Crop Data Analysis of Indian Region using Big Data Techinques for aiding Indian Farmers with Emergencies Alert

By

Ashutosh Anand,

Vineet Rathore,

Aman Sharma

Abstract:

Agriculture or farming is the largest contributor to the GDP of our country. Approximately 70 percent of rural household's primary source of income is agriculture. But still farmers don't get results worth their efforts. As a matter of fact, the condition of farmers in our country is horrible. Even though the agriculture provides whopping 50 percent of the total employment in India, it only constitutes about 15.4 percent of the Gross Domestic Product (GDP). Due to variations in climatic conditions, improper crop selection or irrigation often leads to less yield than expected. On the other note, in the field of Information Technology, Big Data and Machine Learning has come forth as a blazing topic. So by using Big Data approach in analysing the crop data over various factors such as soil type, temperature, water level, humidity, soil pH, fertilizers, we will devise an algorithm that will predict which crop should be sown in which season that both the yield and the profit can be maximized. For fulfilling the above task, we have used Hadoop framework, Machine Learning and NoSQL Database. Furthermore, even if the above factors are kept in mind, we can not deny or predict the forces of nature. One example of such are natural disasters. It is a known fact that most regions of India are prone to natural calamities. Even though farmers do everything right but due to these calamities not only they lose their properties and yield but also their lives. So as an attempt to minimize these circumstances, this paper proposes a very effective and economical way of alerting to all kinds of security emergencies and the good aspect of this is that it does not require any additional efforts. Including this feature with the help of APIs from Indian Meteorological Department has devised another use case, which is, that, it can also provide weather data. Hence, the farmer can also choose an optimal time for sowing a particular crop based upon its water requirements and rain. This security feature is incorporated within the crop analysis and selection project which makes this a complete package for helping our farmers. Thus, it will maximize the yield and profit on one hand and try reducing the loss of lives and properties.

1. Introduction

The Big Data Analysis is the process of inspecting, cleansing and simulating the data with the intension of discovering underlying information and getting the conclusions. It is the procedure of examining, extracting and envisioning the relevant information from the vast data in order to find some pattern. These operations are performed by the companies to turn the raw data of their customer into serviceable information. This analysis can also be used in the field of Agriculture. Most farmers generally rely on their previous experiences on a particular crop to expect a higher yield of that crop in the next harvesting season. But still the they don't get result/income worth their efforts. The above listed problems are the recurring problems that all the farmers have to face. But we also know that during modern times, our environment is affected by global warming, pollution, acid rain, bio-waste and e-waste, so conditions keep changing drastically, and therefore traditional farming techniques and only past experiences are not sufficient. Hence, there is a need new and more reliable farming techniques that can meet the requirements of the country.

As a result, Government is also taking many initiatives to improve farming. One of them is it is making data of previous farming records. But the problem with it is that the record or the dataset has become very large and can not be handled or analysed accurately with old methods. Furthermore doing so is also a very time-consuming process. That's where Big Data comes into play. By using Big Data Analysis, we can analyse large amount of data and find the underlying pattern within it precisely and efficiently. Thus, the Crop Analysis and Selection part of this project is being implemented with the help of Big Data and Machine Learning and the Emergency Alert part of this project is implemented by the information and data provided by the Government Websites, Indian Meteorological Department and others in the form of APIs, text or any other means. The meaningful data is extracted to it and used as per the requirements.

2. Literature Review

In today's world, more and more people are using electronic devices, surfing the web and using the social medias, thereby generating vast amount of data which mainly compromises of semi-structured and unstructured data. But the problem is that the data is highly unstructured and is present in large amount, so analyzing them by traditional means is a tedious and time-consuming task without high chances of errors in outcomes. Fortunately, this problem in modern world can be solved to a great extent by using Big Data. Big Data is more real in comparison to another techniques for processing huge data. Enormous data processing and scaling out of schemas in order to form units which are compatible for big data applications.

The Government of India has created an open data ecosystem for the motive of sharing crops dataset under the National Data Sharing and Accessibility Policy (NDSAP) which was initiated by Open Government Data (OGD) to make the data more transparent and easily accessible to the citizens.

Indian Prime Minister Narendra Modi led government has asked Indian Council of Agriculture Research (ICAR) to prepare a chairmanship of Praveen Rao, comprehensive crop plan for India. A committee under the Vice-Chancellor, Prof. Jayashankar who hails from Telangana State Agriculture University, has been constituted for this purpose, Trilochan Mohapatra, Director General, ICAR said, "The committee is expected to come out with a set of crop recommendations for each geography after considering the climatic conditions, soil health, water stress and the estimated short-term as well as long term demand for the produce within the country and globally".

