

Design and Analysis of a Prediction Model for Crop Yield Production in Agriculture

Dr. Yojna Arora, Ms. Ashima Gambhir

ABSTRACT- Big Data Analytics has its importance in almost all the real time applications ranging from Science & Technology to Market Prediction to Weather forecasting and many more. One such important area where analytics played a crucial role is Agriculture. Understanding the previously available agricultural data and studying the underlying pattern can help in making accurate future predictions. A well-known fact that the majority of population ($\geq 55\%$) in India is into agriculture, the analytics and prediction can be really useful to them. Various factors are to be considered which have direct impact on the production of the crops. The use of technology in agriculture has increased in recent year and data analytics is one such trend that has penetrated into the agriculture field. In this paper, the data of Government Crop Production is analyzed on various parameters which will help in predicting the Crop Yield.

KEYWORDS- Big Data, Data Mining, Data Analytics, Agriculture, Prediction Model, Crop Prediction, Crop Yield

I. INTRODUCTION

For global food production, crop yield prediction is of great importance. To improve national food security, policy makers rely on accurate forecasts to make timely import and export decisions. To breed better varieties, seed companies need to predict the success of new hybrids in different environments. However, because of various complex variables, crop yield prediction is highly challenging. For example, for each plant individual, genotype information is typically represented by high-dimensional marker data, containing many thousands to millions of markers [2]. The impact of genetic markers, which may be subject to interactions with various environmental factors and activities of field management, must be calculated [1].

More recently, machine learning methods, including multivariate regression, decision tree, association rule mining, and artificial neural networks, have been introduced for crop yield prediction [3]. In this paper, a new prediction model is developed for crop yield production in agriculture. Many Crop Yield prediction models are also implemented in literature by various researchers [4], [5], [6]

II. METHODOLOGY

Machine learning mainly consists of three learning methods, namely supervised learning, reinforcement learning and unsupervised learning methods of training a model [7].

Supervised learning is a form of learning that maps existing knowledge. Input has resulted in output mapping from input to output. But in case of unsupervised learning, we will not realize targeted learning in the event of unsupervised learning. In this learning, we should train the model to get production in order to get output required.

A. Dataset Description

The dataset of Crop Production in India (District-wise and State-wise) is analyzed and various outcomes are plotted. The data consists of crop production details of different states and districts containing various fields like state name, district name, crop year, season, type of crop, area of production etc. It contains approximately 2.5 lakhs of data records which are to be analyzed. The data analysis give some results which helps to identify answers to various questions such as how India's each state is involved in agriculture and also to identify which state and crop has highest crop production etc.

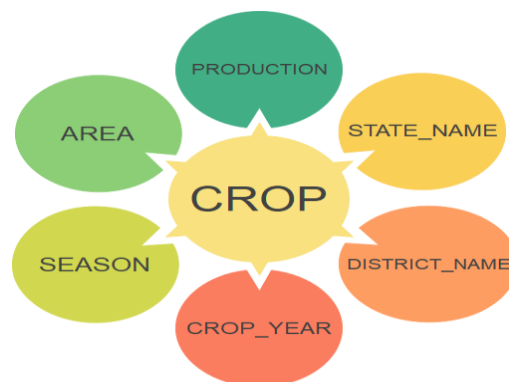


Fig 1: Block Diagram of Crop Production Data

Manuscript received September 20, 2020

Dr. Yojna Arora, Department of Computer Science & Engineering, Amity University, Gurugram, Haryana, India, (email: Yojana183@gmail.com)

Ms. Ashima Gambhir, Department of Computer Science & Engineering Department of CSE, Amity University, Gurugram, Haryana, India

This paper contains the study of various kinds of crop production with their different characteristics. This project is divided into three major groups (subprojects) which have different functionality from one another. The proposed work has been done with the help of PYTHON because each scope of the subprojects has different functionality. That output can be used to get a graphical representation.

Dataset is downloaded from official website of government. This agriculture dataset includes several records which describes the crop production details of various districts of different states from 1997 to 2013. This data set contains Crop Data which is collected from different districts. The dataset includes State_Name, District_Name, Crop_Year, Season, Crop, Area, Production details.

Dataset is analyzed in various aspects using Python based on the following questions:

a) How India's each state is contributing in agriculture?

