**📝 Task 3: Customer Churn Prediction – Final Report**

**📌 Objective**

Build a machine learning model to predict if a customer will **churn** (exit) a banking service using historical data.

**📁 Dataset**

* **Source**: [Kaggle - Bank Customer Churn Dataset](https://www.kaggle.com/datasets/shantanudhakadd/bank-customer-churn-prediction)
* **Size**: 10,000 customers
* **Features**: Age, Geography, Gender, Credit Score, Balance, etc.
* **Target**: Exited (0 = stayed, 1 = churned)

**🛠️ Data Preprocessing**

* Dropped irrelevant fields: RowNumber, CustomerId, Surname
* Encoded:
  + Gender (Label Encoding)
  + Geography (One-Hot Encoding)
* Scaled numerical features using **StandardScaler**
* Split into 80% train / 20% test using **stratification**

**🤖 Model Used**

* **Random Forest Classifier** (100 trees, default settings)

**📊 Model Performance**

| **Metric** | **Result** |
| --- | --- |
| Accuracy | 86.15% |
| Precision (churn) | 77% |
| Recall (churn) | 45% |
| F1-score (churn) | 57% |

**✅ Confusion Matrix**

| **Actual \ Predicted** | **Stayed (0)** | **Churned (1)** |
| --- | --- | --- |
| Stayed (0) | 1538 | 55 |
| Churned (1) | 222 | 185 |

**🔎 Insights**

* Model is **very accurate overall**, but struggles with **recall on churners**.
* Could be improved by:
  + Tuning class weights
  + Using balanced datasets (SMOTE)
  + Exploring other models (XGBoost, LightGBM)

**🔗 Repository**

[GitHub Repo Link – <https://github.com/thiruvarul11/CODSOFT.git>]

**✅ Conclusion**

Successfully predicted customer churn using Random Forests and extracted meaningful insights for improving customer retention strategies.