30/03/21

Project Description

Background:

Hospital-acquired infections account for significant morbidity and mortality in patients in the intensive care unit (ICU): around 20% to 30% of ICU admissions reported incidences of hospital-acquired infections. The infection can be fungal, viral, or bacterial. In case it is bacterial, it can be treated by antibiotics, but several types of antibiotics can be administered, and some bacteria are resistant to certain types. In case the infection is viral or fungal, antibiotics are not useful and even could be damaging to the patient. The gold standard for bacterial infection detection is a culture, where the tissue sample is incubated for 24-72 hours. An interim answer after 24 hours can tell if there is a bacterial infection and after 48 hours the exact species of the bacteria will be detected. For bacterial infections, a final answer obtained after 72 hours is whether the infection is resistant or not to each type of tested antibiotic.

The ability of physicians to diagnose an infection when it is first observed is limited, and therefore the decision of antibiotics initiation (e.g. the treatment for infection) tends to be delayed. Sometimes the physician needs to decide if to administer antibiotics and if so which type of antibiotics to use - before any data from the culture is available. In that situation, physicians rely on their decisions on the patient's clinical presentation and other guidelines. Later on, the data from the culture may indicate that no antibiotics were needed, or that the antibiotic administered was wrong, since the bacteria is resistant to it.

This situation raises two main research questions:

- Can we detect in advance if a culture will be positive (i.e., the infection is bacterial) or negative?
- If the culture was positive, can we detect in advance if the administered antibiotics what appropriate or not?

Project Tasks:

A. Construct a classification model of positive vs. negative cultures of patients with ICU-acquired infections based clinical data that was collected before culture collection time.
B. Construct a classification model of appropriate vs. inappropriate treatment that was given to patients with positive cultures based on the clinical data that was collected until 24 hours after culture collection time.

Project schedule:

Individual meetings (20 min for each group with the workshop team):

- **07/04: Milestone 1** - Present your suggested approaches for the tasks.

  Deliverables:
  - A ~10-minute presentation (oral description and handed in document) including
    - Cohort and data properties
    - Selected features
    - Initial visualizations of the data
    - Project execution plan: outline your planned approaches and the steps that you expect to take in your project.
  - Verbal discussion
- **12/05: Milestone 2** – Present initial results for the two tasks, including the code written. Show different approaches that you tried, including those that did not work well!

  Deliverables:
  - A ~10-minute presentation including
    - The approaches tried
    - Results for each task
    - Execution plan for the next steps
  - A document summarizing your current results (≤7 pages). The document should be self-contained and understandable on its own, independently of the discussion. Make sure figures are precisely labelled with clear captions.
  - Code
  - Verbal discussion

- **16/06: Milestone 3** – Advanced results

  Deliverables:
  - A ~10-minute presentation including
    - The approach(es) you converged to
    - Revised results for each task
    - Execution plan for the next steps
  - A revised document summarizing your current results (≤7 pages)
  - Revised code
  - Verbal discussion
- **31/7: Final Submission** - see guidelines below.
- **1/8: Final oral Presentation** - Final joint meeting of all groups, where each group will have 7 minutes to present its results

Project Evaluation: *(tentative, subject to change)*

Project grade will be composed of the following elements:

- **10% - milestone 1**
- **10% - milestone 2**
- **10% - milestone 3**

You will be judged based on the understanding you show, your presentation, your initiative, and the progress you make. In milestones 2-3 we may check your code, but the quality of the results is *not* a factor in your grade. In milestones 2 and 3 each group will report AUROC and AUPR based on 5-fold cross validation for both tasks on MIMIC III data (we will provide the folds of MIMIC III data set), and the AUROC and AUPR for task A on eICU data (Note that task B is not relevant for the eICU dataset, as the sample size is too small.)

On task A, we recommend that you use the eICU as a validation set and not to use it for the model-training phase.

We will post the results of all groups (anonymized, without your names) in a score board to help you feel where you stand. We trust you to report honestly.

