

COVID-19 USING COGNOS

IMPLEMENTATION.

1. Methodology

We are using Machine Learning to give predictions on the basis of data taken from government website, and then we clean the data by using excel cleaning methods and give prediction by using the algorithm with highest accuracy to predict COVID -ve or +ve on basis on 5 major symptoms.

The process can be explain in following points

:

1. First, Take the dataset, remove redundant data and organise the data according to our needs.
2. Second, Load the dataset on the Jupyter Notebook and apply data visualization techniques to understand the data better.
3. Third, then we calculate accuracy for various algorithms and plot a graph on the basis of accuracy of various algorithms.
4. Finally, using the accuracy graph we finally use the algorithm with best accuracy in this case (Decision Tree Classifier) to predict the person is either -ve or +ve on the basis of symptoms.

2. Description of the Process

We are building our own COVID Prediction System using Jupyter Notebook.

We can describe the process in following steps :

Step 1: Cleaning the dataset

The very first step in our project is to get a reliable and authentic dataset for the prediction and analysis.

Our search for dataset ended on [11] which is govt website which has provided dataset for free use and is absolutely authentic.

Then next thing we did was to clean the dataset and remove unwanted columns from dataset for faster computation.

Step 2: Data Visualization

Here, we use the dataset and check the consistency of the dataset by checking the values out of the dataset randomly.

Then we do data visualization for better understanding of data by the use of various plots, graph and heatmaps.

All this graphs and plots gets us an insight into huge datasets easily.

Step 3: Computing Accuracy

In this step we compute accuracy of all the algorithms by checking the four algorithms mentioned here: Logistic Regression, KNN, Random Forest Classifier, Decision tree Algorithm , we selected these algorithms on the basis of their qualities of regression & classification.

Step 4: Predicting Covid +ve or -ve

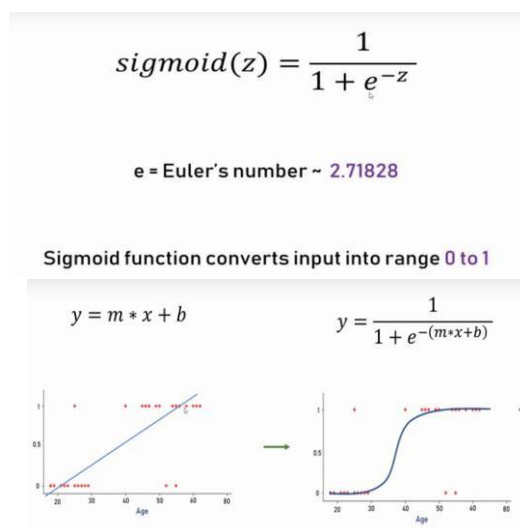
In the last step, all we need to do is plot a graph of accuracy of all the algorithms and use the algorithm with best accuracy to predict whether a person has corona or not.

We take input of 5 symptoms in binary values and using our predictor we predict the person is positive or negative on the basis of these 5 symptoms.

3. Algorithm

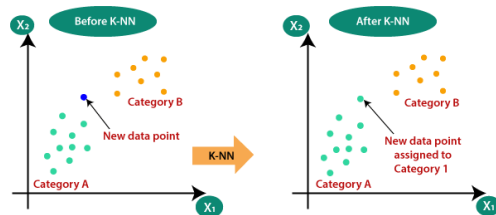
1. Logistic Regression

Logistic Regression is a Classification model, which tries to classify the data based on the probability of it occurring. This algorithm is used in multiple places where classification is required, we have used it to classify if the patient is susceptible to be infected by covid or not. This is one of the classification methods which we have used. It uses Sigmoid function to classify the data.

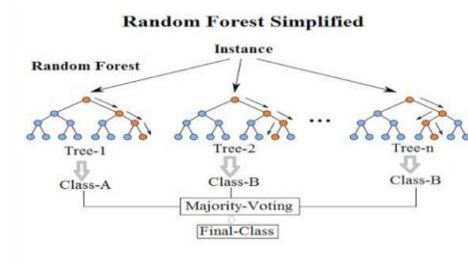


2. KNN

KNN is a supervised machine learning algorithm KNN forms groups based on the criterias and then decides for the incoming data where to put in in which category It can be used for regression and for classification too, but mostly for the classification only its used



3. Random Forest Classifier Random forest is a supervised learning algorithm. The "forest" it builds is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result. Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems



4. Decision tree Algorithm

- Decision Tree is a supervised learning algorithm
- Two nodes which are decision node and leaf node are the ones making the decision
- Repeated if clauses are at work when deciding the classification for the algorithm

