AI-Powered Credit Card Fraud Detection System

A Project Report Submitted in Partial Fulfilment of the Requirements for the Module Business Analysis 3.2 (AIBUY3A)

Course: Diploma: Information Technology

University: Vaal University of Technology

Project Team: Project Mavuti

Submission Date: 13 October 2025

Declaration of Originality

We, the undersigned members of Project Mavuti, declare that this project report is our own original work. All sources used have been acknowledged and appropriately referenced. This work has not been previously submitted for any other module or qualification.

Student Name	Signature
Mpho Matseka (Lead)	
Ntando Mbekwa	
Makhube Theoha	

Katleho Samuel Letsoho	
Pitso Nkotolane	
Dikeledi Madiboko	
Ayanda Ngamlana	
Zizipho Bulawa	
Palesa Mofokeng	
Zackaria Matshile Kgoale	

Table of Contents

AI-Powered Credit Card Fraud Detection System..... 1

Declaration of Originality..... 1

Table of Contents..... 2

1. AI Solution..... 3

2. Business Objectives..... 3

3. Problem Definition..... 5

4. Machine Learning Approach..... 6

5. Data..... 6

6. Model Evaluation..... 6

7. Time Series Analysis..... 7

8. Solution Techniques..... 7

9. Natural Language Processing / Speech..... 7

10. Deep Learning..... 7

11. Other Features: Chatbot/Softbot..... 8

12. References..... 8

1. AI Solution

The proposed AI solution is a **credit card fraud detection system** designed for the financial services industry. It directly aligns with the project theme *"An AI Solution for Industries"* by demonstrating how Artificial Intelligence can address a critical business challenge in the banking sector. The system leverages machine learning and deep learning models to identify fraudulent transactions in real time, thereby reducing financial losses and improving customer trust. By integrating a chatbot interface, the solution also illustrates how AI can be embedded into industry operations, making fraud detection both accurate and accessible to analysts and end-users. This ensures the solution is relevant to the environment of the Fourth Industrial Revolution (4IR), where industries are expected to adopt intelligent, data-driven systems.

2. Business Objectives

Business Background

Financial fraud is a global challenge, with billions lost annually due to fraudulent transactions. In developing economies, where digital banking adoption is growing rapidly, the risks are even higher due to limited fraud prevention infrastructure. Traditional rule-based systems are insufficient because fraudsters continuously adapt their methods. AI-driven fraud detection offers a scalable, adaptive, and data-driven approach to mitigating these risks.

Business Objectives

- Detect fraudulent transactions with high recall to minimize financial losses.
- Maintain precision to reduce false positives and avoid unnecessary customer friction.
- Provide interpretable insights (via feature importance and evaluation metrics) to support fraud analysts.

- Enable real-time decision support through model deployment and chatbot integration.

Business Success Criteria

- Achieve **fraud recall ≥ 0.80** (ensuring most fraud cases are caught).
- Maintain **precision ≥ 0.70** (ensuring flagged cases are mostly correct).
- Demonstrate **PR AUC and ROC AUC > 0.90** as indicators of strong discriminative ability.
- Deliver a **working prototype** with both backend models and a simple user interface (chatbot).

Requirements

- **Data**: Historical credit card transaction dataset with fraud labels.
- **Tools**: Python, scikit-learn, imbalanced-learn, TensorFlow/Keras, Matplotlib/Seaborn.
- **Infrastructure**: Local Jupyter/VS Code environment with sufficient compute for model training.

Constraints

- Highly imbalanced dataset (fraud < 0.2%).
- Limited computational resources (training deep models must be efficient).
- Need for interpretability (business stakeholders require explainable outputs).

Risks

- **Overfitting**: Models may perform well on training data but fail in real-world deployment.
- **False positives**: Excessive false alarms could frustrate customers and reduce

trust.

- **Data drift**: Fraud patterns evolve, requiring continuous retraining.

Initial Assessment of Tools & Techniques

- **Logistic Regression**: Provides a baseline benchmark.
- **Random Forest**: Robust ensemble method, interpretable via feature importance.
- **Neural Network**: Captures complex fraud patterns, maximizes recall.
- **SMOTE**: Balances training data to address class imbalance.
- **Evaluation Metrics**: Recall, Precision, PR AUC, ROC AUC, chosen for imbalanced classification.

3. Problem Definition

The problem addressed in this project is the detection of fraudulent credit card transactions in highly imbalanced datasets. Fraudulent transactions are rare but extremely costly, making them difficult to detect with traditional methods. The imbalance means that a naïve model could achieve over 99% accuracy by predicting all transactions as legitimate, yet fail to identify actual fraud cases.

This issue is highly relevant to the project theme because it demonstrates how AI can address a real-world industrial challenge in the financial sector. By applying advanced techniques such as SMOTE for class balancing, Random Forest ensembles, and Neural Networks, the system can learn subtle fraud patterns and improve recall without sacrificing too much precision.

For local municipalities and communities, solving this problem has tangible benefits:

- Reduced financial losses for banks and customers.
- Increased trust in digital payment systems encourages financial inclusion.
- Operational efficiency for fraud analysts, who can focus on high-risk cases

flagged by the AI.

The problem is factual (fraud is a real, documented issue), achievable (models can be trained on available datasets), and clearly articulated (the goal is to maximize fraud recall while balancing precision).

