International Journal of Gender, Science and Technology UGC Care Group I Journal
..ISSN: 2040-0748 Vol-10 Issue-01 Jan 2021

ANALYSIS OF FARMING USING DATA MINING

PERAM MALLIKARJUNA1, Dr. KUNI VENKATA SUBBAIAH2

#1Assistant Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali
#2Professor, Department of CSE, PBR Visvodaya Institute of Technology and Science, Kavali

ABSTRACT_ Based on the number of people, agriculture is thought to be India's largest economic sector. It is very important to the growth of the country and helps the economy. Here, many crops are grown, with rice and wheat being two of the most important ones. Pulses, potatoes, and other vegetables are some of the other food crops that grow here. Cash crops are also grown here, like sugarcane, oil seeds, cotton, coffee, tea, rubber, and jute. Even though agriculture is a big part of the Indian economy and employs a lot of people, it is very inefficient, not based on science, and can't keep up with the high demand for food in a country with so many people. Even though there have been improvements in this area, most of these problems still exist. These problems can be fixed by doing a good analysis of the agricultural situation and using the information to make suggestions about how to grow crops and what kinds of crops to grow.

1.INTRODUCTION

Farmers and agro-based businesses have to make a lot of decisions every day, and there are a lot of things that affect those decisions. Soil, climate, farming, irrigation, fertilisers, temperature, rainfall, harvesting, and the use of pesticides are some of the things that affect agriculture. By mining the large amount of crop, soil, and weather data that is already out there and analysing the environment, farmers can use this information to help them make important farming decisions. This maximises production and makes farming more resistant to changes in the weather. Information about past crop yields is also important for how companies in industries run their supply chains. These industries use livestock, food, animal feed, chemicals, chicken, fertiliser, pesticides, seed, and paper as raw materials. A good estimate of crop production and risk helps these companies make decisions about the supply chain, like when to schedule production. Estimates of crop production help businesses like those that make seeds, fertiliser, agrochemicals, and agricultural machinery plan their production and marketing.

2.LITERATURE SURVEY

2.1 Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki "Analysis of agriculture data using data mining techniques: application of big

data" Majumdar et al. J Big Data (2017) 4:20 DOI 10.1186/s40537-017-0077-4

In the agriculture sector, farmers and agribusinesses have to make a lot of decisions every day, and the many factors that affect them are very complicated. Accurate yield estimates for the many crops involved in the planning are a key part of agricultural planning. The only way to solve this problem in a practical and effective way is to use data mining techniques. Big data has been used most often in agriculture. Conditions in the environment, differences in soil, input levels, combinations, and prices of goods have made it even more important for farmers to use information and get help when making important farming decisions. This paper is mostly about how to use data mining techniques like PAM, CLARA, DBSCAN, and Multiple Linear Regression to analyse agricultural data and find the best parameters to maximise crop production. Mining the large amount of crop, soil, and climate data that already exists and analysing new, non-experimental data improves production and makes agriculture more resistant to changes in climate.

2.2 Deepak Sharma, Priti Sharma "Rain Fall Prediction using Data Mining Techniques with Modernistic Schemes and Well-Formed Ideas" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019

One of the hardest parts of weather forecasting is figuring out when it will rain. Accurate and timely predictions of rain can help people take good safety precautions for ongoing construction projects, transportation activities, agricultural tasks, flight operations, and flood situations, among other things. By finding hidden patterns in the weather data that are already available, data mining techniques can accurately predict when it will rain. This research helps by giving a critical review and analysis of the most recent data mining techniques used to predict rain. For this research, papers that came out between 2013 and 2017 and were in well-known online search libraries were looked at. This review will help researchers look at the most recent work on predicting rainfall, with a focus on data mining techniques. It will also give a starting point for future directions and comparisons.

2.3 Shalin Paulson "A Survey on Data Mining Techniques in Agriculture" Special Issue - 2015 International Journal of Engineering Research & Technology (IJERT) ISSN: 2278- 0181 Published by, www.ijert.org RTPPTDM-2015 Conference Proceedings

Data mining is a new field of study in agriculture that is growing quickly. It is used to figure out and study how different things

affect crop yield. In this paper, we will look at and try out how data mining can be used in the field of agriculture. Data mining is a unique and important part of making decisions in agriculture about many things. In agriculture, data mining can help predict things like crop yield, weather and rainfall, the quality of seeds and soil, and how many crops will be grown. A method called "predictive data mining" is used to figure out what crops will grow in the future, how much money will be made, and what pesticides and fertilisers will be needed so that crops can grow and work properly. Several data mining techniques, such as kmeans (KM), k-nearest neighbour (KNN), artificial neural network (ANN), and support vector machine (SVM), are used to solve problems and find ways to make agriculture grow better. Each technique for data mining has its own way to show how different problems look. This helps us figure out the best way to solve every problem in agriculture. In this paper, we learned what we needed to know about each data mining technique so that we could use them in all similar situations. It sums up all the information by using all the techniques mentioned and adding some new ones. This helps plan agriculture better, which leads to good growth in agriculture.

