IBM-312 GROUP PROJECT

Team Members Name	Branch	Enrollment Number
Ashutosh Kumar	EPH	19122007
Prince Raj Yadav	EPH	19122025
Smardeep Singh Sandhu	EPH	19311019

Problem Statement:

Crop prediction with respect to temperature and their atmospheric variables

Motivation:

Climate change is very likely to affect food security at the global, regional, and local level. It can disrupt food availability and crop patterns. For example, projected increases in temperatures, changes in precipitation patterns, changes in extreme weather events, and reductions in water availability may all result in reduced agricultural productivity. Our aim is to analyze the differences in types of crop production in a certain region due to varying environmental conditions. We are working with predicted classes of crops according to the given environmental parameters such as temperature, humidity, pH, and rainfall. Using the results of pre and post climate change effects on environments, we are trying to cite the significant changes in type of crop production.

Dataset:

The Dataset used by us has been taken from Kaggle, a popular site for data scientists to hone their skills. It contains the following data fields:

- `N` - ratio of Nitrogen content in soil
- `P` - ratio of Phosphorus content in soil
- `K` - ratio of Potassium content in soil
- `temperature` - temperature in degree Celsius
- `humidity` - relative humidity in %
- `ph` - ph value of the soil
- `rainfall` - rainfall in mm

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   N            2200 non-null   int64  
 1   P            2200 non-null   int64  
 2   K            2200 non-null   int64  
 3   temperature  2200 non-null   float64
 4   humidity     2200 non-null   float64
 5   ph           2200 non-null   float64
 6   rainfall     2200 non-null   float64
 7   label        2200 non-null   object 
dtypes: float64(4), int64(3), object(1)
memory usage: 137.6+ KB
```

For utilizing the ‘label’ column in the dataset for further analysis, we encoded the column using the LabelEncoder object from sklearn.preprocessing. The different classes were encoded as follows:

{'apple': 0, 'banana': 1, 'blackgram': 2, 'chickpea': 3, 'coconut': 4, 'coffee': 5, 'cotton': 6, 'grapes': 7, 'jute': 8, 'kidneybeans': 9, 'lentil': 10, 'maize': 11, 'mango': 12, 'mothbeans': 13, 'mungbean': 14, 'muskmelon': 15, 'orange': 16, 'papaya': 17, 'pigeonpeas': 18, 'pomegranate': 19, 'rice': 20, 'watermelon': 21}

With the modified dataset looking like:

	N	P	K	temperature	humidity	ph	rainfall	label	labelEncoded
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice	20
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice	20
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice	20
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice	20
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice	20

The correlation matrix for the dataset is also illustrated:

	N	P	K	temperature	humidity	ph	rainfall
N	1.000000	-0.231460	-0.140512	0.026504	0.190688	0.096683	0.059020
P	-0.231460	1.000000	0.736232	-0.127541	-0.118734	-0.138019	-0.063839
K	-0.140512	0.736232	1.000000	-0.160387	0.190859	-0.169503	-0.053461
temperature	0.026504	-0.127541	-0.160387	1.000000	0.205320	-0.017795	-0.030084
humidity	0.190688	-0.118734	0.190859	0.205320	1.000000	-0.008483	0.094423
ph	0.096683	-0.138019	-0.169503	-0.017795	-0.008483	1.000000	-0.109069
rainfall	0.059020	-0.063839	-0.053461	-0.030084	0.094423	-0.109069	1.000000

Methodology:

After initial Exploratory Data Analysis(EDA), we proceeded to classify similar crops, first on a very raw basis, using just ‘temperature’, ‘humidity’ and ‘rainfall’ values for each data point and classifying them as ‘Summer’, ‘Winter’ or ‘Rainy’ crops.

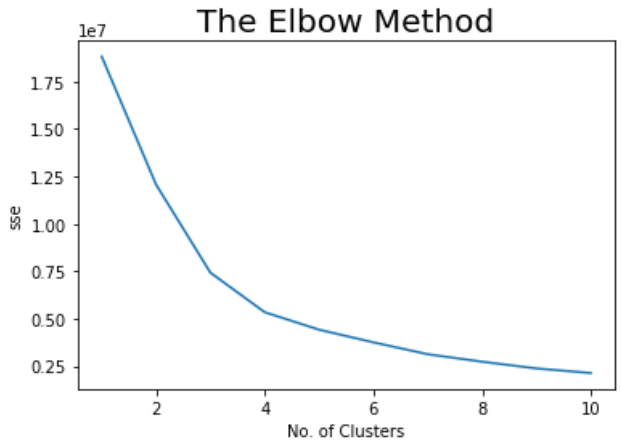
The results were as follows:

