Churn Modelling

A machine learning approach to predicting customer attrition in banking

Yash Chavad (92400584193)

Master of Computer Applications

Marwadi University, 2025-26

Understanding Customer Churn

The Challenge

Banks face significant losses when customers unexpectedly leave. Traditional methods struggle to detect early warning signs in complex, imbalanced datasets.

Key Problem: How can machine learning accurately predict customer churn from imbalanced banking data?

Project Objectives & Scope

Analyze Patterns

Study customer data to identify factors causing churn

Build Models

Train ML algorithms to predict at-risk customers

Handle Imbalance

Apply techniques to manage uneven data distribution

Evaluate Performance

Measure accuracy using comprehensive metrics

Scope: Covers data cleaning, scaling, model training with Python libraries (Pandas, NumPy, Scikit-learn), and evaluation—focusing on supervised classification without real banking deployment.

System Architecture

Data Collection

Load Churn_Modelling.csv with customer features

Preprocessing

Handle missing values, encode categories, scale features

Train-Test Split

Divide data for training and validation

Model Training

Train 6 ML algorithms and compare results

Data Preprocessing Pipeline

01

Missing Value Imputation

SimpleImputer fills numeric gaps with mean, categorical with mode

02

Categorical Encoding

LabelEncoder transforms gender, geography into numeric values

03

Feature Scaling

MinMaxScaler normalizes values between 0 and 1

04

Data Splitting

70-30 train-test split with stratification

Machine Learning Models Tested

Logistic Regression

Linear model for binary classification

K-Nearest Neighbors

Distance-based classification

Decision Tree

Rule-based hierarchical splits

Random Forest

Ensemble of decision trees

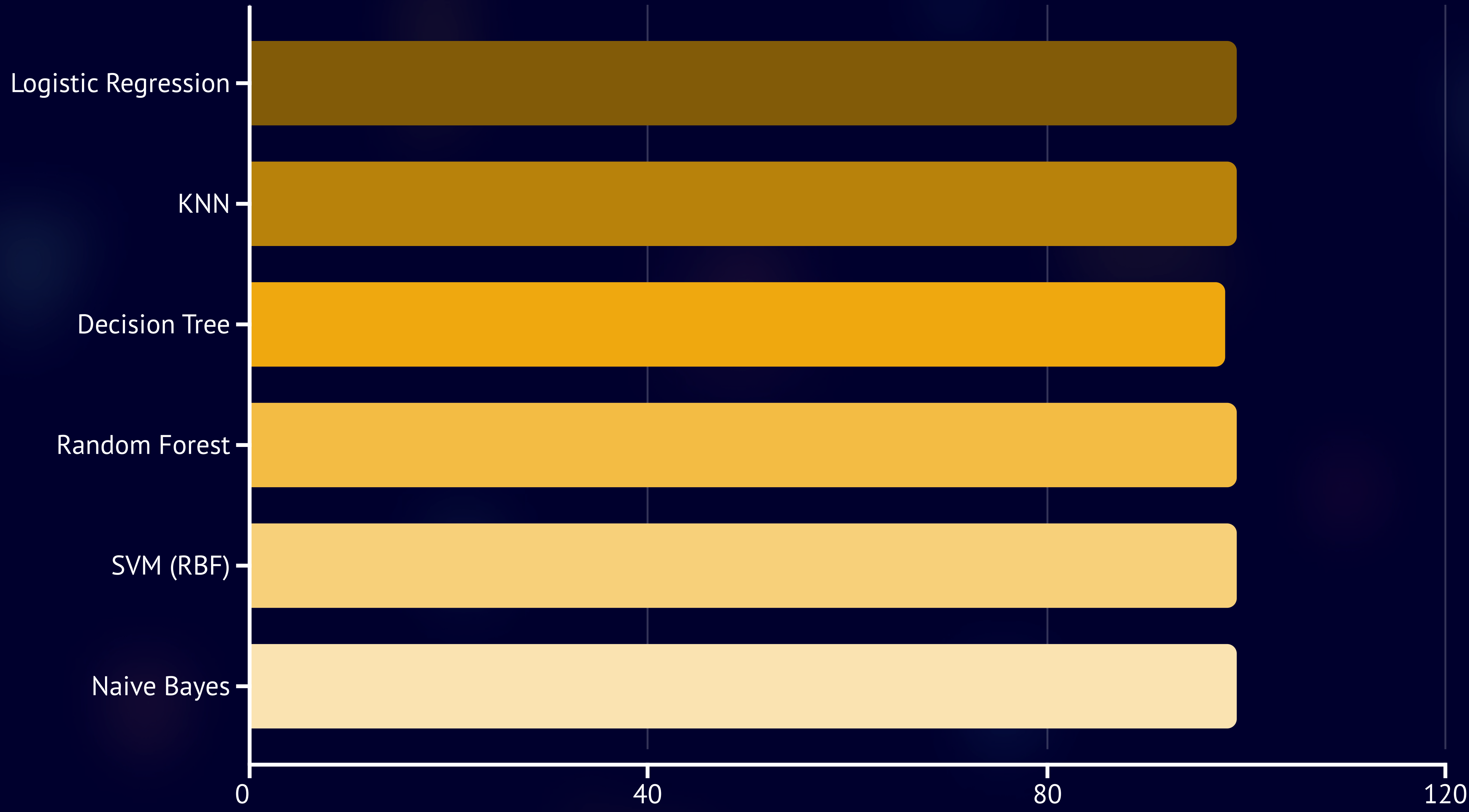
SVM (RBF)

