**Banking Churn Analysis Project Report**

**1. Project Overview**

The **Banking Churn Analysis Project** aims to predict customer churn for a bank using machine learning models and visualize the insights through Power BI dashboards. The project involves data cleaning, preprocessing, model building, evaluation, and visualization.

**2. Dataset Description**

* **Source:** [Kaggle - Banking Customer Churn Dataset](https://www.kaggle.com/datasets/mathchi/churn-for-bank-customers)
* **File:** churn.xlsx
* **Size:** 10,000 customer records with 14 features, including:
  + **Demographics:** Geography, Gender, Age
  + **Financial Information:** CreditScore, Balance, EstimatedSalary
  + **Account Activity:** Tenure, NumOfProducts, HasCrCard, IsActiveMember
  + **Target Variable:** Exited (1 = Churned, 0 = Retained)

**3. Tools & Libraries Used**

* **Data Cleaning & Modeling:** Jupyter Notebook (Python)
  + Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn
* **Visualization:** Power BI
* **Model Files Exported:**
  + predictions.csv
  + model\_metrics.csv
  + feature\_importance.csv
  + roc\_curve\_data.csv
  + confusion\_matrix.csv

**4. Data Preprocessing & Modeling**

**4.1. Data Cleaning**

* Removed irrelevant columns: RowNumber, CustomerId, and Surname.
* Checked for missing values and duplicates:
  + **No missing values** or duplicates were found.
* Converted Geography and Gender into categorical variables for efficient memory usage.
* Separated the dataset into:
  + **Features (X):** All columns except Exited
  + **Target (y):** Exited

**4.2. Data Splitting & Preprocessing**

* Split into **training (80%)** and **testing (20%)** sets.
* Applied transformations:
  + **Numerical columns:** StandardScaler
  + **Categorical columns:** OneHotEncoder (with drop='first')

**4.3. Model Selection & Evaluation**

* **Models Used:**
  + Logistic Regression
  + Random Forest Classifier (with hyperparameter tuning)
* **Model Performance (Best Model: Random Forest):**
  + **Accuracy:** 85.5%
  + **Precision:** 71%
  + **Recall:** 44%
  + **F1-Score:** 54%
  + **AUC:** 80%

**5. Visualization & Insights**

**5.1. Model Performance Overview**

* Displayed metrics:
  + **Accuracy:** 76%
  + **Precision:** 43%
  + **Recall:** 73%
  + **F1-Score:** 55%
  + **AUC:** 80%
* **Confusion Matrix:**
  + **True Positives:** 288
  + **True Negatives:** 1232
  + **False Positives:** 375
  + **False Negatives:** 105

**5.2. Churn Demographics**

* **Churn by Geography:**
  + **France** had the highest churn, followed by **Germany** and **Spain**.
* **Churn by Gender:**
  + Higher churn among **female** customers (45.61%).
* **Churn by Age:**
  + Highest churn rates in the **36-45** age group.

**5.3. Financial Insights**

* **Balance Distribution:**
  + **High Balance Customers** have a lower churn rate.
  + **No Balance Customers** have a higher churn rate.
* **Credit Score Distribution:**
  + Customers with **lower credit scores** are more likely to churn.
* **Estimated Salary:**
  + No significant correlation between salary and churn rate.

**5.4. Customer Activity Patterns**

* **Churn by Number of Products:**
  + Customers with **3 or more products** have higher churn rates.
* **Credit Card Ownership:**
  + **No clear correlation** between credit card ownership and churn.
* **Churn by Tenure:**
  + Customers with **0-2 years of tenure** have a higher churn rate.
* **Activity Status vs. Churn Rate:**
  + **Inactive members** have a higher churn rate.

**6. Key Findings & Recommendations**

* **Top Churn Factors:**
  + **Age**, **Number of Products**, and **Balance** are the most influential features.
  + Customers with **more products** tend to churn more frequently.
  + Inactive customers are more prone to churn.
* **Recommendations:**
  + **Targeted retention strategies** for customers with multiple products.
  + **Increase engagement** with inactive customers to reduce churn.
  + **Customer segmentation** by demographics and financial patterns for better retention campaigns.

**7. Conclusion**

The **Banking Churn Analysis Project** successfully identified key churn drivers, developed predictive models, and visualized insights. The **Random Forest Classifier** performed the best, achieving an **85.5% accuracy** with **0.80 AUC**. The Power BI dashboards provide clear and actionable insights, highlighting customer segments with higher churn risk.