📂 Dataset

The dataset used is Churn_Modelling.csv, containing customer details such as:

- Credit Score
- Age
- Balance
- Gender
- Geography
- Excited (whether the customer churned)

🛠️ Project Workflow

1️⃣ Data Loading & Exploration

- 📥 Load the dataset using pandas.
- 🔎 Explore for missing values, inconsistencies, and general data insights.

2️⃣ Data Visualization

Generate visualizations to understand:

- 📈 Distribution of numerical features (Age, Credit Score, Balance, Estimated Salary).
- 🏷️ Distribution of categorical features (Gender, Geography).
- 🔥 Correlation heatmap to identify relationships between features.
- 🎯 Target variable (Exited) distribution to check for class imbalance.

3️⃣ Feature Engineering

- 🔢 Encode categorical variables (Geography, Gender) using One-Hot Encoding or Label Encoding.
- ✅ Ensure feature cleanliness, relevance, and proper formatting for modeling.

4️⃣ Data Splitting

- ✂️ Split data into 80% training and 20% testing sets using `train_test_split` from scikit-learn.

5️⃣ Handling Class Imbalance

- ⚖️ If class imbalance is detected in the Exited column, apply SMOTE (Synthetic Minority Oversampling Technique) to balance the dataset.

6️⃣ Feature Standardization

- 🎚️ Normalize numerical features to ensure they are on the same scale (mean = 0, standard deviation = 1), which is crucial for models like SVM and KNN.

7️⃣ Model Training & Evaluation

Train the following machine learning models:

- 🔍 K-Nearest Neighbors (KNN)
- 📖 Naive Bayes
- 💡 Support Vector Machine (SVM)
- 🌳 Decision Tree (DT)

Evaluate models using the following performance metrics:

- ✅ Accuracy
- 🎯 Precision & Recall
- 📏 F1-Score
- 📊 ROC-AUC Score

⚖️ Compare all models based on performance metrics and identify the best-performing model.

🎯 Deliverables

- 🤖 A trained model capable of predicting customer churn.
- 📊 A comparative analysis of different machine learning models.
- 🖼️ Visual representations of results and insights from the dataset.

📁 Repository Structure

|-- Customer_Churn_Detection/ |-- data/ |-- Churn_Modelling.csv |-- file.ipynb |-- README.md

🚀 How to Run the Project

🔽 Clone the repository:

- git clone https://github.com/ayaatef11/Churn-Modeling.git
- cd Customer_Churn_Detection

🏗️ Run the preprocessing and training scripts:

python src/data_processing.py python src/model_training.py

📊 View results and model evaluation:

python src/evaluation.py

🛠️ Dependencies

🐍 Python 3.x
📊 Pandas
🔢 NumPy
🏗️ Scikit-learn
📈 Matplotlib
🎨 Seaborn
⚖️ Imbalanced-learn (for SMOTE, if needed)
📈 Results and Findings
📊 The project evaluates four models to determine the best one for predicting customer churn.
🏆 The best-performing model is selected based on accuracy, precision, recall, F1-score, and ROC-AUC score.
📉 Visualizations and charts are provided to illustrate key findings.

👥 Contributors

Aya Atef