ML MODEL FOR DETECTING THE CHANCES OF COVID

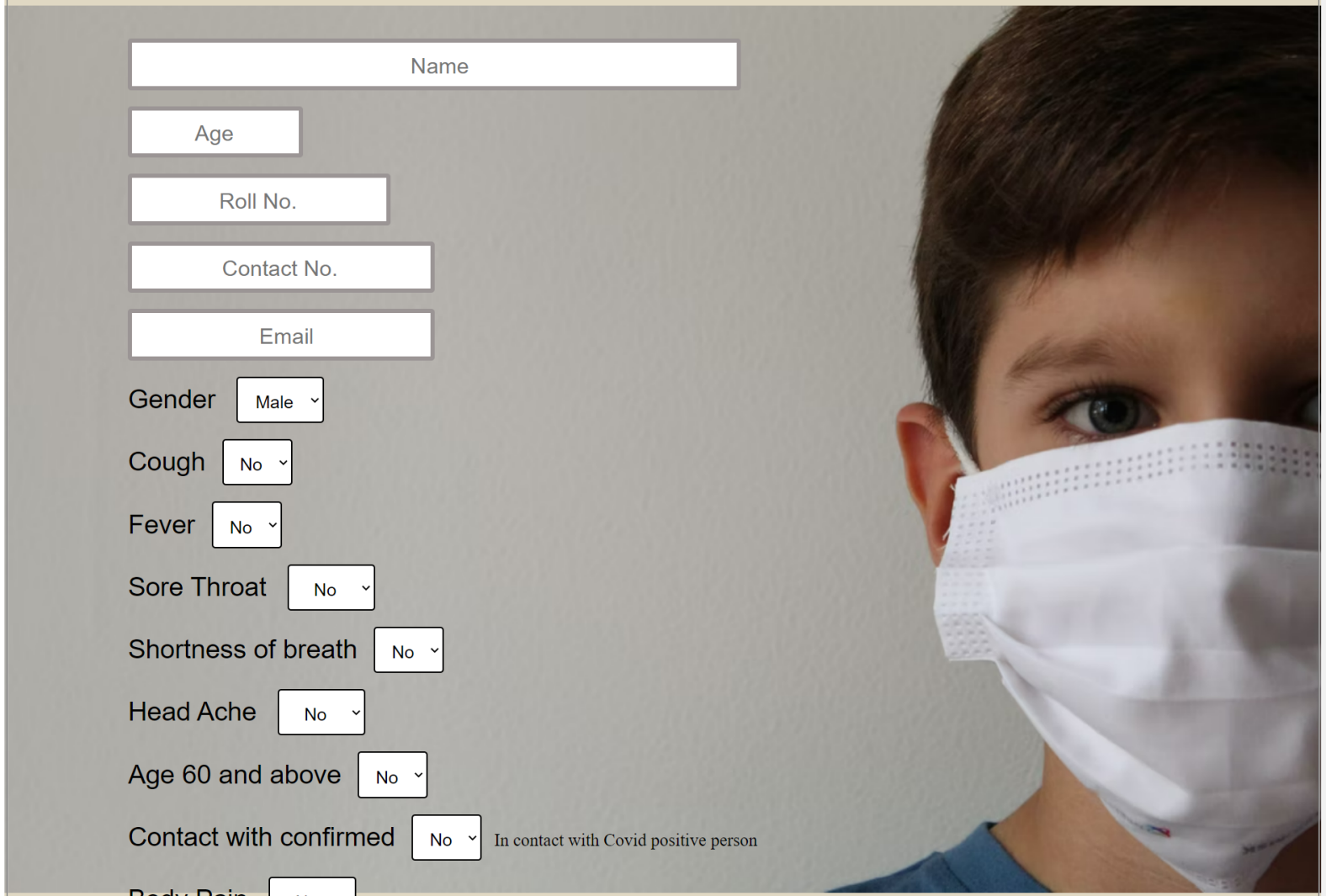
Sachin Gautam (CSE – 1st year)


COVID IITG
BE NEGATIVE IN CORONA , BE POSITIVE IN LIFE
Name
Age
Roll No.
Contact No.
Email
Gender
Male
Cough
No
Fever
No
Sore Throat
No
Shortness of breath
No
Head Ache
No
Age 60 and above
No
Contact with confirmed
No
In contact with Covid positive person
Body Pain
No
Submit

WEB FORM

COVID IITG

NEGATIVE IN CORONA , BE POSITIVE IN LIFE



Name

Age

Roll No.

Contact No.

