

# Project Report: Bank Customer Churn

## Prediction and Retention Offer Generator

### Objective

The goal of this project is to predict whether a customer will churn (leave the bank) based on their profile and transactional behavior. The project also provides personalized retention offers to help the bank retain its customers.

### Dataset Overview

The primary dataset used is 'Churn\_Modelling.csv,' containing 10,000 rows and 14 columns. Key columns include:

- RowNumber: Index of the row
- CustomerId: Unique identifier for each customer
- Surname: Customer's last name
- CreditScore: Customer's credit score
- Geography: Customer's country of residence (France, Spain, Germany)
- Gender: Customer's gender
- Age: Customer's age
- Tenure: Years of banking relationship
- Balance: Account balance
- NumOfProducts: Number of bank products used
- HasCrCard: Ownership of a credit card (1 = Yes, 0 = No)
- IsActiveMember: Active membership status (1 = Yes, 0 = No)
- EstimatedSalary: Customer's estimated salary
- Exited: Whether the customer churned (1 = Yes, 0 = No)

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	
	0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1
	1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
	2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
	3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0
	4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

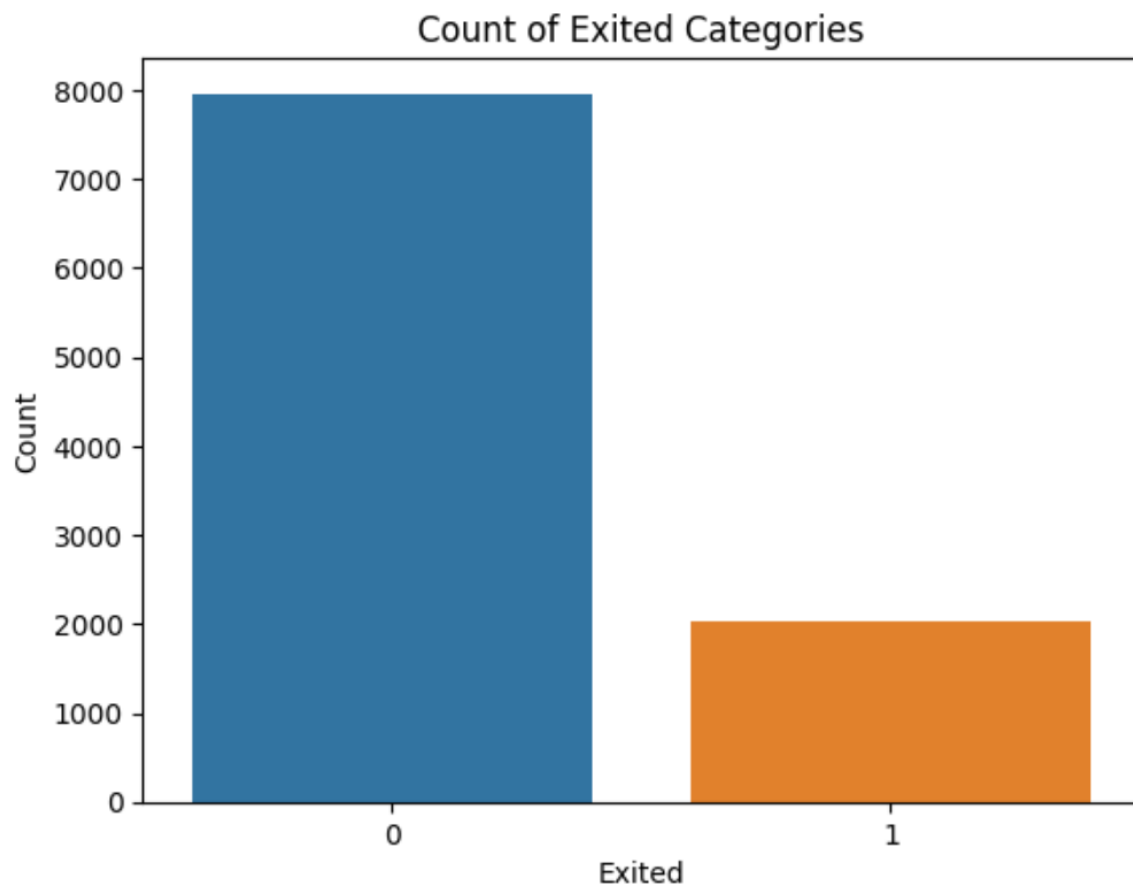
### Role of Data Mining

Data mining played a crucial role in this project by enabling the extraction of meaningful patterns from raw customer data. Core tasks included:

- Data Exploration: Analyzing customer feature distributions and relationships to churn.
- Data Preprocessing: Handling missing values, encoding categorical variables, and balancing the dataset using SMOTE.
- Feature Selection: Identifying key features influencing churn.
- Pattern Identification: Detecting behavioral patterns among customers.
- Predictive Modeling: Building and fine-tuning models to predict churn.
- Knowledge Discovery: Deriving actionable strategies to reduce churn rates.

## Data Preprocessing

The dataset was imbalanced, with a higher proportion of non-churned customers. To address this:



- Balancing: SMOTE (Synthetic Minority Over-sampling Technique) was applied.

Exited		→		Exited	
0	7963			1	6370
1	2037			0	6370

- Feature Scaling: StandardScaler was used to normalize features for model compatibility.

## Models Used

Several classification models were trained, including:

- Random Forest
- Decision Tree
- Support Vector Machine (SVM)
- Logistic Regression

Model	Accuracy	F1 Score
Logistic Regression	0.71	0.50
Decision Tree	0.75	0.45
SVM	0.78	0.57
Random Forest	0.84	0.59

Best Model: Random Forest

## Why Random Forest?

- Ensemble Learning: Combines multiple decision trees for robustness and reduced overfitting.
- Handles Complexity: Effectively manages high-dimensional and imbalanced data.
- Feature Importance: Identifies the most influential factors in churn prediction.
- Non-linear Relationships: Captures intricate relationships within the data.
- Hyperparameter Optimization: Fine-tuned using Random Search for optimal performance.

## Streamlit App

A user-friendly Streamlit web app was developed, offering:

- Churn Prediction: Predicts customer churn likelihood and probability.
- Retention Offers: Personalized strategies to retain customers based on key attributes.

## Retention Offer Examples:

- Youth Advantage Savings Account: For young customers.
- Family Benefit Program: For customers aged 30-50.
- Platinum Membership: For high-balance customers.
- Cross-sell Offers: For customers with fewer products.
- Credit Improvement Assistance: For customers with low credit scores.
- Loyalty Rewards: For active members or customers with high credit scores.

## Power BI Dashboard

A Power BI dashboard ('Churn\_Analysis.pbix') was created to analyze customer churn visually. Key insights include:

- Relationships between features (e.g., age, balance, and credit score) and churn rates.
- Patterns in customer behavior, aiding strategic decision-making.

## Conclusion

This project integrates machine learning and business analytics to address customer churn.

- Predictive Model: The Random Forest classifier ensures accurate predictions.
- Retention Offers: Personalized strategies enhance customer satisfaction and retention.
- Business Value: The Power BI dashboard and Streamlit app empower decision-makers with actionable insights.

By combining data mining techniques, predictive analytics, and targeted offers, the project provides a robust solution to reduce churn and improve customer retention for banks.