3.PROPOSED SYSTEM

Here, many crops are grown, with rice and wheat being two of the most important ones. Pulses, potatoes, and other vegetables are some of the other food crops that grow here. Cash crops are also grown here, like sugarcane, oil seeds, cotton, coffee, tea, rubber, and jute. Even though agriculture is a big part of the Indian economy and employs a lot of people, it is very inefficient, not based on science, and can't keep up with the high demand for food in a country with so many people. Even though there have been improvements in this area, most of these problems still exist. These problems can be fixed by doing a good analysis of the agricultural situation and using the information to make suggestions about how to grow crops and what kinds of crops to grow.

3.1 IMPLEMETATION

3.1. 1 DATA CLEANING

At first, the data needed to be cleaned up. The problems with cleaning up the data are:

1. The data in the databases were from different years and were not the same in each database.

2. Some crop names were not in all of the databases.

3. The database was also missing a lot of information.

4. The data came in different forms.

5. Different databases had different ways of naming crops and states.

6. In different databases, the units of measurement were different.

The databases were changed so that the information was in the right format, and missing values were filled in with the average values from different years. Then the data was ready to be used in any way.

3.1.2 Data Integration

The data from different tables were merged so that it can be analyzed. The tables were also unstacked when required, for proper understanding.

3.1.3 Data Analysis

The data were visualized and several plots were made for statistical analysis. Various data mining algorithms were also used for finding patterns and making predictions. They are described in detail in the following sections.

