Predictive Insights for Banking Customer Churn

Contents

Introduction

- What is customer churn?
- Why studying customer churn in bank is important?
- What are our aims?
  - Aim is to predict churning customers in bank.
- How will we solve banking customer churn problem?
  - Using ML techniques and Tableau Visualizations
- How will our predictions help?
  - It will help marketing team to target customers and retain them.

Problem Statement

Customer churn is a significant concern. It affects:

- Revenue
- Cost of Acquisition
- Reputation
- Market Standing

Methodology

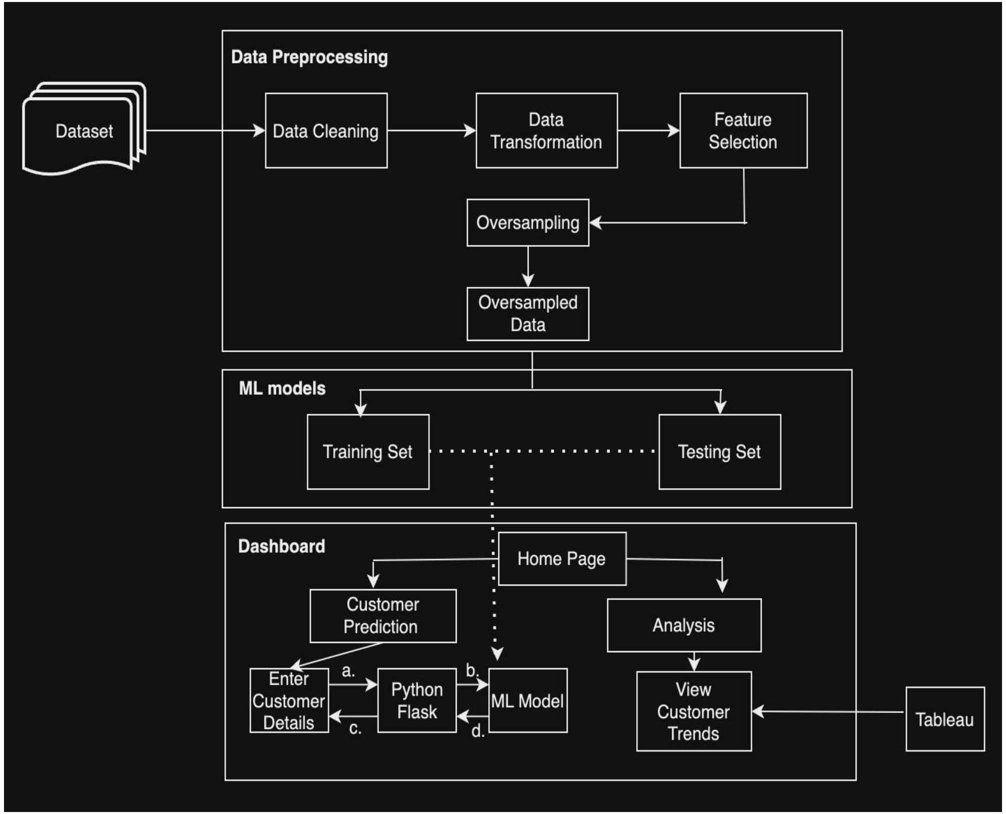


Fig 1: Methodology

Dataset: Churn Dataset from Kaggle.[1] Consists of 10,000 records & 13 features. 'Exit' is the target variable. Dataset loaded using 'pandas'.

Dataset 1

Feature	Description
CustomerId	Unique identifier for the customer holding the account
Surname	Customer Name
CreditScore	Number between 300-850 that depicts a consumer's creditworthiness
Geography	Customer's geographical location
Gender	Customer's gender (Male, Female)
Age	Customer's Age in Years
Tenure	Period of relationship with bank
Balance	The customer account balance
NumOfProducts	Total no. of products held by the customer
HasCrCard	0 or 1 indicates wither a customer has credit card or not (1: Yes, 0: No)
IsActiveMember	0 or 1 indicates wither a customer account is active or not (1: Yes, 0: No)
EstimatedSalary	The customer Estimated salary
Exited	The churn prediction output (0: Will not exit, 1: Will exit)

Fig 2: Dataset description [2]

