

Project title: Healthcare - Persistency of a drug

Group name: DG_team_project_PL-RO-KSA-EGY

Github repo: https://github.com/Omar-Safwat/HealthCare_project

Week: 7

Team members

Name	Specialization	Country	Email
Ms. Larisa Popa	Data Science	Romania	Larisapopa4@gmail.com
Ms. Afshan Hashmi	Data Science	Kingdom of Saudi Arabia	afshanhashmi786@gmail.com
Mr. Omar Safwat	Data Science	Egypt	omarksafwat@gmail.com
Mr. Roger Burek-Bors	Data Science	Poland	roger.burek-bors@hotmail.com

Problem description

Failure to adhere to prescribed-medication regimens is one of the principal reasons patients don't achieve the expected outcomes from their treatment. Solving this challenge has been a major goal for pharmaceutical and healthcare organizations for decades. Studies show that 50 to 60 percent of patients with chronic illnesses miss doses, take the wrong doses, or drop off treatment in the first year. An estimated 125,000 lives are lost annually in the United States and additional healthcare expenditures of 290 billion USD are driven by nonadherence.

In order to gather insights on the factors affecting drug persistency among patients, as per physician's prescription, a dataset is to be collected, investigated and analyzed. To automate this process for a pharmaceutical company, a machine learning classification model is subsequently built and deployed as a WebApp.

Patients in the dataset are suffering from NTM, are environmental organisms that can be found in soil, dust, and water including natural water sources, causing pulmonary diseases that resemble tuberculosis.

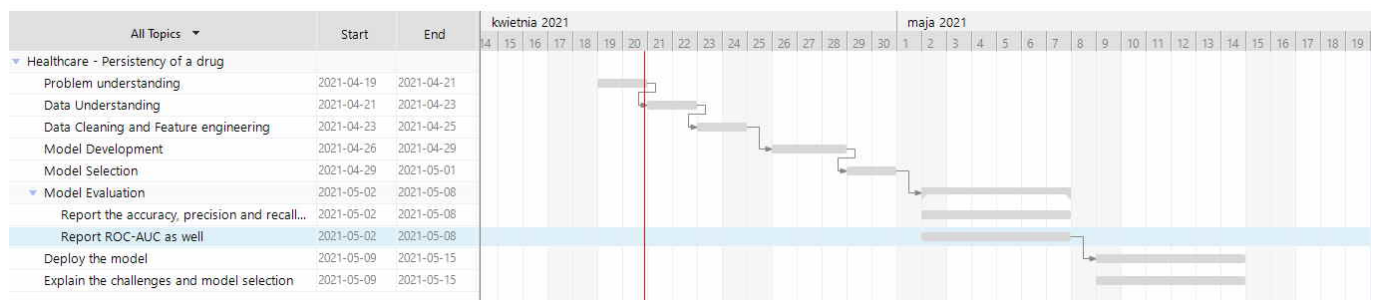
Business understanding

A Pharmaceutical company aims to improve patient adherence to their drug through data-driven insights. The team will analyze the data and provide the company with a WebApp that utilizes Machine Learning technology, thus aiding the company with predicting future patients' adherence to their drug.

The Machine Learning model, built to automate this classification process, is based on a dataset featuring patients infected with NTM. The model will utilize multiple dimensions driving adherence such as:

- Patients' demographics.
- Provider attributes; speciality of the Health Care Provider.
- Clinical Factors pertaining to each patient, e.g. T-Score, Risk Factors, Test results...etc.
- Disease/Treatment factors, e.g. Concomitancy, Comorbidity...etc.

Project lifecycle



As seen in the Gantt Tree above, the project lifecycle is divided into 4 phases, with a submission deadline of **2021-05-15**.

Phase 1

From: 2021-04-19

To: 2021-04-25

This phase focuses on understanding the problem, and setting goals that meet with the client's demands. After which the dataset is preprocessed, and relevant data features are extracted for later phase.

Phase 2

From: 2021-04-26

To: 2021-05-01

In this phase multiple machine learning models will be build and compared.

Phase 3

From: 2021-05-02

To: 2021-05-08

Following the building of several classification models in the previous phase, all models will be evaluated, subsequently, selecting the most accurate and reliable classification model.

Phase 4

From: 2021-05-09

To: 2021-05-15

The final phase will focus on building a Web Application that deploys the chosen Machine Learning model.