CS-351 Introduction to Artificial Intelligence

Semester Project Proposal

Ghulam Ishaq Khan Institute of Engineering Sciences and Technology

Team Members	**Instructor Details**
Salman Shahzad Ali 2023504	Assistant Professor Ayaz Umer
Murtaza Ali Khan 2023472	Faculty of FCSE
Muhammad Huzaifa 2023442	ayaz.umer@giki.edu.pk
Syed Murtaza Salman 2023694	

Fraud Detection System Using Machine Learning

Project Title

Real-Time Credit Card Fraud Detection Using Machine Learning

Problem Statement

Financial fraud costs billions annually, with credit card fraud being one of the most prevalent types. Traditional rule-based fraud detection systems are inefficient and produce high false-positive rates, leading to legitimate transactions being blocked. This project aims to develop a machine learning model that accurately identifies fraudulent transactions in real-time while minimizing false positives.

Motivation (Why This Topic?)

- **Real-world Impact:** Fraud detection directly protects consumers and financial institutions
- **Industry Demand:** Banks and fintech companies actively seek advanced ML solutions for fraud prevention
- **Challenging Dataset:** Highly imbalanced datasets (fraud cases are rare) present a realistic ML problem
- **Practical Learning:** Covers data preprocessing, model selection, and performance evaluation on imbalanced data

Dataset Source

Credit Card Fraud Detection Dataset from Kaggle (publicly available):

- Contains 284,807 transactions with 31 features
- Includes PCA-transformed features and transaction amounts
- Highly imbalanced (~0.17% fraud cases), reflecting real-world scenarios

Proposed Methodology

1. Data Preprocessing: Handle class imbalance using SMOTE and undersampling techniques

2. Feature Engineering: Analyze transaction patterns and normalize features

3. Model Development: Implement and compare multiple algorithms:

- Logistic Regression (baseline)
- Random Forest Classifier
- Support Vector Machines (SVM)
- Gradient Boosting (XGBoost)

4. Evaluation: Use precision, recall, F1-score, and ROC-AUC (accuracy alone is misleading for imbalanced data)

5. Optimization: Tune hyperparameters for maximum detection accuracy with minimal false positives

Expected Outcomes

- A trained model with >95% recall and >90% precision on fraud detection
- Comparison report showing which algorithm performs best
- Analysis of key features that indicate fraudulent transactions
- Insights into fraud patterns and detection limitations

Tools & Libraries

- **Python 3** (primary language)
- **Scikit-learn** (ML algorithms)
- **Pandas & NumPy** (data manipulation)
- **Matplotlib & Seaborn** (visualization)
- **XGBoost** (advanced boosting)
- **Imbalanced-learn** (SMOTE for handling imbalance)
- **Flask or Streamlit** (web UI for model interaction)
- **GitHub** (code repository)