Methodology

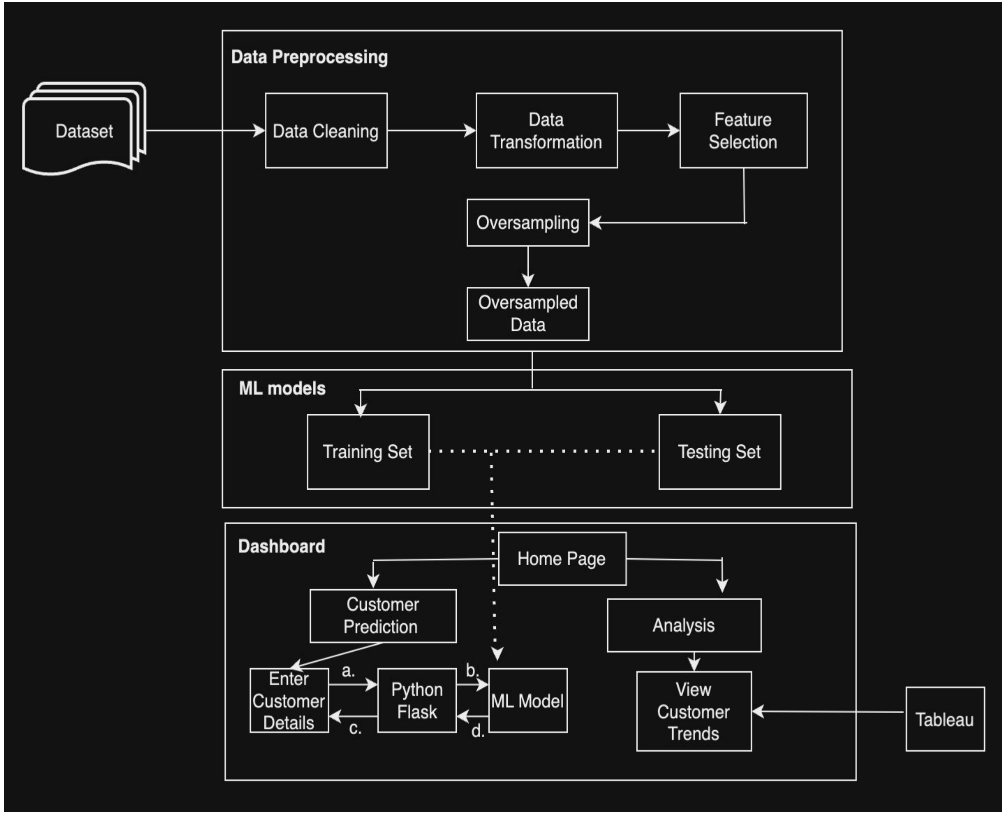


Fig 1: Methodology

Data Preprocessing:

Data Cleaning	Irrelevant features removed: (RowNumber, CustomerId, and Surname)
Data Transformation	Performing Data Encoding (Gender, Geography) and Feature Scaling using MinMaxScalar
Feature Engineering	Added balanceSalaryRatio, TenureGivenAge, CreditScoreGivenAge for EDA, removed in feature selection
Feature Selection	Important features selected
Oversampling	Why need of oversampling? Data balanced using SMOTE.

Methodology - Need of Oversampling

- Because the data was imbalanced:
- We are interested in data of ‘Exited’ rather than ‘Retained’
- Exited data only 20.4%
- We need to balance data
- Used SMOTE technique to balance it

Proportion of customer churned and retained

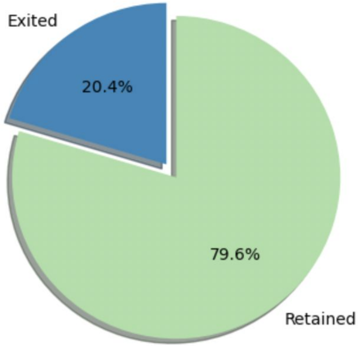


Fig: Before Data Balancing

6.Handling the problem of Imbalanced dataset

```
y.value_counts()
```

```
0    7963
1    2037
Name: Exited, dtype: int64
```

```
from imblearn.over_sampling import SMOTE

s = SMOTE(sampling_strategy='minority')
X_smote, y_smote = s.fit_resample(X, y)

y_smote.value_counts()
```

```
1    7963
0    7963
Name: Exited, dtype: int64
```