- **30% - Final Submission**

Here we will test your code for readability and effectiveness. You will also be judged based on your understanding, the presentation, code quality, and the quality of the results.

- **30% - Performance on an external validation set**

Up to 30 points for the performance of your methods on new data that you were not previously given. We will run your code on that data.

In each score and dataset, the groups are ranked by their performance, where 1 is the highest rank, 2 is second etc. Ties will be scored in accordance to the House of Hillel. Here is the way the ranks for score $s \in \{AUPR, AUROC\}$ of a group are weighted:

	Rank for score *s* on task	Final rank for score *s*
Task A	δ^s	$\rho^s = 0.7\delta^s + 0.3\,\gamma^s$
Task B	γ^s	

The final score is:

$$31.5 - (0.6 * \rho^{AUPR} + 0.4 * \rho^{AUROC}) * 1.5$$

- **10% - Final oral presentation**
- **Bonus - up to 10%** for a group that will conduct additional integrations or develop a new computational method. Please elaborate about your special plans on Milestone I.

Technical Guidelines

Data Sources reminder:

- **MIMIC III** - a dataset developed by the MIT Lab for Computational Physiology, comprising de-identified health data on ~60,000 intensive care unit admissions. It includes demographics, vital signs, laboratory tests, medications, and more.
- **eICU** – a dataset from many critical care units throughout the United States. It covers de-identified health data on ~200,000 patients who were admitted to critical care units in 2014 and 2015.

Inclusion Criteria:

- All patients with blood cultures that were not contaminants or were not cancelled.
- Only patients admitted directly to Emergency dept. or ICU
- (1) ICU-acquired infection:
  - All patients who were hospitalized **at least 48 hours** in the ICU.
  - Patients with a **first culture that was collected in the ICU** and **after at least 48 hours** from ICU admission.
  - We will use only the first culture that was collected in the ICU.
- (2) Suspected ICU acquired infection:
  - Patient whose first culture in the **entire hospitalization** was collected in the ICU.
  - The culture was collected at least **24 hours after** hospital admission.

Positive class is defined if there was a growth of pathogen in the culture.

*Notice that the patient could be included both in groups (1) and (2) (for example – a patient whose first culture in the entire hospitalization was collected 2 days after > 48 hours in the ICU).

MIMIC III Dataset

Model A

1. First, please download from *moodle* the CSV file “model_a_mimic_cohort”.
2. Load this CSV file to a table based on the queries on ‘MODEL A - Data Loading’ in the ‘MIMIC_Queries.sql’ file.
3. Use queries ***q4_a*** and ***q4_b*** to create a table with all the features that will be collected.
4. Use querires **q*5*_*a***, ***q5_b*** and ***q5_c*** collects all the features mentioned in section 4 for patients from section 3. In ***q5_d*** we collect only features that were collected before the

culture collection time (target_time) and add some demographic data (age, gender). ***q5_e*** save the data for a CSV file.

Positive Class size – N = 165

Negative Class size – N = 2,939

Model B

1. First, please download from *moodle* the CSV file "model_a_mimic_cohort".
2. Load this CSV file to a table based on the queries on 'MODEL B - Data Loading' (notice that the target time is different from Model A since we add 24 hours to it) in the 'MIMIC_Queries.sql' file.
3. Use the querires ***q4_**** and **q*5_**** in the same manner that was mentioned on Model A section to collect the relevant features. Notice that you should change 'model_a_mimic_cohort' -> 'model_b_mimic_cohort' in queries ***q5_a*** and ***q5_b***.

