



# Assignment 1

COSC 2673-2793 | Semester 1 2021  
(computational) Machine Learning

# Key Dates and Important Information

---

- **Assignment Type:** Individual
- **Due:** Friday 16<sup>th</sup> April 2021 (Week 6)
- **Marks:** 30%
- **Specifications:** on canvas – Specification PDF & Marking rubric
- **Late policy:** After the due date, you will have 5 days to submit your assignment as a late submission. Late submissions will incur a penalty of 10% per day. After these five days, Canvas will be closed and you will lose ALL the assignment marks.

Clarifications/updates may be made via announcements/relevant discussion forums.

# Task

---

The machine learning **task** we are interested in is:

*“Predict if a given patient (i.e. newborn child) will be discharged from the hospital within 3 days (class 0) or will stay in hospital beyond that - 4 days or more (class 1).”*

# Dataset

---

The data set for this assignment is available on Canvas.

- `README.md`: **Description of dataset.**
- `train_data.csv`: This data is to be used in developing the
- `test_data.csv`: You need to make predictions for this data and submit the prediction via canvas.
- `s1234567_predictions.csv`: expected format for your predictions on the unseen test data. **Any deviation from this format will result on zero marks for the results part.**

# Restrictions

---

- You must **NOT** explicitly perform **manual feature selection**. That is, your models should have all features (attributes) as input (except the “ID” and “Health Service Area” fields which are not attributes).
- You are **only** allowed to use **techniques taught in class up to week 5 (inclusive)** for this assignment. That is, you are NOT allowed to use ML techniques such as: Neural networks or SVM for this task.

# Deliverables

---

- **The PDF version of the python notebook** used for the model development including critical analysis of your approach and ultimate judgement.
- A **set of predictions** from your ultimate judgement. Should be in CSV format. If your model predicts the patient will be discharged from the hospital within 3 days, the associated “LengthOfStay” value in CSV should be 0 (1 otherwise).
- **Your code** (Jupyter notebooks) used to perform your analysis. Should be a ZIP file containing all the support files. will be used for plagiarism checking – notebook should match PDF.

# Marking Guidelines

---

- Approach 50%;
- Ultimate Judgment & Analysis 20%;
- Performance on test set (Unseen data) 20%;
- Implementation 10%;

**Rubric attached on canvas**

Practice the typical machine learning process which includes:

- Selecting the appropriate ML techniques and applying them to solve a real-world ML problem.
- Analysing the output of the algorithm(s).
- Research how to extend the modelling techniques that are taught in class.
- Providing an ultimate judgement of the nal trained model that you would use in a real-world setting.

# Academic Integrity and Plagiarism

---

You code and report will be screened using plagiarism checking software.

- **PDF Notebook:** Turnitin
- **Code:** CodeQuiry

See section 6 on assignment specifications for more details.