

## Executive Summary

The rapid expansion of India's Unified Payments Interface (UPI) has transformed the nation's digital economy, enabling seamless peer-to-peer and merchant transactions. However, this exponential growth has introduced challenges in forecasting transaction demand and detecting fraudulent activities. This project, titled "**Predictive Analytics in the UPI Ecosystem: Demand Forecasting and Fraud Detection**," applies data-driven modeling to address these dual challenges.

The study employs two predictive analytics frameworks:

1. **Supervised Learning Regression (SLR)** for **UPI demand forecasting**, and
2. **Logistic Regression** for **fraud detection**.

Both models utilize cleaned and preprocessed datasets sourced from **RBI**, **NPCI**, **MoSPI**, and **Kaggle**, combining economic, technological, and behavioral variables.

For **UPI Demand Forecasting**, the regression model achieved an exceptionally high explanatory power ( $R^2 = 0.998$ , Adj  $R^2 = 0.997$ ). The analysis identified **Smartphone Penetration (p = 0.000)**, **Internet Users (p = 0.005)**, and **PMJDY Accounts (p = 0.017)** as the most significant predictors of digital payment growth. The forecast predicts approximately **272 million UPI transactions in 2025**, underscoring the vital role of digital infrastructure and financial inclusion in sustaining growth. Business recommendations emphasize expanding smartphone access, improving broadband connectivity, and promoting UPI-linked Jan Dhan accounts to enhance adoption in underserved regions.

For **Fraud Detection**, the **Logistic Regression model** demonstrated an **accuracy of 83%**, with strong precision (78%) and recall (80%) for fraudulent transactions. Key predictors included **Failed Login Attempts**, **Transaction Amount**, and **Location Risk Score**, which collectively explain **33% of the variation in fraud probability (Pseudo R<sup>2</sup> = 0.33)**. The model effectively identified 170 users with a fraud probability above 70%, offering a practical tool for real-time fraud risk monitoring. Recommended strategies include integrating geospatial risk scoring, deploying adaptive authentication, and continuous model retraining to counter evolving fraud techniques.

Overall, this project illustrates the potential of predictive analytics in optimizing **digital payment operations and security**. The findings support policy formulation for the **Reserve Bank of India (RBI)**, **NPCI**, and **fintech firms** by enabling data-backed decisions to drive financial inclusion, strengthen fraud prevention, and sustain UPI's momentum in India's transition toward a cashless economy.