4.RESULTRS AND DISCUSSION

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

Table 1: Year wise crop production

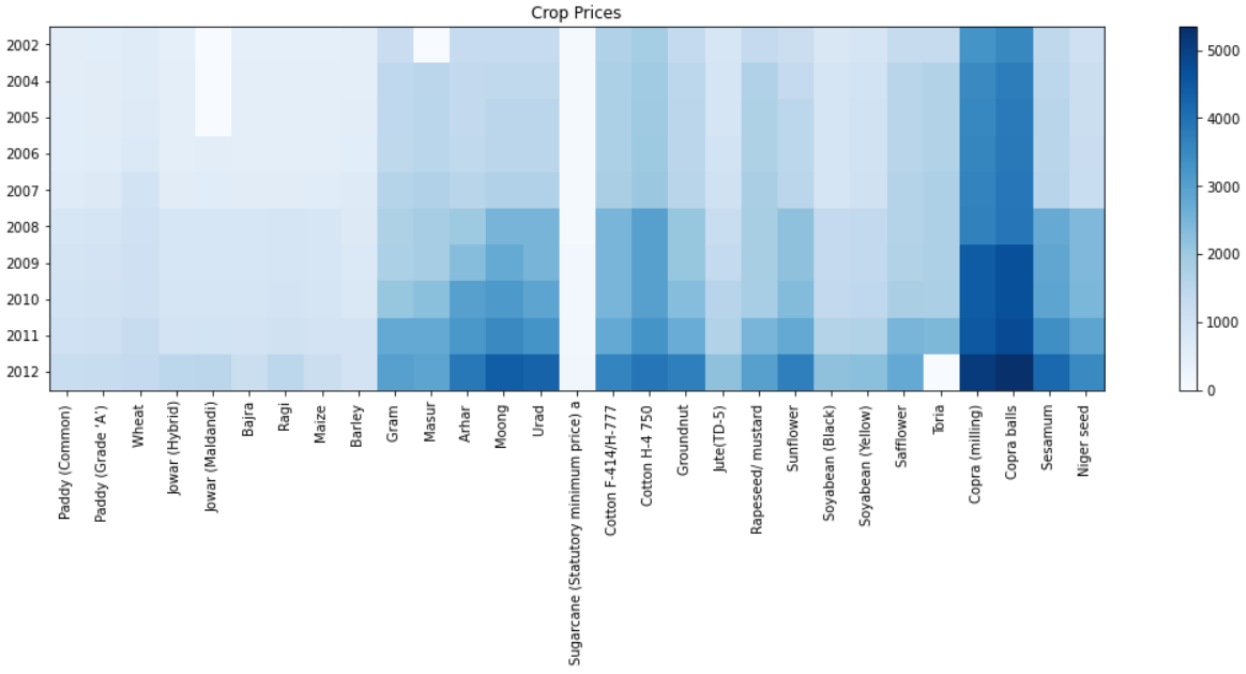

Figure 1: Crop prices of various crops in different year in Rs/quintal

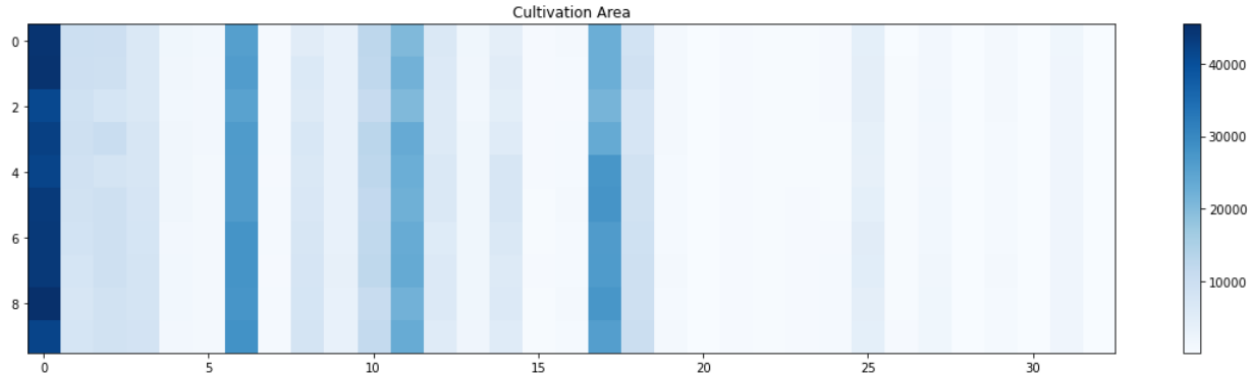

Figure 2: Cultivation area of various crops in hectares

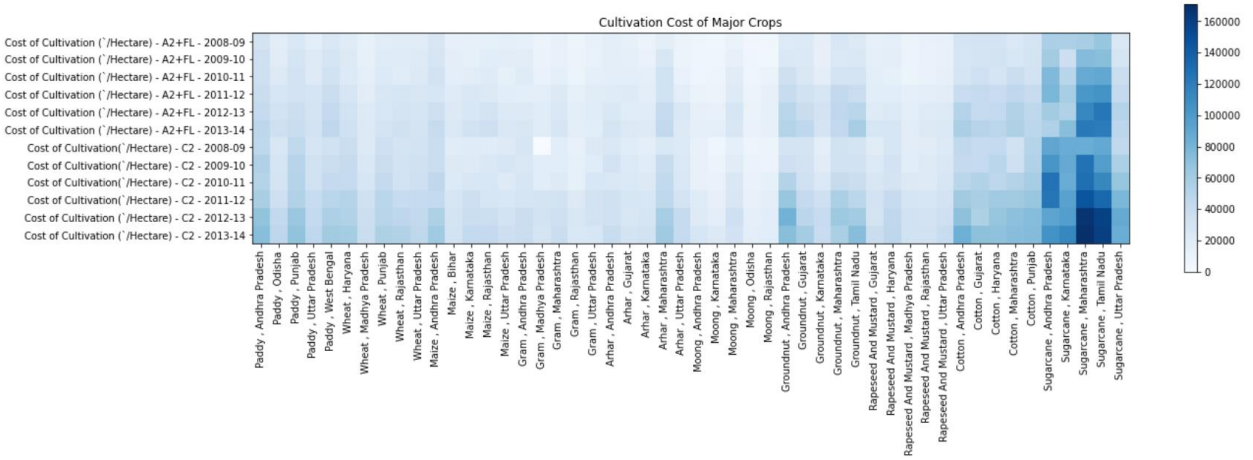

Figure 3: Cultivation cost by area of major crops in respective states in Rs/hectare

