Case – Exploring Electronic Health Records

Please prepare a presentation based on the scenario presented below. We would like to focus on your skills in analytics, communication, and problem-solving methodology. Put code and results in your public git repository and send a link to **ernst.ahlberg@molnlycke.com** **AND** **jessica.martin@molnlycke.com**

In this case you are instructed to explore a synthetic dataset comprised of Electronic Health Records (EHR). EHR data is generally a collection of medical-related records at patient-level where it is not uncommon to find components such as: *demographic data* (patient gender, age, etc.), *medications* (which medications and prescribed and if these are dispensed from the pharmacy), *medical history* (e.g. diagnose codes, treatment codes and where and when the patients visited the hospital), *lab results* and *vital signs*.

A sample database of the synthetically generated dataset called Synthea can be downloaded here:

https://synthetichealth.github.io/synthea-sample-data/downloads/synthea_sample_data_csv_apr2020.zip

Your list of tasks is the following, for the 1k dataset:

- Load the data
- Explore the data and make a visualization of a *single patient trajectory* as she transitions through the medical care system over time
- Explore the data to find and present patterns of patients with the same conditions such as:
    - Which are the three most common conditions (present graphs and numbers)?
    - Are there similarities in how the three conditions are treated? Showcase examples.
    - What other common pattern characteristics can be found for the three groups of conditions?
- Formulate three other questions that could be interesting from a machine learning perspective using this data (that could potentially be used in a clinical setting to improve care)
- Compile your methods, results/answers and conclusions in a presentation and prepare to be able to present this during 10 minutes

Useful links

- Description of the Synthea method (journal article) used for generating synthetic patients and electronic health records https://academic.oup.com/jamia/article/25/3/230/4098271
- The 1K patients dataset (subset) of the synthetically generated data: https://synthetichealth.github.io/synthea-sample-data/downloads/synthea_sample_data_csv_apr2020.zip