Churn Problem for Bank Customer

Machine learning models that predict if the customers will leave the bank or no

Introduction:

The competition in the banking industry is fast increasing, So the banks are required to implement customer retention strategies and try to increase their market share by acquiring new customers at the same time to stay in this industry .

Improving customers retention rate by up to 5 % will increase your bank's profit by up to 85 %. Also, attracting new customers costs more than retaining the old customers who are more likely to produce more profit. Thus, banks may maintain their competitive advantage by using machine learning models to predict customer churn.

Data:

It consists of 10000 observations and 14 variables.

Independent variables contain information about customers.

The dependent variable refers to customer abandonment.

Features:

- RowNumber: Row Number
- CustomerId: Customer ID
- Surname: Surname
- CreditScore: Credit score
- Geography: Country (Germany / France / Spain)
- Gender: Gender (Female / Male)
- Age: Age
- Tenure: How many years of customer
- Balance: Balance
- NumOfProducts: Bank product used
- HasCrCard: Credit card status (0 = No, 1 = Yes)
- IsActiveMember: Active membership status (0 = No, 1 = Yes)
- EstimatedSalary: Estimated salary
- Exited: Abandoned or not? (0 = No, 1 = Yes)
- The link for data https://www.kaggle.com/mathchi/churn-for-bank-customers/code

Tools:

Data processing: Pandas and NumPy.

Modeling: Scikit-learn.

Visualization: Matplotlib and Seaborn.

MVP:

The goal of this project is to better understand which factors are most important in customer abandonment, and how those factors relate to customer abandonment.

To achieve this goal, The data will be loaded and cleaned. Different models will be built, like logistic regression, KNN, SVM, and random forest. Then these models will be measured and evaluated using different matrices. In the end, any needed improvement will be done.