	YEAR	Suicides	Area	Production	Annual_Rain	Avg_Temperature	Price	Cost_per_Hectare	Cost_per_quintal	Export
YEAR	1.000000	-0.072314	0.018936	-0.021838	-0.068967	-0.844673	-0.579313	-0.089484	-0.268659	0.520077
Suicides	-0.072314	1.000000	0.569297	0.053705	-0.153630	0.078271	0.153719	0.629253	0.649622	-0.174800
Area	0.018936	0.569297	1.000000	-0.141455	-0.474405	-0.018899	0.174528	0.611186	0.529003	-0.068576
Production	-0.021838	0.053705	-0.141455	1.000000	0.338707	0.000170	-0.180207	-0.091603	-0.115446	-0.156324
Annual_Rain	-0.068967	-0.153630	-0.474405	0.338707	1.000000	0.011724	-0.221371	-0.394261	-0.348268	0.116749
Avg_Temperature	-0.844673	0.078271	-0.018899	0.000170	0.011724	1.000000	0.820575	0.108782	0.382569	-0.269068
Price	-0.579313	0.153719	0.174528	-0.180207	-0.221371	0.820575	1.000000	0.400618	0.589711	-0.067872
Cost_per_Hectare	-0.089484	0.629253	0.611186	-0.091603	-0.394261	0.108782	0.400618	1.000000	0.735738	0.036654
Cost_per_quintal	-0.268659	0.649622	0.529003	-0.115446	-0.348268	0.382569	0.589711	0.735738	1.000000	-0.096512
Export	0.520077	-0.174800	-0.068576	-0.156324	0.116749	-0.269068	-0.067872	0.036654	-0.096512	1.000000

