DR. B.R. AMBEDKAR NATIONAL INSTITUTE OF TECHNOLOGY JALANDHAR, PUNJAB-144008

Department of Computer Science and Engineering

Mini Project Report File

For

Data Mining and Analytics (CSDC0335)

B.Tech (CSE) 5th Semester (Sec-B)

Prepared during (July-Dec 2025) By

V Sai Kiran Reddy-23103156

Uppalapati Navya -23103155

Currency Usage Patterns - A Survey-Based Analysis

1. Introduction

In recent years, digital payment systems have transformed the way individuals manage their finances. From UPI transactions to cryptocurrency adoption, the shift towards a cashless economy has been rapid, particularly among the youth.

This project, titled **"Currency Usage Patterns"**, aims to analyze how people spend, save, and make payments in the digital age. The study explores behavioral factors like trust in online banking, awareness of cryptocurrencies, savings habits, and susceptibility to financial fraud.

Data was collected via a **Google Form survey** targeting students and working individuals to capture diverse monetary habits and perceptions about digital finance.

2. Objective of the Study

The main objectives of this study are:

- To analyze spending and saving patterns among different user groups.
- To study awareness and adoption of digital payment methods and cryptocurrencies.
- To assess users' perception of online banking security.
- To predict the likelihood of being a financial fraud victim based on behavioral factors.

3. Dataset Overview

- **Total responses:** 198
- **Attributes (columns):** 17
- **Source:** Google Form Survey
- **Major features include:**
    - Age Range
    - Gender
    - Occupation
    - Payment Method
    - Average Monthly Digital Spending

- Saving Percentage
- Familiarity with Digital Currency
- Security Rating (1–5)
- Fraud Experience (Yes/No)

A preview of the dataset revealed that most respondents belong to the **18–25 age group**, and the dominant mode of payment was **UPI/Digital Wallets**.

4. Data Preprocessing

Step 1: Data Loading & Exploration

The raw Excel file was imported using the `readxl` package in R. Initial exploration included checking data dimensions, data types, missing values, and duplicates.
The dataset contained **198 rows × 17 columns** with a few missing values and no duplicates.

Step 2: Data Cleaning & Standardization

- **Removed irrelevant columns:** Timestamp, Email address
- **Handled missing data:** Replaced categorical NAs with "Unknown" and numeric NAs with the column median.
- **Standardized text:** Converted all categorical text to lowercase and trimmed extra spaces.
- **Manual merging:** Unified similar entries (e.g., "phone pay" → "PhonePe", "google pay" → "Google Pay").After cleaning, the dataset was free from missing or inconsistent values.

Step 3: Encoding Categorical Data

All categorical attributes were converted into **numeric dummy variables** using the `fastDummies` package. After encoding, the dataset expanded to **49 numeric columns** suitable for machine learning algorithms.

Step 4: Data Transformation / Normalization

Min–Max normalization was applied to all numeric columns using the `caret` package to bring values into a **0–1 range**.This ensured equal weightage among all variables during model training.

5. Data Visualization & Insights

5.1 Demographics

- Majority of respondents (≈ 82%) belong to the **18–25** age group.
- **71%** of participants were male, and most identified as **students**.

5.2 Payment Preferences

- **UPI/Digital Wallets (≈ 90%)** emerged as the dominant payment mode.
- Cash and card payments accounted for less than 10%.
- Among UPI apps, **PhonePe** and **Google Pay** were most used.

5.3 Savings & Money Personality

- Over **60%** of users save **less than 5%** of their income.
- "**The Spender**" personality type dominates, followed by "The Saver".
- Students generally reported lower saving rates compared to working professionals.

5.4 Digital Awareness & Cryptocurrency

- Around **55%** of respondents were **"slightly familiar"** with digital currencies.
- Only **15%** had ever purchased or used cryptocurrency.
- Higher familiarity correlated with higher online security ratings.

5.5 Fraud & Security Analysis

- About **10%** of respondents had experienced financial fraud.
- Those who faced fraud tend to rate online banking platforms as **less secure**.
- Correlation heatmap showed that **security perception**, **digital familiarity**, and **fraud experience** are weakly but meaningfully related.

6. Predictive Modeling

6.1 Model Chosen

To predict whether a user had ever been a victim of financial fraud, a **Logistic Regression Model** was built using normalized behavioral and awareness variables.

Target Variable:
`Have you ever been a victim of financial fraud or scams?`

Predictor Variables:

- Security rating
- UPI/Digital wallet usage
- Cryptocurrency familiarity
- Cryptocurrency usage
- Saving percentage

6.2 Model Implementation

The model was trained using **80% training data** and tested on **20% unseen data**.
Key packages used: `caret`, `pROC`, `ggplot2`.

6.3 Results

Metric	Value	Interpretation
Accuracy	89.74%	High accuracy, model predicts most non-fraud cases correctly
AUC (ROC)	0.63	Moderate discrimination ability
Sensitivity	100%	All non-fraud cases identified correctly
Specificity	0%	Could not identify rare fraud cases due to class imbalance

Inference:
The model performs well in recognizing safe users but struggles with fraud prediction because the dataset contains very few fraud cases.
Despite this, behavioral attributes such as **digital familiarity**, **security perception**, and **UPI usage** show meaningful predictive power.

6.4 Visual Results

- **ROC Curve:** Demonstrates a moderately strong model (AUC = 0.63).
- **Predicted Probability Distribution:** Distinguishes users with low vs. high likelihood of fraud involvement.

7. Discussion & Inferences

1. Majority of respondents rely heavily on digital payment methods.
2. Low savings and frequent UPI usage indicate impulsive spending behavior among youth.
3. Awareness of cryptocurrency remains limited, but positively linked to perceived online security.
4. Logistic regression highlights that **trust and awareness factors** influence fraud exposure.
5. Imbalanced dataset limits the model's ability to predict fraud precisely — future data collection should ensure balanced sampling.

8. Conclusion

The "Currency Usage Patterns" analysis successfully identifies behavioral and financial trends among digital-age users.Findings reveal that while users are enthusiastic about digital payments, **awareness about security and digital finance** is still developing.The predictive model, though moderately effective, demonstrates potential in using behavioral data to anticipate financial fraud vulnerability.

9. Tools and Technologies Used

Tool	Purpose
R Programming	Data cleaning, preprocessing, visualization, and modeling
ggplot2	Data visualization
caret	Model training and normalization
pROC	ROC/AUC evaluation
Google Forms	Data collection

10. Github References

--> Github Repository Link