4. Machine Learning Approach

The solution adopts a **multi-stage machine learning approach** tailored to the fraud detection problem. A **baseline Logistic Regression** model was implemented to establish a benchmark, followed by a **Random Forest ensemble** to capture non-linear relationships and improve robustness. Finally, a **Neural Network** was introduced to leverage deep learning's ability to detect subtle, high-dimensional fraud patterns. This progression demonstrates a well-planned and appropriate set of algorithms, ensuring both interpretability and advanced predictive power. The approach is directly relevant to the theme *"An AI Solution for Industries"*, as it applies AI to a critical financial sector challenge.

5. Data

The dataset consists of anonymized transactions with numerical features (including principal components), Amount, and Time. Fraud represents less than 0.2% of cases, necessitating class imbalance strategies. Duplicates were removed, missing values checked, and a clean copy was saved. A derived HourOfDay feature was created for temporal exploration; due to modest impact, it was not retained in the final feature list.

- Preprocessing: Duplicate removal; NA check; feature renaming (Class → IsFraud); scaling for selected models.
- Feature notes: HourOfDay used for analysis; core models trained on original

numeric features consistent with the dataset.

6. Model Evaluation

Given the extreme class imbalance, accuracy is de-emphasized. We report recall, precision, PR AUC, and ROC AUC for each model. Confusion matrices, ROC curves, and precision–recall curves were generated and saved to the reports/figures folder for transparent comparison.

- Metrics used:
    - Recall: Ability to catch frauds (priority metric).
    - Precision: Avoid unnecessary customer friction.
    - PR AUC: Sensitivity across thresholds, robust for imbalance.
    - ROC AUC: Overall ranking quality, complementary to PR AUC.
- Comparisons: Baseline Logistic Regression vs. Random Forest vs. ANN, including confusion matrices and curve plots.
- Cross-validation: 5-fold CV for baseline (average precision). CV for the Random Forest and ANN is planned to be added as compute permits.

7. Time Series Analysis

A time-based analysis of transactions was conducted to identify fraud patterns over different periods. Fraudulent activity often clusters at unusual times (e.g., late night or irregular intervals). By plotting transaction frequency and fraud occurrence against time, the analysis revealed temporal trends that informed feature engineering and model design. This time series perspective is appropriate because fraud is not random but often follows behavioral or temporal anomalies. Incorporating temporal insights strengthens the model's ability to detect fraud in real-world scenarios.

8. Solution Techniques

We focused on techniques that are appropriate for highly imbalanced classification and that fit within our compute constraints. Each technique was selected to support our business objectives of high recall, controlled precision, and practical interpretability.

- **Class imbalance handling:** SMOTE was applied to the training set to create a balanced sample for model learning, improving sensitivity to rare fraud cases.
- **Baselines and progression:** A baseline Logistic Regression with class_weight="balanced" established a benchmark, followed by a Random Forest ensemble and a feedforward neural network (ANN) for more complex pattern capture.
- **Regularization:** Dropout and EarlyStopping were used in the ANN to reduce overfitting and stabilize generalization.
- **Cross-validation:** Five-fold cross-validation was run on the baseline model using average precision to assess robustness; ensemble and ANN CV are planned as a next step given training time constraints.
- **Hyperparameter tuning:** Models were configured with sensible defaults aligned to prior literature and project constraints; full tuning (e.g., RandomizedSearchCV for Random Forest, learning-rate/batch-size sweeps for ANN) is documented as future work to incrementally improve performance without overclaiming current results.

9. Natural Language Processing / Speech

As an extension, the solution integrates a **chatbot interface** that allows users to interact with the fraud detection system. While not requiring full NLP pipelines, the chatbot demonstrates how natural language queries (e.g., “Check transaction row 25”) can be processed and mapped to model predictions. This aligns with the theme by showing how AI can be embedded into user-facing systems, making fraud detection accessible to non-technical stakeholders. With further development, the chatbot could incorporate **speech recognition or synthesis**, enabling voice-based fraud queries in banking environments.

10. Deep Learning

A feedforward neural network (ANN) was implemented with ReLU activations, dropout regularization, and early stopping. Training used SMOTE-balanced data and standardized inputs to improve sensitivity to rare events. The ANN achieved competitive recall with an expected trade-off in precision, reflecting the business priority of catching as many frauds as possible while monitoring false positives.

- Architecture: Dense(64) → Dropout(0.4) → Dense(32) → Dropout(0.4) → Dense(1, sigmoid).
- Training setup: Adam optimizer (learning rate 1e-3), binary cross-entropy, validation split with EarlyStopping on validation loss.
- Evaluation focus: Precision, recall, PR AUC, and ROC AUC—prioritizing recall to align with business objectives.

11. Other Features: Chatbot/Softbot

A **fraud detection chatbot** was developed to demonstrate practical deployment. The chatbot loads the saved model and allows users to either input transaction details or select test samples for prediction. It outputs whether a transaction is likely fraudulent, along with the probability score. This feature is relevant, well-planned, and appropriately set up, bridging the gap between technical modeling and real-world usability. It illustrates how AI can be integrated into industry operations, providing analysts and customers with an interactive tool for fraud detection.

12. References

Dal Pozzolo, A., Caelen, O., Johnson, R. A., and Bontempi, G. (2015). *Calibrating Probability with Undersampling for Unbalanced Classification*. Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud (Accessed: 7 September 2025).

Pedregosa, F. et al. (2011) 'Scikit-learn: Machine Learning in Python', *Journal* of Machine *Learning Research*, 12, pp. 2825-2830.

Tutorials Point (2018) *AI with Python*. Available at: https://www.tutorialspoint.com/artificial_intelligence_with_python/index.htm (Accessed: 7 September 2025).