The need to realign India's agricultural practices and other related systems was one of the major recommendations by the high-level committee which was given the task of searching various ways to increase the farmers' income at least by two-fold. The technical, scientific and economical feasibility of the various crops will be looked at a microscopic level because of the agro-climatic diversity of the country, Mohapatra said.

According to him, water, the most critical element for any farming exercise, is fast becoming a scarce resource, and hence calls for urgent interventions at every level of agriculture. We have taken the dataset from the government website data.gov.in/crop. Our dataset consists of the data of all the states which consist of their districts and the production of crops with their respective seasons. The dataset further consists of area of the districts.

	A	B	C	D	E	F	G
1	State_Nam	District_Na	Crop_Year	Season	Crop	Area	Production
2	Andaman a	NICOBARS	2000	Kharif	Arecanut	1254	2000
3	Andaman a	NICOBARS	2000	Kharif	Other Khar	2	1
4	Andaman a	NICOBARS	2000	Kharif	Rice	102	321
5	Andaman a	NICOBARS	2000	Whole Yea	Banana	176	641
6	Andaman a	NICOBARS	2000	Whole Yea	Cashewnu	720	165
7	Andaman a	NICOBARS	2000	Whole Yea	Coconut	18168	65100000
8	Andaman a	NICOBARS	2000	Whole Yea	Dry ginger	36	100
9	Andaman a	NICOBARS	2000	Whole Yea	Sugarcane	1	2
10	Andaman a	NICOBARS	2000	Whole Yea	Sweet pota	5	15
11	Andaman a	NICOBARS	2000	Whole Yea	Tapioca	40	169
12	Andaman a	NICOBARS	2001	Kharif	Arecanut	1254	2061
13	Andaman a	NICOBARS	2001	Kharif	Other Khar	2	1
14	Andaman a	NICOBARS	2001	Kharif	Rice	83	300
15	Andaman a	NICOBARS	2001	Whole Yea	Cashewnu	719	192
16	Andaman a	NICOBARS	2001	Whole Yea	Coconut	18190	64430000
17	Andaman a	NICOBARS	2001	Whole Yea	Dry ginger	46	100
18	Andaman a	NICOBARS	2001	Whole Yea	Sugarcane	1	1

Figure 1: Initial Dataset

One of the most important aspect of analyzing a data requires cleaning of dataset. We removed "crop year" column because of some redundancy. We used the "pandas" which is a famous library of python that takes data in various formats (like a CSV or TSV file, or a SQL database, depending upon how the data are separated) and builds an object of python with rows and columns called dataframe that looks very similar to a table in a statistical software (or as generally seen in an excel sheet). It is mainly used for both data manipulation and data analysis.

We used group by function of pandas to group our states with their respective districts and seasons. We used aggregate function to retrieve the average of the area and the production of the crop in a particular season of a district. Further we added a new column named as "P/A" which shows about production per area of a particular of a respective season with districts.

1	State_Nan	District_Na	Season	Crop	Area	Production	PA	Price
2	Andaman a	NICOBARS	Autumn	Rice	3.5	10	2.857143	19412
3	Andaman a	NICOBARS	Autumn	Sugarcane	13.4	41.75	3.115672	19886
4	Andaman a	NICOBARS	Kharif	Arecanut	1254	2030.5	1.619219	14506
5	Andaman a	NICOBARS	Kharif	Other Khar	2	1	0.5	14347
6	Andaman a	NICOBARS	Kharif	Rice	80.205	217.7733	2.715209	14489
7	Andaman a	NICOBARS	Rabi	Arecanut	944	1610	1.705508	13490
8	Andaman a	NICOBARS	Rabi	Black pepp	23	8.5	0.369565	13748
9	Andaman a	NICOBARS	Rabi	Cashewnu	1000.5	260.5	0.26037	15457
10	Andaman a	NICOBARS	Rabi	Dry chillies	12	25	2.083333	19666
11	Andaman a	NICOBARS	Rabi	Dry ginger	7	9.64	1.377143	18108
12	Andaman a	NICOBARS	Rabi	Maize	3.84	18.22	4.744792	19853
13	Andaman a	NICOBARS	Rabi	Moong(Gr	1.5	1.1	0.733333	19731
14	Andaman a	NICOBARS	Rabi	Sweet pota	22	208	9.454545	13589
15	Andaman a	NICOBARS	Rabi	Turmeric	2	0.5	0.25	14898
16	Andaman a	NICOBARS	Rabi	Urad	1.5	1.16	0.773333	12672
17	Andaman a	NICOBARS	Whole Yea	Arecanut	1095.074	1222.863	1.116694	15717
18	Andaman a	NICOBARS	Whole Yea	Banana	219.2029	1295.589	5.910455	15031

