AI and Machine Learning Theory and Practice

Project 1

Deadline For Submission: Oct 31, 2022

Hand-in Instructions: The completed report and code must be submitted online.

Instructions for completing assessment: The project will be assessed by means of a formal report that has a cover page. Each student must submit an individual work.

Instructor: Baha Rababah

Instructions:

- This project contributes 25% to your overall unit mark.
- Each student must submit an individual work.
- The completed report must be submitted online by Oct 31, 2022.
- Late submissions will attract the standard academic penalty.
- The assignment also has scheduled formative assessments.
- The work will be checked for possible plagiarism.

Introduction:

In this project you will work on two tasks (task 1 and task 2) from "Diagnostic Questions: The NeurIPS 2020 Education Challenge" competition, that aimed to accurately predict students' answers to assessments, determine question quality, and identify a personalized sequence of questions for each student that best predicts the student's answers.

In order to get the data and the tasks instructions, Students are required to use the data and instructions from the following competition (task 1 and 2):
NeurIPS Education Challenge (eedi.com)

Details about the competition: https://arxiv.org/pdf/2007.12061.pdf

Project Objectives:

1- exploring the data.

2- applying visualization techniques.

3- applying data cleaning techniques.

4- applying preprocessing techniques.

5- building classification/regression models.

Projects Phases:

Phase 1:

In this phase, you will write two sections for your report:

Introduction (1/2 page)

Background (1/2 page)

- Machine learning techniques

- Education with predicting students performances

Phase 2:

There are many submitted codes for the competition, **do not copy and paste**. You should include the following in your report:

- Explore the data and fill the empty fields.
- Use visualizations to explore the data.
- Drop non-useful columns/features.
- Rank the features with any a technique of your choice.
- Use Scikit-learn to perform the following:
    - SVM (3 different predications using different kernels and hyperparameters)
    - Naive Bayes classifier
    - Logistic regression
    - Decision tree (3 different trees with different hyperparameters)
    - Random Forest (use Grid search to find best parameters:)
        - ✓ For the criterion (Gini or entropy)
        - ✓ Max feature
        - ✓ Max depth
        - ✓ Number of trees
    - Gradient Boosting Classifier (GridSearchCV for Hyperparameter Tuning)
    - Bagging Classifier (one with Naïve bayes and one with decision trees)
    - AdaBoosting (GridSearchCV for Hyperparameter Tuning)
    - Add your own method for predication that you believe it outperform the above predications
    - Use Ensembling method to improve the results
- Compare the above classifiers methods according to Accuracy, Precision, Recall, F1-Score, and ROC.
- Plot the comparisons!
- Complete the report (submit your code with the report)
- Setup an appointment with me to discuss your work!

Report Guidelines:

a. Introduction with background (1 pages)
c. Data Description (1 pages)
d. Describe the architecture of the successful model (1 pages).
e. Discuss the results and the evaluation and all the fine-tuning to get the best model (2-3 pages)
f. Conclusion (1/2 page)
g. References

Submissions:
1- Your code and all the figures & plots
2- A PDF format