Table 2: Suicide rate included in analysis

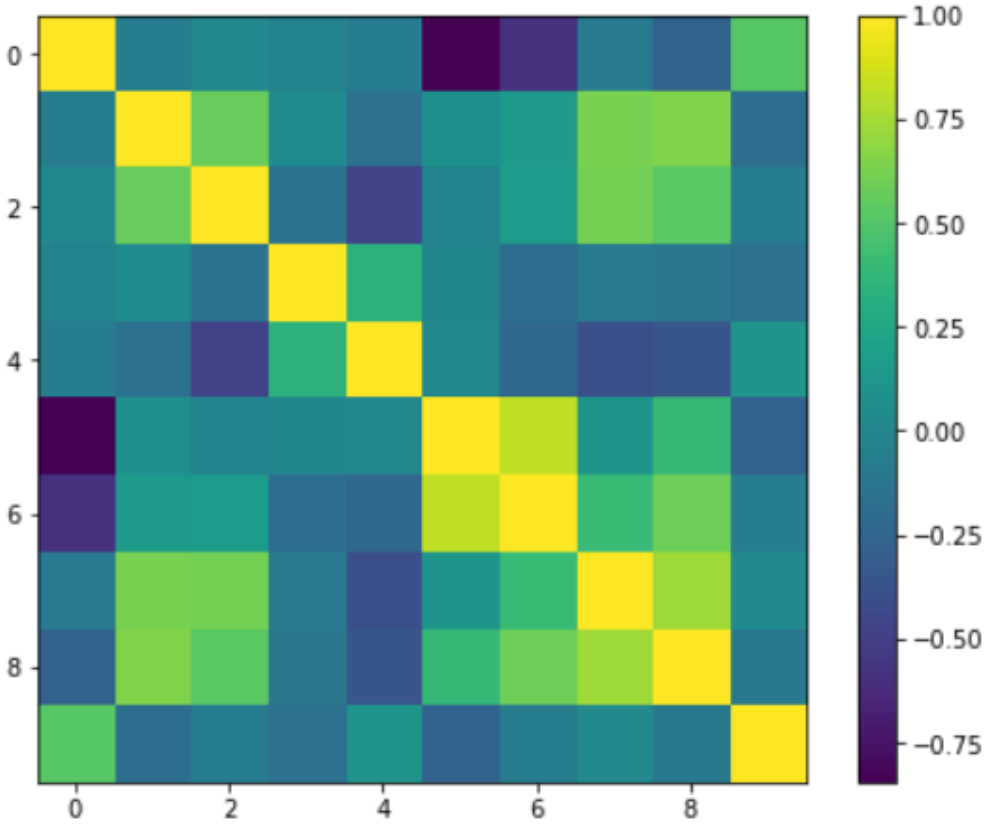

Figure 4: Correlation between different features

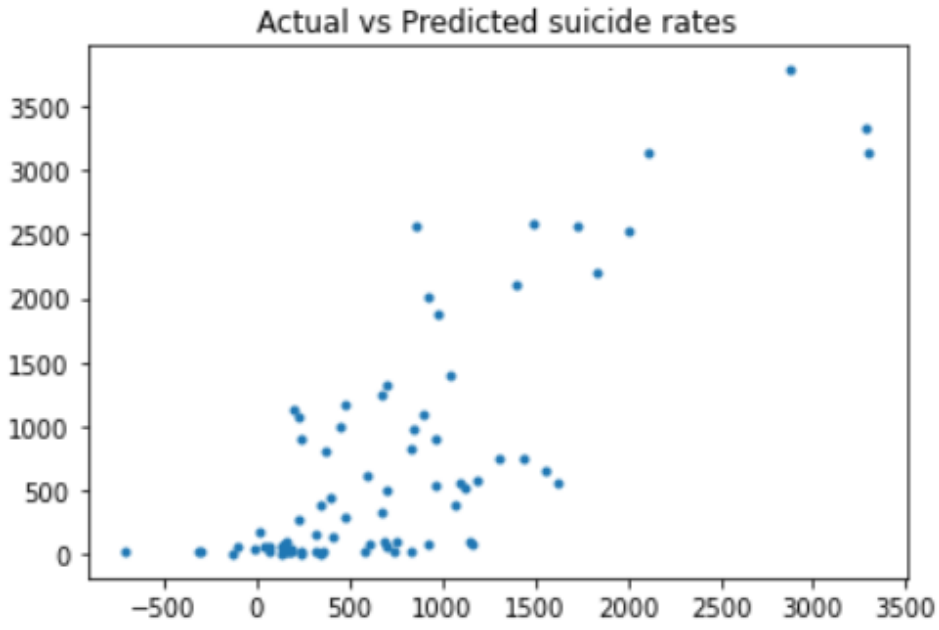

Figure 5: Actual vs Predicted suicide rates

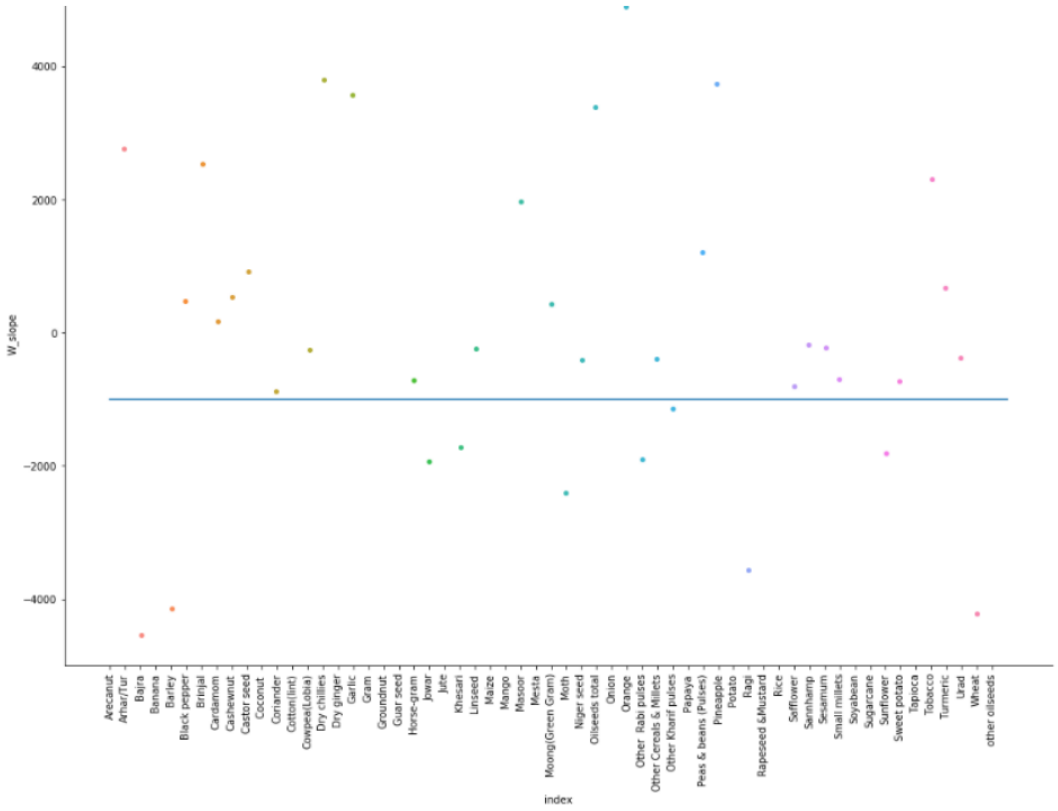

Figure 6: Crops with reduction in production over the years

	index	W_slope
2	Bajra	-4537.162480
4	Barley	-4141.496114
18	Groundnut	-12034.886469
19	Guar seed	-12780.917262
21	Jowar	-1927.821737
22	Jute	-54720.777614
23	Khesari	-1717.234226
28	Mesta	-5094.770567
30	Moth	-2411.072500
35	Other Rabi pulses	-1908.905869
37	Other Kharif pulses	-1140.953895
42	Ragi	-3564.839019
43	Rapeseed &Mustard	-6685.204140
51	Sunflower	-1812.156509
53	Tapioca	-11008.462271
57	Wheat	-4216.058051

