COVID-19 Vaccines Analysis

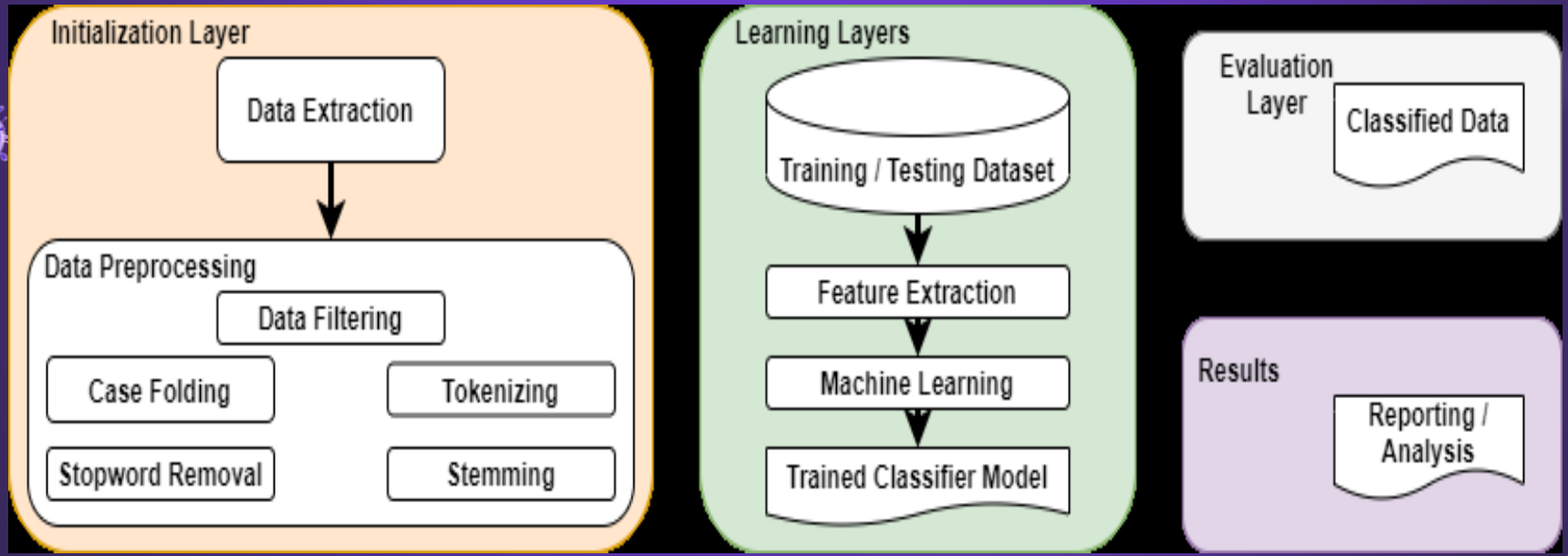
Name-D.SACHIN KUMAR
Gmail-sachinkumarmtp333@gmail.com
Collage name-Sri Ranganathar Institute of Engineering and Technology(SRIET)
Course Name-Applied data science

