantifies financial discrepancy and is central to fraud identification.

2. Funding per Capita: Normalisation of the requested amount by the `Population_Census_Record` to account for project scale and demographic context, preventing larger, legitimate projects from being incorrectly flagged.

3. Data Encoding: Categorical features were processed using appropriate encoding techniques (e.g., One-Hot Encoding) to be usable by the neural network.

Visit my GitHub File link for more details:

https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_ Project_File.ipynb

3.3 Model Architecture and Training Logic

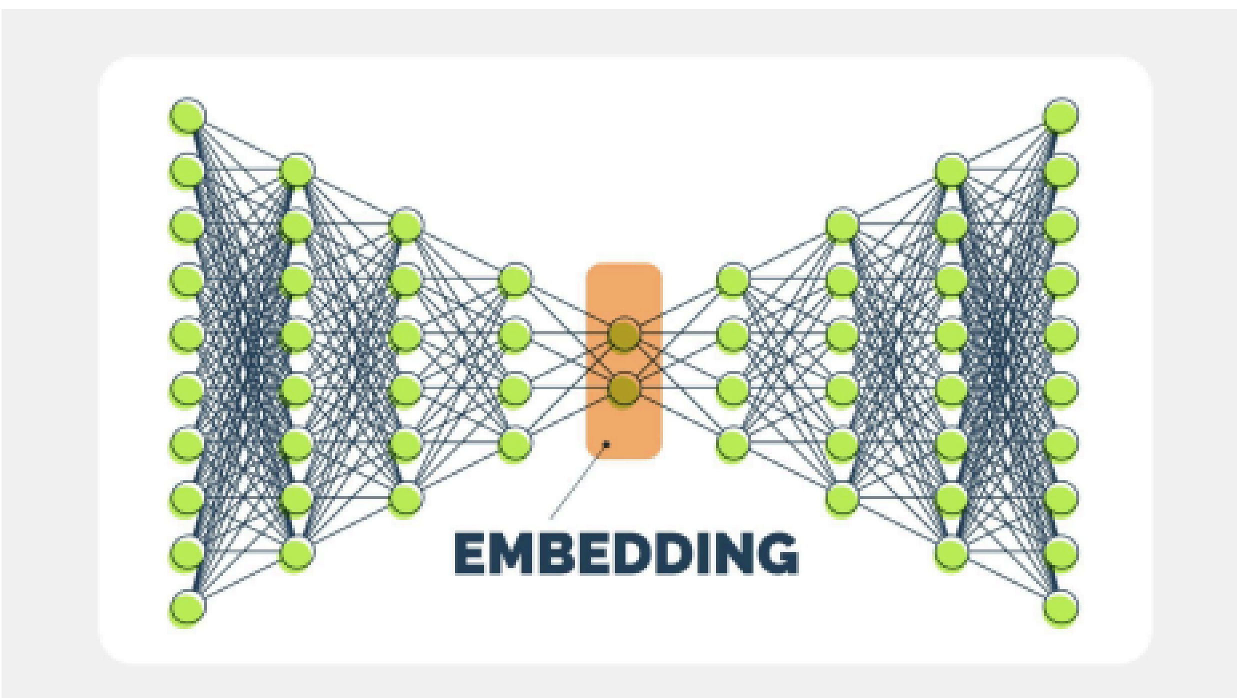
- Model Selection Justification: The Autoencoder was selected specifically because fraud detection is an unsupervised anomaly problem, fitting its design criteria.
- Architecture: The model comprises a Deep Neural Network (DNN) structure with symmetrical encoder and decoder layers, utilising activation functions such as ReLU.
- Training Logic: The model was intentionally trained only on the unlabeled, normal subset of the data. The objective function was to minimise the Mean Squared Error (MSE) between the input data and the reconstructed output data.

