

Python Programming

Midterm and Final Project

Midterm Part

Please choose one dataset and explain the details of it in one paragraph. Please explain what you decide to do with your data set (Classification or Regression). Then, choose 3 of the following algorithms and explain their working principles in detail. Finally, upload the semi report to classroom. **(Deadline: 05.05.2023)**

Final Part

Please test your dataset by training it with each algorithm you choose. Then, explain the evaluation parameters mentioned below and calculate these parameters of the algorithms you use. Explain all the libraries you use and why you use them. At the end of the report, add the codes for each algorithm as an appendix. Finally, upload the source codes of your project together with the full report to classroom. **(Deadline: Final week)**

Algorithms

1. Random Forest
2. K-Nearest Neighbors
3. Decision Tree
4. Multilayer Perceptron (MLP)
5. Support Vector Machine

Evaluation Parameters

For regression: You need to use 3 metrics below to evaluate your models. (MAE, MSE, R^2)

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

Where,

\hat{y} – predicted value of y
 \bar{y} – mean value of y

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

For classification: You need to use 4 metrics below to evaluate your models.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Useful Links

- <https://builtin.com/data-science/regression-machine-learning>
- <https://stackabuse.com/overview-of-classification-methods-in-python-with-scikit-learn/>
- https://scikit-learn.org/stable/supervised_learning.html
- <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>

Good Luck!