# **Non-neural machine learning (MIB-ITLB358)**

# **Home Assignment**

## **Conditions Contributing to covid 19**

*Your task is to explore and prepare data and build classical ML models.* Ideally, you would optimise the models' hyperparameters and evaluate models quantitatively and qualitatively from several aspects.

Data location for the task (single CSV file zipped): <https://drive.google.com/file/d/1Qnp0WeRAlycJLGsr4ghyB-TzS4cQpaLl/view?usp=drive_link>

(Original data source: <https://catalog.data.gov/dataset/conditions-contributing-to-deaths-involving-coronavirus-disease-2019-covid-19-by-age-group>)

*Data preparation steps for everyone* (if you encounter issues in the data preparation step, you can download the prepared data for modelling from <https://drive.google.com/file/d/1WZOXBPz99WfHOriHiYxRVlOpgqDunCVM/view?usp=drive_link> --- of course, you will not get points for the data preparation step this way):

* Restrict the dataset to the rows where Start Date is "01/01/2020" and the End Date is "06/24/2023".
* Remove rows not belonging to specific age groups.
* Remove rows with the United States as State.
* Investigate missing values, and deal with them if you find any.
* Encode age groups as you see fit as numeric columns.
* Create a Broad Condition Group variable, which should be the same as the Condition Group variable for the two most frequent condition groups, but has the value "other" for all other condition groups.
* Label-encode the Broad Condition Group variable.

### **Regression group:**

* Your target is the COVID-19 Deaths variable, and your input variables are Condition and Age Group.
* Choose at least two model types, then train and evaluate them. Pay attention to using an appropriate train-test split.
* Try to reason about the models' performance, given what you know about the data and the relevant model types.

### **Classification group:**

* Your target is the Broad Condition Group variable, and your input variables are COVID-19 Deaths and Age Group.
* Choose at least two model types, then train and evaluate them. Pay attention to using an appropriate train-test split.
* Try to reason about the models' performance, given what you know about the data and the relevant model types.

## **Please observe the following**

●  You must use a single standalone Jupyter Notebook to solve the task and submit the .ipynb file. **Upload your .ipynb file to Moodle.**Note for those working on Google Colab: a link to your notebook will not suffice: you have to download and submit the file itself.

●  Follow the **principle of literate programming**, and make use of the markdown cells of the notebook.

○  Don't just code. Explain and motivate the steps you take.

○  Don't just display results and plots. Interpret and evaluate them.

●  For the **deadline**, please refer to the Moodle page of the module.

## **Assessment**

The assignment will be assessed based on the following criteria (see the marking grid on Moodle):

●  Specification fulfilment and conceptual grounding (60%)

●  Literate programming and markdown cells (20%)

●  Coding (20%)

The **resit arrangement** for the assignment is the same as above; you may resubmit the same paper, with corrections, that you submitted by the original deadline. The resubmission deadline will be specified on Moodle after the grades for the original submission are published.

## **Academic conduct notice**

Where the Academic Conduct Officer has reason to suspect that a piece of work submitted by a student was wholly or in part written by someone other than the student who submitted it, and this has not been disclosed by the student, they may call for the student to defend the work in viva or a written comprehension test. The burden of proof in such a viva or test will be upon the student to demonstrate to the examination panel’s satisfaction his/her full comprehension of the work s/he has submitted. Failure to appear without satisfactory explanation will result in immediate failure of that assessment, with consequences of academic misconduct and application of sanctions.