Phase-II

Innovation

Framework

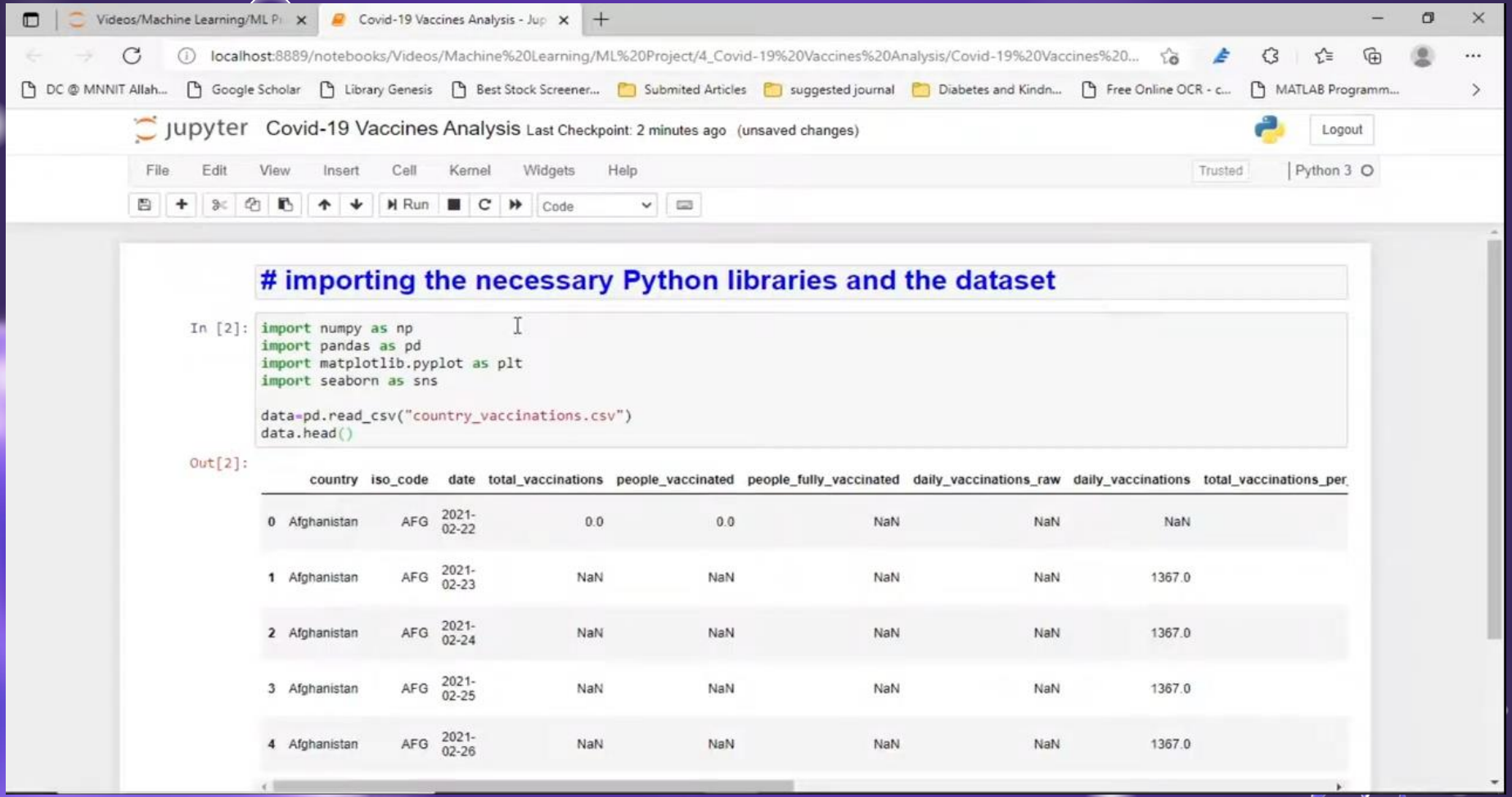
This research presents a framework for sentiment analysis of COVID-19 vaccines. We have used python as a programming language and several libraries for text mining that will be explained.



The innovation to solve the problem

- Importing the necessary Python libraries and the dataset
- Pre process the Data
- Prepare the Data
- Machine learning Techniques.

Importing the necessary Python libraries and the dataset



	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	
2	Afghanistan	AFG	2021-02-24	NaN	NaN	NaN	NaN	1367.0	
3	Afghanistan	AFG	2021-02-25	NaN	NaN	NaN	NaN	1367.0	
4	Afghanistan	AFG	2021-02-26	NaN	NaN	NaN	NaN	1367.0	

Videos/Machine Learning/ML P | Covid-19 Vaccines Analysis - Jup

localhost:8889/notebooks/Videos/Machine%20Learning/ML%20Project/4_Covid-19%20Vaccines%20Analysis/Covid-19%20Vaccines%20...

DC @ MNNIT Allah... | Google Scholar | Library Genesis | Best Stock Screener... | Submited Articles | suggested journal | Diabetes and Kindn... | Free Online OCR - c... | MATLAB Programm...

jupyter Covid-19 Vaccines Analysis Last Checkpoint: 8 minutes ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run | Code