Non-linear kernel classification

Naive Bayes

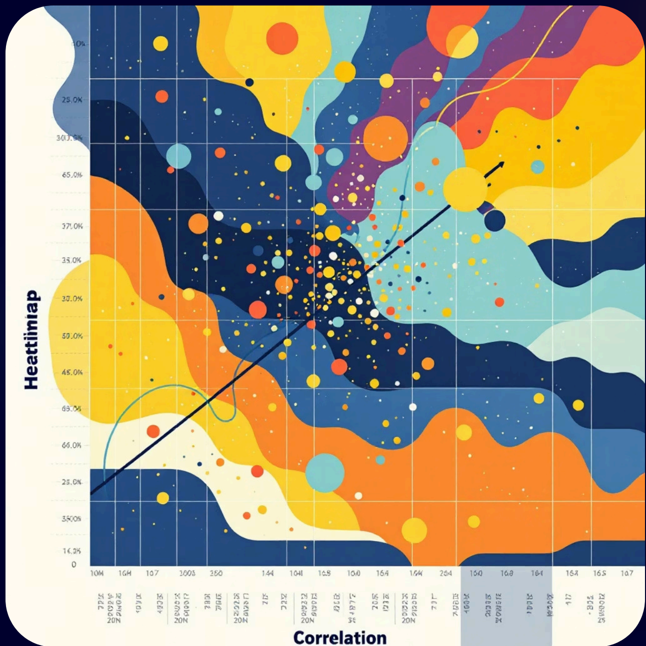
Probabilistic classifier

Model Performance Results



Five models achieved **99% accuracy**, demonstrating strong predictive capability. Decision Tree showed slightly lower performance at 97.86%.

Key Insights & Visualizations

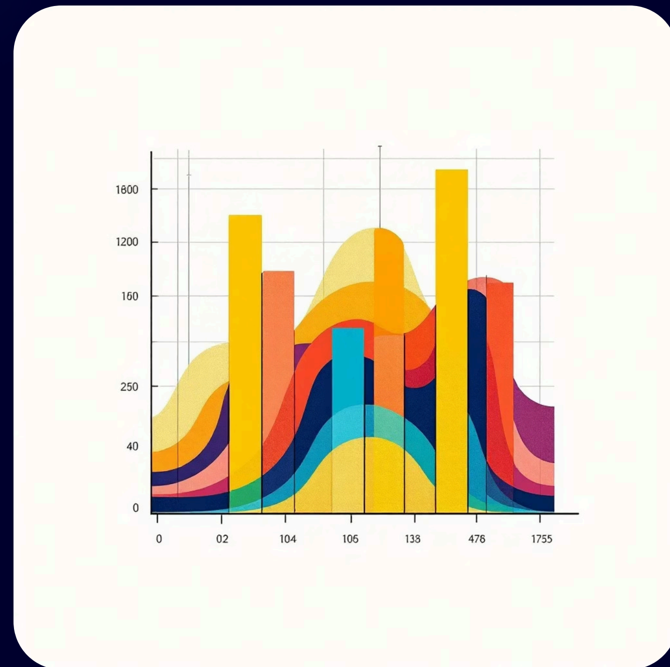


Feature Correlations

Heatmap reveals relationships between customer attributes

Transaction Patterns

Amount vs. MerchantID shows fraud distribution



Distribution Analysis

Box plots compare features across churn classes

Challenge Addressed: The dataset exhibited class imbalance with fewer churned customers. Stratified sampling and evaluation metrics beyond accuracy helped ensure robust model performance.

Future Enhancements

Deep Learning Integration

Implement LSTM for sequence analysis and autoencoders for anomaly detection

Real-Time Prediction

Deploy streaming analytics using Apache Kafka and Spark for live fraud detection

Web/Mobile Deployment

Build API with Flask/FastAPI and deploy to AWS/GCP/Azure for bank integration

Explainable AI

Use SHAP or LIME to provide transparent model explanations for auditors

Conclusion & Learning Outcomes

Project Success

This project successfully demonstrates machine learning's effectiveness in predicting customer churn. After preprocessing and handling class imbalance, trained models achieved 99% accuracy with strong recall.

Key Learnings

- Real-world data handling and preprocessing techniques
- Model evaluation using comprehensive metrics
- Practical Python and ML library implementation
- Solving business problems with data science

Impact: Machine learning provides reliable, automated fraud detection—helping financial systems reduce risks and improve security.