B.Tech Project Report for Mid Sem Evaluation

(CSIR 32)

on

Intelligent Risk Analytics in Modern Financial Infrastructure

by

Tejasvi Murmu (12112238)

Under the Supervision of

Dr. Santosh Kumar

Assistant Professor

Department of Computer Engineering
National Institute of Technology, Kurukshetra
Haryana-136119, India

Jan-May 2025

Certificate

We hereby certify that the work presented in this B.Tech Project (CSPE40) report, entitled *"Risk Management in Financial Sector Using Artificial Intelligence"*, in partial fulfillment of the requirements for the Mid-Semester evaluation of the **Bachelor of Technology (Computer Engineering)** program, under the supervision of **Dr. Santosh Kumar, Assistant Professor, Computer Engineering Department, National Institute of Technology, Kurukshetra, India**.

The matter presented in this project report has not been submitted for the award of any other degree elsewhere.

Signature of Candidate

Tejasvi Murmu (12112238)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date: 01.05.2025

Signature of Supervisor Faculty Mentor
Dr. Santosh Kumar
Asst. Prof.

Table of Contents

Abstract

The financial sector has undergone a significant transformation in fraud detection with the integration of Artificial Intelligence (AI) and Machine Learning (ML), as digital transactions surge and fraudsters develop increasingly sophisticated methods that render traditional rule-based systems less effective. AI-driven solutions enhance security by leveraging real-time anomaly detection, predictive analytics, and automated risk evaluation to minimize financial losses. However, these advancements come with challenges, requiring a structured risk management approach to ensure model accuracy, data security, and regulatory compliance. The adoption of AI in fraud detection is driven by the need for improved accuracy, efficiency, and adherence to regulations, ultimately enhancing fraud prevention while streamlining operations, reducing costs, and minimizing human errors. As AI technology evolves, financial institutions must implement comprehensive risk management strategies to fully capitalize on its benefits while mitigating potential threats.

Important Keywords:

Finance Sector ; Fraud Detection ; Artificial Intelligence(AI) ; Machine Learning(ML) ; Synthetic Minority Over-sampling Technique(SMOTE) ; Artificial Neural Network(ANN) ; Convolutional Neural Network (CNN) ; Least Absolute Shrinkage and Selection Operator(LASSO) ; Extreme Gradient Boosting(XGBoost)

1. Introduction

The rise of digital transactions has transformed the financial sector, offering convenience and accessibility while simultaneously increasing the risk of sophisticated fraud schemes that challenge traditional security measures. Conventional rule-based fraud detection systems struggle to keep pace with rapidly evolving threats, prompting financial institutions to adopt [3] Artificial Intelligence and Machine Learning for more adaptive and intelligent fraud detection. AI-powered systems utilize real-time anomaly detection, predictive analytics, and automated risk assessment to identify suspicious activities with greater accuracy and efficiency, helping financial institutions mitigate financial losses and enhance security. However, integrating AI into fraud detection presents challenges, including model accuracy, data security, algorithmic bias, and regulatory compliance, all of which must be managed to ensure reliability and fairness. To maximize the benefits of AI while mitigating risks, financial institutions must implement robust risk management strategies that involve continuous monitoring, model refinement, and adherence to strict regulatory standards. As AI technology evolves, financial institutions must stay ahead of emerging threats by refining their fraud detection frameworks, ensuring both security and compliance in an increasingly digital financial landscape.

2. Motivation

The rise of digital transactions has created new opportunities for financial growth but has also increased the risk of fraudulent activities. Traditional rule-based fraud detection methods are becoming less effective as fraudsters develop more sophisticated techniques. This evolving threat landscape has driven financial institutions to explore AI and ML-powered solutions that can detect fraud in real time, analyze vast datasets for patterns, and improve accuracy in identifying suspicious activities. The ability of AI to adapt and learn from new fraud techniques makes it a crucial tool in safeguarding financial systems.

Implementing AI-driven fraud detection is not just about enhancing security but also about improving operational efficiency. [4] Manual fraud detection processes are time-consuming and prone to human error, whereas AI automates risk evaluation, reducing costs and increasing accuracy. Additionally, regulatory bodies require financial institutions to strengthen their fraud prevention measures, further motivating the adoption of AI-based solutions. By leveraging AI, financial organizations can ensure secure transactions, maintain customer trust, and stay ahead of emerging financial threats.

