

# Customer Transaction Prediction Project Documentation

## Introduction

Customer Transaction Prediction focuses on identifying whether a customer is likely to make a transaction in the future based on historical behavioral and transactional data. By leveraging machine learning techniques, this project aims to uncover hidden patterns in customer data and provide predictive insights that support proactive business strategies.

## Business Case

Financial institutions and businesses lose significant revenue due to missed opportunities when potential customers are not identified early. Predicting customer transactions enables organizations to:

- 1 • Improve targeted marketing and personalized offers
- 2 • Increase customer engagement and conversion rates
- 3 • Optimize resource allocation
- 4 • Enhance overall revenue and customer lifetime value

## Task

The primary task of this project is to build a supervised machine learning model that accurately predicts whether a customer will make a transaction. This involves analyzing historical data, engineering meaningful features, training multiple models, and selecting the best-performing model based on evaluation metrics.

## Domain Analysis

This project falls under the Banking and Financial Services domain. Customer transaction behavior is influenced by multiple factors such as spending patterns, engagement levels, and historical activity. Understanding these factors through data analysis helps institutions make data-driven decisions and reduce uncertainty in customer behavior prediction.

## Dataset

- 1 • 200,000+ observations
- 2 • 200+ numerical features representing customer behavior
- 3 • One unique customer identifier
- 4 • One binary target variable indicating whether a transaction occurred

The dataset is highly dimensional and requires careful preprocessing and feature handling.

## Data Preprocessing and Feature Engineering

- 1 • Handling missing and inconsistent values
- 2 • Scaling and normalizing numerical features
- 3 • Removing redundant or low-variance features
- 4 • Transforming raw variables into meaningful representations

- 2 • Improving data quality for better model learning
- 3 • Ensuring compatibility with machine learning algorithms

## Feature Selection

Given the high dimensionality of the dataset, feature selection was essential to reduce noise and overfitting, improve model performance, and decrease training time. Statistical techniques and model-based importance methods were used to identify the most impactful features contributing to transaction prediction.

## Model Creation and Selection

- 1 • Logistic Regression
- 2 • Decision Trees
- 3 • Random Forest
- 4 • Gradient Boosting (XGBoost)

Models were compared using evaluation metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The final model was selected based on its ability to generalize well and deliver consistent performance on unseen data.

## Conclusion

This project demonstrates how machine learning can be effectively applied to predict customer transactions using large-scale data. The solution provides actionable insights that can help businesses improve decision-making, enhance customer targeting, and drive measurable business impact.