Appropriate Class size – N = 100

Inappropriate Class size – N = 32

eICU dataset

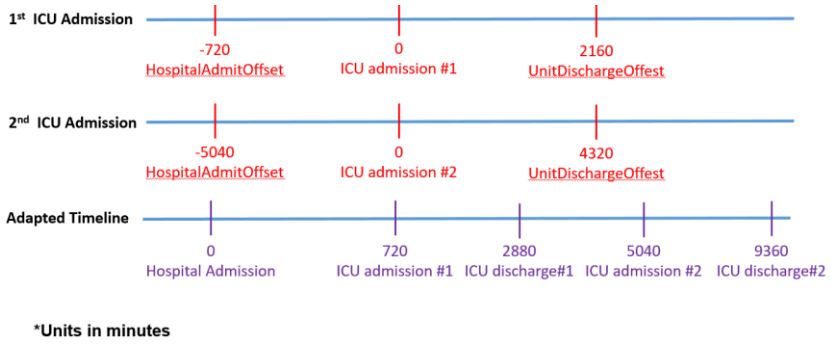
Model A

1. First, please download from *moodle* the CSV file "model_a_eicu_cohort".
2. Load this CSV file to a table based on the queries on 'MODEL A - Data Loading' 'eICU_Queries.sql' file.
3. Use queries ***q1_**** and ***q2_**** to load the data for the cohort.
   Positive Class size – N = 10
   Negative Class size – N = 118

Some more points:

- The sample size in model B is smaller than the size of the positive class for Model A, since only part of the patients got antibiotics after the culture was collected, and in case a patient did not get an antibiotic, we could not asses if the antibiotic was appropriate or not.
- In the US, when patients are in the emergency department (ED) they may not yet be admitted to the hospital. The ED is considered an outpatient clinic. Many patients will go home and never be admitted as inpatients. Hence for part of patients lab tests will be available before their admission time.

- In eICU the timestamps are a bit complicated, and the timeline does not start at the time of hospital admission. A patient (*patienthealthsystemstayid*) can have multiple ICU admissions and each ICU admission (*patientUnitStayID*) started at time 0, and each measurement is measured with a time offset relative to this admission time. In order to create the 'adapted timeline' you'll need to add the *hospitalAdmitOffset* for each timestamp. Please see this link. The following illustration will assist you with the understanding.



*Units in minutes

External validation set

We will evaluate the final submitted models by running them on disjoint datasets from your training data, but with similar characteristics. The file formats will be identical to the formats of 'model_a_mimic_cohort' and 'model_b_mimic_cohort' files. The exact instructions for final code submission will be provided later.

Final Submission Requirements:

A. A **word/PDF File with detailed project** description, its main analyses, and results:
   (1) **Project Introduction**
   (2) **Cohort Description** – An explanation of the datasets - their characteristics, size, missing rates, etc. Add a table of characteristics that compare the two groups in in each data set.
   (3) **Methods:**
      a. **Data Preprocessing** – Explain the feature engineering process – normalization, transformations, shapelets, etc.
      b. **Models** - Succinctly describe your models. For example, suppose you select only continuous features that were significantly different between the 'Positive-Culture' vs. 'Negative Culture' groups on the training set based on t-test. Then, mention how you compared the groups (=t-test), but do not explain or repeat what is a t-test and how it is computed. If you are unsure – provide a reference to the method. The same goes for algorithms – when you use existing models, such as random forest, XGBoost, Logistic

Regression, mention and reference them but do not need explain their theory. If you develop something new – describe it in detail, in an appendix if needed.

c. **Evaluation Process** - how did you evaluate your project. K-fold validation? Bootstraping?

(4) **Results:**

a. Describe your results with relevant figures. The figures should be informative and you should add a legend to describe each of them.

b. The results should be reported without any reference to the External Validation set.

(5) **Discussion:**

a. Describe your project results, your model's limitations, and your main conclusions.

* It is highly recommended to use many figures in order to validate your conclusions.

* The size of the doc should include ~8-10 pages of text (without figures and tables).

* This document should be detailed up to the level that someone could replicate your work based on it.

B. **Python Code**

* You are not required to implement code from scratch – if the code that you need already exists in a specific library, import and use it. **Exact submission guidelines will be published later.**

Submission Date: 31/7/2021 by email to Dan Coster (dancoster@gmail.com).

You are more than welcome to consult with us throughout the project (especially at the beginning!) and ask for consultation meetings besides the milestones.