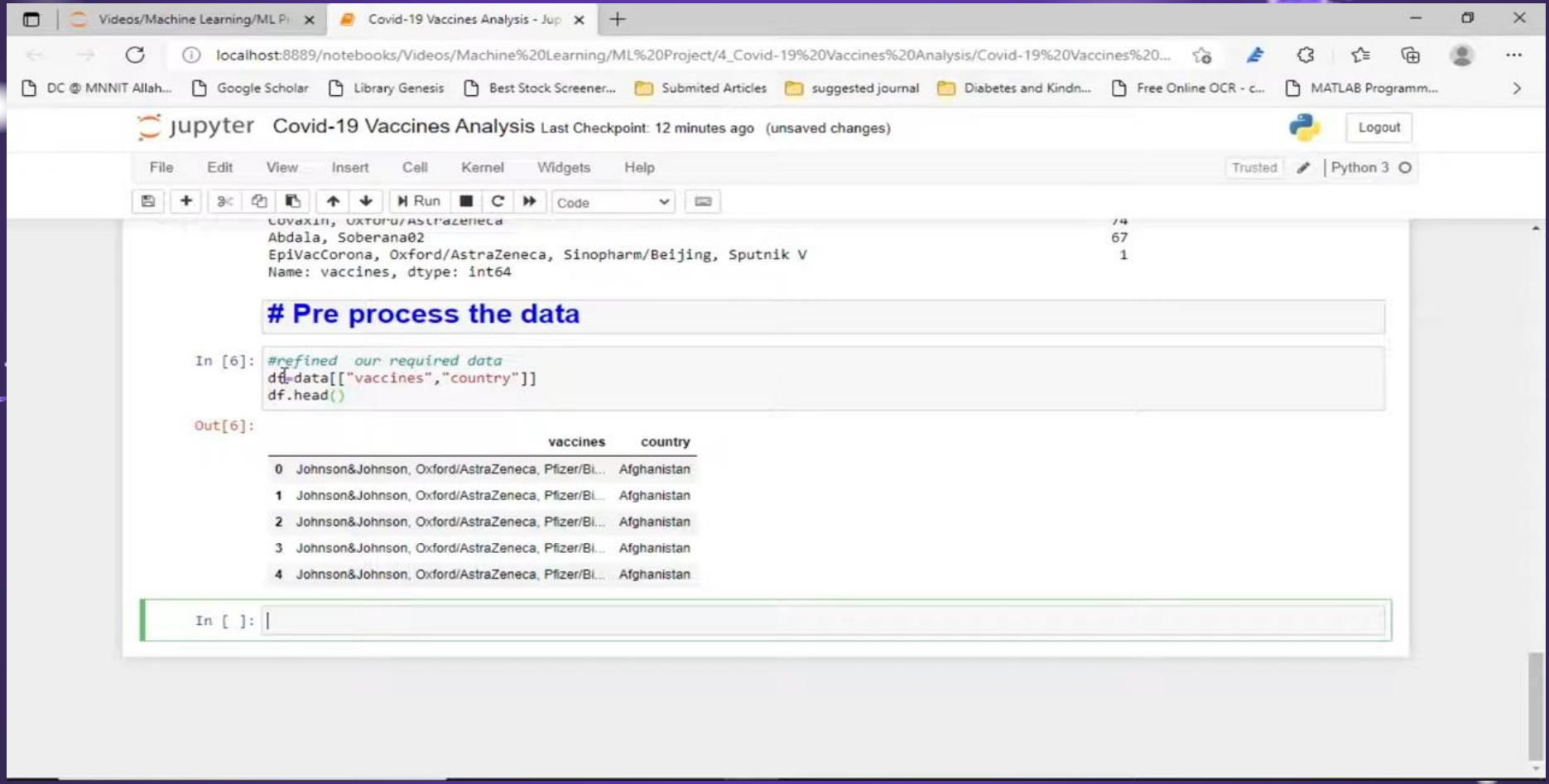
25%	1.379360e+05	1.110202e+05	5.200275e+04	4.507500e+03	8.410000e+02	3.310000
50%	9.802550e+05	7.017030e+05	4.002405e+05	2.213300e+04	6.451500e+03	16.390000
75%	4.722680e+06	3.128880e+06	1.853223e+06	9.859150e+04	3.682175e+04	49.110000
max	1.565872e+09	6.220000e+08	2.232990e+08	2.474100e+07	2.242429e+07	232.720000

