ML V/S COVID19

Problem Statement

Classifying the patient into one of the 3 categories based on the immediate medical help required.

OR

Prioritizing the use of limited medical testing kits to test for COVID'19

Datasets

1. **IndividualDetails.csv:**
   It contains the personal details of each individual patient like gender, location, current condition, etc...
2. **Patient_symptoms.csv:**
   It contains the symptoms the patient suffered from while testing.
3. **Categories.csv:**
   This is the file created by manually giving the categories to each patient.

4. **Dataset.csv:**
   The data required some preprocessing.
   a. The missing values were filled in with the maximum occurring value (mode).
   b. The ages were converted into age groups. Each age group was a bucket of 10 years.
   c. The days required for the status to change i.e. to recover, to be declared as deceased or hospitalised was recorded as bucket of 5 days.
   d. The symptoms were categorized as safe, Suspect, High Risk and so were the notes which had information on travel history and health history.

5. All this conversion is done with the help of **dataset.py**

Model Used

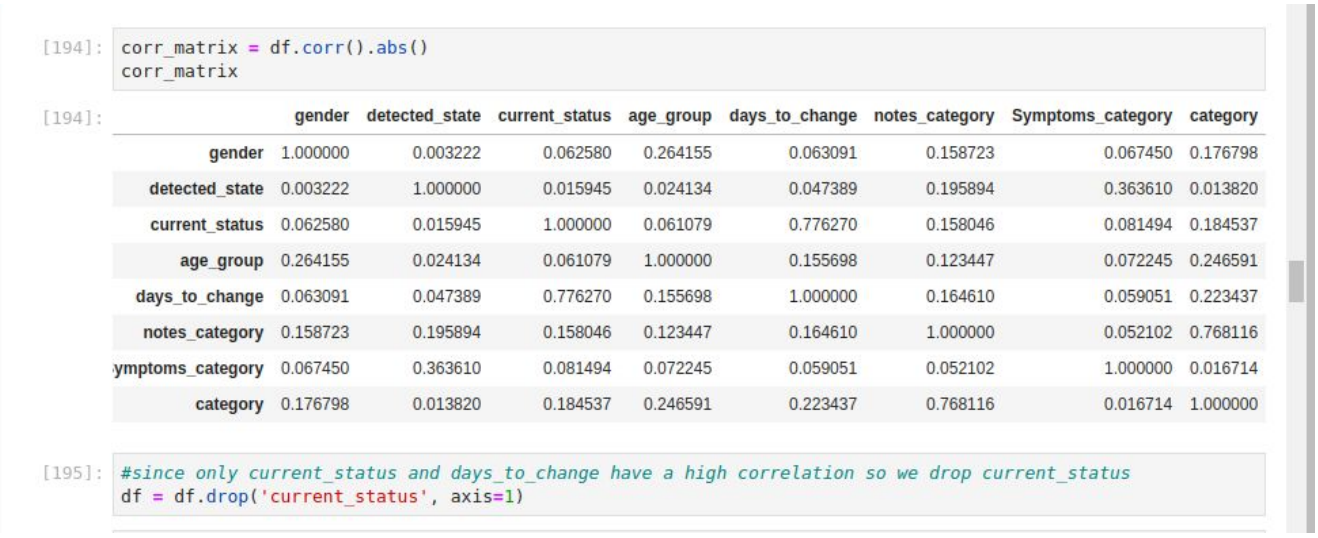
Multinomial Naive Bayes from scikit learn: Link to API docs

We have taken care of the case that p(Xi/y) i.e the probability distribution for each random variable can be a multinomial distribution as well. Hence we have used the implementation of Multinomial Naive Bayes classifier using scikit learn library of python.

Note Xi is the feature variable which in our case includes gender, age,notes_category, etc… and y variable denotes the target category.