Figure 2: Updated Dataset

Data visualization is very important to represent the features of data in graphical form to understand complicated relationship in data. Standardizing the data is essential need before visualization. We used MinMaxScaler of sci-kit learn library to standardize the dataset. We plotted bar graphs using seaborn library.
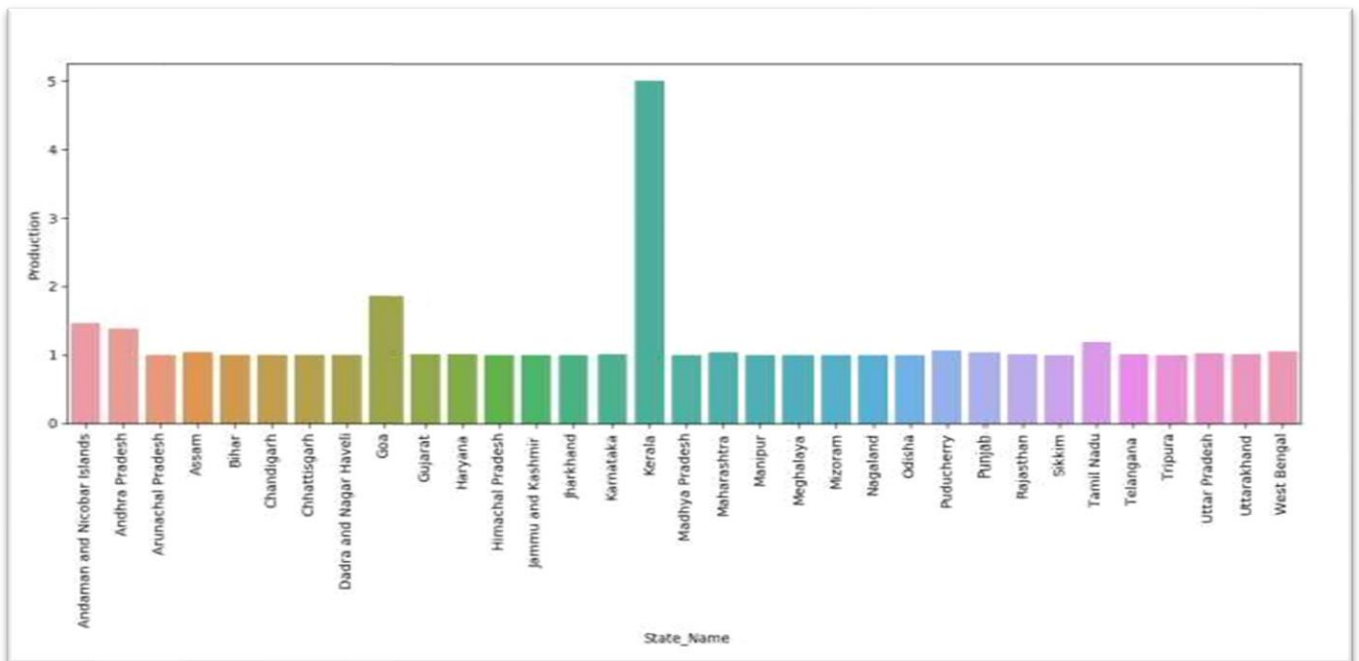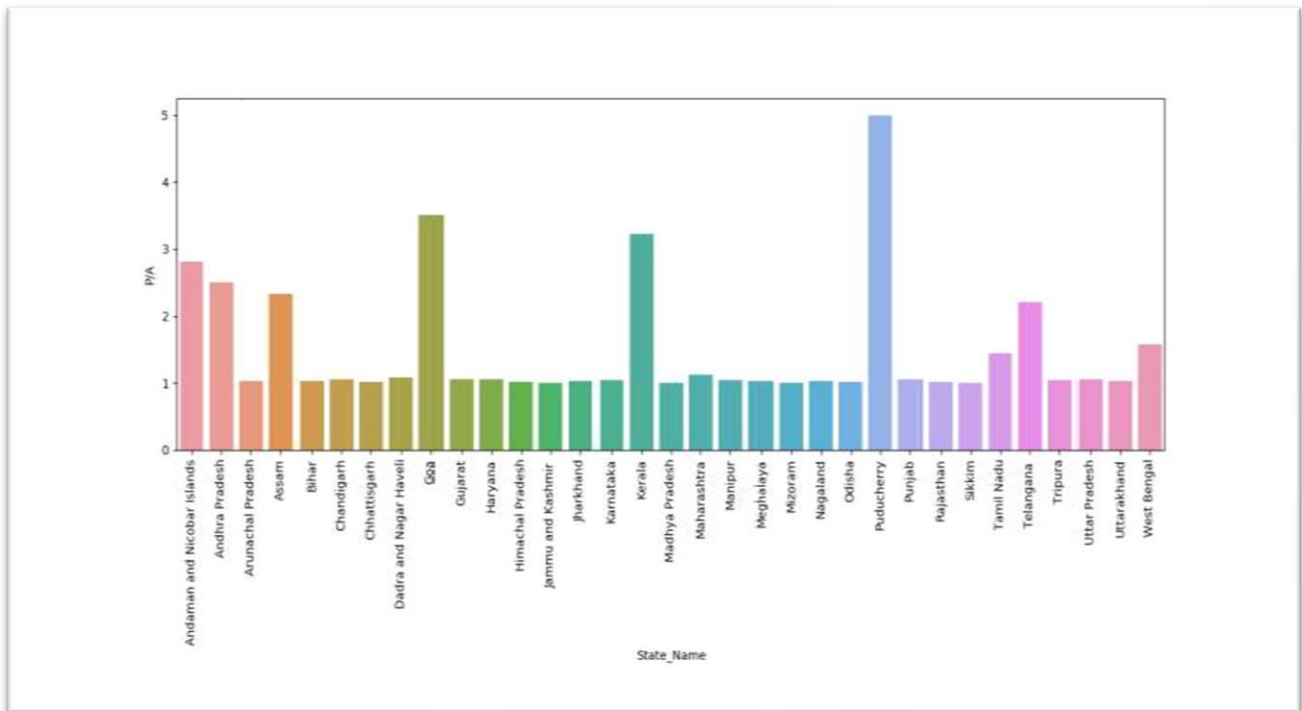