In [4]:
```
data.country.value_counts()
```

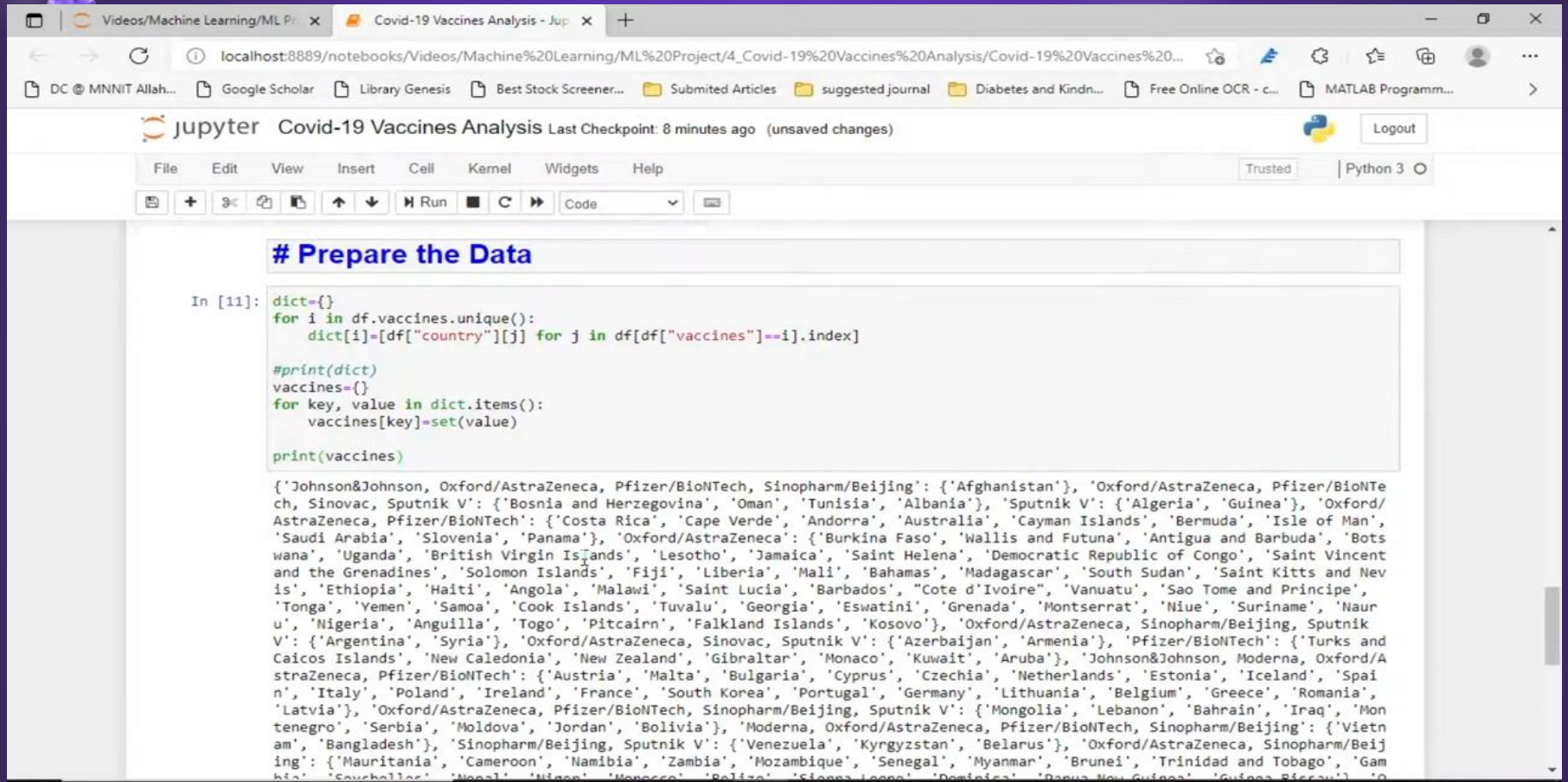
Out[4]:
```
Norway                          236
Latvia                          235
England                         230
Scotland                        230
Wales                           225
                               ...
Niue                             16
Haiti                             7
Pitcairn                          7
Bonaire Sint Eustatius and Saba   1
Turkmenistan                      1
Name: country, Length: 219, dtype: int64
```

In [ ]:

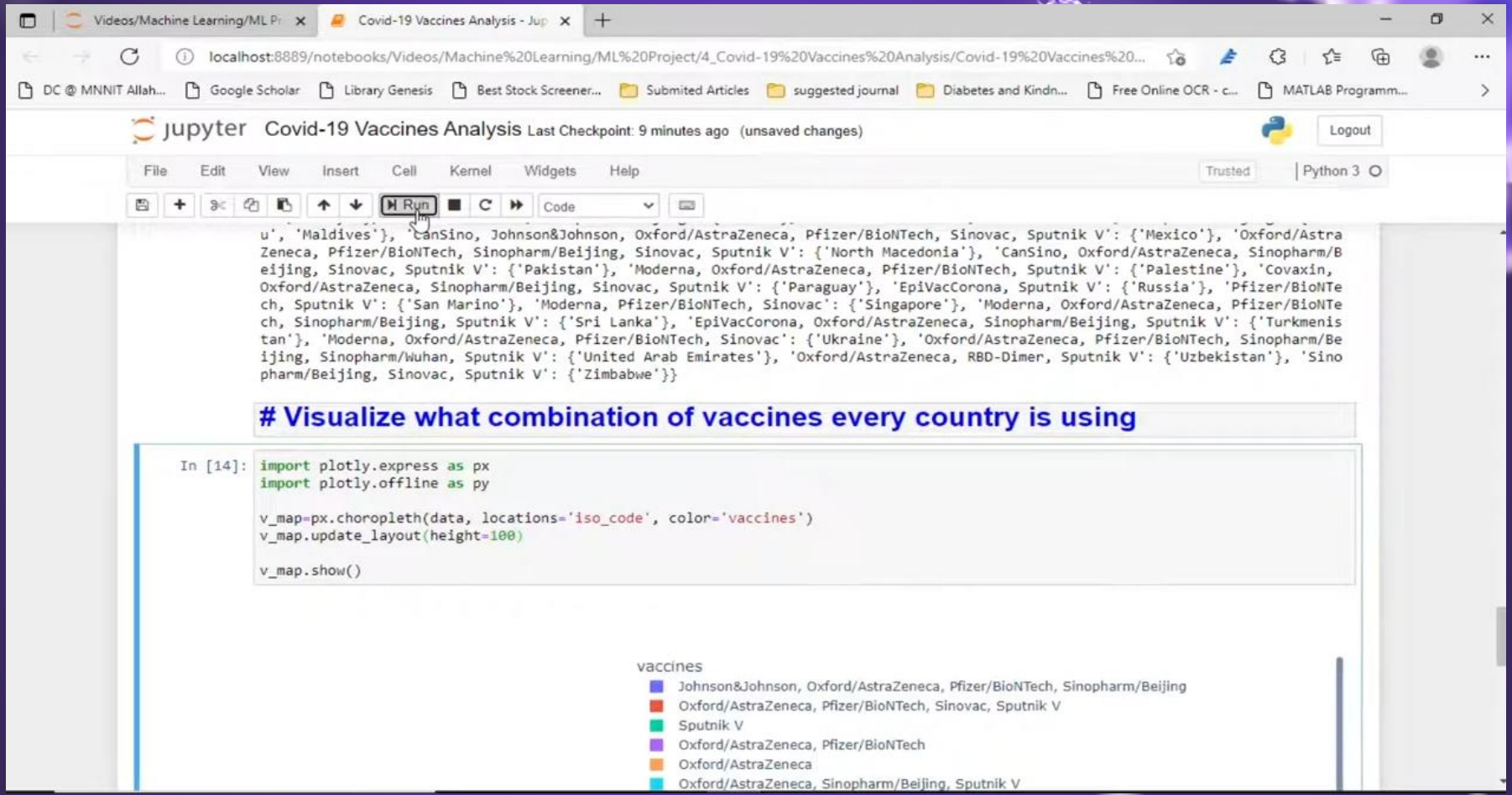
Pre process the data



Prepare the Data



Visualize what combination of vaccines every country is using



Videos/Machine Learning/ML Pr | Covid-19 Vaccines Analysis - Jup

localhost:8889/notebooks/Videos/Machine%20Learning/ML%20Project/4_Covid-19%20Vaccines%20Analysis/Covid-19%20Vaccines%20...

