

Task 3: Customer Churn Prediction (Bank Customers)

Objective

The objective of this task is to identify bank customers who are likely to leave the bank (customer churn).

Predicting churn helps banks take preventive measures to improve customer retention.

Dataset

The Churn Modelling Dataset contains customer demographic and financial information such as credit score,

geography, gender, age, tenure, balance, number of products, credit card status, active membership, and salary.

The target variable is **Exited**, where 1 indicates churn and 0 indicates retention.

Data Cleaning and Preparation

The dataset was loaded using Pandas. Missing values were checked and none were found.

Irrelevant columns such as RowNumber, CustomerId, and Surname were removed because they do not contribute to churn prediction.

The target variable (Exited) was separated from the feature set.

Encoding Categorical Variables

Categorical features were converted into numerical form for model compatibility.

Gender was encoded using Label Encoding.

Geography was encoded using One-Hot Encoding to avoid ordinal relationships and dummy variable trap.

Feature Scaling

Numerical features were scaled using StandardScaler to ensure that all features contribute equally during model training.

This prevents features with large values from dominating the learning process.

Model Training

The dataset was split into training (80%) and testing (20%) sets.

A Random Forest Classifier was trained using the training data to learn patterns associated with customer churn.

Model Evaluation

The model performance was evaluated using accuracy score, confusion matrix, and classification report.

These metrics helped measure the effectiveness of the churn prediction model.

Feature Importance Analysis

Feature importance was extracted from the Random Forest model.

Key influential features included Age, Balance, Number of Products, Geography, and Active Membership.

This analysis helps understand the main factors influencing customer churn.

Conclusion

The customer churn prediction model was successfully developed using machine learning techniques.

The project demonstrates effective data preprocessing, classification modeling, and feature importance analysis.

The results provide valuable insights for banking decision-making and customer retention strategies.