Table 1: State Wise Crop Production

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Areca nut	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
5	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168.0	65100000.0
6	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36.0	100.0
7	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1.0	2.0
8	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5.0	15.0
9	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40.0	169.0
10	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Areca nut	1254.0	2061.0
11	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2.0	1.0
12	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83.0	300.0
13	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Cashewnut	719.0	192.0
14	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Coconut	18190.0	64430000.0

b) Production details of various crops

Table 2: Crop Production Data

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Areca nut	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
5	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Coconut	18168.0	65100000.0
6	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Dry ginger	36.0	100.0
7	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sugarcane	1.0	2.0
8	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Sweet potato	5.0	15.0
9	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Tapioca	40.0	169.0
10	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Areca nut	1254.0	2061.0
11	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Other Kharif pulses	2.0	1.0
12	Andaman and Nicobar Islands	NICOBARS	2001	Kharif	Rice	83.0	300.0
13	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Cashewnut	719.0	192.0
14	Andaman and Nicobar Islands	NICOBARS	2001	Whole Year	Coconut	18190.0	64430000.0

c) Which crop has highest and lowest production and cultivation cost from 1997-2013?

Table 3: Cost of Production & Cultivation

Crop	State	Cost of Cultivation (Hectare) A2+FL	Cost of Cultivation (Hectare) C2	Cost of Production (Quintal) C2	Yield (Quintal Hectare)	
0	ARHAR	Uttar Pradesh	9794.05	23076.74	1941.55	9.83
1	ARHAR	Karnataka	10593.15	16528.68	2172.46	7.47
2	ARHAR	Gujarat	13468.82	19551.90	1898.30	9.59
3	ARHAR	Andhra Pradesh	17051.66	24171.65	3670.54	6.42
4	ARHAR	Maharashtra	17130.55	25270.26	2775.80	8.72

d) Which state has highest and lowest production and cultivation cost in India?

III. RESULTS AND DISCUSSION

The dataset includes several records which are extracted for the analysis. The dataset includes various crops such as wheat, maize, cotton, paddy, rice etc. Therefore the crop production analysis has been done in which state the crops has been produced.

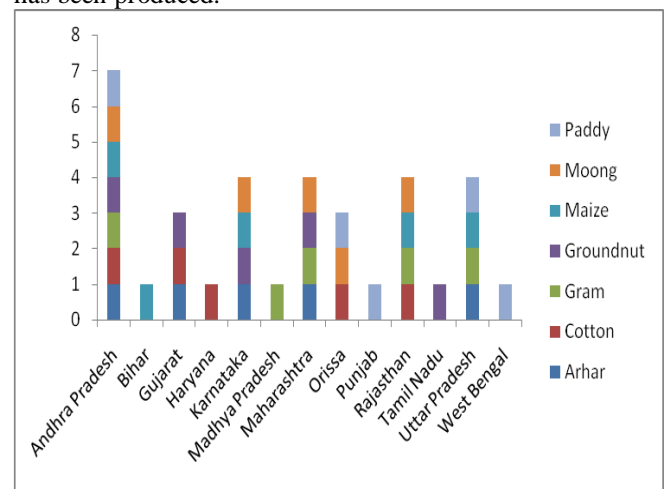


Fig 2: State wise crop production

Through the graphical representation it can be easily analyzed that which crop has been produced in which state. 10 crops are produced around the states. Following outcome can be drawn from the analysis:

- Andhra Pradesh is the region in which the rapeseed, mustard and wheat is not produced.
- Bihar is the only state in which single crop has been produced from 1997-2013 which was maize.
- Mustard and cotton is produced in Haryana from 1997-2013.
- Also, West Bengal is one of the region in which single crop has been produced namely paddy.
- Andhra Pradesh is the only state which has the maximum crop production from 1997-2013.

The crop production details of various crops are also analyzed using Python. Dataset includes production details from 1997-2013.

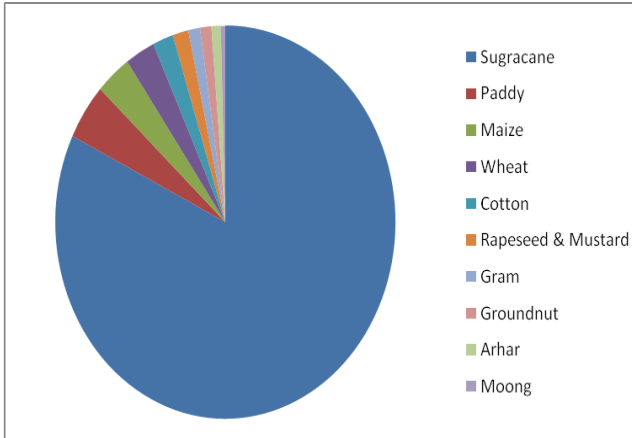


Fig 3: Crop wise Production

In the figure 2, the percentage of the crop production is mentioned that how much crop has been produced. The crop is measured in hectares or quintals.

