Capstone Project

Title :- Suggesting best trading option for a perishable food crop for a state among the states across India.

Data 606

Instructor:- Dr.Chaojie Wang

Importance of this issue

- Indian economy mostly depends on agriculture
- Perishable food crops are the crops that must be transported and used with in 2-3 weeks of harvesting.
- Farm price of a particular crop varies in different seasons in a particular sate and also varies in different states in a particular season
- Effective trade system is mandatory to regulate the perishable food crop prices

How I am going to address this Issue

- I want to design a machine learning model which can suggest me best set of sates to trade with for a particular crop in a particular season.
- In case of importing a perishable food crop it should suggest the set of sates so that the import price is less.
- In case of exporting a food crop, I want the model to suggest me best set of states to trade with so that farmers will get higher farm price.

Other Issues that can be addressed using this model

1.Predicting most profitable crop for that season

2.Helping traders to make high profit margins

Factors that effect the farm price

1) Climatic Conditions:-
   - Sunlight
   - Precipitation
   - Temperature
   - Wind Speed

2) Soil Conditions

About Datasets

1) **)Monthly data of different perishable cops at different states across the country.**

- Note: - In my project I am taking top 10 most produced perishable crop datasets from January 2011 till December 2020.
- Source: - National Horticulture Board of India
- Link: - http://nhb.gov.in/OnlineClient/MonthlyPriceAndArrivalReport.aspx?enc=3ZOO8K5CzcdC/Yq6HcdIxJ4o5jmAcGG5QGUXX3BlAP4=
- Link to Dataset in my Google Drive(Anyone from UMBC group can access it): - https://drive.google.com/drive/folders/1OxtPZU1gt_ehgOiGhl_hpdmPWlvjO2ZW?usp=sharing
- Attributes of the datasets: - Each individual dataset consists of 6 attributes, namely
- Centre Name, Year, Month, W.sale avg.price, Retail Avg.price and Total Arrival

2)Monthly historical weather and soil report of each sate in India from January 2010 till December 2020.

- Source: - NASA's Data Access Viewer: - Its Provides solar and meteorological data sets from NASA research for support of renewable energy, building energy efficiency and agricultural needs.
- Link: - https://power.larc.nasa.gov/data-access-viewer/ 2.3) Link to Dataset in my Google Drive(Anyone from UMBC group can access it):- https://drive.google.com/drive/folders/15zXCuLByVlfpdg5eg_yfT2qBmqdDHeif?usp=sharing
- Attributes of the Datasets: -
  - All Sky Surface PAR Total
  - Dew/Frost Point at 2 Meters
  - Temperature at 2 Meters Maximum
  - Temperature at 2 Meters Minimum
  - Specific Humidity at 2 Meters
  - Precipitable Water
  - Wind Speed at 2 Meters Maximum
  - Wind Speed at 2 Meters Minimum
  - Surface Soil Wetness
  - Root Zone Soil Wetness Profile Soil Moisture

ML Models:-

- I will be using supervised learning techniques in my algorithm. As of now I think of two different techniques to design the model.

1) Using Multi-Class Multi Label Classification Algorithms

   - Implementation:- Under this Multi-Class classification I plan to divide the farm price of each crop into 3 or 4 different slabs such as high profit margin, low profit margin or loss and predict under which slab a particular crop can fall given all the features.
   - Algorithms:- I will try to train my model using multi-class classification machine learning algorithms such as k-Nearest Neighbor's, Decision Trees, Naive Bayes etc...

   - Performance Metrics :- Precision and Recall values and F1 score

Regression Methods

2) Using Regression Methods

- Using regression algorithms to train my model and predict the farm price of a perishable food crop.

- Algorithms:- Neural Networks or LASSO regression(As I feel there is good co-relation between by features)

- Performance Metrics:- Mean Square Error or Mean Absolute Errors

EDA of Crops Dataset:-

- Combined 110 different datasets

											...						
6	NaN	NaN	January	NaN	NaN	February	NaN	NaN	March	NaN	...	NaN	October	NaN	NaN	November	NaN
7	S.No.	Center Name	W.sale Avg.Price	Retail Avg. Price	Total Arrival	W.sale Avg.Price	Retail Avg. Price	Total Arrival	W.sale Avg.Price	Retail Avg. Price	...	Total Arrival	W.sale Avg.Price	Retail Avg. Price	Total Arrival	W.sale Avg.Price	Retail Avg. Price
8	1	AHMEDABAD / अहमदाबाद	0	0	0	0	0	0	1307	3238	...	806	960	2500	167	800	2000
9	2	AMRITSAR / अमृतसर	0	0	0	0	0	0	0	0	...	68	1076	2476	55	0	0
10	3	BANGALURU / बेंगलुरू	1410	2083	82	1073	1796	91	965	1694	...	197	1716	2458	54	1563	2258
11	4	BARAUT / बड़ौत	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0
12	5	BHOPAL / भोपाल	0	0	0	0	0	0	2546	3032	...	191	1035	1346	368	1608	1917
13	6	BHUBANESHWAR / भुवनेश्वर	2540	3130	52	2339	2911	33	1630	2189	...	53	1926	2621	67	1635	2208