Bias :

This model assumes that the data has features that are independent of each other. Hence we performed the Pearson Correlation test and got the following output.
This shows all the variables are independent of each other except for days_to_change and current_status.
Hence we drop current_status.



```
[194]: corr_matrix = df.corr().abs()
       corr_matrix
```

[194]:

	gender	detected_state	current_status	age_group	days_to_change	notes_category	Symptoms_category	category
gender	1.000000	0.003222	0.062580	0.264155	0.063091	0.158723	0.067450	0.176798
detected_state	0.003222	1.000000	0.015945	0.024134	0.047389	0.195894	0.363610	0.013820
current_status	0.062580	0.015945	1.000000	0.061079	0.776270	0.158046	0.081494	0.184537
age_group	0.264155	0.024134	0.061079	1.000000	0.155698	0.123447	0.072245	0.246591
days_to_change	0.063091	0.047389	0.776270	0.155698	1.000000	0.164610	0.059051	0.223437
notes_category	0.158723	0.195894	0.158046	0.123447	0.164610	1.000000	0.052102	0.768116
ymptoms_category	0.067450	0.363610	0.081494	0.072245	0.059051	0.052102	1.000000	0.016714
category	0.176798	0.013820	0.184537	0.246591	0.223437	0.768116	0.016714	1.000000

```
[195]: #since only current_status and days_to_change have a high correlation so we drop current_status
       df = df.drop('current_status', axis=1)
```

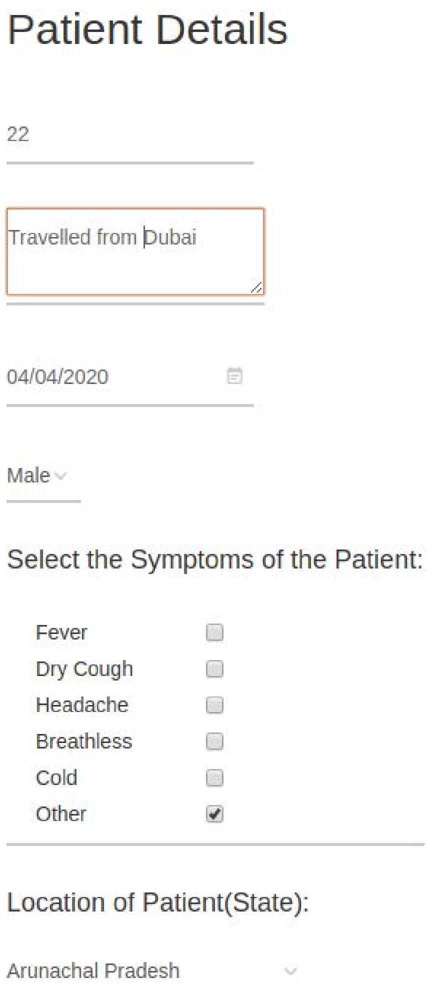
The Classifier

Bayes_Classifier/bayes_classifier.py: It has the code for training and testing the model. Multinomial Naive Bayes is **trained using 80%** of the total data and **testing** is done using the remaining **20%**.
The **accuracy** turns out to be **82.3%.**

My_dumped_classifier.pkl: It is the dump created after training the model. Hence this will be loaded when we need to predict.

The Platform

The web platform is built using Flask (python web framework).
The user enters his/her details and is allotted the category based on which the user can proceed appropriately.



Patient Details

22

Travelled from Dubai

04/04/2020

Male

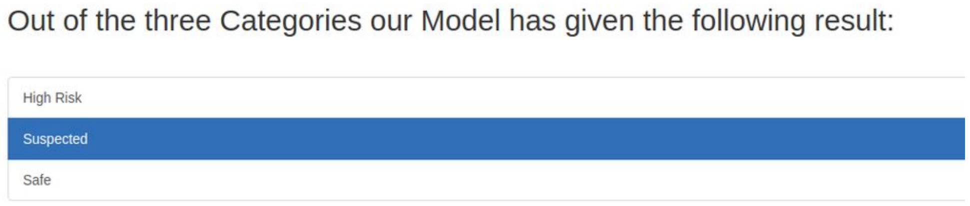
Select the Symptoms of the Patient:

Fever	□
Dry Cough	□
Headache	□
Breathless	□
Cold	□
Other	☑

Location of Patient(State):

Arunachal Pradesh

Final Results are shown as follows:



Out of the three Categories our Model has given the following result:

High Risk

Suspected

Safe

Installation guide

To execute using the web Platform

1. Install the prerequisite python packages.
    a. Flask
    b. Sci-kit learn

2. Clone the repository from GitHub:
   https://github.com/ritwik-deshpande/ML-vs-COVID-19

3. Navigate to the ML v/s COVID19 folder
4. Run the **server.py** file
5. Log on to the http://127.0.0.1:5000/ to access the web platform

To execute directly from the terminal

1. Navigate to the ML-vs-COVID-19 directory to the directory where naive_bayes is stored.
2. Run the **naive_bayes.py** file.
3. Enter the required input. You can take help from the Sample Inputs file.