

## LAB 3 – Supervised Learning

### Requirements

- Jupyter Notebook + Python (sk-learn, pandas, etc.) + (Optional) Keras-Tensorflow

### Methodology

1. Select a 6-month date-range within the total duration of the Covid-19 pandemic (since 2020).
2. Load the Malaysian Covid-19 datasets from [1] and [2], performing any conversions which may be necessary. These (and following) steps should all be done in a Jupyter notebook, which will be part of your submission.
3. Define a set of features and a single target (e.g. hospitalization, ICU utilisation, infection count). The features and the target can have an offset date range.
4. Define a training, testing, and validation data set.
5. Create and train at least two supervised learning techniques. Parameters should be chosen well.
6. Observe the performance and training speed.

### Report

- Your report should present your design decisions, your results, and analysis of your results.
- You should also justify your design decisions, especially feature/target selection.
- Be sure to state clearly your selected date range and features/target.
- There is no page limit, but be sensible. Aim for 7-8 pages (less is better than more).
- Remember to include the standard lab title page (this doesn't count towards your 7-8 pages).

### Submission

- Your submission comprises of ONE (1) Jupyter notebook and ONE (1) PDF document (Word files not accepted).
- Submission deadline is 7 days after your lab finishes (11.59pm one week after your lab).
- Name the documents Lab3\_P<X>\_<Group Number>.<file extension> where 'X' is your practical number (P1, P2, P3 until P8) before submission to WBLE.

### Groups

- Groups should be formed for every session, containing only students from the same session (e.g. Tuesday and Friday sessions cannot combine). 4 members per group. At most one/two groups should have less than 4 members (with lecturer's permission).

### Selecting Features and Date Range

- You are free to select any features and date ranges (must be 6 consecutive months). Make sure your date ranges are not identical with any other groups in your session (a Google sheet[3] is provided to help facilitate this). Some date ranges may prove harder than others, so feel free to change to another date range if this is the case.

[1] – MoH Malaysia C-19 data <https://github.com/MoH-Malaysia/covid19-public/tree/main/epidemic>

[2] – C-19 Immunisation Task Force (CITF) data <https://github.com/CITF-Malaysia/citf-public>

[3] – <https://tinyurl.com/AI2022SupervisedCoordination>

## LAB 3 – Supervised Learning Marking Rubric

### Graded Components Weightage

- Presentation and Formatting (15%)
- Problem and Model Design (25%)
- Results and Analysis (30%)
- Code Quality (30%)

### Presentation and Formatting Rubrics

- 0 Unreadable report.
- 1 Difficult to read, with obvious errors in formatting, grammar etc.
- 2 Acceptable, with some errors in formatting, grammar etc.
- 3 Good readability, appropriate use of graphics/tables. Minimal grammatical and formatting errors.
- 4 Outstanding presentation and formatting, no errors at all.

### Problem and Model Design Rubrics

- 0 Not included.
- 1 Design parameters incompletely presented.
- 2 Design parameters completely presented.
- 3 Design parameters presented with reasoning/justification.
- 4 Design parameters presented with complete reasoning/justification.

### Results and Analysis Rubrics

- 0 Not provided.
- 1 Incomplete/inaccurate results with minimal analysis.
- 2 Results report and some brief analysis provided.
- 3 Results reported in well organised/summarised manner along with clear analysis.
- 4 Results attractively presented along with convincing analysis/justification.

### Code Quality Rubrics

- 0 Not submitted.
- 1 Very poor code (no cells, hard to read etc.) or provided code does not work.
- 2 Working code.
- 3 Code is well organised and commented.
- 4 Code is easy to read because it is very well organised, showing proper planning.

### Tabulation of Marks

Each graded component receives a mark based on the above rubrics. This assigned mark N is then divided by the maximum mark for the rubric M and multiplied by the weightage W. So the sum of your report marks S will be:-

$$S = \sum_{i=1}^n \frac{N_i}{M_i} \times W_i$$