- Sugarcane is highest produced crop since 1997-2013 about 82.2%.

In the figure 3, the cultivation cost is mentioned. Through which it can be easily analyzed that how much cultivation cost is required for the production of crop.

- The maximum cultivation cost is required for sugarcane and the minimum cultivation cost is required for moong.

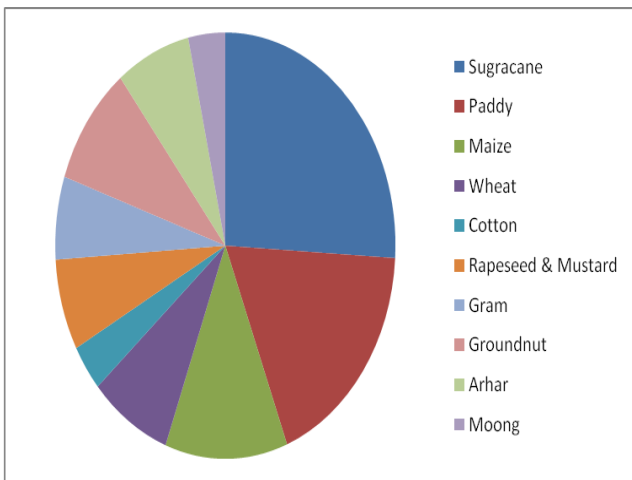


Fig 4: Crop wise Cultivation Cost

In the figure 4, the production cost is mentioned. Through which it can be easily analyzed that how much cost is required to produce the crop.

- The maximum production cost is required for moong and the lowest production cost is required for wheat since 1997-2013.

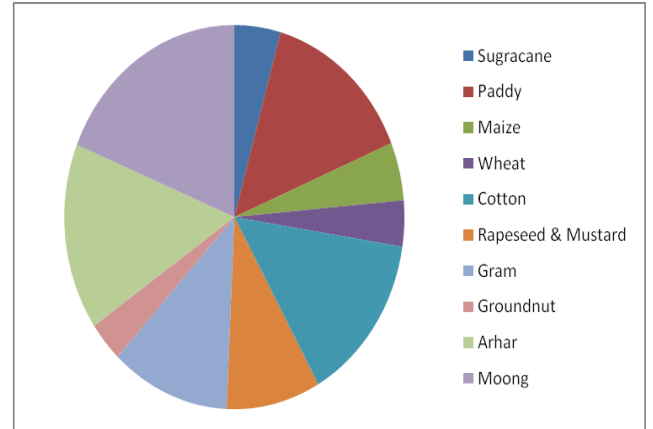


Fig 5: Crop wise Production Cost

Lastly, the highest and lowest production and cultivation cost was analyzed from the dataset. Following observations were made.

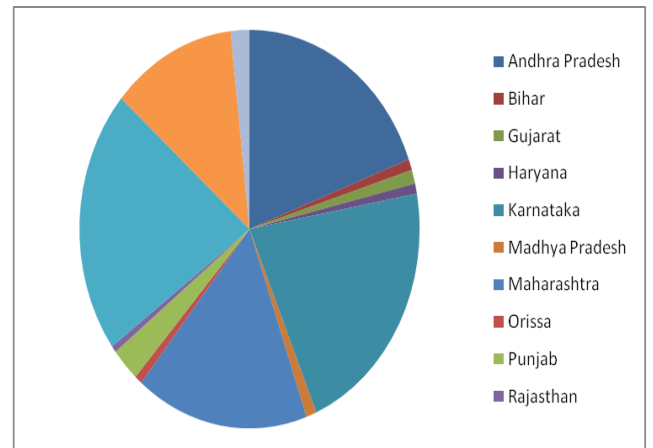


Fig 6: State wise Crop Production

In the figure 6, the production of crop is mentioned according to several states. The crop production of various crops is analyzed in different states to infer the following:

- Karnataka and Tamil Nadu have the maximum crop production in India. Both the states have the equal proportion of crop production, whereas Rajasthan has the lowest crop production in India since 1997-2013.