Email

Gender Male

Cough No

Fever No

Sore Throat No

Shortness of breath No

Head Ache No

Age 60 and above No

Contact with confirmed No In contact with Covid positive person

Body Pain No

Submit

MOBILE-VERSION

DATATSET FOR ML MODEL

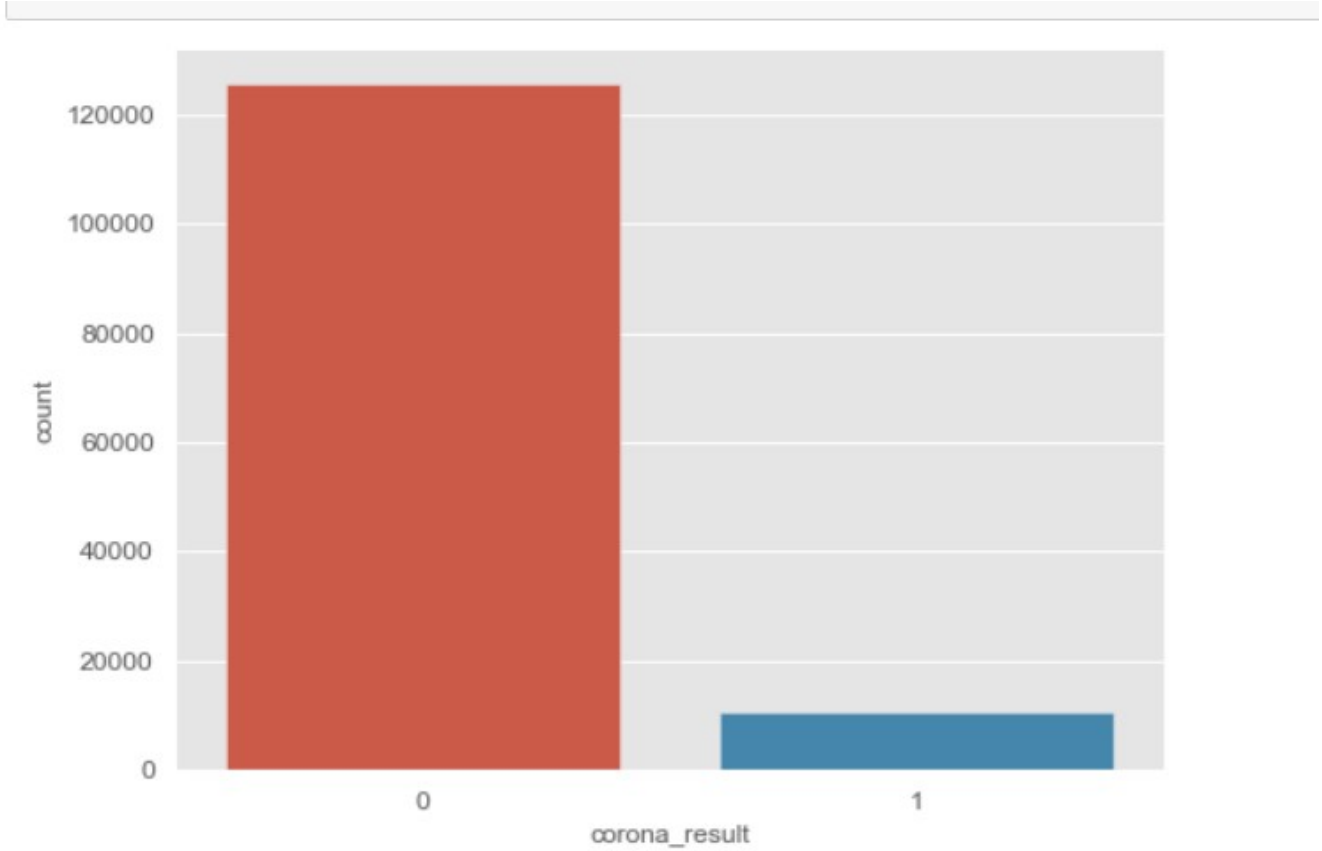
We use dataset of 2,00,000 people (included in file) person.
First we use Supervised Machine Learning Algorithms to train our data