3. Related Work

Artificial Intelligence is revolutionizing financial risk management by leveraging machine learning and deep learning to enhance risk assessment and predictive accuracy. [5] AI-driven models can identify market trends, detect anomalies, and optimize financial decision-making, reducing uncertainty in volatile markets. However, key challenges such as regulatory compliance, data security, and the transparency of AI models must be addressed to ensure reliability and trust. The study also explores the role of neural networks in improving financial stability by refining risk evaluation strategies. While AI-powered risk management enhances financial resilience, it is crucial to establish ethical guidelines and regulatory frameworks to ensure responsible and secure implementation in the financial sector.

The integration of blockchain technology in financial fraud detection is transforming the security landscape by providing decentralized and tamper-proof record-keeping systems. [6] This approach enhances transparency, security, and efficiency in financial transactions, reducing vulnerabilities to fraud. The study explores real-world applications where blockchain improves fraud prevention, identity verification, and transaction monitoring. Additionally, it examines challenges such as regulatory uncertainties, scalability constraints, and the complexities of integrating blockchain with traditional banking systems. The research concludes that while blockchain has significant potential to strengthen fraud detection and financial security, its widespread adoption depends on technological advancements, regulatory clarity, and industry-wide collaboration to ensure seamless and secure implementation.

This study explores the ethical implications of artificial intelligence in the healthcare sector. AI-driven systems have revolutionized medical diagnostics and treatment planning, but concerns about privacy, bias, and accountability persist. [7] The research presents case studies demonstrating both the benefits and ethical dilemmas of AI applications in medicine. It suggests that improving algorithm transparency, setting ethical standards, and incorporating human oversight can enhance trust and reliability. The study concludes that while AI has the potential to transform healthcare, ethical safeguards and responsible implementation are crucial to ensuring fair and patient-centric medical solutions.

The role of big data analytics in business decision-making. It highlights how predictive modeling, machine learning, and real-time data processing improve efficiency, customer insights, and risk management. [8] The study presents case studies from various industries to showcase the benefits of data-driven strategies. However, challenges such as cybersecurity risks, high implementation costs, and the need for skilled professionals are also explored. The research emphasizes that while big data enhances business performance, ethical data governance and technological investment are necessary to maximize its potential responsibly and effectively.

The growing impact of social media on consumer behavior and marketing strategies. It investigates how platforms like Instagram, Twitter, and Facebook shape purchasing decisions**,** brand image, and customer engagement. [9] The study evaluates marketing techniques such as influencer collaborations, targeted advertisements, and content personalization. Ethical concerns, including misinformation**,** data privacy, and evolving regulatory standards, are also addressed. The research highlights that while social media is a powerful tool for digital marketing, businesses must adopt responsible advertising practices and adapt to regulatory changes to maintain consumer trust and brand credibility.

3.1 Limitations of Existing Work

1. High False Positives: Many fraud detection systems incorrectly flag legitimate transactions as fraudulent, leading to inconvenience for users.

2. Imbalanced Data: Fraudulent transactions are much fewer than genuine ones, making it difficult for models to learn effectively.

3. Adaptability Issues: Traditional models struggle to detect new fraud patterns as fraudsters continuously change their tactics.

4. Computational Complexity: Some advanced models require significant processing power, making real-time detection challenging.

4. Lack of Interpretability: Many machine learning models, especially deep learning-based ones, act as black boxes, making it hard to understand why a transaction is classified as fraud.

5. Data Privacy Concerns: Sharing transaction data for fraud detection may raise privacy and security issues.

3.2 Objective

- **Need for Advanced Fraud Detection**: Traditional fraud detection methods are insufficient against evolving fraudulent tactics, necessitating AI-driven solutions.
- **Machine Learning for Enhanced Accuracy**: ML models improve fraud detection by identifying complex patterns and minimizing false positives.
- **Real-Time Fraud Detection & Prevention**: AI enables instant identification of suspicious transactions, preventing fraud before financial losses occur.
- **Data-Driven Insights for Proactive Risk Management**: Analyzing transaction data helps predict and mitigate emerging fraud risks effectively.
- **Improving Operational Efficiency**: Automation reduces manual intervention, lowers costs, and enhances compliance with regulatory standards.

