Applied Machine Learning

Project Report and Code

Brief)

The aim of this assignment is to create two machine learning models using a dataset.

Using Python and associated ML libraries, implement the machine learning model as well as the appropriate tasks such as cross-validation, feature engineering and hyperparameter tuning.

Write a report that documents the process, describes and visualises the results, and gives a comparative analysis between the two models for the same prediction problem and a baseline predictor as well as conclusions.

Task)

- Select a dataset from any source: E.g.
  - Our Data | FiveThirtyEight
  - Home - UCI Machine Learning Repository
  - Kaggle: Your Home for Data Science
- Identify a prediction problem given this dataset. Either Regression or Classification.
- Train **two** machine learning models for this prediction task.
- Train a baseline predictor for comparison.
- Generate performance metrics and visualisations.

Report)

Introduction:

- Describe the prediction problem and the dataset.

Data Engineering:

- Discuss any data cleaning, transforming or normalising.

Model Selection:

- Discuss the chosen two models and your justification for using them.

- Discuss the baseline model used.

Feature Selection

- How did you choose the variable input for this model?

Cross-Validation:

- Describe the cross-validation method used to test the performance of the models.
- Describe the hyperparameter tuning process for each model.
- Present the final hyperparameters.

Results:

- Present the results of each model and compare them.

Conclusion:

- Describe the advantages and disadvantages of your approach and an interpretation of your results.

Submit both your code and report. Report should be 1400 words approx.