

1 Task Description

During the last week, the hospital of Lechi suffered a viral epidemic. The virus has affected both, patients and hospital stuff. In order to better apply contingency measures you must develop an early detection procedure.

In this assignment, you will use machine learning to predict the potential patients.

The task is formulated as a binary classification problem where you have two classes to predict: infected or not-infected. You will be evaluated with the F1-score metric.

2 Dataset Description

The dataset contains a list of patient/workers of the hospital together with some personal information.

The target attribute information is described below.

Number of instances: 13742

Number of attributes: 12

2.1 Target Class:

0- not-infected, 1- infected

2.2 Attribute Information:

Item	Attribute	Type	Values
1	age: age	Numerical	Year
2	speciality-years: years since speciality department opening	Numerical	Years
3	speciality: speciality department	Categorical	Aesthetic, Anaesthesiology, Cardiology, Emergency, Endocrinology, Gastroenterology, Gynecology/Urology, Nephrology, Neurosciences, Oncology, Ophthalmology, Orthopedic, Pediatrics, Psychiatry, Pulmonary, Radiology
4	wing: building wing	Categorical	?, E, N, NE, NW, S, SE, SW, W
5	entry-level: gravity entry status	Categorical	high, medium, medium-high, minor, minor-medium
6	occupation	Categorical	Cafeteria, Delivery, Dietitian, ER doctors, Gift Shop, Lab, Medical Assistants, Nurse, Patient, Physician, Surgeons, Techs, Therapist
7	intervention: intervention procedure	Categorical	Anesthesia, Diagnostic, Other, Propaedeutic, Surgical, Therapeutic
8	ethnicity	Categorical	Other ,Amer-Indian-Eskimo ,Asian-Pac-Islander ,Black ,White
9	sex	Categorical	Male, Female
10	hours: hours since infection	Numerical	Hours
11	country: country of origin	Categorical	country

3 Prediction File Submission

3.1 Model Submission: 04.12.2019, 11:59am

You are asked to kindly submit the following supporting information:

1. A brief description of the step by step methodology (i.e. pre-processing, visualization, training, testing, etc.) that you have followed to do the assignment, with the aim of illustrating the motivation behind your selected approach.

- File Format: .pdf
- Filename: **6-digit** student code (e.g. 123456.pdf)

2. The python code that you used to do the assignment, with comments within the code to ensure that they can be clearly understood.

- File Format: .ipynb, .py
- Filename: **6-digit** student code (e.g. 123456.ipynb or 123456.py)

The model can not be modified after submission.

3.2 Test Publication: 04.12.2019, 14:00pm

3.3 Prediction Submission: 06.12.2019, 11:59am

You are kindly requested to strictly follow the described submission guidelines:

- File Format: .csv
- Filename: **6-digit** student code (e.g. 123456.csv)
- Column Format: 1 Column named **"target"**
- Row Format: Your predictions (0 or 1) with the **same number of rows and in the same order as the test set** provided to you

4 Important:

- The model can not be modified after submission.
- Any submission out of the guidelines (out of deadline, empty file, wrong student code) will not be graded.
- Check your uploaded files.