BAN 620 Data Mining

INTERIM PROJECT REPORT

Team -2

NAME	NET ID:
Veera Venkata Sai Kalyan Kaparaju	ss4785
Adrian Padilla	bu7569
ARPIT GHATORIYA	op9580
Praneeth Reddy Gaddam	id8973

Instructor: Balaraman rajan

INTERIM PROJECT REPORT: CREDIT CARD FRAUD DETECTION

Introduction:

Our project is focused on spotting tricky transactions in credit card data. This dataset gives us details about transactions, and we're using it to build some predictive analysis, that can find out any suspicious activity. The aim is to make sure credit card transactions are safe and secure by understanding and preventing fraudulent transactions with the help of this detailed dataset.

Source Link: link.

Dataset Overview:

The dataset encompasses credit card transaction details, including indicators for fraudulent transactions. This dataset helps us get insight of the credit card transactions happening around USA. Key features include -

1. Merchant details - name, category, latitude, longitude, transaction number
2. Transaction details – date & time, amount, id, place
3. Customer details – first name, last name, date of birth, gender, address, job, population of city

Data Available:

Total Records: 1,048,575
Variables: 22

Objective:

The objective behind studying this dataset lies in the potential to improve fraud detection, enhance consumer protection & gain valuable business insights. We can analyze these patterns and prevent it from happening in the future by using the historical data available.

- ***Credit Card Spending Behavior Analysis:***
  Understanding how people use credit cards is vital for banks. Using regression analysis, we aim to predict transaction amounts based on factors like age, state, and job roles. Our target is the transaction amount ("amt"), and we measure accuracy through Root Mean Squared Error (RMSE). This helps us see how age, location, and job types influence credit card spending.

- ***Location-Based Transaction Analysis:***
  Geographical insights are crucial for preventing fraud. Employing KNN classification, this analysis predicts location and category using demographic and transaction features. The accuracy metric, Classification Accuracy, helps us understand spending variations across regions. This allows for tailored anti-fraud measures based on location.

- ***Fraud Data Pattern Analysis:***
  Identifying patterns linked to fraudulent transactions is essential for beefing up security. Using a Classification approach, specifically Decision Tree or Random Forest Classifier, we predict fraud instances based on transaction features. Our focus is on minimizing false negatives, ensuring a solid strategy for spotting potential fraud cases tied to specific merchants.

- ***Prevention of Fraud:***
  Pooling insights from spending behavior, location patterns, and merchant-related fraud data, we're crafting both rule-based and model-based fraud prevention strategies. Target variables include implementing rules derived from spending behavior, geographical patterns, and specific merchant fraud data. Evaluation metrics encompass the reduction in fraud instances, false positives, and false negatives, ensuring the effectiveness of our fraud prevention measures.

Our study on credit card transactions looks at different aspects to make things safer and easier to understand. We're checking out how people use credit cards by looking at age, state, and jobs to guess how much money they spend. We're also using some analysis techniques to predict spending in different areas. To keep things secure, we're looking closely at fraud patterns, especially those linked to specific stores. Our plan to prevent fraud is like a mix of smart rules and tools to catch and stop fraud, making sure credit card transactions are safe from potential problems.