CHANDNI KUMARI
4NM18IS030

PROJECT REPORT

COVID-19 Vaccination in India

India owns the record as the country having the largest vaccine manufacturing company in the world, the Serum Institute of India. Amidst the development of covid-19 vaccine, it was even reported that SII was targeted by a cyber-espionage group for exfiltration recently. Being a highly populated and developing country, it is important to study the ongoing impact of the pandemic and how well India is able defend it by the largest vaccination drive in the world that began on 16 January 2021. The role of engineers is significant in learning the huge data effectively and this is indeed important to bring the pandemic to an end.
The objective is to study the vaccination in India in detail. I have collect relevant data in national level constituting the states/districts/constituencies and analysed the data and put forward important findings.

Data

There are open source data available, one relevant source of data is cowin.gov.in The data you use should help in the study of development of vaccination program in India. The data may or may not be specific to: gender, occupational level, age category, people with lung related diseases, other major health risks, diabetes patients, people with hypertension, people with chronic kidney disease, people with cardiac ailments etc., Are cancer patients given vaccines or can they be? What is the effect of diet in the efficacy of vaccine in a person? Besides, you should definitely include the data related to the number of people who received the vaccine, 2 doses, just 1 dose, per cent of population vaccinated in the country, states, districts, constituencies or in any other relevant scale. Cowin vaxin data has been used.

The dataset has been collected from the link given below which gives the latest dataset
http://api.covid19india.orrg/csv/latest/cowin_vaccine_data_districtwise.csv

Algorithm

- Data collection
- Data visualization
- Data cleaning
- Algorithm used to implement the model
- Testing
- Inference and Conclusion

Implementation

Data Collection

First, we will import libraries and the dataset. The dataset we'll use for this project- we'll call it cowin-2.csv .

Required packages:(pandas ,matplotlib,seaborn)

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

Basic Analysis

Next step is to analyze the data using functions like head(), describe(), info(), etc.

Data Cleaning

Removing columns which have no significance in the prediction process like ID or S.No .

#Removing S No and COWIN ID because they have no

significance in the analysis

Data Visualization (Note that the graphs are plotted for values of a single day)

1. Which state has the highest number of vaccinated people?
2. Gender wise distribution of vaccinations

What are the different vaccines used by different districts?

Linear Regression Algorithm

After cleaning and visualizing the data we will move onto the selection, training, and testing of the

algorithm- Linear regression .

Training(or learning part):

Fit the training data to the algorithm

Model_Name.fit(Training)

Testing

(check the efficiency of the algorithm):

This step includes -

1.Predicting the outcomes of new data

Model_Name.predict(Features-of-testing-set)

2.Checking the accuracy of the

algorithm(testing set)

Model_Name.score(Arrays-of-testing-set)

Inferences and Conclusion

:like which state got more vaccines

Screen Shot

```
In [11]: valid_data
```

Out[11]:

	16/01/2021.1	17/01/2021.1	18/01/2021.1	19/01/2021.1	20/01/2021.1	21/01/2021.1	22/01/2021.1	23/01/2021.1	24/01/2021.1	25/01/2021.1	...	20/05/2021.1	21/0
1	0	0	0	0	0	0	0	3	5	7	...	350	
2	0	0	0	0	0	6	11	11	15	15	...	1460	
3	2	5	9	12	16	16	18	18	18	21	...	3802	
4	28	87	122	197	284	364	443	497	517	529	...	8400	
5	63	92	136	184	296	492	552	646	741	754	...	7130	
...	...	...	...	...	...	...	...	...	...	...	...	...	
751	1	16	17	27	55	67	69	84	92	94	...	15660	
752	0	5	10	10	12	18	20	24	30	30	...	1054	
753	6	14	25	25	45	47	54	61	65	78	...	7393	
754	NaN	7	16	22	35	46	43	50	50	62	...	3740	
755	2	8	14	14	59	76	84	84	84	87	...	2277	

755 rows × 134 columns

```
In [12]: valid_data = valid_data.dropna()

In [13]: # Spiit data into independent and dependent variable
         X = valid_data.iloc[:, :-1]
         y = valid_data.iloc[:, -1]
```