- Deleting unwanted rows, Deleting unwanted columns, Renaming the header values, Reshaping the data frame, Fill null values, Resetting index

S.No.	enter_Nam	e_in_Janu	e_in_Febr	ce_in_Mar	rice_in_Ap	rice_in_Ma	rice_in_Jur	rice_in_Jul	ce_in_Aug	_in_Septe	e_in_Octo	_in_Nover	_in_Decer	Crop	Year
1	Chhattisga	1126.889	1126.889	1307	1119	751	1729	1164	1173	1139	960	800	1126.889	BITTER GO	2011
2	Andhra Pra	1031.5	1031.5	1031.5	1031.5	1179	640	942	1083	1269	1076	1031.5	1031.5	BITTER GO	2011
3	Rajasthan	1410	1073	965	932	1546	1660	1589	1542	1659	1716	1563	1539	BITTER GO	2011
4	Jhammu &	1086	1086	1086	1370	1018	750	1231	1061	1086	1086	1086	1086	BITTER GO	2011
5	Maharasht	1673.8	1673.8	2546	1907	1298	1462	2262	1576	1448	1035	1608	1596	BITTER GO	2011
6	Jhammu &	2540	2339	1630	847	795	1498	1723	1648	2105	1926	1635	1480	BITTER GO	2011
7	Chhattisga	1219.5	1219.5	1219.5	1219.5	1488	869	996	1148	1359	1457	1219.5	1219.5	BITTER GO	2011
8	Andhra Pra	1908	1387	1222	1013	1276	1248	1546	1447	1486	1500	1378	1420	BITTER GO	2011

EDA of weather datasets

- Combined 29 states weather datasets

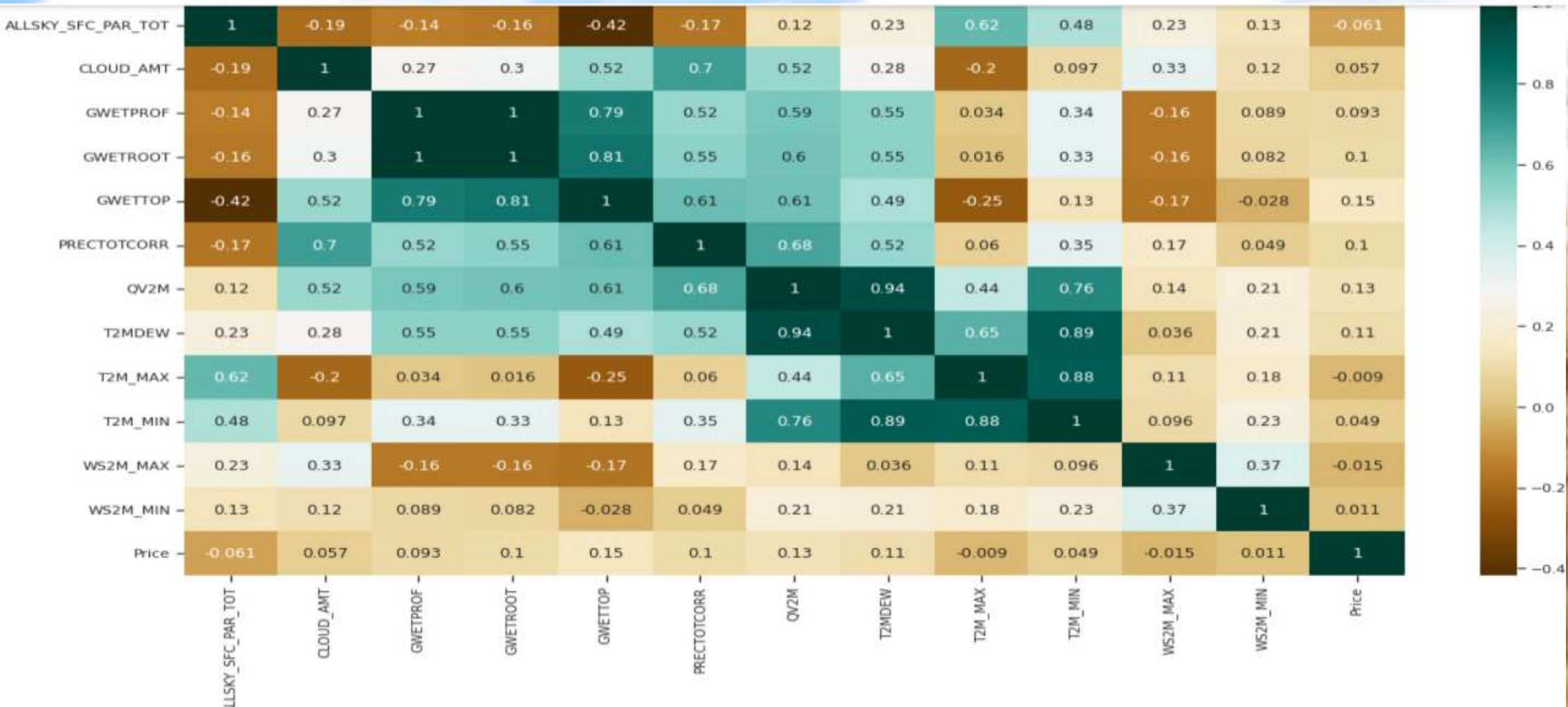
16	CLOUD_AMT CERES SYN1deg Cloud Amo...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
17	PRECTOTCORR MERRA-2 Precipitation C...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	ALLSKY_SFC_PAR_TOT CERES SYN1deg All Sky S...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19	-END HEADER-	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	PARAMETER	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANN
21	QV2M	2011	13.24	13.85	14.22	16.78	17.15	16.3	16.91	17.76	17.82	17.33	15.2	14.16	15.93
22	QV2M	2012	13.43	12.45	15.14	16.78	16.48	15.75	17.21	17.21	17.88	16.97	15.62	14.53	15.81
23	QV2M	2013	13.67	13.55	14.34	16.97	17.09	17.21	17.09	18.07	18.62	18.49	16.05	12.94	16.17
24	QV2M	2014	13.37	13	13.61	15.81	17.88	16.66	17.09	17.64	18.01	17.15	15.08	13.98	15.81

