

Summative Assignment

Artificial Intelligence COMP2261 – Machine Learning 2020/2021

Deadline for submission: 05 May 2021 14:00 (GMT)

Background

- Coronavirus (COVID-19) is spreading fast. Since first reported in December 2019, by mid-January 2021, it has affected more than 95 million people and killed more than 2 million people worldwide. To aid the analysis and inform public health decision making, machine learning models trained on real data can be very useful.
- In this assignment, you are to build and compare predictive models using machine learning algorithms and epidemiological data from the COVID-19 outbreak.

Task Specification

- To explore the dataset and define ONE problem you want to solve, with motivations.
- To choose any 3 machine learning algorithms (not necessarily from the lectures in this module) that are sophisticated and appropriate to solve the problem.
- To build 3 predictive models and make comparisons between them.
- To write a **scientific report** (in a style of an academic paper).

Dataset

- Download https://github.com/beoutbreakprepared/nCoV2019/tree/master/latest_data
- Description <https://www.nature.com/articles/s41597-020-0448-0>

Submission (2 files) via DUO

- **bannerID.pdf** – the scientific report
 - Articulate the problem you plan to solve, and the methodology you plan to use.
 - Provide necessary tables and charts to summarise and support the comparisons.
 - Discuss the results and draw conclusions from your experimentation.
 - Up to 4 pages, inclusive of appendices and figures, exclusive of references.
- **bannerID.py (or .ipynb)** – the source code
 - Make it clear in the initial comments how to run your code (if necessary).
 - You may use libraries, e.g., Numpy, Pandas, Matplotlib, SciKit-Learn, etc.
 - This is not to be marked, but plagiarism and collusion will be checked.

Marking Scheme

- [10%] Problem framing.
 - [35%] Experimental procedure.
 - [25%] Presentation, visualisation, interpretation and comparison of the results.
 - [20%] Discussions of chosen models, experimental procedure, and limitations.
 - [10%] Conclusions and lessons learnt from the assignment.
- N.B.** marks are generally awarded for good experimental procedure that supports the results, NOT for achieving the best prediction performance.