Table 3: Crops with reducing productions over the years

	index	production var	price var
0	Barley	-129.610556	51.277056
1	Jute	-43011.146430	125.248918
2	Niger seed	-144.009443	249.534632
3	Safflower	-1031.251935	122.445887
4	Sunflower	-4511.467222	223.906926

Table 4: Crops with Reducing productions and Increasing prices

	index	production var	price var
0	Arhar/Tur	5237.572915	260.551948
1	Groundnut	1758.235163	199.339827
2	Jowar	3648.323679	234.956710
3	Jute	-43011.146430	125.248918
4	Moong	941.797058	308.993506
5	Niger seed	-144.009443	249.534632
6	Safflower	-1031.251935	122.445887
7	Sesamum	1494.291172	279.404762
8	Sunflower	-4511.467222	223.906926
9	Urad	2670.533797	284.469697

Table 5: Crops with slow increase in productions but high increase in prices

	index	production var	price var
0	Arhar/Tur	5237.572915	260.551948
1	Bajra	13476.838875	69.145022
2	Barley	-129.610556	51.277056
3	Cotton(lint)	98938.700867	379.372294
4	Gram	17651.118445	169.123377
5	Groundnut	1758.235163	199.339827
6	Jowar	3648.323679	234.956710
7	Jute	-43011.146430	125.248918
8	Maize	49300.778390	68.593074
9	Moong	941.797058	308.993506
10	Niger seed	-144.009443	249.534632
11	Ragi	433.893412	92.521645
12	Rice	176762.226287	150.562771
13	Safflower	-1031.251935	122.445887
14	Sesamum	1494.291172	279.404762
15	Soyabean	63821.050069	239.870130
16	Sugarcane	642131.800127	10.367338
17	Sunflower	-4511.467222	223.906926
18	Urad	2670.533797	284.469697
19	Wheat	198027.175392	81.904762

Table 6: Crops with various trends in productions and price

	index	cost_per_hec var	price var	grow_ratio
5	Moong	571.901298	308.993506	0.540292
0	Arhar/Tur	1155.433382	260.551948	0.225501
2	Gram	865.233108	169.123377	0.195466
1	Cotton(lint)	2477.784265	379.372294	0.153109
3	Groundnut	1623.278151	199.339827	0.122801
6	Rice	1253.114850	150.562771	0.120151
4	Maize	868.141613	68.593074	0.079011
8	Wheat	1122.494264	81.904762	0.072967
7	Sugarcane	3281.720563	10.367338	0.003159

Table 7 : Crops with lower rate of increase in cost per hectare than price

Crop	Apple	Arcanut (Processed)	Arecanut	Arhar/Tur	Ash Gourd	Atcanut (Raw)	Bajra	Banana	Barley	Bean	...	Turmeric	Turnip	Urad
STATES														
ANDAMAN & NICOBAR ISLANDS	0.0	0.0	48035.81	104.0	0.0	0.0	0.0	97424.65	0.0	0.0	...	1593.03	0.0	475.0
ANDHRA PRADESH	0.0	0.0	3784.00	1429055.0	0.0	0.0	1043457.0	15041631.00	0.0	0.0	...	1568134.00	0.0	3921961.0
ARUNACHAL PRADESH	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.00	0.0	0.0	...	34933.00	0.0	0.0
ASSAM	0.0	0.0	845536.00	72565.0	0.0	0.0	0.0	10559119.00	0.0	0.0	...	158551.00	0.0	338637.0
BIHAR	0.0	0.0	0.00	643465.0	0.0	0.0	47543.0	1913108.00	277542.0	0.0	...	24375.00	0.0	256575.0