4. Proposed Work

4.1 Conceptual Design Diagram

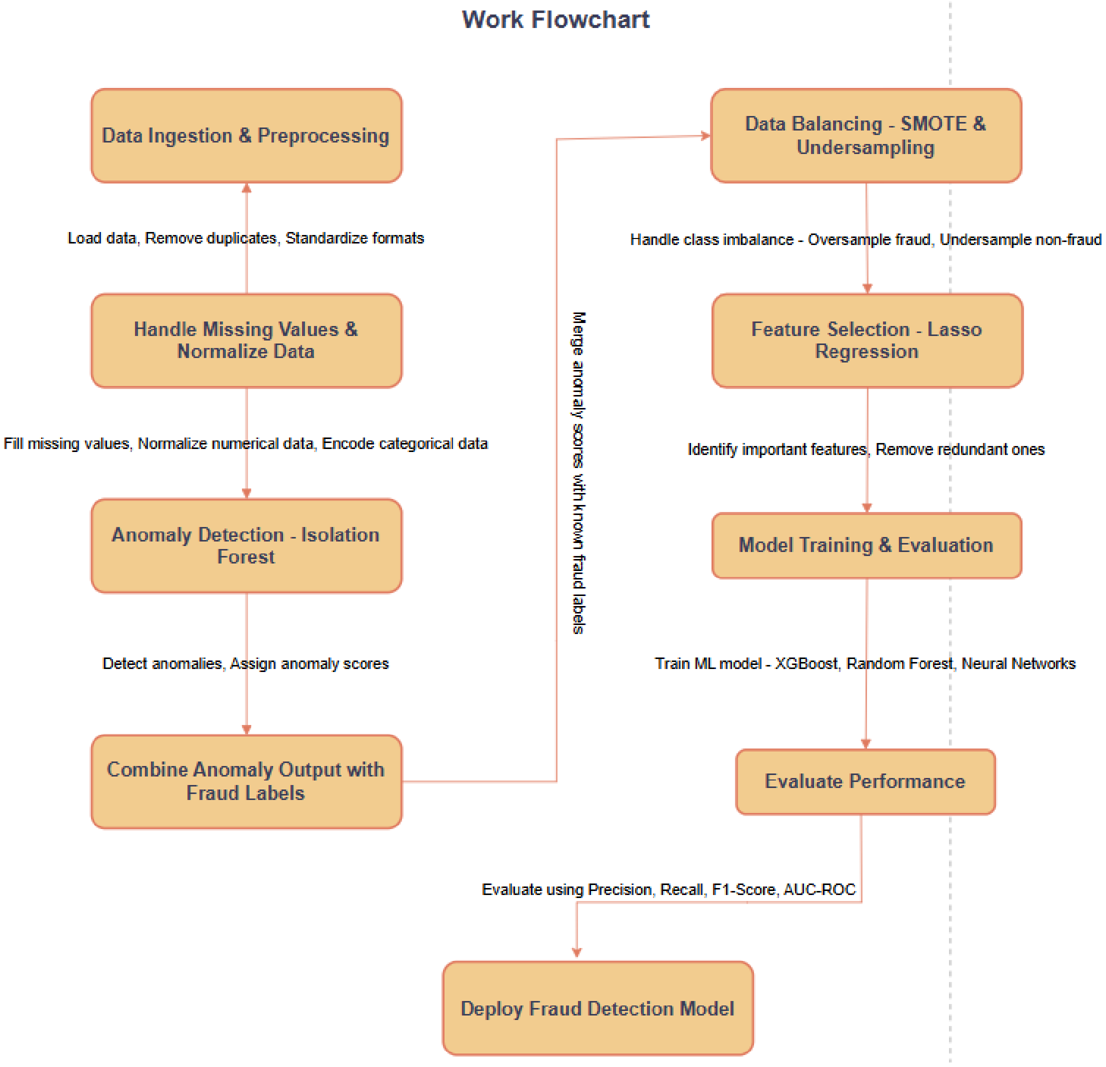
Work Flowchart
Data Ingestion & Preprocessing
Load data, Remove duplicates, Standardize formats
Handle Missing Values & Normalize Data
Fill missing values, Normalize numerical data, Encode categorical data
Anomaly Detection - Isolation Forest
Detect anomalies, Assign anomaly scores
Combine Anomaly Output with Fraud Labels
Merge anomaly scores with known fraud labels
Data Balancing - SMOTE & Undersampling
Handle class imbalance - Oversample fraud, Undersample non-fraud
Feature Selection - Lasso Regression
Identify important features, Remove redundant ones
Model Training & Evaluation
Train ML model - XGBoost, Random Forest, Neural Networks
Evaluate Performance
Evaluate using Precision, Recall, F1-Score, AUC-ROC
Deploy Fraud Detection Model

Figure : 1. Workflow Chart of the Proposed Scheme.

