**Model Choice, Feature Engineering, and Training Process**

**1. Model Choice**

* **Model:** XGBoost Classifier
* **Parameters:**
  + **learning\_rate=0.01:** Controls the step size at each iteration.
  + **n\_estimators=500:** Number of boosting rounds or trees.
  + **max\_depth=3:** Maximum depth of trees.
  + **alpha=1.0:** L1 regularization term on weights.

**2. Feature Engineering**

* **Data Cleaning:**
  + **Handling Missing Values:** Remove rows with null values.
  + **Encoding Categorical Variables:** Use LabelEncoder for Gender and pd.get\_dummies for other categorical variables.
  + **Dropping Redundant Columns:** Remove original categorical columns and CustomerID.
* **Feature Scaling:**
  + **Standard Scaling:** Apply Standard Scaler to numerical features to standardize them.

**3. Training Process**

* **Data Preparation:**
  + Split dataset into training and test sets (20% reserved for testing).
* **Model Training:**
  + Train XGBoost model with specified hyperparameters.
* **Model Evaluation:**
  + **Prediction and Probability Scores:** Evaluate model on test set.
  + **Metrics Calculation:** Compute accuracy, precision, recall, F1 score, and ROC AUC score.
* **Visualization and Reporting:**
  + **Confusion Matrix:** Save heatmap.
  + **ROC Curve:** Plot and save ROC curve.
  + **Histograms and Density Plots:** Generate and save comparison plots.
  + **Feature Importance:** Plot and save feature importance.

**4.API Chatbot**

* **Chat Functionality:**
  + **Endpoint:** /chat
  + **Description:** Handles user inputs for normal chatting and prediction requests.
  + **Normal Chat:** Uses the LLaMA 3 model to respond to user queries.
  + **Prediction Requests:** Detects keywords like "prediction" or "predict" to invoke the prediction model.

**5.Pipeline Diagram**

A diagram of a software system

Description automatically generated with medium confidence