Table 8: Analysis of crop production in various states

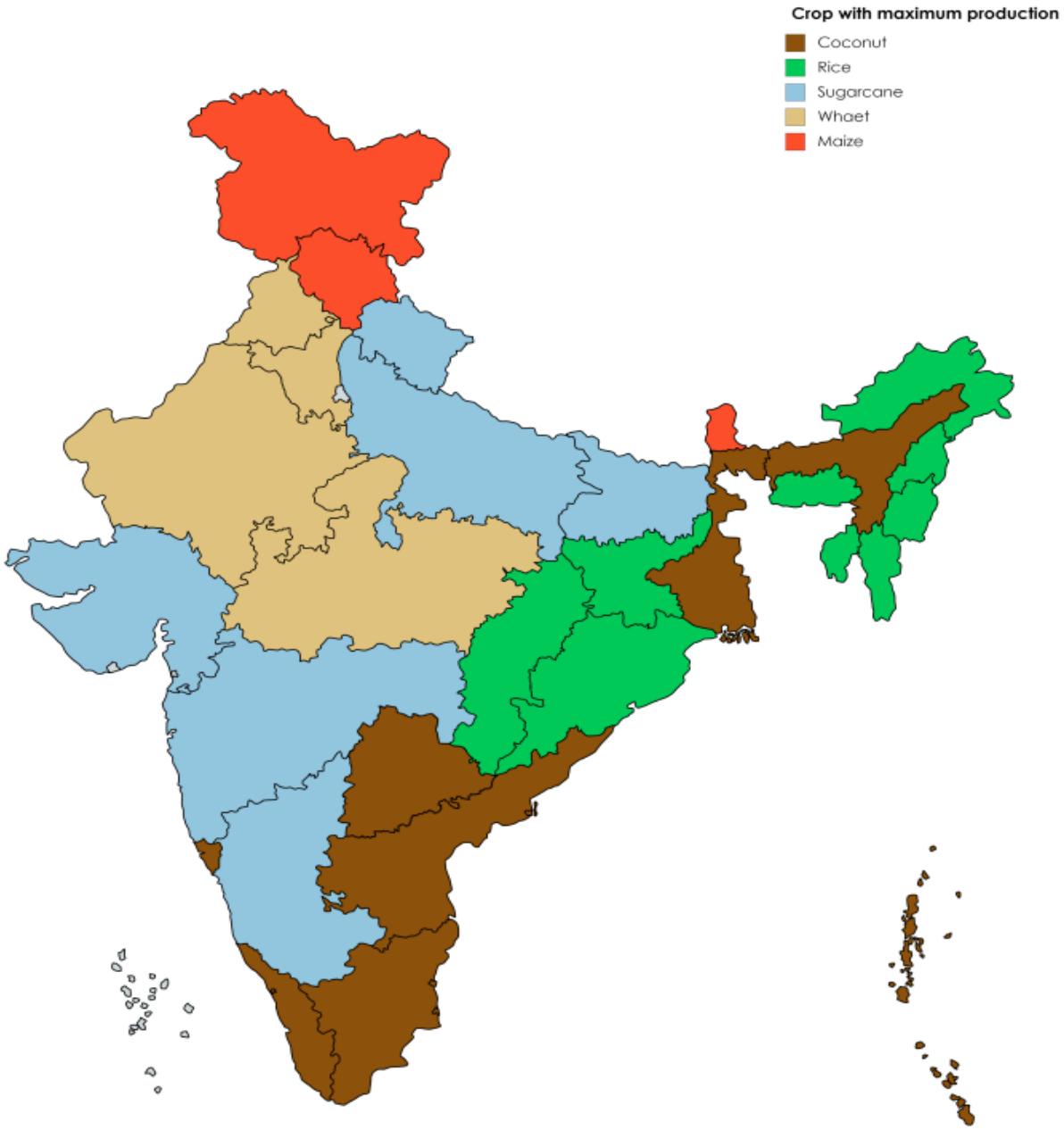

Figure 7: State wise maximum producing crops.

Maximum producing crops of various states are shown in above figure.As expected states like Uttar Pradesh,Gujarat, Maharashtra are major producer of sugarcane.While crop like rice are majorly producer in areas with high rainfall like eastern states and states like Odisha, West Bengal.Coconut is centered in coastal areas like Tamil Nadu, Kerala, while most of northern states are major producer of wheat.

5.CONCLUSION

There are many things that affect a country's agriculture, and it's very helpful to study them all. In this project, we've tried to get agricultural data and put it in a way that makes it easy to analyse. Visualizing the

data sets helps people understand them better, so they are done that way. The datasets are put together and looked at to see how different things affect crop production. Simple statistical inferences help us see how patterns change over time, which makes us want to find out why these patterns change.

