Introduction to Machine Learning - WFHBOICT.MCL.21

Productbeoordeling

For this 'productbeoordeling' you will be required to solve a case by answering 5 questions. The data for this case is available as .csv file on ELO.

One major difference between the Coronavirus and previous epidemics is the abundance (or lack of) of Data. The enormous flow of data from health centers across many countries is very attractive for Data scientists, who are building ML models to understand the combination of factors that could potentially lead to higher death rates.

Attached find a dataset with multiple variables emerging from diverse countries around the world. To summarize the data set consists of data organized by country and date since the beginning of 2020 until about the 1st week of October. The daily new cases during this period is presented along with the daily number of deaths. The data also presents demographic factors such as population, population density, median age, aged 65 older, aged 70 older, gdp per capita, extreme poverty, cardiovasc death rate, diabetes prevalence, female smokers, male smokers, handwashing facilities, hospital beds per thousand, life expectancy and human development index.

Based on this data set please provide your answer to the following 5 questions:

Question 1: Which variables can we intuitively accept as independent and which ones as dependent? Motivate your answer in max 250 words. (10 points)

Question 2: What are your observations on the completeness of the dataset? Which mitigating actions do you plan to take to work around the incompleteness of the dataset? Max. 250 words. (10 points)

Question 3: How do you dissect the data into learning data and test data? Motivate your answer in max. 250 words. (10 points)

Question 4: Build a supervised learning model in Python in a Jupyter Notebook environment based on your answers in questions 1, 2 and 3 above. Test your model and report its performance. (50 points)

Question 5: Visualize your insights using graphs. Motivate the insights in max 250 words. (20 points)

In groups of 2 students you are expected to make three file submissions: (1) a Python notebook containing the code, (2) a Pdf document motivating your insights and answers and (3) a 5 minute video explaining the code and your top 3 insights. To qualify for a grade each student must individually upload the 3 files to ELO.