Hackathon Challenge ITADATA 2024: Predicting Customer Creditworthiness

Welcome to the ITADATA 2024 Hackathon! In this challenge, you'll dive into a rich dataset of bank customers spanning five years, with data recorded quarterly. Your mission is to develop a robust model to predict whether a customer will be a "good" or "bad" client based on the "repays_debt" feature.

Challenge Overview

Objective Level 1: Analyze the provided dataset of bank customers and build a predictive model to determine the likelihood of a customer repaying their debt. (First 2 days)

Dataset Description

The dataset contains customer information over a period of five years, with data collected every quarter (20 trimesters in total). Key features include:

- **product8, product10, product13, product12, product11, product4, product17, product2, product3, product1, product7, product6, product5, product14, product15, product16, product9:** Various product features associated with the bank's offerings to the customer.
- **has_products:** Boolean indicating if the customer holds any products with the bank.
- **balance:** Current balance in the customer's account.
- **left_bank:** Boolean indicating if the customer has left the bank.
- **joined_bank:** Boolean indicating if the customer has joined the bank.
- **wire_transfers2_amt_inbound, wire_transfers1_amt_inbound:** Amount of inbound wire transfers (two different types).
- **wire_transfers2_amt_outbound, wire_transfers1_amt_outbound:** Amount of outbound wire transfers (two different types).
- **counter_amt_inbound:** Amount of inbound counter transactions.
- **counter_amt_outbound:** Amount of outbound counter transactions.
- **securities_bought_amt:** Amount of securities bought by the customer.
- **securities_sold_amt:** Amount of securities sold by the customer.
- **wire_transfers2_num_inbound, wire_transfers1_num_inbound:** Number of inbound wire transfers (two different types).
- **wire_transfers2_num_outbound, wire_transfers1_num_outbound:** Number of outbound wire transfers (two different types).
- **counter_num_inbound:** Number of inbound counter transactions.
- **counter_num_outbound:** Number of outbound counter transactions.
- **securities_operations:** Number of securities operations.
- **securities_bought:** Boolean indicating if the customer bought securities.
- **securities_sold:** Boolean indicating if the customer sold securities.
- **counter_amt_tot:** Total amount of counter transactions.
- **counter_num_tot:** Total number of counter transactions.
- **period:** Time period of the record (from 1 to 20).

- **client_id:** Anonymized client ID.
- **Category**: if the client is a firm account, sole proprietorship account or a personal account
- **LABEL: repays_debt:** Boolean feature indicating if the customer repays their debt (1 for yes, 0 for no).

Training Set = clients history for the 20 periods
Test Set = clients history until a certain period, the last one must be predicted using the knowledge until that period (you must replace the '??' with a 0 or a 1).
Expected_output_template = how the result of the prediction must be presented (subsituting the ?? with the predicted values).

Requisiti Tecnologici

- Notebook Python ri-eseguibile sulla macchina virtuale
- Notebook commentato per fasi
- No modelli pre-addestrati

Goals and Tasks

1. **Data Exploration and Preprocessing:**
    - Perform exploratory data analysis (EDA) to understand the structure, distribution, and relationships within the data.
    - Clean the data by handling missing values, outliers, and any inconsistencies.
    - Engineer new features that could enhance the predictive power of your model.
2. **Model Development and Explainability:**
    - Choose appropriate machine learning algorithms to build your predictive model.
    - Train your model using historical data and validate its performance using suitable evaluation metrics.
    - Measure the F1 measure on the test set provided.
    - Analyze the feature importance to understand which features contribute most to predicting whether a customer will repay their debt.
3. **Presentation:**
    - Prepare a compelling presentation that showcases your approach, findings, and the performance of your model.