CSAI 801 Artificial Intel & Mach Learn

COVID-19 Outcome Prediction

Ahmed Mohamed Moussa Mohamed

**Project overview**

COVID-19 is the disease caused by the emerging coronavirus called SARS-CoV-2. There are many symptoms that accompany the disease, but they differ from one person to another according to several factors such as age. The most common symptoms of COVID-19 are:

• Fever

• dry cough

• stress

Other less common symptoms that may affect some patients include:

• loss of taste and smell,

• Nasal congestion,

• Conjunctivitis (also known as red eyes), etc.

What determines whether a person will die from COVID-19 or not, there are many factors such as age or severity of symptoms and whether or not he has a chronic disease or not. It turned out that the death rate among those over 80 years old reached 25%. Doctors discovered that the main cause of death for people infected with the Corona virus is pneumonia that affects the lungs together, which prevents the body from obtaining the necessary amount of oxygen, and as a result, the body’s organs stop performing their functions, and thus the death of the infected person.

**Problem statement**

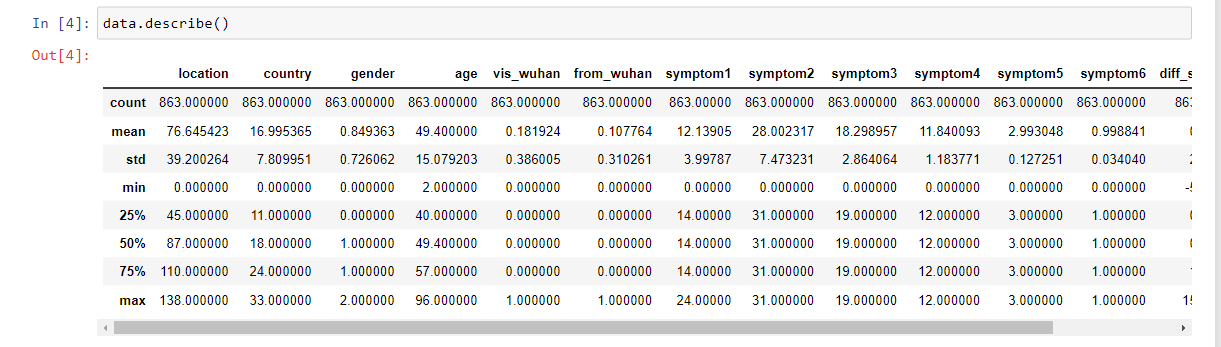
The provided dataset is available from 22 Jan, 2020. And contains the reported daily level confirmed information on the number of affected cases, deaths and recovery from 2019 novel coronavirus. The data is mainly used to make a prediction of whether a person is going to recover from corona virus symptoms or not with the consideration of the pre-defined standard symptoms acquired by the given guidelines by the World Health Organization.

**Data description**

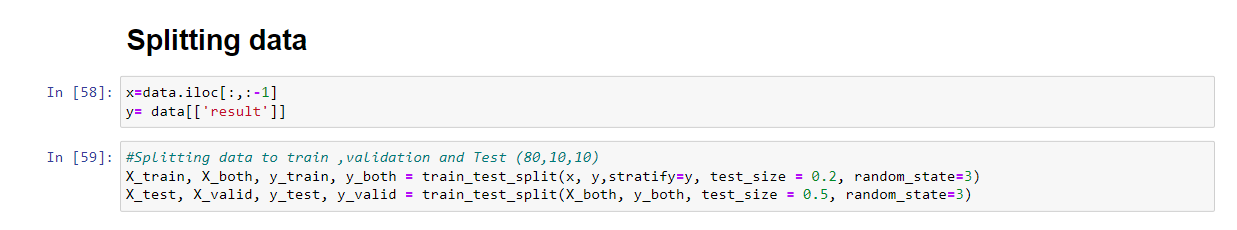
The dataset contains 14 major variables (13 features that shows symptoms and one label Column) that will be having an impact on whether someone has recovered or not. The description of each variable are as follows: 1. Country: where the person resides 2. Location: which part in the Country 3. Age: Classification of the age group for each person, based on WHO Age Group Standard 4. Gender: Male or Female 5. Visited\_Wuhan: whether the person has visited Wuhan, China or not 6. From\_Wuhan: whether the person is from Wuhan, China or not 7. Symptoms: there are six families of symptoms that are coded in six fields. 13. Time\_before\_symptoms\_appear: 14. Result: death (1) or recovered (0)

**Model building, tuning and testing**

The first step in the project is importing the necessary libraries , loading the dataset (“data.csv”) and describing it then doing some visualization on the data



Second step is Splitting data to train, validation and Test (80,10,10) and using stratify field to solve the imbalance in the data