In the figure 7, the cultivation cost is analyzed. Observations inferred are as follows:

- Andhra Pradesh needs the maximum cultivation cost and Bihar has the minimum cultivation cost from 1997-2013.

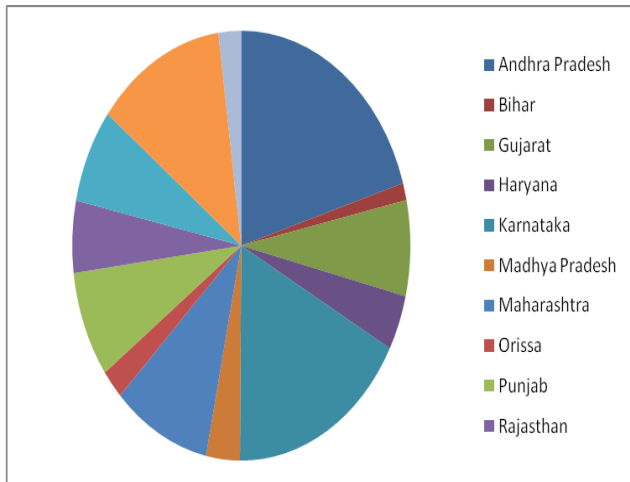


Fig 7: State wise Cultivation Cost

In the third diagram, the production cost is analyzed to compare the cost production cost.

- Andhra Pradesh has the maximum production cost 18.31% whereas the states which has 0.5% production cost has the minimum production cost.

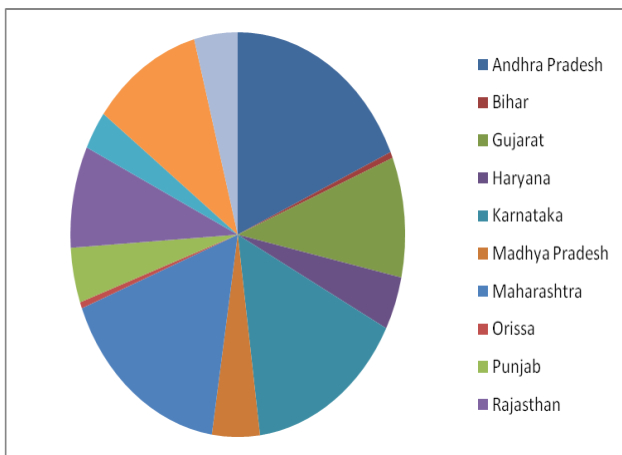


Fig 8: State wise Production Cost

IV. CONCLUSION

In a nut-shell Effective business decisions are mainly focused on accurate agriculture data and reliable analysis. The Farmers and other stakeholders have started using technology to get benefits such as cost savings, reduction in fertilizer, yield optimization, etc., In this paper, Agriculture data of different districts of different states is analyzed using Python based on various aspects such as percentage of crop produced in various season, Production details of various crops, which state has highest and lowest yield crop production cost, crop production in different years from 1997 to 2013. The dataset is also analyzed to list out the low production crops, average production crops and high production crops using Python. This business intelligence will help in better decision making and understanding in te field of Agriculture.

REFERENCES

- [1] Abhishek, K., Singh, M., Ghosh, S., and Anand, A. (2012). Weather forecasting model using artificial neural network. *Procedia Technol.* 4, 311–318. doi: 10.1016/j.protcy.2012.05.047.
- [2] Saeed Khaki, LizhiWang,” Crop Yield Prediction Using Deep Neural Networks, *Industrial and Manufacturing Systems Engineering*, Iowa State University, Ames, IA, United States.
- [3] Liu, J., Goering, C. E., and Tian, L. (2001). A neural network for setting target corn yields. *Trans. ASAE* 44, 705. doi: 10.13031/2013.6097
- [4] Michael gurstein, “A decision support system to assist the rural poor in Bangladesh”, *IEEE TECHNOLOGY AND SOCIETY MAGAZINES*, September 2013.
- [5] Luke Bornn, James V. Zidek, “Efficient stabilization of crop yield prediction in the Canadian Prairies”, *Elsevier*, vol. 0, no. 0, pp. 223-232, 2012
- [6] Yung-Hsing peng, Chin-Shun Hsu, and Po-Chuang Huang, “Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets” Nov. 20-22, 2015.
- [7] Yojna Arora, Abhishek Singhal & Dr. Abhay Bansal,”A Study of Applications of RBF Network”, *International Journal of CmputerApplications*, Volume 94, Issue 2, 2014.