DC @ MNNIT Allah... | Google Scholar | Library Genesis | Best Stock Screener... | Submited Articles | suggested journal | Diabetes and Kindn... | Free Online OCR - c... | MATLAB Programm...

jupyter Covid-19 Vaccines Analysis Last Checkpoint: 12 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [16]:
```
import plotly.express as px
import plotly.offline as py

v_map=px.choropleth(data, locations='iso_code', color='vaccines')
v_map.update_layout(height=500)

v_map.show()
```

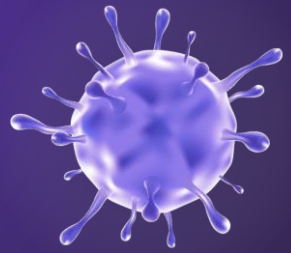
Qazvac, Sinopharm/Hayatvax, Sputnik V
Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V
Johnson&Johnson, Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V
Pfizer/BioNTech, Sinopharm/Beijing
Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
CanSino, Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V
Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinovac, Sputnik V
CanSino, Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V
Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sputnik V
Covaxin, Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V
EpiVacCorona, Sputnik V
Pfizer/BioNTech, Sputnik V
Moderna, Pfizer/BioNTech, Sinovac
Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sputnik V
EpiVacCorona, Oxford/AstraZeneca, Sinopharm/Beijing, Sputnik V
Moderna, Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac
Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing, Sinopharm/Wuhan, Sputnik V
Oxford/AstraZeneca, RBD-Dimer, Sputnik V
Sinopharm/Beijing, Sinovac, Sputnik V

Machine learning

Machine learning (ML) is a popular use of artificial intelligence since it automates the system and allows it to learn and improve from diverse experiences without being programmed. Computer programs can teach how to learn by giving them access to data and allowing them to utilize it for learning in ML. The learning process in ML begins with seeing the data through examples or instructions that humans offer; these observations enable ML to look for patterns in order to make the best predictions. Five different ML models were used to train the classifier and evaluate classification performance using the test dataset. These are discussed below.

Machine learning Techniques

- ❑Random Forest
- ❑Naive Bayes
- ❑Decision Tree
- ❑Logistic Regressions
- ❑Support Vector Machine



Random Forest

The RF model is an ensemble model that generates high-precision predictions by combining the results obtained from several sub-trees. The supervised ML method known as RF may be used for both classification and regression analysis

An RF can be represented as:

$$RF = mode\{tR_1, tR_2, tR_3, \cdots, tR_n\}$$

$$RF = mode\{\sum_{i=1}^{n} tR_i\}|$$

where tR1, tR2, tR3,... , tRn represent the Decision Trees in RF and n denotes the number of trees.

Naive Bayes

The Bayes Theorem's premise of class conditional independence is used in the NB classification technique. This indicates that the existence of one characteristic in the likelihood of a certain event has no bearing on the presence of another, and each predictor has an equal impact on the outcome. Multinomial NB, Bernoulli NB, and Gaussian NB are the three kinds of NB classifiers. Text categorization, spam detection, and recommendation systems are all applications of this technology.

An NB can be represented as:

$$P(A\backslash B)=\frac{P(B\backslash A)\ \ P(A)}{P(B)}$$

Decision Tree

DTs are a technique for non-parametric supervised learning that may be used for classification and regression. DT is a model for ML that may be used for the problem solving process of regression as well as classification. The purpose of this project is to build a model that can accurately forecast the value of a target variable by gleaning fundamental decision rules from the features of the data. A DT with multiple branches of varying sizes is used in conjunction with partitioning the dataset into an incremental method of construction.

Logistic Regression

Logistic Regression is a statistical approach to data analysis in which one or more variables are utilized to determine the outcome. When the target variable is categorical, the optimum learning model to utilize is LR, which is the regression model that was used to estimate the likelihood of class members. Linear Regression uses a logistic function to estimate probabilities for the association between the categorical dependent variable and one or more independent variables.

Support Vector Machine

A support vector machine(SVM), which was created by Vladimir Vapnik, is a supervised learning model that can be used to both classify and regress data [51]. On the other hand, the most popular use for it is in the realm of classification problems; in this context, it is used to generate a hyperplane on which the distance between two classes of data points is maximized.