Methodology

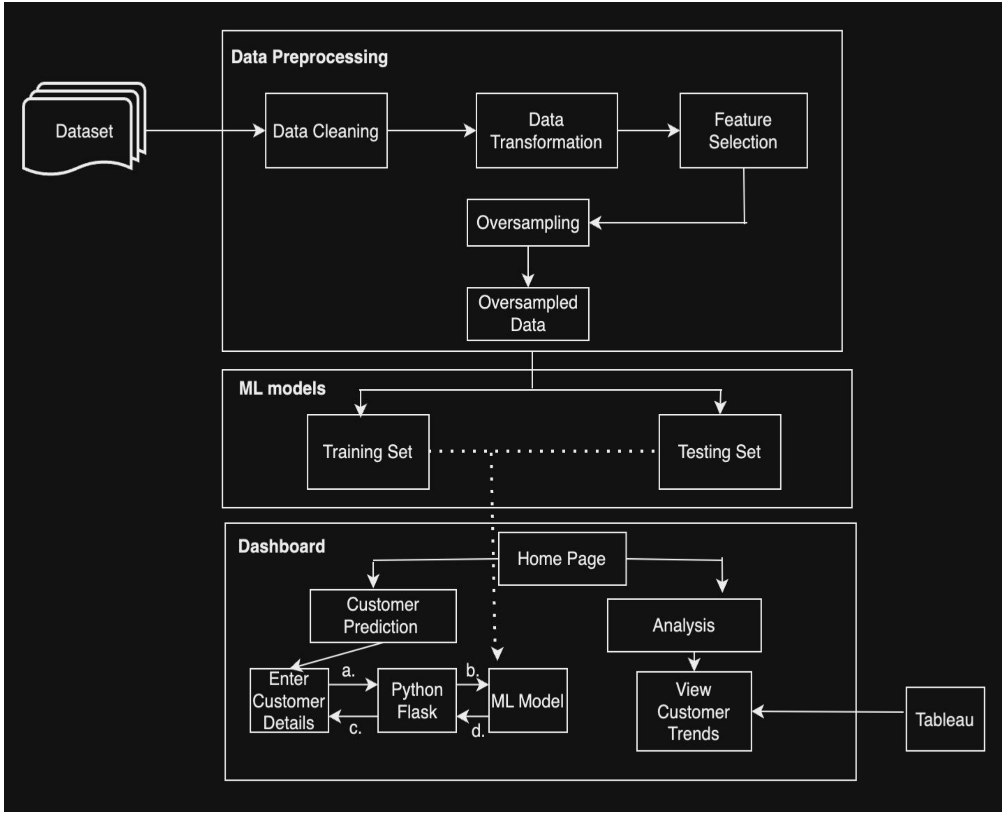


Fig 1: Methodology

ML Models:

- Data split: 70% for training, 30% for testing.
- 8 Machine Learning Algorithms used:
  - Logistic Regression
  - Logistic Regression with 2nd degree polynomial
  - SVM with Rbf and poly kernels
  - Random Forest Classifier
  - Extreme Gradient Boosting Classifier
  - Artificial Neural Network
- 8 models created and compared.
- Best model chosen for dashboard: XGBoost.
- XGBoost model exported as pickle file for dashboard use.