4.2 Dataset Description

A credit card fraud detection dataset consists of transaction records labeled as fraudulent or legitimate, helping machine learning models identify suspicious activities. The dataset is typically highly imbalanced, with fraudulent transactions being significantly fewer than legitimate ones. To handle this imbalance, techniques like *oversampling (e.g., SMOTE), undersampling, or anomaly detection methods* are often used to ensure the model does not become biased towards the majority class. The dataset includes various transaction-related and behavioral features, some of which are transformed using Principal Component Analysis (PCA) for anonymization. These transformed features are labeled as *V1, V2, ..., V28*, ensuring user privacy while preserving important data patterns. Below is a structured table of key features present in the dataset:

Table 1. Key Features in Dataset

Feature	**Description**
Time	The timestamp of the transaction, measured in seconds since the first transaction in the dataset.
V1-V28	Anonymized numerical features obtained using PCA transformation to protect sensitive information.
Amount	The transaction amount in monetary units, which can help detect unusually high or low transactions.
Class	The target variable indicating fraud status: 0 for legitimate transactions, 1 for fraudulent transactions.

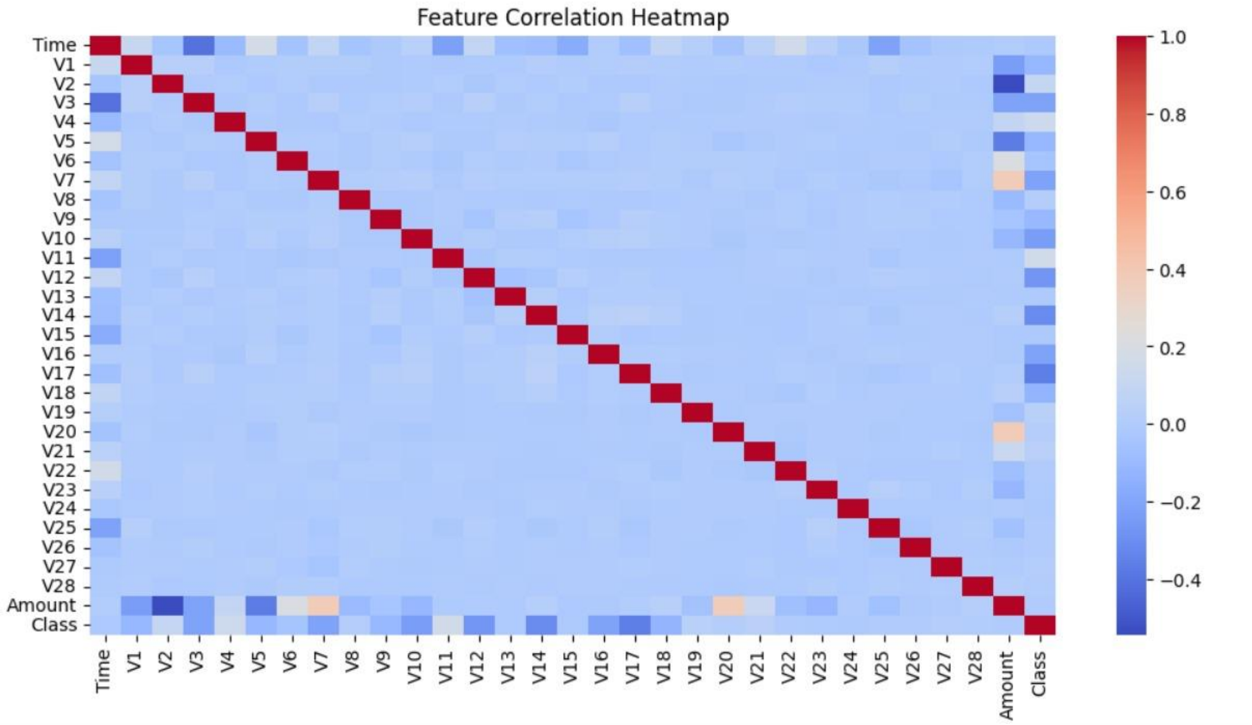

Figure: 2. Feature Correlation Heatmap

4.2.1 Handling Imbalanced Data

Since fraudulent transactions make up a very small portion of the dataset, models trained without balancing techniques may struggle to detect fraud effectively. Common solutions include Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic fraudulent samples, undersampling the majority class, and using anomaly detection methods that focus on identifying rare cases. Additionally, evaluation metrics like Precision, Recall, F1-score, and AUC-ROC are preferred over simple accuracy, as they better reflect model performance in imbalanced datasets.