- Appended individual data frames, No nulls, Renaming the header values, Reshaping the data frame.

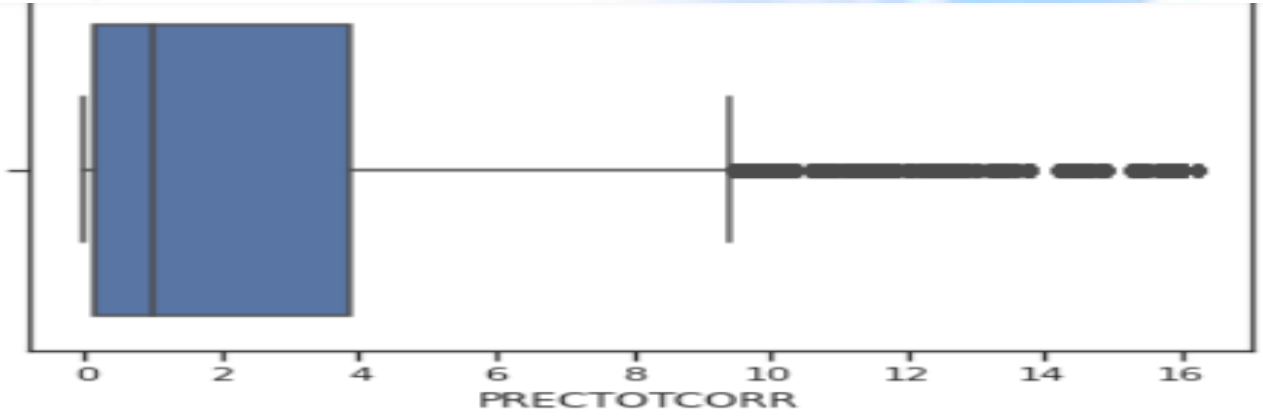
	PARAMETER	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANN	STATE
0	QV2M	2011	13.24	13.85	14.22	16.78	17.15	16.3	16.91	17.76	17.82	17.33	15.2	14.16	15.93	Andhra Pradesh
1	QV2M	2012	13.43	12.45	15.14	16.78	16.48	15.75	17.21	17.21	17.88	16.97	15.62	14.53	15.81	Andhra Pradesh
2	QV2M	2013	13.67	13.55	14.34	16.97	17.09	17.21	17.09	18.07	18.62	18.49	16.05	12.94	16.17	Andhra Pradesh
3	QV2M	2014	13.37	13	13.61	15.81	17.88	16.66	17.09	17.64	18.01	17.15	15.08	13.98	15.81	Andhra Pradesh
4	QV2M	2015	12.63	12.63	14.65	16.48	17.03	17.76	16.97	18.07	18.31	17.03	16.72	14.59	16.11	Andhra Pradesh

Combining both data frames

- Reshaped the data frames for merging
- Checked for highly co-related column values



- Removed Outliers for each individual columns if any
- Code Snippet




```
1 df_final[int_columns[5]].describe() # Viewing statistical details
```

```
count     28764.000000
mean          2.744346
std           3.835773
min           0.000000
25%           0.150000
50%           1.010000
75%           4.000000
max          25.790000
Name: PRECTOTCORR, dtype: float64
```

```
[674]  1 # Defining inter quartile range based on box plot
       2 Q1 = df_final1[int_columns[5]].quantile(0.02)
       3 Q3 = df_final1[int_columns[5]].quantile(0.75)
       4 IQR = Q3-Q1
```

```
[675]  1 Q3+1.5 * IQR
```

```
9.65
```


Thank You
Goutham Kakani