Methodology

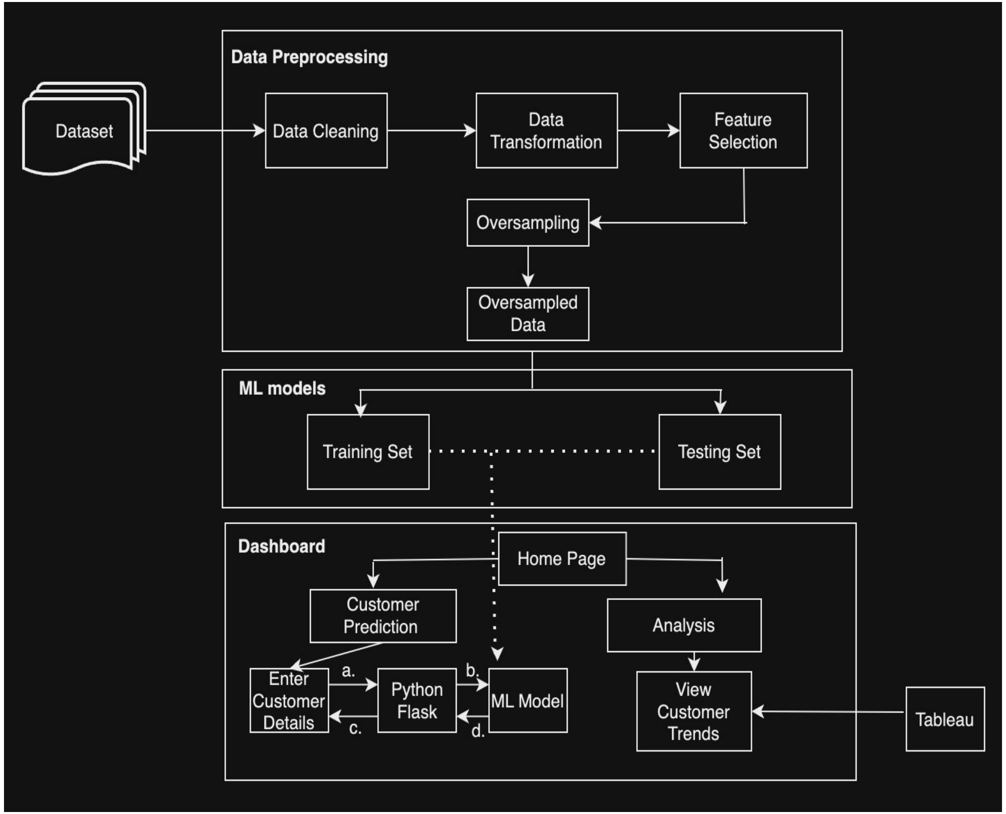


Fig 1: Methodology

Dashboard:

- Offers a user-friendly interface for interacting with the XGBoost model and viewing predictions.
- User Interaction: Users input customer details via the dashboard.
- Request Process: Dashboard triggers a request to Flask server's API for prediction.
- Flask Server: Processes the request, loads XGBoost model from pickle file, applies it to input data.
- Prediction Result: Flask server sends prediction outcome back to dashboard for display.
- Additional Insights: Dashboard provides customer trends via Tableau integration on the website.

Demo

Results

- Compared ML in terms of metrics before and after balancing data
- Improved Performance after balancing: Precision, recall, f-measure
- Accuracy declined
- XGBoost best model

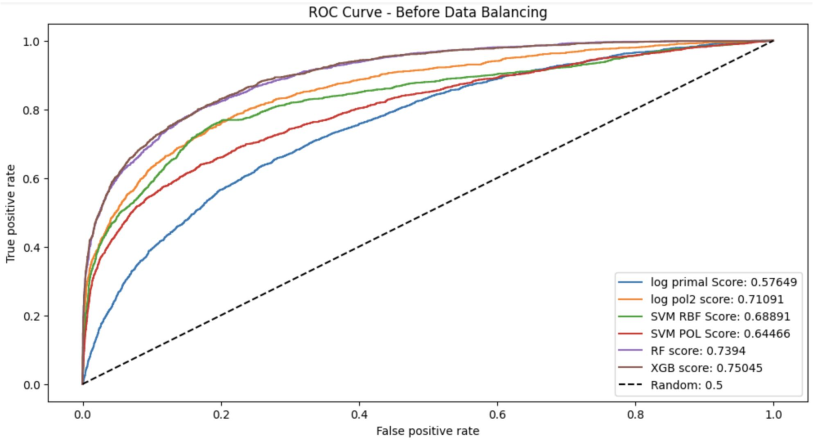
TABLE II : ML Results for test dataset

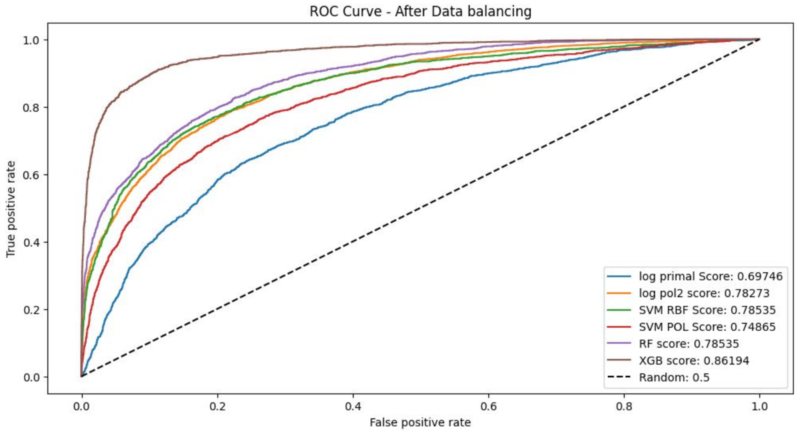
Algorithm	Values Before Balancing Data				Values After Balancing Data			
	Accuracy	*Precision*	*Recall*	*F-Measure*	*Accuracy*	*Precision*	*Recall*	*F-Measure*
Logistic Regression	0.81	0.58	0.18	0.27	0.7	0.7	0.69	0.69
Logistic Regression with 2nd Degree Polynomial	0.85	0.72	0.43	0.54	0.78	0.78	0.79	0.78
SVM with Rbf	0.85	0.76	0.39	0.51	0.80	0.80	0.80	0.80
SVM with Poly Kernels	0.85	0.82	0.31	0.45	0.76	0.77	0.74	0.75
Random Forest Classifier	0.86	0.73	0.44	0.55	0.86	0.88	0.84	0.86
XGBoost	0.86	0.71	0.47	0.57	**0.91**	**0.93**	**0.89**	**0.91**
Artificial Neural Network	0.84	0.70	0.30	0.44	0.78	0.81	0.75	0.78
Decision Tree	0.84	0.64	0.45	0.53	0.74	0.77	0.68	0.73