4.3 Proposed Algorithm: Methodology

4.3.1 Data Loading and Preprocessing

Step 1 Loading the Dataset
The dataset is loaded from a CSV file using the Pandas library. After loading, basic checks are performed to understand the structure and quality of the data. The .info() function provides details about the number of entries, column names, data types, and any missing values. The. describe() function generates statistical summaries such as mean, standard deviation, and percentiles for numerical features, helping in data exploration.

Step 2 Feature and Label Separation
The dataset consists of multiple numerical features along with a target variable, labeled as "Class". The features (X) include all columns except "Class," as they provide the input attributes for fraud detection. The target label (y) is extracted separately from the "Class" column, where transactions are classified as either 0 (legitimate transactions) or 1 (fraudulent transactions). This separation ensures that the fraud detection model is trained using relevant feature inputs while predicting the correct labels.

Step 3 Handling Missing Values
Missing values in the dataset can lead to inconsistencies in model training and reduce accuracy. To address this, all missing values are replaced with the median of their respective columns. The median is chosen over the mean because it is less sensitive to extreme values and outliers, ensuring a more balanced data distribution. This step maintains data integrity and prevents any missing values from affecting model performance.

Step 4 Feature Scaling
Since different features in the dataset may have varying ranges and magnitudes, standardization is applied to normalize the data. Z-score normalization is used, where each feature is transformed to have a mean of zero and a standard deviation of one. This ensures that no feature dominates others due to differences in scale. Standardization is particularly important for machine learning models that rely on distance-based calculations, as it helps improve model efficiency and convergence. The StandardScaler from the Scikit-Learn library is used to achieve this transformation.

4.3.2 Handling Imbalanced Data Using SMOTE and Under-Sampling

Step 1 Issue of Imbalanced Data
In credit card fraud detection, the dataset is highly imbalanced, meaning that the number of legitimate transactions is significantly higher than the number of fraudulent ones. Since machine learning models tend to favor the majority class, they may struggle to correctly identify fraudulent

transactions. To address this issue, a combination of SMOTE (Synthetic Minority Over-sampling Technique) and Random Under-sampling is applied to balance the dataset.

SMOTE (Synthetic Minority Over-sampling Technique)
SMOTE is an oversampling technique used to artificially increase the number of fraudulent transactions. Instead of simply duplicating existing fraud cases, SMOTE creates new synthetic samples by interpolating between real fraud cases. This helps the model learn more diverse fraud patterns and improves its ability to detect fraudulent transactions.

Random Under-sampling
While SMOTE increases the number of fraudulent transactions, the dataset can still remain imbalanced due to the overwhelming number of legitimate transactions. To further balance the dataset, Random Under-sampling is applied, which randomly removes some legitimate transactions. This prevents the model from being biased toward normal transactions while maintaining a representative dataset.

Step 2 Combining SMOTE and Under-sampling
To ensure an optimal balance between fraudulent and legitimate transactions, *a resampling pipeline is created* that applies both techniques in the following order:
1. **SMOTE** increases the number of fraudulent transactions to 30% of the total dataset.
2. **Random Under-sampling** then reduces the number of legitimate transactions so that the final dataset maintains a balanced ratio.

By combining these two methods, the dataset becomes more evenly distributed, helping the machine learning model to recognize fraudulent transactions more accurately while preventing bias toward normal transactions.

4.3.3 Feature Selection Using LASSO Regression

Step1 Importance of Feature Selection
In machine learning, selecting the most relevant features improves model accuracy and efficiency. A dataset may contain unnecessary or less significant features that do not contribute much to fraud detection. Removing such features helps reduce complexity, prevents overfitting, and improves the model’s generalization.

To achieve this, LASSO Regression (Least Absolute Shrinkage and Selection Operator) is used, which is a type of regression technique that applies L1 regularization. This method shrinks the coefficients of less important features to zero, effectively eliminating them from the dataset.

Step 2 Applying LASSO for Feature Selection
Feature Scaling: Before applying LASSO, all features are standardized using Z-score normalization to ensure fair comparison across different scales. This ensures that no single feature dominates the selection process.

Fitting the LASSO Model: A logistic regression model with L1 regularization is trained on the dataset. The regularization parameter (C) controls how aggressively features are removed.

Step 3 Selecting Important Features: After training, only the features with nonzero coefficients are retained, while the rest are discarded. These retained features are considered the most important for detecting fraud.

