Explainable Fraud Alert System

Domain: Finance & Banking

Abstract:

In today's digital banking environment, fraud detection systems must not only identify suspicious transactions accurately but also explain why they were flagged. This project focuses on building an Explainable Fraud Alert System that uses machine learning and Natural Language Processing (NLP) to both detect and explain anomalous financial transactions.

The system models user behaviour based on transaction patterns such as time, amount, location, and merchant category. An anomaly detection algorithm identifies deviations from this pattern, and a Large Language Model (LLM) is used to generate simple, user-friendly explanations for each fraud alert. To improve user experience and trust, the system includes a chatbot-style conversational interface where users can interact and understand why a transaction was flagged. Additionally, the system provides a risk score for each transaction to support decision-making.

This explainable approach not only enhances trust in AI-powered banking systems but also supports transparency and accountability in fraud prevention.