The things that affect crop production don't always work on their own, so the lack of just one of them can have a huge effect on crop production. Using these numbers, it's also possible to make accurate predictions about important things like suicide rates, which helps with planning, taking preventative steps, and making insurance policies. We've made predictions using both linear regression and decision trees. 83 percent of the time, decision trees are right. Decision trees make it easier to understand how each of the factors affects the prediction because they show how they all work together. Finding crops that change in unusual ways over time, like having a sudden drop in production, can help us figure out why they change the way they do.

We can also use these numbers to suggest new crops that could be grown in places with the right climate and economy. If similar studies are done and followed, it can stop land from being wasted and increase production, which will help us meet people's needs and boost our economy..

FUTURE WORK

Large data sets with a lot of attributes can be added to the study. The data can be used to learn more about the different parts of each state. As features, crop production, cultivation cost, crop yield, area under cultivation, rate of farmer suicide, growth rate of production, temperature, and rainfall are used in the above analysis.

More things can also affect how well crops grow. For instance, we've used data on rain and temperature to show how they affect crop production, price, etc. For much better results, you can also think about things like the type of soil, its PH value, the weather (winds, humidity), and the fertilisers you use. We can also improve the predictions and grouping by using some other algorithms.

Aside from these, you can also make predictions about other things, like how crops will grow the next year. If we can predict the things that affect crop production, we can make better plans for farming..

REFERENCES

[1] Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity —A review article. In: IJCST. 2011; 2(1).

[2] Gleaso CP. Large area yield estimation/forecasting using plant process models.paper presentation at the winter meeting American society of agricultural engineers palmer house, Chicago, Illinois. 1982; 14–17

[3] Majumdar J, Ankalaki S. Comparison of clustering algorithms using quality metrics with invariant features extracted from plant leaves. In: Paper presented at international conference on computational science and engineering. 2016.

[4] Jain A, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv. 1999;31(3):264–323.

[5] Jain AK, Dubes RC. Algorithms for clustering data. New Jersey: Prentice Hall; 1988.

[6] Berkhin P. A survey of clustering data mining technique. In: Kogan J, Nicholas C, Teboulle M, editors. Grouping multidimensional data. Berlin: Springer; 2006. p. 25–72.

[7] Han J, Kamber M. Data mining: concepts and techniques. Massachusetts: Morgan Kaufmann Publishers; 2001.

[8] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Paper presented at International conference on knowledge discovery and data mining. 1996

[9] Ramesh D, Vishnu Vardhan B. Data mining techniques and applications to agricultural yield data. In: International journal of advanced research in computer and communication engineering. 2013; 2(9).

[10] MotiurRahman M, Haq N, Rahman RM. Application of data mining tools for rice yield prediction on clustered regions of Bangladesh. IEEE. 2014;2014:8–13.

[11] Verheyen K, Adrianens M, Hermy S Deckers. High resolution continuous soil classification using morphological soil profile descriptions. Geoderma. 2001;101:31–48.

[12] Gonzalez-Sanchez Alberto, Frausto-Solis Juan, Ojeda-Bustamante W. Predictive ability of machine learning methods for massive crop yield prediction. Span J Agric Res. 2014;12(2):313–28.

[13] Pantazi XE, Moshou D, Alexandridis T, Mouazen AM. Wheat yield prediction using machine learning and advanced sensing techniques. Comput Electron Agric. 2016;121:57–65.

[14] Veenadhari S, Misra B, Singh D. Machine learning approach for forecasting crop yield based on climatic parameters. In: Paper presented at international conference on computer communication and informatics (ICCCI-2014), Coimbatore. 2014.

[15] Rahmah N, Sitanggang IS. Determination of optimal epsilon (Eps) value on DBSCAN algorithm to clustering data on peatland hotspots in Sumatra. IOP

conference series: earth and environmental. Science. 2016;31:012012.

[16] Forbes G. The automatic detection of patterns in people's movements. Dissertation, University of Cape Town. 2002.

[17] Ng RT, Han J. CLARANS: A Method for Clustering Objects for Spatial Data Mining. In: IEEE Transactions on Knowledge and Data Engineering. 2002; 14(5).

[18] Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. Wiley. 1990. doi:10.1002/9780470316801.

[19] Multiple linear regression - http://www.originlab.com/doc/Origin-Help/Multi-Regression-Algorithm. Accessed 3 July 2017.

[20] Elbatta MNT. An improvement for DBSCAN algorithm for best results in varied densities. Dissertation, Gaza (PS): Islamic University of Gaza. 2012

[21] Kirkl O, De La Iglesia B. Experimental evaluation of cluster quality measures. 2013. 978-1-4799-1568-2/13. IEEE.

[22] Meila M (2003) Comparing clustering. In: Proceedings of COLT 2003.