```
Summer Crops
['pigeonpeas' 'mothbeans' 'blackgram' 'mango' 'grapes' 'orange'
'papaya']

Winter Crops
['maize' 'pigeonpeas' 'lentil' 'pomegranate' 'grapes' 'orange']

Rainy Crops
['rice' 'papaya' 'coconut']
```

PS. Various crops may belong to multiple seasons because according to the dataset, these crops could be grown in different conditions, pertaining to different seasons.

Once this raw classification was done to identify the different seasonal patterns in crops, it was time to cluster the different crops by using the unsupervised learning technique of KMeans.

We used the Kmeans object from sklearn.cluster. Using the ‘Elbow method’ for finding the optimal number of clusters, which gave us the result of 4.



Code Snippet for Elbow Method:

```
sse = []
for i in range(1, 11):
    km = KMeans(n_clusters = i, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
    km.fit(x)
    sse.append(km.inertia_)

plt.plot(range(1, 11), sse)
plt.title('The Elbow Method', fontsize = 20)
plt.xlabel('No. of Clusters')
plt.ylabel('sse')
plt.show()
```

Using the KMeans object, with 4 clusters, the clustering resulted in the following outputs:

```
Crops in First Cluster: ['maize' 'chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans' 'mungbean'
 'blackgram' 'lentil' 'pomegranate' 'mango' 'orange' 'papaya' 'coconut']

Crops in Second Cluster: ['maize' 'banana' 'watermelon' 'muskmelon' 'papaya' 'cotton' 'coffee']

Crops in Third Cluster: ['grapes' 'apple']