Step 4 Splitting the Dataset for Training and Testing
Once the most important features are selected, the dataset is divided into training (80%) and testing (20%) sets. The training set is used to train the final model, while the testing set is used to evaluate its performance.

4.3.4 Feature Selection Using LASSO Regression

Fraud detection is a challenging task due to the highly imbalanced nature of transaction data. To effectively identify fraudulent activities, different machine learning models can be applied, each with its unique strengths. This project explores three powerful models: Artificial Neural Networks (ANN)**,** eXtreme Gradient Boosting (XGBoost), and Convolutional Neural Networks (CNN), which leverage different approaches to detect fraudulent transactions accurately.

Artificial Neural Networks (ANN) are inspired by the human brain and consist of multiple layers of interconnected neurons that learn patterns in the data. The model contains an input layer that takes transaction features, hidden layers that process complex relationships, and an output layer that classifies transactions as legitimate or fraudulent. Each neuron in the ANN computes a weighted sum of inputs and applies an activation function such as ReLU for hidden layers and sigmoid for binary classification. The output is obtained using the sigmoid function.

$$y^{\wedge} = \frac{1}{1+e-^{Z}} \qquad (1)$$

Z is the weighted sum of inputs. The ANN model is trained using backpropagation and gradient descent, adjusting the weights to minimize classification errors. ANNs are particularly useful in capturing non-linear patterns in transaction data, making them highly effective for fraud detection.

XGBoost (eXtreme Gradient Boosting) is a robust machine learning algorithm based on decision trees. It works by combining multiple weak learners (decision trees) in a sequential manner, where each new tree corrects the mistakes made by the previous ones. XGBoost optimizes performance using gradient boosting, which minimizes classification errors while preventing overfitting through regularization techniques. The objective function of XGBoost consists of two parts: the loss function, which measures the error between actual and predicted values, and the regularization term, which penalizes model complexity.

$$l(\theta) = \sum_{i=1}^{n} l(y_i - y^{\wedge}{}_{i)} + \sum_{k=1}^{k} \Omega(f_{k)} \qquad (2)$$

Convolutional Neural Networks (CNNs), although traditionally used in image processing, can be adapted for fraud detection by treating transaction data as structured input. CNNs excel at capturing local dependencies and hierarchical relationships between features. The model applies convolutional layers that use filters (kernels) to detect patterns, followed by pooling layers that reduce the dimensionality while retaining important information. The extracted features are then passed through fully connected layers for classification. The convolution operation is mathematically defined as:

$$S(i,j)=(X * K)(I\ ,\ j)=\sum m \sum n\ \ X(i-m,j-n)K(m,n) \qquad (3)$$

X. The final classification probability is obtained using the softmax or sigmoid activation function**.** CNNs are particularly useful when analyzing transaction sequences or detecting hidden relationships between features, making them a powerful tool for fraud detection.

5. Performance Evaluation

5.1 Environment Setting (HW/SW requirements)

To successfully implement and execute the proposed project, the following hardware and software resources were utilized:

Hardware Requirements

- Laptop Model: HP Laptop
- Processor: (Specify if needed, e.g., Intel Core i5/i7, AMD Ryzen)
- RAM: (Specify RAM size, e.g., 8GB/16GB)
- Storage**:** (Specify storage capacity, e.g., 256GB SSD/1TB HDD)

Software Requirements

- Platform: Google Colab (for coding and execution)
- Programming Language: Python
- Libraries Used:
  - NumPy – for numerical computations
  - Pandas – for data handling and preprocessing
  - Matplotlib & Seaborn – for data visualization
  - Scikit-learn – for machine learning model development
  - TensorFlow/PyTorch – for deep learning models

5.2 Performance test

1. Accuracy

Accuracy is the ratio of correctly predicted instances (both true positives and true negatives) to the total instances.

2. Precision

Precision measures the accuracy of the positive predictions made by the model, which in this case refers to how many of the emails classified as spam are truly spam. It is defined as:

$$\text{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \times \mathbf{100} \qquad (1)$$

3. Recall

Recall (or sensitivity) is the ratio of true positives to the sum of true positives and false negatives. It measures the model's ability to identify all relevant instances.

$$\text{Recall} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \times \mathbf{10} \qquad (2)$$

4. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both. A good F1-score of indicates a good balance between precision and recall.

$$\text{F1 Score} = \mathbf{2} \times \frac{\boldsymbol{Precision \times Recall}}{\boldsymbol{Precision+Recall}} \qquad (3)$$

5. AUC-ROC

The AUC-ROC (Area Under the Receiver Operating Characteristic Curve) is a crucial metric that evaluates the model's performance across all classification thresholds. The ROC curve plots the true positive rate (recall) against the false positive rate (1-specificity). The AUC score represents the area under this curve, with a value closer to 1 indicating a better-performing model.