DATA ANALYSIS

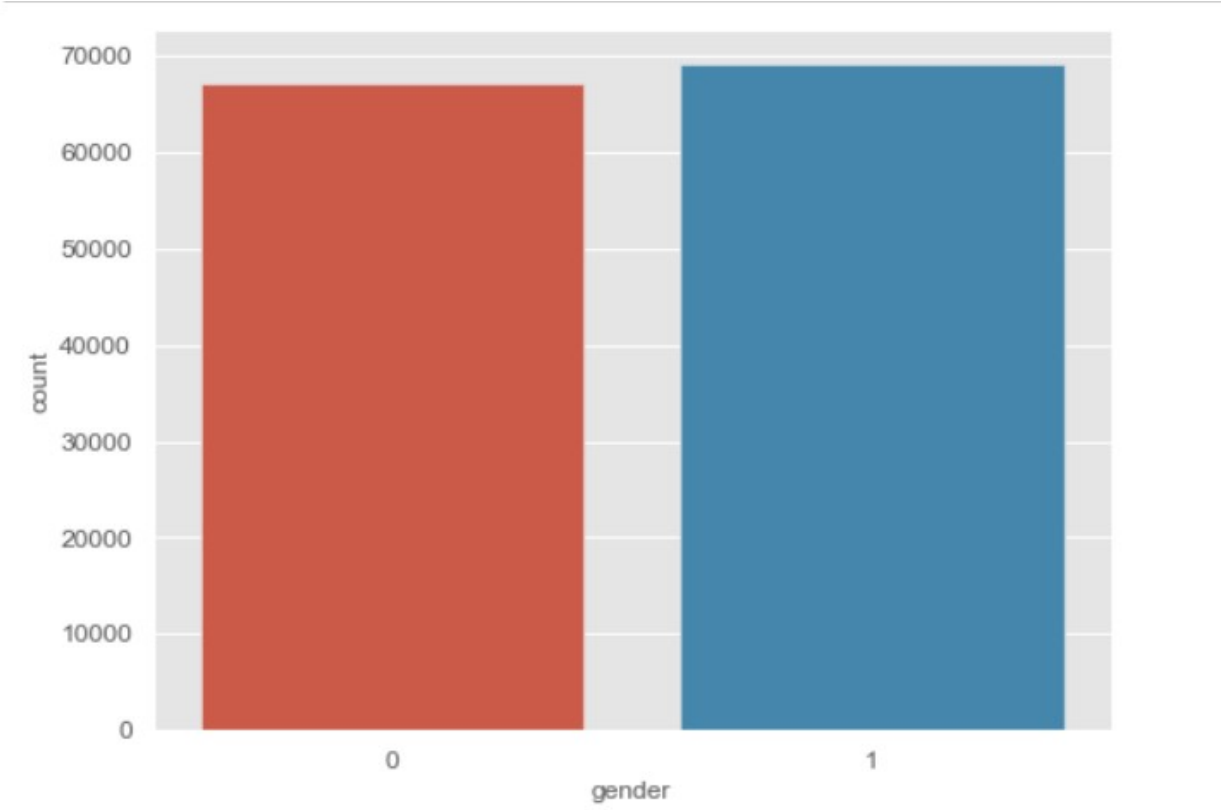
For making a ML it is imp of analyse the data carefully , because sometime due to poor quality of data model would not perform accurately as we wanted.
For making data useful for our ML model we used Data wrangling technique to clean our data.

We tackle this problem as a CLASSIFICATION problem.

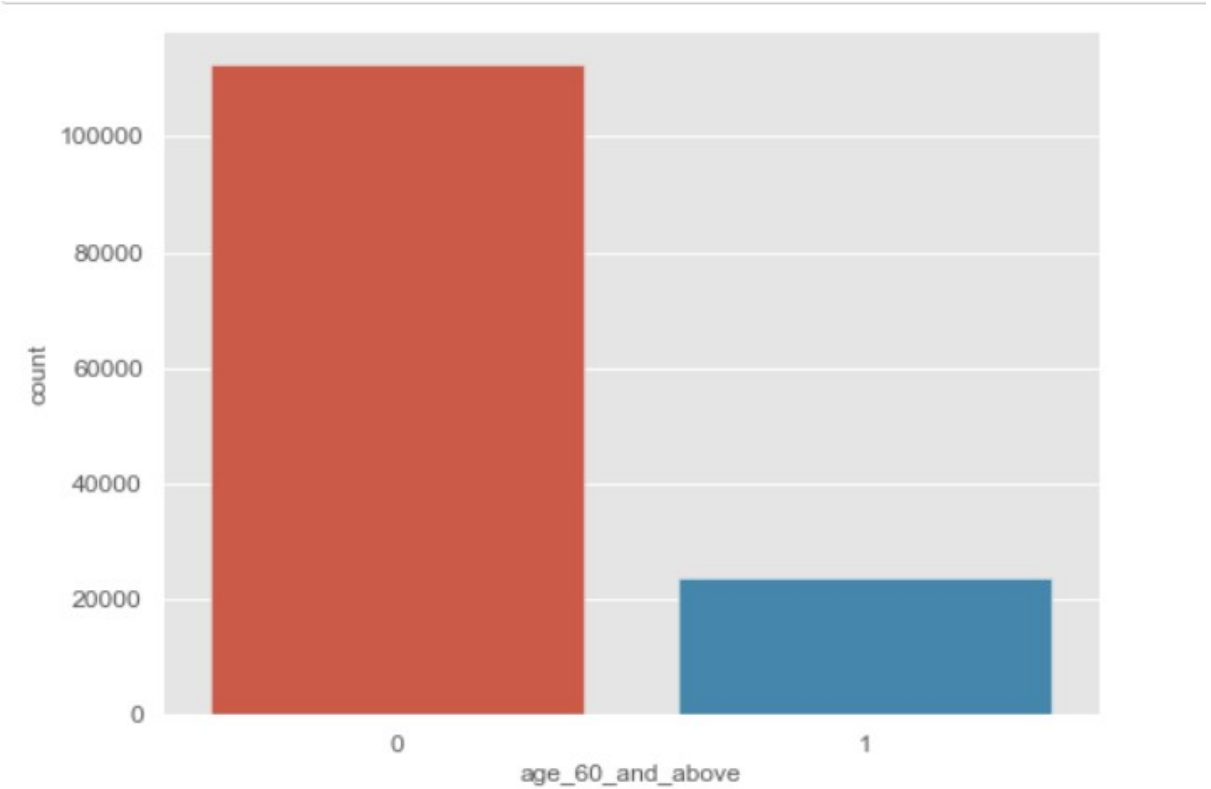
PLOT OF COVID VS NO OF PERSON (FROM DATASET)



