

Customer Churn Prediction Project Summary

1. Business Situation

A major bank is facing a high churn rate in its credit card customer base. The management is looking for a data-driven solution to proactively identify customers likely to churn, so they can take preventive actions and retain them.

2. Key Problems and Objective

The key problem is the increasing number of customers leaving the credit card services. The objective of this project is to build a predictive model that can accurately classify whether a customer is likely to churn or not, enabling targeted retention strategies.

3. Dataset Overview

- Source: Kaggle (Customer Transaction Behavior - Synthetic)
- Records: 10,127 customers
- Features: 21 features including demographics, credit usage, and customer activity
- Target Variable: Attrition_Flag (Existing vs. Attrited Customer)
- Churn Rate: 16.07% (Imbalanced dataset)

4. Tools and Techniques Used

- Programming: Python
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, shap, flask, joblib
- Visualization: seaborn, matplotlib, SHAP
- Model Deployment: Flask Web API with HTML forms
- Model Persistence: joblib
- Environment: Jupyter Lab, Anaconda

Customer Churn Prediction Project Summary

5. Data Preprocessing

- Dropped irrelevant column: CLIENTNUM
- Handled ordinal features with category ordering (Education_Level, Income_Category)
- Applied one-hot encoding to categorical features
- Performed feature selection based on correlation with target
- Standardized numerical features using StandardScaler

6. Exploratory Data Analysis

- Distribution analysis of churn vs. non-churn
- Count plots of categorical variables against churn
- Box plots of numerical variables against churn
- Correlation heatmaps for feature relationships
- Target-wise percentage breakdown per category for insights

7. Model Building

Trained and evaluated four classification models:

- Logistic Regression
- Naive Bayes Classifier
- Random Forest Classifier
- XGBoost Classifier (final model)

All models trained on a stratified 80/20 train-test split.

8. Evaluation Metrics

- Accuracy

Customer Churn Prediction Project Summary

- Precision, Recall, F1-Score
- ROC-AUC Score
- XGBoost outperformed all models with the best F1 and ROC-AUC scores
- Feature importance and SHAP values used for interpretability

9. Key Takeaways

- Churn is rare (16%) (imbalance techniques may be needed)
- Transaction-related features (Total_Trans_Ct, Amt_Change, Utilization) are key churn indicators
- XGBoost provides the best balance of accuracy and interpretability
- Flask app allows easy real-time predictions via user input

10. Resources

- Dataset: <https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers>
- XGBoost: <https://xgboost.readthedocs.io>
- SHAP Interpretability: <https://shap.readthedocs.io>
- Flask Documentation: <https://flask.palletsprojects.com/>