

CMPE 407 - Machine Learning

Project

Deadline: 30.05.2023 23.50

RULES

- Codes without documentation (comments) will not be evaluated.
- Report and code are both required. In case of missing one, the project will not be evaluated.
- Plagiarism is strictly prohibited. If it is detected either in codes or reports, involved students will get zero.
- Project can be conducted either individually or in pairs. Note that the pair can be **at most 2** people. There will not be any exceptions.
- The reports which have similarity score **above than 30%** will not be evaluated.

SUBMISSION

- One submission from one of the group members is enough.
- You are asked to prepare a detailed report about your experiments. Upload **PDF version** of your report.
- As well as your report, you have to upload your codes. You will find another submission slot dedicated for code submissions in Learn. Codes without report or vice versa will not be evaluated.
- Late submissions will not be accepted. Submission system will be closed after the deadline. Submissions via e-mail will not be accepted.

REPORT

There is no page or time restriction for reports. However, you are asked to convince that you fully understand the subject. Therefore, the report without sufficient descriptions can be penalized. Recommended page length is 5 without cover and references pages.

Provide at least the below sections in your report. You can add more sections if you need.

- **Introduction**

What was your aim for this project? Which dataset(s) did you find in order to achieve this aim? Which method did you use (classification, regression, clustering etc.)?

- **Dataset**

Explain your dataset in details. Use plots, statistics or any kind of visualization technique. Be sure that you analyzed the dataset comprehensively.

- **Preprocessing**

Explain what you did with your dataset samples to prepare them to feed models (label encoding, feature selection, feature extraction etc.).

- **Experiments**

Use at least 3 different algorithm (k-means, Linear Regression, Decision Tree etc.) for your selected method. Be careful that these algorithms has to be suitable which method you selected. For example, you cannot use k-means algorithm for regression.

Explain the algorithms briefly, and justify why you select this algorithms. Provide best hyper-parameters for each algorithm with respect to your experiment results.

Explain which metrics that you used in your model evaluations. Be sure that you made cross validation for your experiments.

- **Results**

Create a table that includes results of all models. Discuss why the best algorithm provided better results than the others; and worst algorithm got worst scores compared to others. Analyse the errors on the best model in order to enhance the further studies.

- **Conclusion**

Summarize the report. Indicate the best model for your aim with respect to your experiments. Provide what else that can be done in order to improve results.

QUESTION

1. Find an aim and dataset excluding the ones that we made experiment through the semester on.
2. Make experiments, and report your experiments.

Bonus. (20 points) Select a benchmark dataset used in literature and provide information about related works and results in your report.

Note: Ensure that your codes are fully documented, using comments. Codes without report or vice versa will not be evaluated.