

# Predictive Modeling for Bank Customer Churn

## 1. Introduction

- **Objective:** Develop and evaluate predictive models to identify customers likely to churn. The aim is to provide actionable insights that can help in reducing churn rates.
- **Dataset Description:** The dataset includes customer information with features such as age, balance, tenure, and churn status. It consists of 10,000 records with 12 columns.
- **Feature Engineering:**
  - **Categorical Variables:** Gender, Geographical Region.
  - **Numerical Variables:** Customer id, Age, Balance, Tenure, Credit Score, Products Number, Credit Card, Active Member, Estimated Salary, Churn
  - **Target Variable:** Churn.
- **Data Preprocessing:**
  - **Encoding Categorical Variables:** Used one-hot encoding for Gender and Geographical Region.
  - **Normalization:** Scaled numerical features using Min-Max normalization.
  - **Train-Test Split:** Split the dataset into 80% training and 20% testing subsets.

## 3. Model Selection

- **Models Evaluated:**
  - **Random Forest:** Ensemble method with improved accuracy.

## 4. Model Training

- **Random Forest:**
  - **Parameters:** Default settings.
  - **Accuracy:** 86.35%.

## 5. Model Evaluation

- **Confusion Matrix:**

- **Random Forest:**

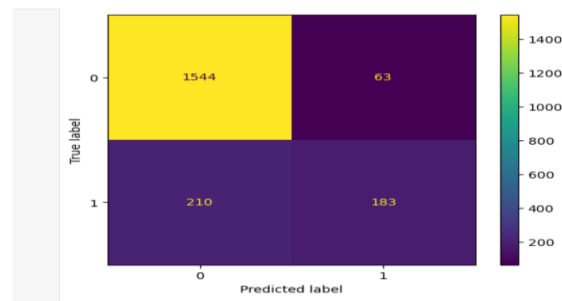
- TP: 1,544
    - FP: 63
    - TN:183
    - FN:210

- **ROC Curve and AUC:**

- **Random Forest:** AUC = 0.71

- **Precision, Recall, and F1-Score:**

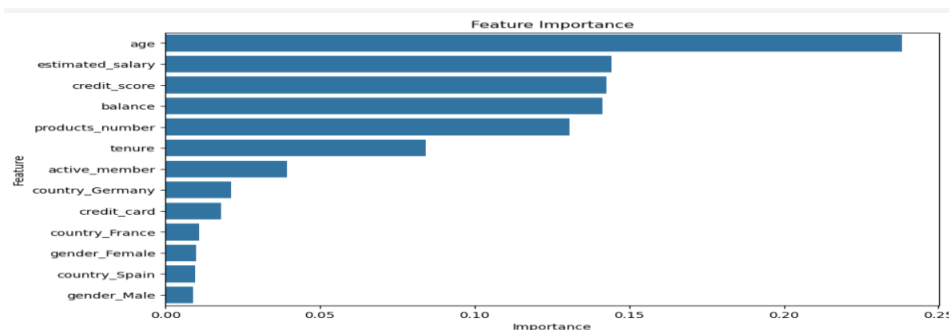
- **Random Forest:** Precision =0.75 Recall = 0.48, F1-Score = 0.58



## 6. Insights and Recommendations

- **Feature Importance:**

- The most important feature is Age, indicating that age significantly impacts churn prediction.
  - Estimated Salary is also significant, suggesting that financial status plays a crucial role.
  - Credit Score and Balance are important but less so than age and salary.
  - Gender is the least important feature, implying that gender has minimal impact on the model's predictions.



## 7. Conclusion

- **Summary:** The Random Forest model offers the best performance in predicting customer churn, with high accuracy and AUC. Implementing this model can help the bank in proactively managing customer relationships and reducing churn rates.
- **Future Work:** Explore additional features, such as transaction history, and experiment with other advanced modeling techniques.