Crops in Fourth Cluster: ['rice' 'pigeonpeas' 'papaya' 'coconut' 'jute' 'coffee']
```

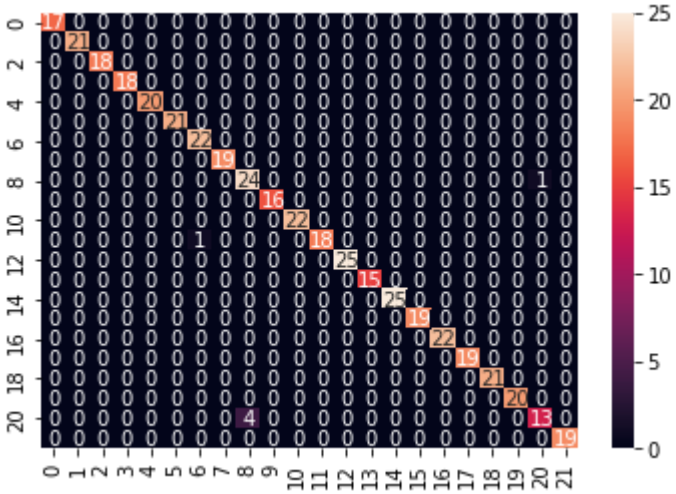
PS. Various crops may belong to multiple clusters because according to the dataset, these crops could be grown in different conditions, pertaining to different clusters.

Now comes the interesting part, the LogisticRegression model for predicting which crops to be grown for a certain set of parameters. The model was trained on 80% of data and tested on the remaining 20%. The following were the results obtained:

Result on Training Data:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	17
1	1.00	1.00	1.00	21
2	1.00	1.00	1.00	18
3	1.00	1.00	1.00	18
4	1.00	1.00	1.00	20
5	1.00	1.00	1.00	21
6	0.96	1.00	0.98	22
7	1.00	1.00	1.00	19
8	0.86	0.96	0.91	25
9	1.00	1.00	1.00	16
10	1.00	1.00	1.00	22
11	1.00	0.95	0.97	19
12	1.00	1.00	1.00	25
13	1.00	1.00	1.00	15
14	1.00	1.00	1.00	25
15	1.00	1.00	1.00	19
16	1.00	1.00	1.00	22
17	1.00	1.00	1.00	19
18	1.00	1.00	1.00	21
19	1.00	1.00	1.00	20
20	0.93	0.76	0.84	17
21	1.00	1.00	1.00	19
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

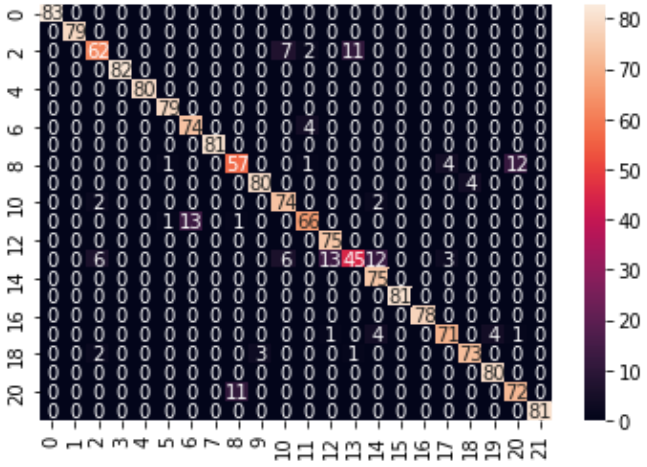
Train Accuracy Score: 0.9863636363636363



Result on Testing Data:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	83
1	1.00	1.00	1.00	79
2	0.86	0.76	0.81	82
3	1.00	1.00	1.00	82
4	1.00	1.00	1.00	80
5	0.98	1.00	0.99	79
6	0.85	0.95	0.90	78
7	1.00	1.00	1.00	81
8	0.83	0.76	0.79	75
9	0.96	0.95	0.96	84
10	0.85	0.95	0.90	78
11	0.90	0.81	0.86	81
12	0.84	1.00	0.91	75
13	0.79	0.53	0.63	85
14	0.81	1.00	0.89	75
15	1.00	1.00	1.00	81
16	1.00	1.00	1.00	78
17	0.91	0.88	0.89	81
18	0.95	0.92	0.94	79
19	0.95	1.00	0.98	80
20	0.85	0.87	0.86	83
21	1.00	1.00	1.00	81
accuracy			0.93	1760
macro avg	0.92	0.93	0.92	1760
weighted avg	0.92	0.93	0.92	1760

Test Accuracy Score: 0.925



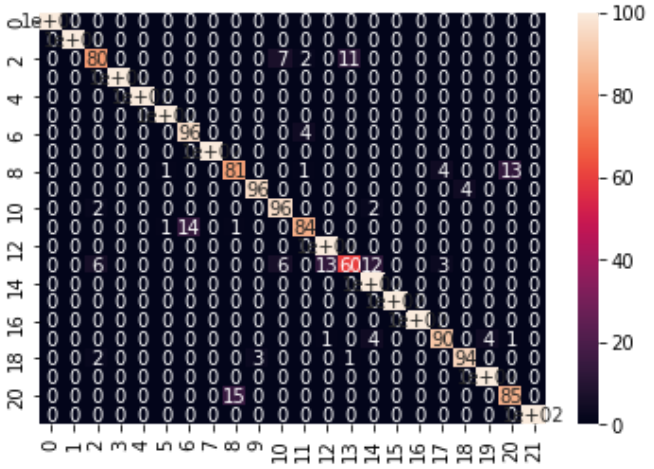
from HumanActivites import GlobalWarming:

To replicate the results of Global Warming on our data, we modified the temperature values in our dataset, to have an average increase of 2.5 degrees, and temperature increasing within the range 1 to 5 degrees.
We then used our pre-existing model to predict the new crops, and if the accuracy is lowered significantly, this may suggest that temperature rise due to global warming will have a huge impact on how and what we grow for food.
