Introduction to Machine Learning
T1508M/T1608M
Project 2024

Project Submission Deadline: 27th November
Project Presentation: 11th December
(Each group will be given 15min and each student will be given 5 min to present)

Instructions:

The project should be carried out in group of 3 students.

In the Jupyter Notebook, in the markdown give following details of student in begining:

Group Name :

Student # 1:
Full Name:
Student Number/ID:
Student # 2:
Full Name:
Student Number/ID:
Student # 3:
Full Name:
Student Number/ID:

The code, report and presentation of the project will be done in Jupyter Notebook. Use the code part for coding. Use markdown for writing description related to various steps you are required to follow in the project. Also give your report findings and presentation in same notebook .

Project Goals and Objective:

You are working as a data analyst. You have to work on a real-world dataset. The goal is to apply different machine learning algorithms and to explore new techniques even though not covered in this course.

Total Marks: 30% / 30pts

Bonus points: 3pts

Step 1: Dataset Selection

Select a dataset that is given in the project. They are as follows:

Dataset A: Churn Data

Dataset B: data200

Dataset C: Weather_Data

Dataset D: House_Data

OR

Select a dataset that you are really passionate about. This can be a dataset you have always wanted to analyze. Your chosen dataset should be appropriate dataset to be worked on, as per requirement of project.

[Bonus Marks:1 pt]

Step 2: Scenario/About Dataset:

You have to define the problem statement or state the situation because you want to make sure that the machine learning solution you are providing is aligned with your client's needs.

Identify weather the project tackles a problem of regression, clustering or classification?

[Marks: 3pts]

Step 3: Data Loading:

Brief description of the data set and a summary of its attributes

[Marks: 3pts]

Step 4: Data Wrangling or Data Pre-processing

Perform data preprocessing tasks, including handling missing values, scaling features, encoding categorical variables.

Actions taken for data cleaning and feature engineering.

[Marks: 5pts]

Step 5: Exploratory Data Analysis:

Apply techniques of data exploration to understand the data set, identify relevant features

Key Findings and Insights, which synthesizes the results of Exploratory Data Analysis in an insightful and actionable manner. Visualization Charts

[Marks: 5pts]

Step 6: Model Development:

Apply multiple machine learning algorithms

[Marks: 5 pts]

If you choose to use two ML algorithms (e.g. Decision tree classifier, random forest classifier, logistic regression, and so on) not taught in class you get bonus points.

[Bonus pts: 2]

Step 7: Model Evaluation

Implement and compare different metrics to evaluate the quality of the model. Discuss the results. You may use various plots also.

[Marks: 4 pts]

Step 8: Model Refinement:

Clearly present the findings, insights and recommendations

Compare the the results obtained from the different applied methods and discuss the limitation of the models before presenting areas for future improvement

[Marks:5pts]

Your code, report of results and presentation will be in same Jupyter notbook file.

Once you complete your project, you will have to share it. You can download the notebook by navigating to "File" and clicking on "Download" button. This will save the (.ipynb) file on your computer. Once saved, you can upload this file.