ML Classification Assignment

Introduction

Yojo.com is one of the main actors in the online shopping industry. One of the key aspects of its success is the continuous monitoring of their client satisfaction by performing satisfaction surveys which measure different aspects of the purchase process.

The objective of the company is to develop a process to predict customer satisfaction based on both the historical opinion of the customer and the characteristics of the purchases.

In this assignment, you will use Machine Learning techniques in order to predict if the clients of the Yojo platform are satisfied after their purchase.

Dataset Description

You will receive two datasets containing a list of purchased with the information related to the client and the purchase itself.
There is a total of 70000 records and 20 explanatory variables divided into two datasets.

• `model.csv`: the dataset contains the information of 50000 client with the respective target variable. You must use this data to create and evaluate your model.

• `predictions.csv`: the dataset contains the information of 20000 clients **without** the target variable. You are requested to provide the predictions for this set of records.

Target Class:

The target attribute is binary: Satisfied, Not satisfied.

The task is formulated as a binary classification. Your grade will be based on the **F1-score** metric and on the modeling process presented in the **report**.

Attribute Information:

n	Variable	Description
1	id	Client ID
2	Gender	Client gender: Male, Female
3	Customer Type	Client type: Premium, Standard
4	Age	Client age
5	Price	Purchased item price in dollars
6	New/Used	Item status: New/Used
7	Category	Item category: Entertainment, Home/Kitchen, Technology, Other
8	Product description accuracy	Level of satisfaction on product description
9	Website navigation	Level of satisfaction on the website navigation experience
10	Manufacturer sustainability	Level of satisfaction on the manufacturing sustainability process
11	Packaging quality	Level of satisfaction on packaging
12	Additional options	Level of satisfaction on extra options
13	Reviews and ratings	Level of satisfaction on reviews and rating information
14	Integrity of packaging	Level of satisfaction on packaging state
15	Check-out procedure	Level of satisfaction on payment procedure
16	Relevance of related products	Level of satisfaction on related product suggestion
17	Costumer insurance	Level of satisfaction on insurance options
18	Shipping delay in days	Delay of shipping in days
19	Arrival delay in days	Arrival delay on days
20	Satisfaction	Target: Satisfied, Not Satisfied

Submission Instructions

1. Model Training Data Release: 05 November 2021, 19:00.

2. Description of analysis on the training set and model identification: 12 November 20:00.
You are kindly asked to submit the following supporting information in the WeBeep page of the course:

a) A **brief report** of the step-by-step methodology (i.e., pre-processing, visualization, training, testing, etc.) that you have followed to develop your model, this document must illustrate the motivation behind your selected approach.

- File Format: .pdf
- Filename: 6-digit student code (e.g., 123456.pdf)

b) **The commented python code** that you used in your model. Comments in the code must ensure that the code is easy to follow.
- File Format: .ipynb, .py
- Filename: 6-digit student code, e.g., 123456.ipynb or 123456.py (note that this is **not** your 8-digit POLIMI personal code)

3. Prediction Data Release: 12 November 21:00.

4. Prediction Submission: 14 November 20:00.
You are kindly requested to strictly follow the described submission guidelines:
- File Format: .csv
- Filename: 6-digit student code (e.g., 123456.csv)
- Column Format: **A single** column named "target"
- Row Format: Your predictions (1- Satisfied, 0 - Not satisfied) with **the same number of rows** and in the same order as the **prediction** test set.

Example:

Target
1
0
1
1
0
0

Further Instructions

- **Verify** the integrity and coding of your uploaded files in platform.
- The assignment can be developed in groups with a maximum number of three participants.
- Nevertheless, **submission is individual**, therefore each student must upload his/her own submission files (even if they are the same for all participants in the group).
- **Verify** the integrity and coding of your uploaded files in platform.

- **Any submission that does not respect the guidelines (submission after deadline, empty file, wrong student code) will not be graded**.