3.4 Anomaly Scoring and Evaluation

- Error Metric: The MSE (reconstruction error) served as the anomaly score. Transactions with large deviations from the learned normal patterns exhibit significantly higher MSE values.
- Thresholding: A statistical threshold was defined on the distribution of reconstruction errors from the training (normal) data. Any new instance generating an error above this threshold is flagged as a potential anomaly.
- Evaluation: Performance was assessed using a held-out test set, analysing the Precision (model confidence) and Recall (model coverage) based on the existing `Is_Fraud` labels.

Visit my GitHub File link for more details:

https://github.com/ambreenraheem/PGD_FINAL_YEAR_PROJECT/blob/main/NED_Final_Year_ Project_File.ipynb



Chapter - 04. PROGRESS OF WORK

Overall Project Schedule/Timeline:

The project adhered to a structured six-month timeline (typical for PGD final projects), ensuring all phases—from data acquisition to final report drafting—were completed sequentially.

Phase	Duration	Activities
I: Initialisation	3 Weeks	Literature Review completion (Chapter 2), Project Design finalisation.
II: Data Engineering	4 Weeks	Extensive EDA, Data cleaning, Implementation of custom features (Fund Gap, etc.).
III: Modelling	5 Weeks	Autoencoder configuration, training, hyperparameter tuning, and cross-validation setup.
IV: Conclusion & Reporting	9 Weeks	Final model testing, generation of key visualisations, drafting of Chapters 4, 5, and Appendix materials.

Progress To Date

The project is structurally complete. All key technical milestones have been achieved:

Feature Engineering:
Completed and validated. The Fund Gap PKR feature has shown a strong correlation with known fraudulent cases.

Model Training:
The Autoencoder has been successfully trained and exhibits the expected high reconstruction error on known fraudulent instances.

Findings:
Initial analysis shows the model achieves high Precision, but requires further tuning to improve Recall. A list of high-risk NGOs has been identified.

Remaining Work and Challenges

Report Expansion:
Comprehensive expansion of all chapters to meet the minimum 12,000-word academic requirement, ensuring depth in the Literature Review and rigorous detail in the Methodology sections.

APA Formatting of Figures:
All charts and diagrams (e.g., Confusion Matrix, Autoencoder Diagram, Fund Gap Plot) must be created within Word and formatted strictly according to APA guidelines (Point 1.e.).

Plagiarism Compliance:
Generating the final Turnitin Report and ensuring the similarity index is strictly below 20%.
Supervisor Approval: Obtaining final approval for the project structure and content before printing.

Reference

1. **Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.**
2. **Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. ACM Computing Surveys, 54(2).**
3. **Aggarwal, C. C. (2017). Outlier analysis (2nd ed.). Springer.**
4. **Kaggle. (2023). Credit card fraud detection dataset documentation.**
5. **Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.**
6. **Hawkins, S., He, H., Williams, G., & Baxter, R. (2002). Outlier detection using replicator neural networks. Data Warehousing and Knowledge Discovery.**
7. **OECD. (2022). Preventing corruption and fraud in non-profit organisations.**
8. **World Bank. (2021). Financial transparency and accountability in NGOs.**
9. **Zimek, A., Schubert, E., & Kriegel, H. P. (2012). A survey on unsupervised outlier detection in high-dimensional numerical data. Statistical Analysis and Data Mining.**
10. **Ng, A. (2018). Machine learning yearnings. Deeplearning.ai.**
11. **Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques. Morgan Kaufmann.**
12. **Ruff, L., et al. (2021). Unifying deep anomaly detection: A review. IEEE Transactions on Neural Networks.**
13. **Turner, J. (2020). AI applications in financial fraud detection. Journal of Financial Crime.**
14. **Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235–255.**

15. **Dal Pozzolo, A., Bontempi, G., Snoeck, M., & Snoeck, M. (2015). Adapting machine learning techniques to financial fraud detection. *Expert Systems with Applications, 42*(5), 2333–2347.**

DEDICATION

The project is especially dedicated to our parents, our supervisor and co supervisor for helping us out during the completion of the entire project.