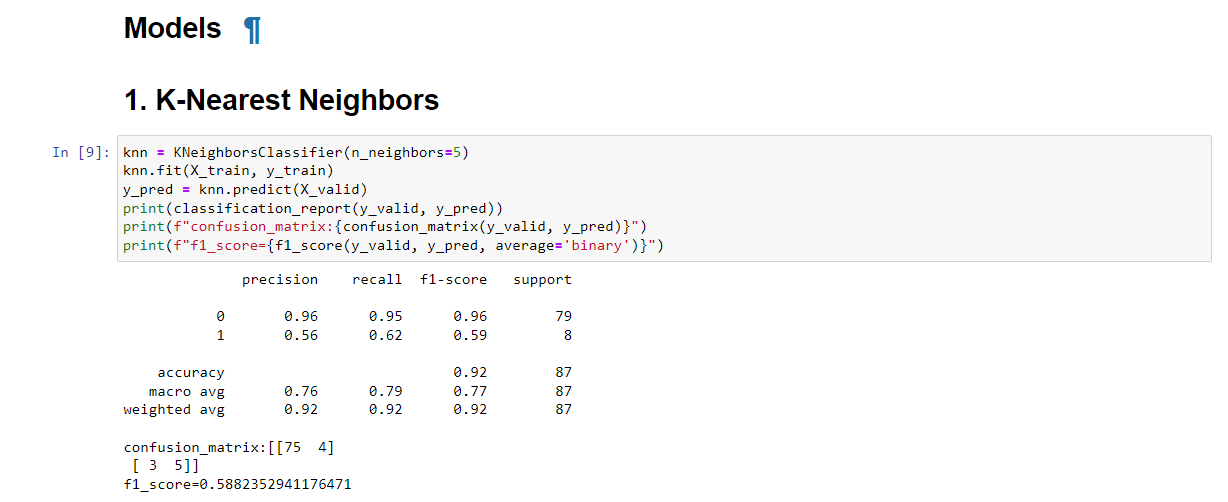


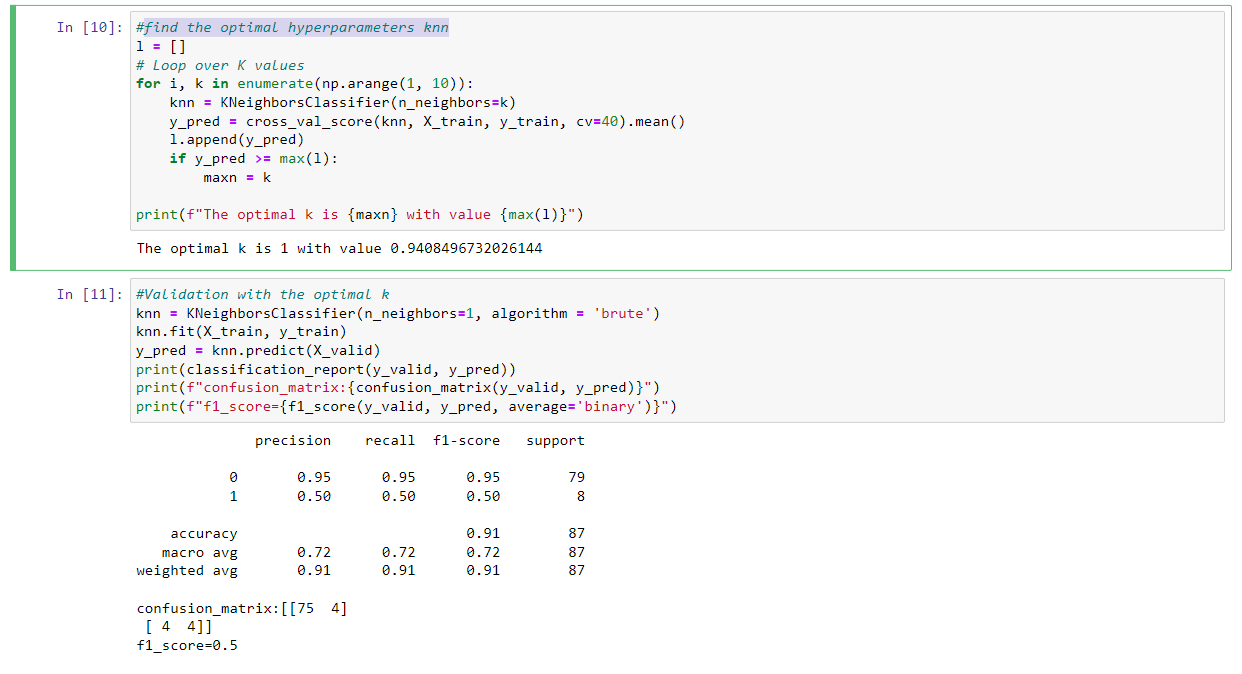
Then start training the KNN model with random k equal 5 and observe its performance using some different metrices precision ,recall ,f1-score and accuracy

Then find the optimal hyperparameters for the model and validate it with it optimal hyperparameters