The results obtained were as follows:

Result on Normal Data:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	100
1	1.00	1.00	1.00	100
2	0.89	0.80	0.84	100
3	1.00	1.00	1.00	100
4	1.00	1.00	1.00	100
5	0.98	1.00	0.99	100
6	0.87	0.96	0.91	100
7	1.00	1.00	1.00	100
8	0.84	0.81	0.82	100
9	0.97	0.96	0.96	100
10	0.88	0.96	0.92	100
11	0.92	0.84	0.88	100
12	0.88	1.00	0.93	100
13	0.83	0.60	0.70	100
14	0.85	1.00	0.92	100
15	1.00	1.00	1.00	100
16	1.00	1.00	1.00	100
17	0.93	0.90	0.91	100
18	0.96	0.94	0.95	100
19	0.96	1.00	0.98	100
20	0.86	0.85	0.85	100
21	1.00	1.00	1.00	100
accuracy			0.94	2200
macro avg	0.94	0.94	0.94	2200
weighted avg	0.94	0.94	0.94	2200

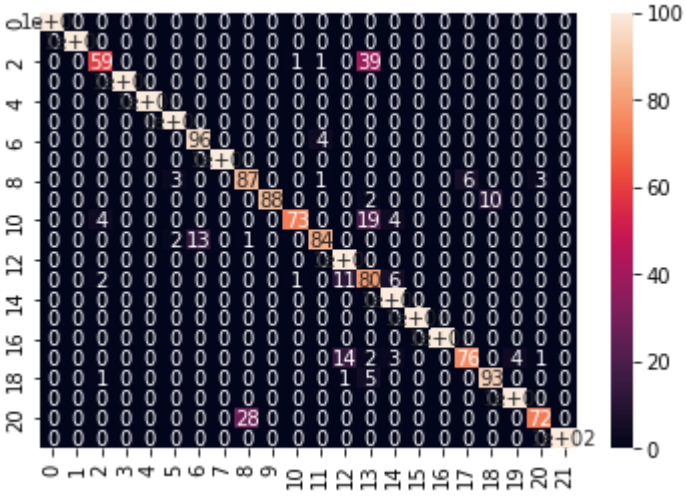
Accuracy Score(Normal): 0.9372727272727273



Result on Increased Temperature(because of Global Warming) Data:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	100
1	1.00	1.00	1.00	100
2	0.89	0.59	0.71	100
3	1.00	1.00	1.00	100
4	1.00	1.00	1.00	100
5	0.95	1.00	0.98	100
6	0.88	0.96	0.92	100
7	1.00	1.00	1.00	100
8	0.75	0.87	0.81	100
9	1.00	0.88	0.94	100
10	0.97	0.73	0.83	100
11	0.93	0.84	0.88	100
12	0.79	1.00	0.88	100
13	0.54	0.80	0.65	100
14	0.88	1.00	0.94	100
15	1.00	1.00	1.00	100
16	1.00	1.00	1.00	100
17	0.93	0.76	0.84	100
18	0.90	0.93	0.92	100
19	0.96	1.00	0.98	100
20	0.95	0.72	0.82	100
21	1.00	1.00	1.00	100
accuracy			0.91	2200
macro avg	0.92	0.91	0.91	2200
weighted avg	0.92	0.91	0.91	2200

Accuracy Score(Increased Temp): 0.9127272727272727



Results

According to NOAA's 2020 Annual Climate Report the combined land and ocean temperature has increased at an average rate of 0.13 degrees Fahrenheit (0.08 degrees Celsius) per decade since 1880; however, the average rate of increase since 1981 (0.18°C / 0.32°F) has been more than twice that rate.(Source)

The increase of rate of Global warming will change the types of crops that we grow, which is what we showed in our analysis as well. The accuracy on the same data using the same parameters drops by more than 3%, highlighting a significant change between the crops that

were grown and that would need to be grown, if we don't try to revert the catastrophe of Global Warming now.

Link to the notebook used:
https://colab.research.google.com/drive/1gxnJ4k1ycFxM4OzqydR6avoQkxDaJ0de?usp=sharing

Link to the Dataset used:
https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

_____Thank You_____