CPC251: Machine Learning and Computational Intelligence
Academic Session: Semester 2, 2024/2025
School of Computer Sciences, USM, Penang

Mini Project

Description

The project assignment is divided into **two** (2) parts.

Form a group of **four** (4) members.

This project must be implemented using Python.

Choose one dataset, register your group and specify your preferred dataset. The dataset and the registration form are available on eLearn@USM. Note that each dataset will be limited to a certain number of groups. Thus, you will be assigned the following preferred dataset in the list if the most preferred dataset has reached the limit.

- Climate Model Simulation
- Seismic
- Dermatology
- Cervical Cancer

Part 1

Part 1 contributes 10% to your overall grade.

You are tasked to build two (2) predictive models to predict the target variable of the dataset. One of the predictive models must be either Decision Tree or Support Vector Machine. Perform a comparison between the predictive models.

Report the accuracy, recall, precision and F1-score measures as well as the confusion matrix (classification problem).

Report the R2 score, mean squared error and mean absolute error (regression problem).

Document the codes.

Note: The focus of the project is on the fine-tuning of the model. Therefore, it is acceptable if the model's performance is not high due to an imbalanced dataset.

Submission requirements

- Due date: **16th May 2023 (Friday)**, 11:59 p.m. (Week 8).
- Use the given Jupyter notebook template and do not change the steps.
- The Jupyter notebook must be executed to **show** the outputs.
- Submission must be made in **ipynb** format (submitted online). Do **not** submit the dataset.
- The filename **must** follow these naming conventions.
    - <CPC251_Project_ Part1_GroupNo.ipynb>
- Plagiarism (using other people's ideas and text without proper acknowledgment and using them as your own) is a serious academic offense. The consequences for plagiarism are severe.

Part 2

Part 2 contributes 10% to your overall grade

You are tasked to build two (2) predictive models to predict the target variable of the dataset. One of the predictive models must be either Neural Network or Fuzzy Logic System. Perform a comparison between the two predictive models.

Report the accuracy, recall, precision and F1-score measures as well as the confusion matrix (classification problem).

Report the R2 score, mean squared error and mean absolute error (regression problem).

Document the codes.

Submission Requirements

- Due date: **5th July 2023 (Friday)**, 11:59 p.m. (Week 15).
- The jupyter notebook must be executed to **show** the outputs.
- Submission must be made in **ipynb** format (submitted online). Do **not** submit the dataset.
- The filename **must** follow these naming conventions.
    - <CPC251_Project_ Part2_GroupNo>
- Plagiarism (using other people's ideas and text without proper acknowledgment and using them as your own) is a serious academic offence. The consequences for plagiarism are severe.

Rubric for Part 1

Component	10-9 (Excellent)	8-6 (Good)	5-2 (Average)	1-0 (Poor)	Weightage
Feature selection	The process of feature selection is clearly implemented. The chosen features are clearly justified based on the implementation.	The process of feature selection is fairly implemented. The chosen features are fairly justified based on the implementation.	The process of feature selection is minimally implemented. The chosen features are minimally justified based on the implementation.	The process of feature selection is not implemented. The chosen features are not justified.	1
Model construction	Two models are trained. The models' parameters are completely fine-tuned.	Two models are trained. The models' parameters are fairly fine-tuned.	Two models are trained. The models' parameters are minimally fine-tuned.	The model construction and selection are poorly or not implemented, and the model performance is absent.	3
Model evaluation	Models are completely evaluated using appropriate performance metrics. Model comparison is performed and justified.	Models are fairly evaluated using appropriate performance metrics. Model comparison is performed and justified.	Models are minimally evaluated using appropriate performance metrics. Model comparison is performed.	Model evaluation is incomplete or not performed. No comparison is performed.	3
Runtime and Algorithm	Executes without errors. The algorithm and outputs are correct.	Executes without errors. The algorithm and/or outputs have minor errors.	Executes without errors. The algorithm and/or outputs are partially correct.	Does not execute due to error. Algorithm is incorrect and no output.	2
Documentation	The source codes are well documented and commented.	The source codes are partially documented and commented.	The source codes are minimally documented and commented	The source codes are not documented and commented	1

Rubric for Part 2

Component	10-9 (Excellent)	8-6 (Good)	5-2 (Average)	1-0 (Poor)
Model construction and selection	The model construction and selection are clearly explained i.e. parameters, fine-tuning and selected in terms of performance metrics. The best suited model is clearly discussed and justified.	The model construction and selection are fairly explained i.e. parameters, fine-tuning and selected in terms of performance metrics. The best suited model is fairly discussed and justified.	The model construction and selection are minimally explained i.e. parameters, fine-tuning and selected in terms of performance metrics. The best suited model is minimally discussed and justified.	The model construction and selection are poorly or not presented, and discussion of the model performance is absent. The best suited model is not discussed and justified.
Results and Discussion	The performance metrics are comprehensively reported. The results are clearly compared and discussed.	The performance metrics are comprehensively reported. The results are fairly compared and discussed.	The performance metrics are minimally reported. The results are minimally compared and discussed.	The performance metrics are not reported. No comparison and the discussion is absent.