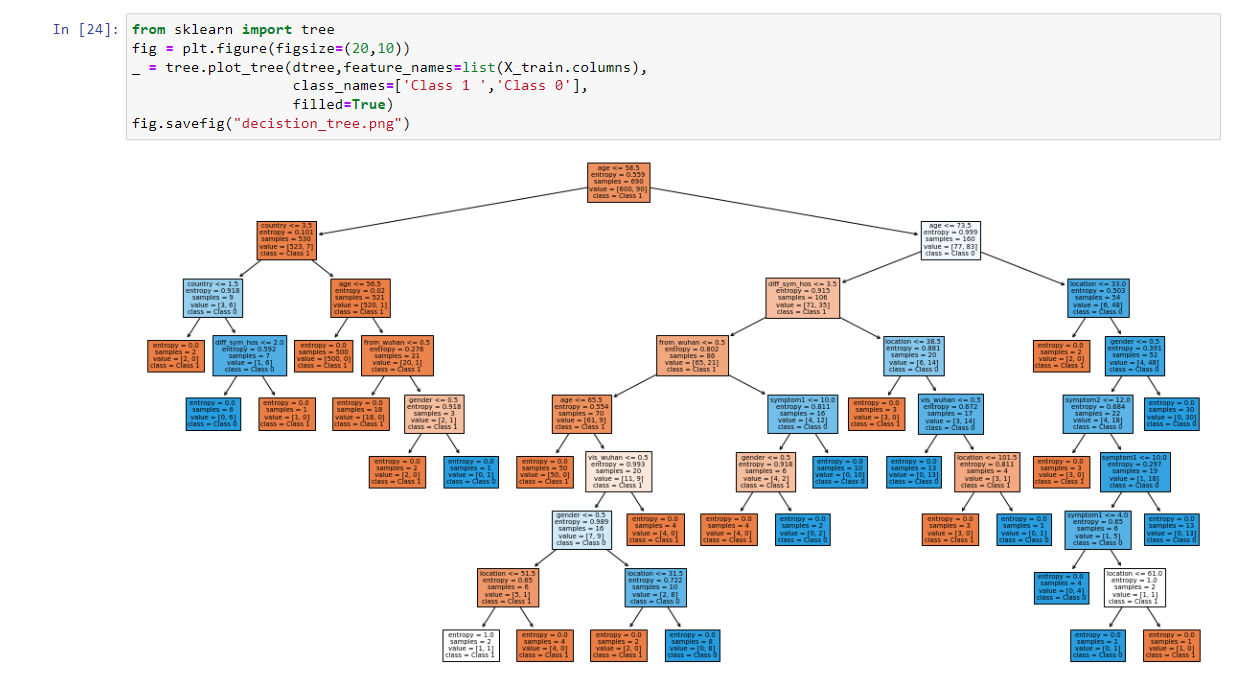
Repeating these steps with the other models (Naïve Bayes , Logistic Regression , Decision Trees , Support Vector Machines ) and validate each

Model before and after finding its optimal hyperparameters

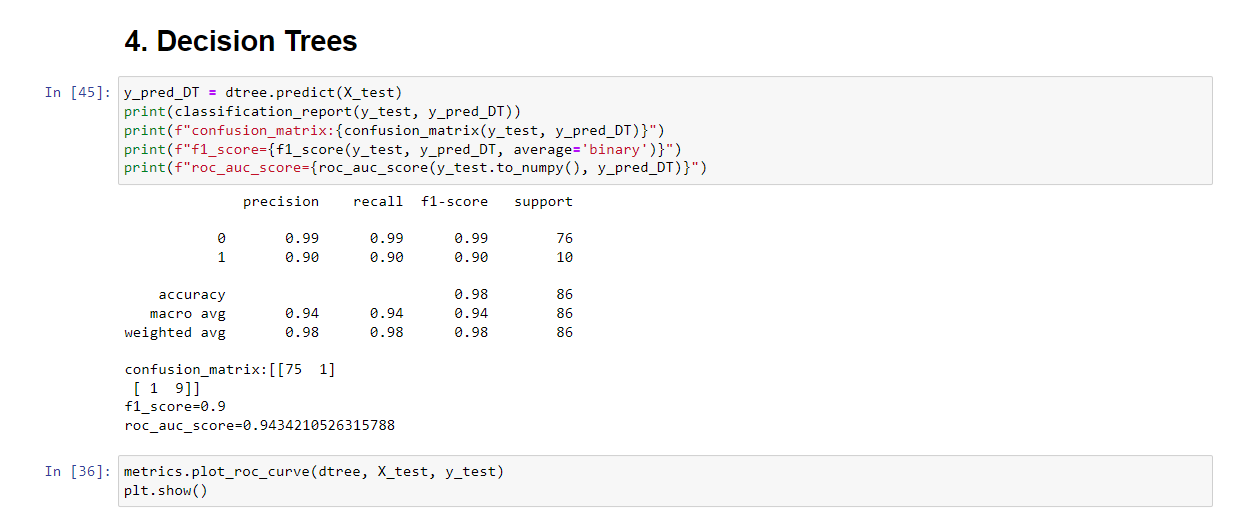




**Plotting the Decision tree**



Now start testing each model with the optimal hyperparameters using the test set and measure its performance with the different metrices



**Comparative analysis:**