CPC251: Machine Learning and Computational Intelligence
Academic Session: Semester 2, 2021/2022
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 can be implemented using any programming language, but preferably Python.

Choose one dataset and register your group and specify your preferred dataset. The dataset and the registration form are given 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 preferred dataset has reached the limit.

- QSAR biodegradation
- Cardiotocography
- Anuran species
- Steel plates fault

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.

Submission requirements

- Create an Adobe Spark page describing the solution to the problem. Refer to the given example.
- Due date: **29th May 2022 (Sunday)**, 11:59 p.m. (Week 8).
- Submission must be made in softcopy (submitted online).
- Export Adobe Spark page to pdf
- Compress all files (adobe spark in pdf, adobe spark link, and source code) into a zip file. The filename **must** follow these naming conventions.
    - <CPC251_Project_ Part1_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.

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.

Submission Requirements

- Create an Adobe Spark page describing the solution to the problem. Refer to the given example.
- Due date: **17th July 2022 (Sunday)**, 11:59 p.m. (Week 15).
- Submission must be made in softcopy (submitted online).
- Export Adobe Spark page to pdf
- Compress all files (adobe spark in pdf, adobe spark link, and source code) into a zip file. 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)	5-3 (Average)	2-1 (Poor)
Background study	Background study demonstrates deep understanding of the topic. Problem definition is clear, concise, and easy to understand. Aim is clearly stated.	Background study demonstrates adequate understanding of the topic. Problem definition is adequate and generally easy to understand. Aim is stated.	Background study minimal. Problem definition is not clear. It may be verbose or utilize a lot of field-specific jargon. Aim is vaguely stated.	Background study is missing. It is unclear what is being defined. Aim is not stated.
Feature selection	The process of feature selection is clearly described. The chosen method is clearly justified.	The process of feature selection is fairly described. The chosen method is fairly justified.	The process of feature selection is minimally described. The chosen method is minimally justified.	The process of feature selection is not described. The chosen method is not justified.
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.

Rubric for Part 2

Component	10-9 (Excellent)	8-6 (Good)	5-3 (Average)	2-1 (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.