Table 2. Final Test Results in Percentage

Evaluation	**ANN**	**CNN**	**XGBoost**
Accuracy	0.9862	0.9868	0.9925
Precision	0.9710	0.9609	0.9773
Recall	0.9884	1.0000	1.0000
F1-Score	0.9796	0.9801	0.9885
AUC-ROC	0.9992	0.9990	0.9998

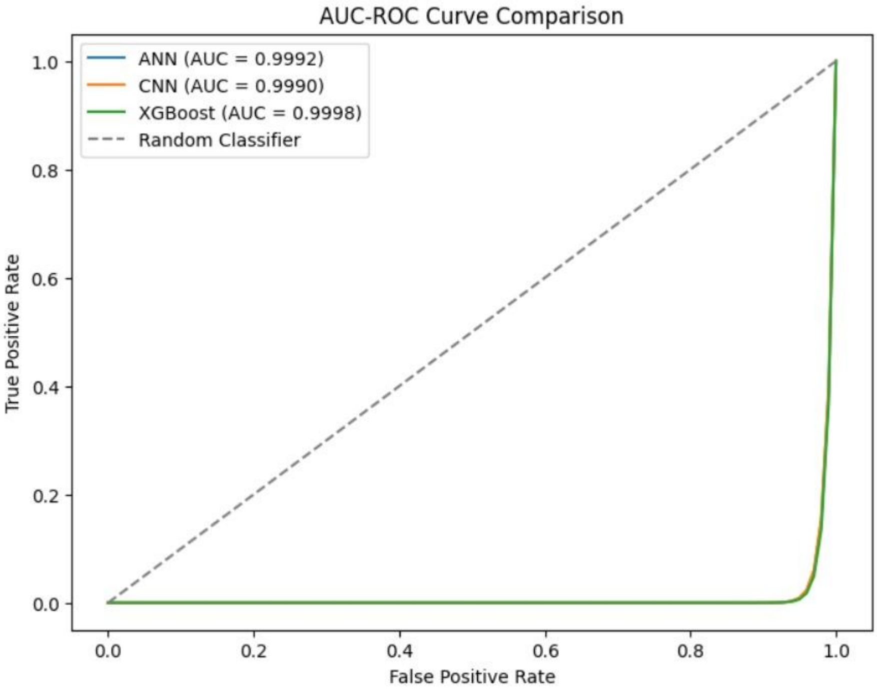

Figure: 3. AUC-ROC Curve

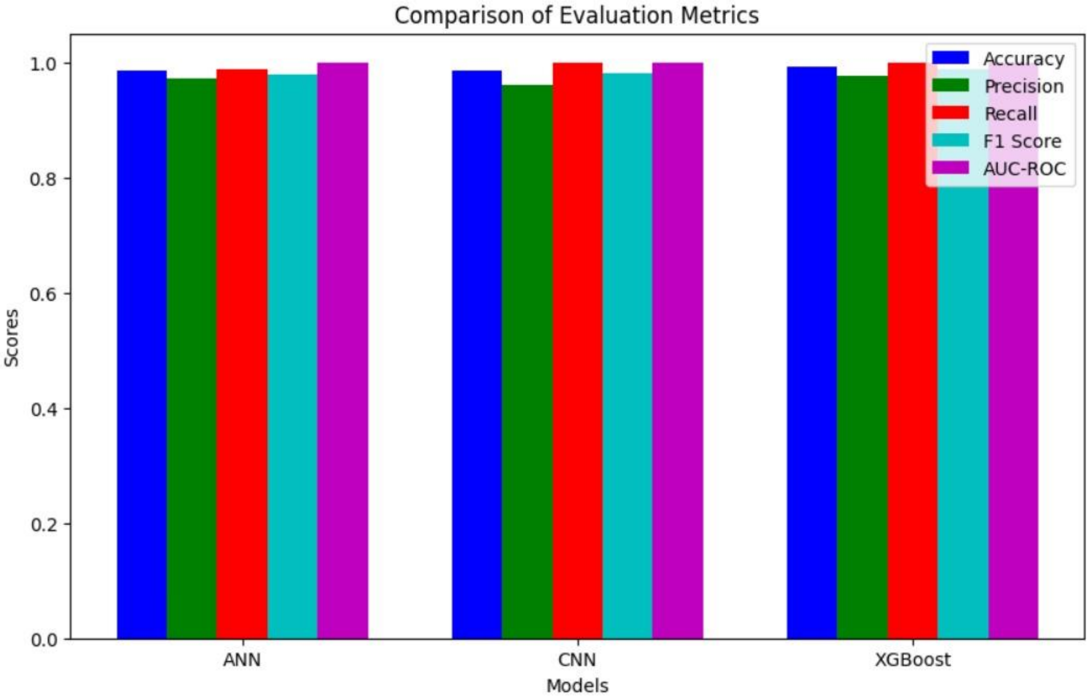

Figure: 4. Graphical Representation of Evaluation Metrics

6. Current Status and Future Plans of Project Work

Current Status

The project has successfully developed an AI-powered credit card fraud detection system**.** The key accomplishments include:

- Implementation of Machine Learning Models**:** Advanced models like Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and XGBoost have been utilized for detecting fraudulent transactions.
- Data Preprocessing & Feature Engineering: Techniques such as feature selection, anomaly detection, and data balancing were employed to enhance model performance and reduce bias.
- Performance Optimization: The system has demonstrated high detection accuracy while effectively minimizing false positive rates.

Future Work

To further refine fraud detection capabilities, the project aims to:

1. Develop Real-Time Detection Mechanisms: Implementing a system capable **of** instant transaction analysis to prevent fraud before transactions are completed.
2. Enhance Model Complexity with Advanced Architectures: Exploring Recurrent Neural Networks (RNNs) and Transformers to detect sophisticated fraud patterns.
3. Improve Transparency with Explainable AI (XAI)**:** Enhancing model interpretability to help financial institutions understand AI-driven decisions and ensure compliance with regulatory standards.
4. Expand Data Diversity: Incorporating global transaction data and cross-border transactions to improve adaptability to varying fraud patterns worldwide.

7. Conclusion