MALE AND FEMALE IN DATASET



PEOPLE AGE >=60 IN THE DATASET



CONCLUSION

After analysing dataset precisely .

We use different algorithm to create a model which predict the accurate and precise chances if the person is suffering from COVID or not.

Featured that are used to predict the result are

1-Cough 2-Fever 3-Sore Throat 4- Shortness of breath

5-Headache 6-Age 60 and above 7-Gender

8-Conatact with confirmed(It states that a person is in Contact with the person which is COVID positive)

9-Body Pain

TRAINING MODEL

For getting greater accuracy and precision of the model .

We divided the dataset in 4:1 ratio 4 for part training purpose and 1 part for testing.

We use multiple algorithm for achieving a highly accurate model.

Dataset	Algorithm	Precision	Recall	F1 Score	AUC	Log Loss
Age (0–20)	XGBoost	0.89	0.94	0.91	0.85	4.14
	GBM	0.89	0.96	0.92	0.86	3.68
	SVM	0.92	0.98	0.95	0.91	2.30
	Random Forest	0.90	0.92	0.91	0.86	4.14
	Decision Tree	0.88	0.89	0.89	0.85	3.68
Age (21–60)	XGBoost	0.95	0.92	0.93	0.87	3.57
	GBM	0.98	0.86	0.91	0.89	4.35
	SVM	0.98	0.86	0.91	0.89	4.45
	Random Forest	0.95	0.91	0.93	0.87	3.74
	Decision Tree	0.94	0.93	0.93	0.89	3.74
Age (61–96)	XGBoost	0.87	0.90	0.88	0.82	5.42
	GBM	0.90	0.87	0.88	0.84	5.25
	SVM	0.93	0.80	0.86	0.84	5.93
	Random Forest	0.89	0.88	0.89	0.84	5.08
	Decision Tree	0.88	0.89	0.88	0.83	5.25
Age (0–96)	XGBoost	0.93	0.91	0.92	0.85	4.08

ALGORITHM OF THE FINAL MODEL

We use a combination of RandomForest with Adaboost Classifier

which has a accuracy up to 95%

* TRAINED MODEL FILE is included
* AdaBoost is used to boost the performance of ML Model

WEBPAGE

Link

https://covidiitgapp.herokuapp.com/ - Mobile version

https://covidiitg.herokuapp.com/ -Desktop/PC version

Web-form are Live now .

Click on the links to go there.

HOW WEBPAGE WORKS

We create a website form using Flask and python .

We deploy our Machine Learning(ML) model with it.

Every time you go to the website it take input(features)

and in the background model runs and predict the results.

HOW IT IS USEFUL

People with mild symptoms who are otherwise healthy , on average it takes

5-6 days from when someone is infected with virus for symptoms to show ,

However it can take up to 14 days.

By daily checking on webform people get their daily possibility that they are

suffering from COVID or not , from regular and daily result people can contact to the doctors

who needs by pre-predictions from the web-form.

And also in beginning stage people know that they are suffering from COVID help

them to cure as fast as possible.

ANOTHER PROTOTYPE ON WHICH WE ARE WORKING

We create a database where the data of the input information from the Web-form is get collected .

Everyone in the collage needs to fill the web-form daily.

We classify the person on the basis of this data which one need treatment fast .

** After we start collecting data from the collage we use this data to train the model.

After getting data everyday we regularly train the data

Day by day model become more accurate and precise.

And classify the people who need treatment fast.

On the basis of their symptoms(features)

DATABASE USED

We are trying to create SQL database for storing the daily data.

** SQL DATABASE ML PROGRAMME IS INCLUDED IN FILE

(WORK IN PROGRESS)