Figure 1: Dataset

```
In [46]: sns.stripplot(data['16/01/2021.1'], data['17/01/2021.1'])
Out[46]: <matplotlib.axes._subplots.AxesSubplot at 0x1a5762f6908>
```

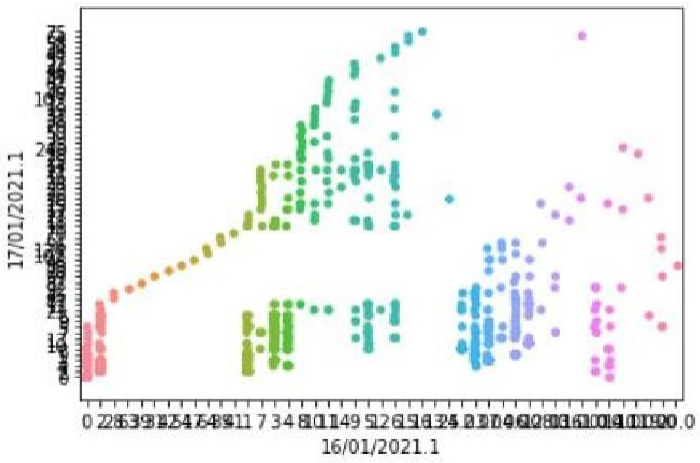

Figure 2: Data Visualization

```
In [47]: sns.countplot(data['State_Code'])

Out[47]: <matplotlib.axes._subplots.AxesSubplot at 0x1a578ac1d88>
```

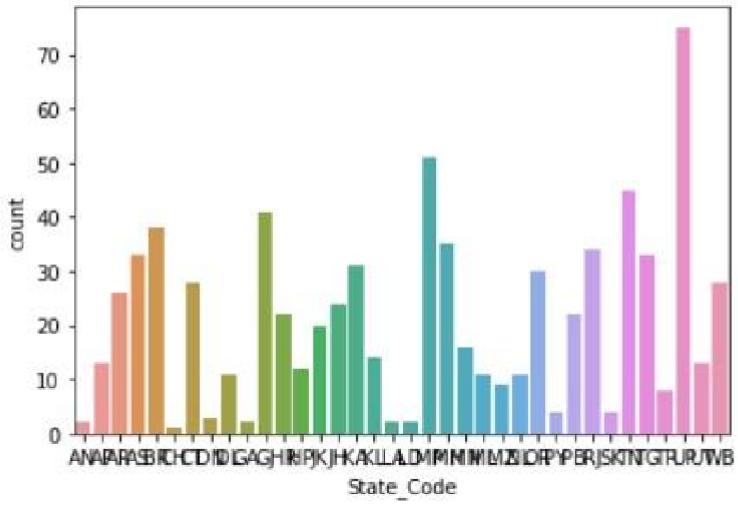

Figure 3: Data Visualizaion-2

In [35]:
```
# Total CoviShield Administered
x = []
y = 15
for i in range(139):
    x.append(y)
    y += 10
valid_data = data.iloc[:, x]
valid_data.drop('Unnamed: 1355', axis=1, inplace=True)
valid_data.drop('Unnamed: 1365', axis=1, inplace=True)
valid_data.drop('Unnamed: 1375', axis=1, inplace=True)
valid_data.drop('Unnamed: 1385', axis=1, inplace=True)
valid_data.drop('Unnamed: 1395', axis=1, inplace=True)
valid_data.drop(0, axis=0, inplace=True)
valid_data = valid_data.dropna()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\pandas\core\frame.py:3997: SettingWith
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stabl
rsus-a-copy
  errors=errors,
```

In [36]:
```
X_train = valid_data.iloc[:, :-2]
y_train = valid_data.iloc[:, -2]
X_test = valid_data.iloc[:, 1:-1]
y_test = valid_data.iloc[:, -1]
```

In [37]:
```
linreg.fit(X_train, y_train)
```

Out[37]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [38]:
```
my_pred = linreg.predict(X_test)
```

In [39]:
```
my_score = linreg.score(X_test, y_test)
```

In [40]:
```
print(my_score)
```

```
0.9869614772895904
```

Figure 4: Accuracy of model

Observation

We are able to get the accuracy of about 98.69% in predicting what number of people will get vaccinated in future day. This will helpful in such a way that, government can prepare that much of vaccine for next day. This is very helpful in this current situation.