

FIT3077 Assignment 2 bonus mark question

Machine Learning

Written by Megan Ooi Jie Yi (30101670) & Hew Ye Zea (29035546)

The machine learning model is created using R. To view our classification model and the accuracy of our model, please open the machineLearning.Rmd file.

Instructions

1. Run the get_data.py script to poll the server and get patients' data.

The script gets data for every 50 pages and sleeps for 40s to not overload the server.

2. patient_data.csv stores patient data obtained from the server.

3. Run data_wrangling.ipynb.

This file is responsible for cleaning up the data set; imputing missing values and removing unused data. It also splits the data set into train and test data sets for machine learning.

4. Run machineLearning.Rmd.

This file is responsible for generating the machine learning model and calculating the predicted values to determine the accuracy of the machine learning model.