Types of SVM

-> Linear SVM

-> Nonlinear SVM

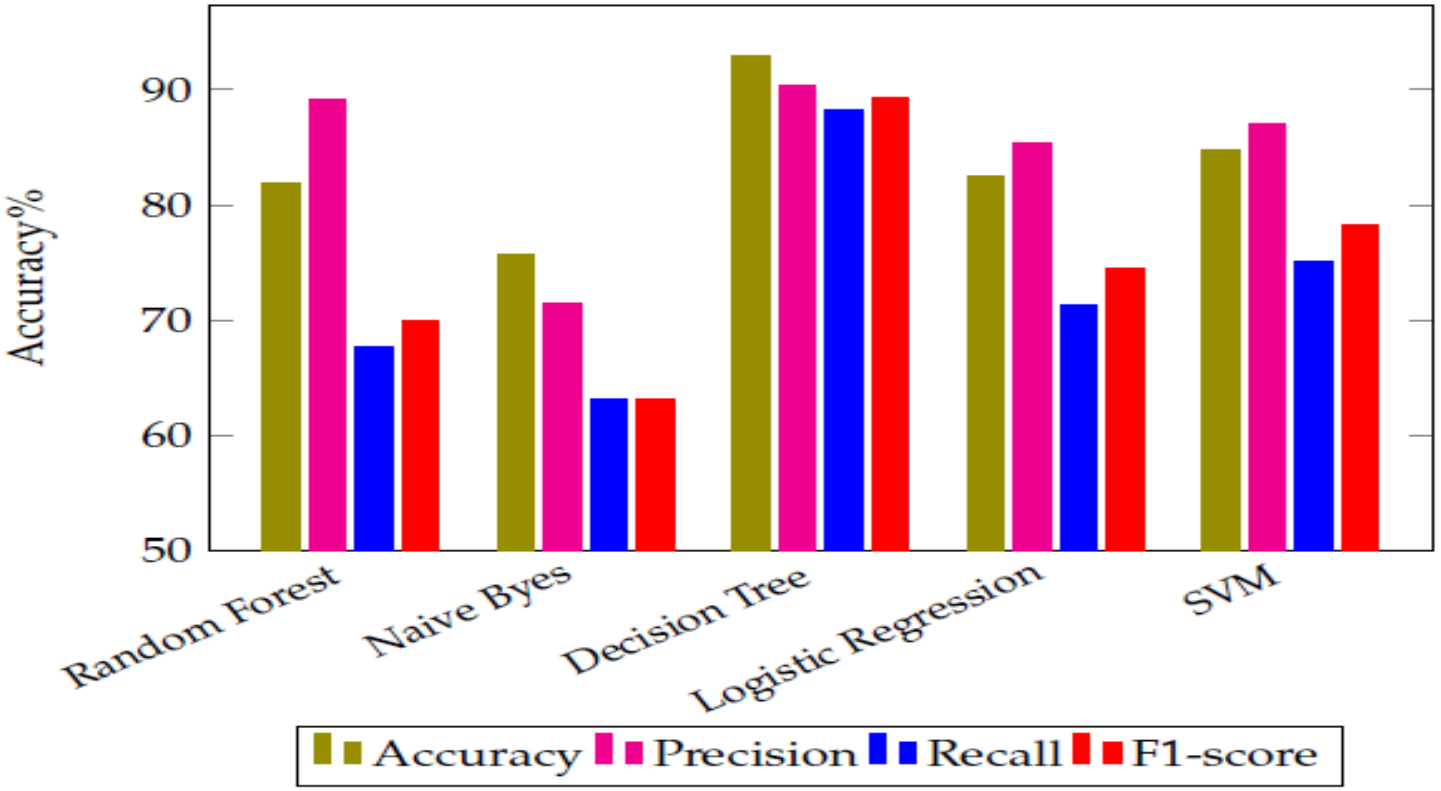
Results and Discussion

This section presents the accuracy results of sentiment analysis carried out using five distinct methods applied to two distinct datasets, with the second dataset being further subdivided into five distinct vaccination datasets. The accuracy, precision, recall, F1 score, and support measurement are derived from the Random Forest, Naive Bayes, Decision Tree, Logistic Regression, and Support Vector Machine (SVM).

Machine learning Performance on COVID-19 vaccine analysis

Classifier Name	Accuracy%	Precision%	Recall%	F1-Score%
Random Forest	81.94	89.18	67.76	69.9
Naive Bayes	75.67	71.55	63.19	63.2
Decision Tree	93.0	90.43	88.27	89.24
Logistic Regression	82.5	85.35	71.36	74.47
SVM	84.78	87.0	75.05	78.31

Bar Chart


Accuracy%
90
80
70
60
50
Random Forest
Naive Byes
Decision Tree
Logistic Regression
SVM
Accuracy
Precision
Recall
F1-score