Results

ROC Curve - Before & After Data Balancing:

- What is ROC Curve?
- After data balancing: Improved Scores
- XGBost: highest curve

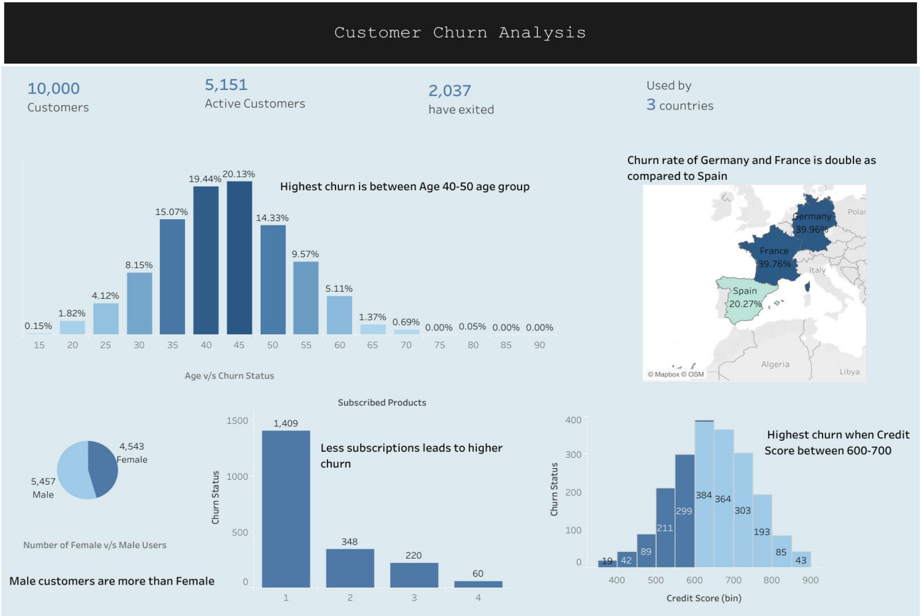


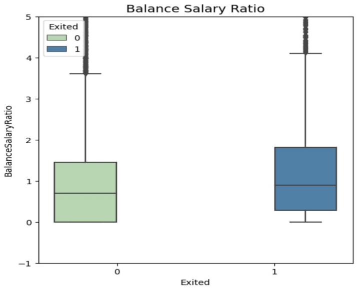


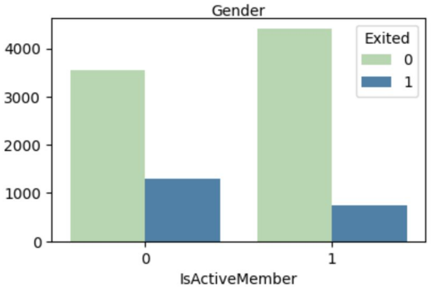
Discussion

Attention should be paid to:

- Age 40-50 group
- Credit Score between 600-700
- Less number of subscribed products
- Females
- Germany customers followed by France
- Inactive customers
- High Balance Salary Ratio







Future Work

- Future Work - Algorithm Expansion
  - Include advanced algorithms like Recurrent Neural Networks, LSTM, and Deep Learning Models
- Future Work - Dashboard Enhancement:
  - Enhance the dashboard's flexibility and adaptability.
  - Make it suitable for diverse data and business scenarios.

Conclusion

- Effective Solution:
  - Successfully addressed banking's customer churn using XGBoost and Tableau.
- Impressive Results:
  - XGBoost achieved 96% accuracy on training and 91% on testing, with the highest ROC curve.
- Dashboard Utilization:
  - Chose XGBoost for dashboard, combining predictive power with Tableau insights.
- Future Enhancements:
  - Advanced algorithm, dashboard enhancement, cloud
- Long-term Perspective:
  - Future focus on holistic customer retention strategies to ensure banking industry success and stability.

References

Adam (2018) *Predicting churn for Bank customers*, *Kaggle*. Available at: https://www.kaggle.com/datasets/adammaus/predicting-churn-for-bank-customers (Accessed: 23 August 2023).

Dalbah, L.M., Ali, S. and Al-Naymat, G. (2022) ‘An interactive dashboard for predicting bank customer attrition’, *2022 International Conference on Emerging Trends in Computing and Engineering Applications (ETCEA)* [Preprint]. doi:10.1109/etcea57049.2022.10009818.

Thank you!!