# Bank Loan Prediction

## Final Report

## Introduction

Banks often struggle with inefficient loan targeting, leading to unnecessary resource allocation and missed revenue opportunities. This project aims to develop a predictive model to identify customers most likely to accept a personal loan offer. By leveraging machine learning, the model enhances marketing efficiency, reduces costs, and improves customer targeting.  
  
The objective is to analyze customer demographic, financial, and behavioral data to derive insights into loan acceptance patterns. The predictive model will help banks optimize decision-making, ensuring personalized marketing efforts and improved financial risk management.

## Business Impact

Analyzing personal loan data has significant implications for both the bank and its customers:  
1. Enhanced Customer Targeting – Insights into demographics and financial behavior allow personalized marketing strategies.  
2. Increased Loan Approvals – Identifying key predictors of loan acceptance streamlines the approval process and reduces loan disbursal time.  
3. Improved Risk Management – Understanding credit card spending and income levels enables better risk assessment and safer loan product design.  
4. Revenue Growth – Targeting high-income groups and high-acceptance areas maximizes loan-related revenue.  
5. Customer Satisfaction – Personalized offerings based on education level and family size improve customer loyalty.

## Dataset

### Primary Dataset

Name: Bank Personal Loan Dataset  
Source: Kaggle  
Size: 5000 rows × 14 columns

### Supplementary Dataset

Name: Economic Indicators Dataset  
Source: World Bank  
Size: 120 rows × 3 columns (Interest Rate, Inflation Rate, Unemployment Rate)

## Data Analysis & Computation

### Data Wrangling & Cleaning

- Missing Values: 5% missing data in Mortgage, handled using median imputation.  
- Outliers: Detected in Income and Credit Card Average Spend, treated by capping at 1st and 99th percentiles.  
- Data Transformation: One-hot encoding for Education, Min-Max scaling for numerical features.

### Exploratory Data Analysis (EDA)

1. Age Distribution: Customers between 30-50 years form the primary demographic for loan marketing.  
2. Income & Loan Acceptance: Higher income positively correlates with loan acceptance.  
3. Education & Income: Advanced education levels correspond to higher income and a higher likelihood of accepting loans.  
4. Family Size Impact: Smaller families (1-2 members) show higher acceptance rates.  
5. ZIP Code Trends: Certain regions exhibit high loan acceptance rates.  
6. Correlation Analysis: Strong correlations exist between Income, Credit Card Spend, and loan acceptance.

## Predictive Modeling

Three machine learning models were trained and compared:  
1. Logistic Regression: Simple and interpretable, with decent accuracy.  
2. Random Forest: Provides high accuracy and handles non-linearity well.  
3. XGBoost: Optimized for performance, delivering the highest accuracy.

### Model Evaluation & Performance Metrics

- Precision & Recall: Ensures a balance between false positives and false negatives.  
- F1 Score & ROC-AUC: Helps determine the best-performing model.  
- Best Model: XGBoost achieved 92% accuracy, outperforming others in predictive power.

## Challenges and Solutions

1. Data Imbalance: The dataset was imbalanced, with only 10% positive cases. Solution: Used SMOTE to balance data.  
2. Feature Redundancy: Some features had high multicollinearity. Solution: Removed redundant variables after VIF analysis.  
3. Outliers in Financial Data: Extreme income values impacted model predictions. Solution: Applied winsorization and log transformation.  
4. Model Overfitting: The complex models showed signs of overfitting. Solution: Applied cross-validation and regularization techniques.

## Dashboard Description

### Use Case

A Tableau dashboard was developed to visualize insights and provide real-time loan acceptance predictions.

### Dashboard Features

1. Loan Prediction Tool: Allows users to enter demographic and financial data to get an instant prediction.  
2. Visualizations:  
- Heatmaps show feature correlations.  
- Bar charts illustrate loan acceptance trends by family size and education level.

## Conclusion & Future Work

### Key Findings

- Customers aged 30-50 years, with higher incomes and advanced education levels, are most likely to accept personal loans.  
- Income and average credit card spending are strong predictors of loan acceptance.  
- The model successfully identifies high-potential customers, optimizing loan targeting strategies for banks.

### Future Enhancements

1. Real-Time Data Integration: Incorporate live economic indicators to enhance predictions.  
2. Feature Expansion: Explore additional behavioral attributes.  
3. Deep Learning Implementation: Evaluate neural networks for complex decision-making.  
5. A/B Testing: Implement in a real banking environment to measure the impact of predictive insights.

## References

Data Sources:  
- Kaggle: Bank Personal Loan Dataset  
- World Bank: Economic Indicators