Project for Machine Learning

This document describes the project available for Machine Learning, which will contribute to 50% of your total grade in the class. You can work in groups of 4 people.

Instructions

When to submit your project.
You must submit your project by April 27 2022.

What to submit for your project.
Each group must submit one PDF document describing the project and its working code. It must be possible to easily run the code to test the solution delivered.
The PDF document should be 4-6 pages (excluding figures and references) and should include each of the following:
- Title and Team members
- [Section 1] Introduction – Describe briefly your project.
- [Section 2] Methods – Describe your proposed ideas (e.g., features, algorithm(s), training overview, design choices, etc.) and your environment so that:
  - A reader could understand why you made your design decisions, and the reasons behind any other choice related to the project
  - A reader should be able to recreate your environment (e.g., conda list, conda env export, etc.)
  - It may help to include a figure illustrating your ideas, e.g., a flowchart illustrating the steps in your machine learning system(s)
- [Section 3] Code Description – Describe your code (at a high level) so that:
  - A reader could easily follow up your implementation
- [Section 4] Experimental Design – Describe any experiments you conducted in order to demonstrate/validate the target contribution(s) of your project; indicate the following for each experiment:
  - Main purpose: 1-2 sentence high-level explanation
  - Baseline(s): describe method(s) that you used to compare your work to
  - Evaluation Metrics(s): which ones did you use and why?
- [Section 5] Results – Describe the following:
  - Main finding(s): report your results and what you might conclude from your work
  - Include at least one placeholder figure and/or table for communicating your findings
  - All the figures containing results should be generated from the code
- [Section 6] Conclusions – List some concluding remarks. In particular:
  - Summarize in one paragraph what is the take-away point from your work.
  - Include one paragraph to explain what questions may not be fully answered by your work as well as natural next steps for this direction of future work
- [Section 7] Bibliography (Optional) – Include any bibliographic reference needed

How to submit your project.

For this project, you are required to send your running code and your PDF report by email to gitaliano@luiss.it by the given deadline.

Churn Prediction

Introduction.
StayWithMe (SWM) Bank is very worried about some of its customers leaving their credit card service. They hired you to predict who is going to leave, as this is information could be very valuable to offer better services and conditions to customers that are thinking to leave the bank.

Dataset.
The dataset is available on https://www.dropbox.com/s/7nwimmta836si5f/churn.csv?dl=0.
It consists of 10,127 observations of 17 variables:

- *Basic info*:
    - CLIENTNUM : Unique identifier for the customer holding the account.
- *Target*:
    - Attrition_Flag: Specifies whether the account was closed (Attrited Customer).
- *Demographic Variables*:
    - Customer_Age: Demographic variable - Customer's Age in Years.
    - Gender: Demographic variable - M=Male, F=Female.
    - Dependent_count: Demographic variable - Number of dependents.
    - Education_Level: Demographic variable - Educational Qualification of the account holder (example: high school, college graduate, etc.).
    - Marital_Status: Demographic variable - Married, Single, Divorced, Unknown.
    - Income_Category: Demographic variable - Annual Income Category of the account holder (< 40K, 40K - 60K, 60K–80K, …).
- *Variables (Product)*:
    - Card_Category: Product Variable - Type of Card (Blue, Silver, Gold, Platinum).
    - Months_on_book: Period of relationship with bank.
    - Total_Relationship_Count: Total no. of products held by the customer.
    - Months_Inactive_12_mon: No. of Months in the last 12 months.
    - Contacts_Count_12_mon: No. of Contacts in the last 12 months.
    - Credit_Limit: Credit Limit on the Credit Card.
    - Total_Trans_Amt: Total Transaction Amount (Last 12 months).
    - Total_Trans_Ct: Total Transaction Count (Last 12 months).
    - Avg_Utilization_Ratio: Average Card Utilization Ratio.

Specific Tasks.
- Analyze the data (data visualization is appreciated).
- Generate a training set and a test set.
- Create different models to perform predictions (the dependent variable is given by the flag specifying if the user has closed the account).
- Compare the models and evaluate them properly.
- Discuss your results.