Credit card fraud detection is a critical challenge in the financial sector, and this project has demonstrated the effectiveness of AI-based models in identifying fraudulent transactions. By utilizing machine learning techniques, such as artificial neural networks (ANN), convolutional neural networks (CNN), and XGBoost, the system was able to detect fraudulent activities with high accuracy while minimizing false positives. The combination of feature selection, anomaly detection, and data balancing techniques improved model performance and ensured reliable fraud detection. Implementing these AI-driven approaches enhances financial security, reduces financial losses, and protects customers from fraudulent transactions. However, as fraudsters continue to develop sophisticated techniques, fraud detection models must continuously adapt and improve to stay ahead of emerging threats.

8. Reference

[1] Jain, Jitender. (2022). Leveraging Advanced AI and Cloud Computing for Scalable Innovations in Fintech Systems This work is licensed under CC BY-NC-SA 4.0. 10.6084/m9.figshare.28450010.

[2] Titilola, Abayomi & Olutimehin, Abayomi. (2025). Article no.JERR.131308Original Research Article Olutimehin.

[3] Odufisan, Oluwaseun & Abhulimen, Osekhonmen & Ogunti, Erastus. (2025). Harnessing Artificial Intelligence and Machine Learning for Fraud Detection and Prevention in Nigeria.. Journal of Economic Criminology. 7. 100127. 10.1016/j.jeconc.2025.100127.

[4] Malik, Mubashir & Lali, Hina. (2025). The Future of Modern Finance: AI-Driven Fraud Detection and Energy Market Forecasting.

[5] Hafez, I.Y., Hafez, A.Y., Saleh, A. et al. A systematic review of AI-enhanced techniques in credit card fraud detection. J Big Data 12, 6 (2025). https://doi.org/10.1186/s40537-024-01048-8

[6] Zou, Y., & Cheng, D. (2025). Effective high-order graph representation learning for credit card fraud detection. arXiv preprint arXiv:2503.01556.

[7] Bonde, L., & Bichanga, A. K. (2025). Improving Credit Card Fraud Detection with Ensemble Deep Learning-Based Models: A Hybrid Approach Using SMOTE-ENN. Journal of Computing Theories and Applications, 2(3), 384.

[8] Ileberi, E. (2023). Improved Machine Learning methods for enhanced credit card fraud detection. University of Johannesburg (South Africa).

[9] Wang, S. X. (2024). The Application of Artificial Intelligence Based Risk Management Models in Financial Markets. Open Journal of Social Sciences, 12, 274-284. https://doi.org/10.4236/jss.2024.1211019