Figure 3: Graph of state name v/s production

Figure 4: Graph of state name v/s "P/A"

Here, P/A represents production of crop per unit area

The trend of social media is increasing rapidly. Almost all the peoples are active on social media platforms like Facebook, Instagram, Twitter etc. On these platforms we post text messages, images, audio files, video files. More than millions of people use these platforms it means more than billon of messages and to store these messages there is a database which is present. These messages can be stored in RDBMS but RDBMS is not efficient to store and process the data which is more than 1 million and this type of data is a unstructured data and RDBMS is not able to store unstructured data so here NOSQL databases comes into picture. NOSQL databases are efficient to store and process the unstructured, semi structured as well as structured data also.

MongoDB is used to store the data comprising of more than 33,000 documents.

Here, we are connecting our Mongoose database with node.js with the help of mongoose.js library and modeling (ODM) library that provides a meticulous training environment for the data, imposing schemas as per the requirement while still preserving the flexibility that makes MongoDB powerful and robust. After that we embed mongoDB queries in nodeJS by which it will process the input data given by the user and will print the desired output from database by taking the given input into account.

Additionally, to implement the other aspect of the project we took help of the information provided by the Government with respect to the weather, climate, etc. on their website/portals and mainly their APIs. A department of our Government only works on weather forecast, predicting calamities and so on.

The Indian Meteorological Department is an organization of the Ministry of Earth Sciences which comes under the Government of India. It is the chief corporation in charge of the meteorological observations, weather forecasting, seismology and other similar activities. By using the data provided by these agencies, we have incorporated this security alert feature. Whenever a farmer selects his region for crop selection, he will be notified about the conditions of his region within the same webpage. It would have been useful if was present during Karnataka Floods, Kerala Floods, Cyclone Fani, Bihar Floods and many more calamities.

Currently, it the midst of corona pandemic, it alerts the farmers how safe their zone is. Moreover, they are informed about various safety and measure and precautions. Also, whenever such a condition arises, Government announces various programs to help the people and they can be notified about it all in the same place. Similarly, it can work as same for any other calamity. As of now we are able to predict tsunamis, cyclones, etc. ahead of time and if any farmer is situated in its area of impact then they can be notified about the same. Another utilization of this project can be as weather forecast. This can not only predict calamities but also rain and sunshine. Farmers can know when its likely to rain in their region and when it's going to be sunny and thus can choose the time to sow seeds, plough fields, etc. accordingly.
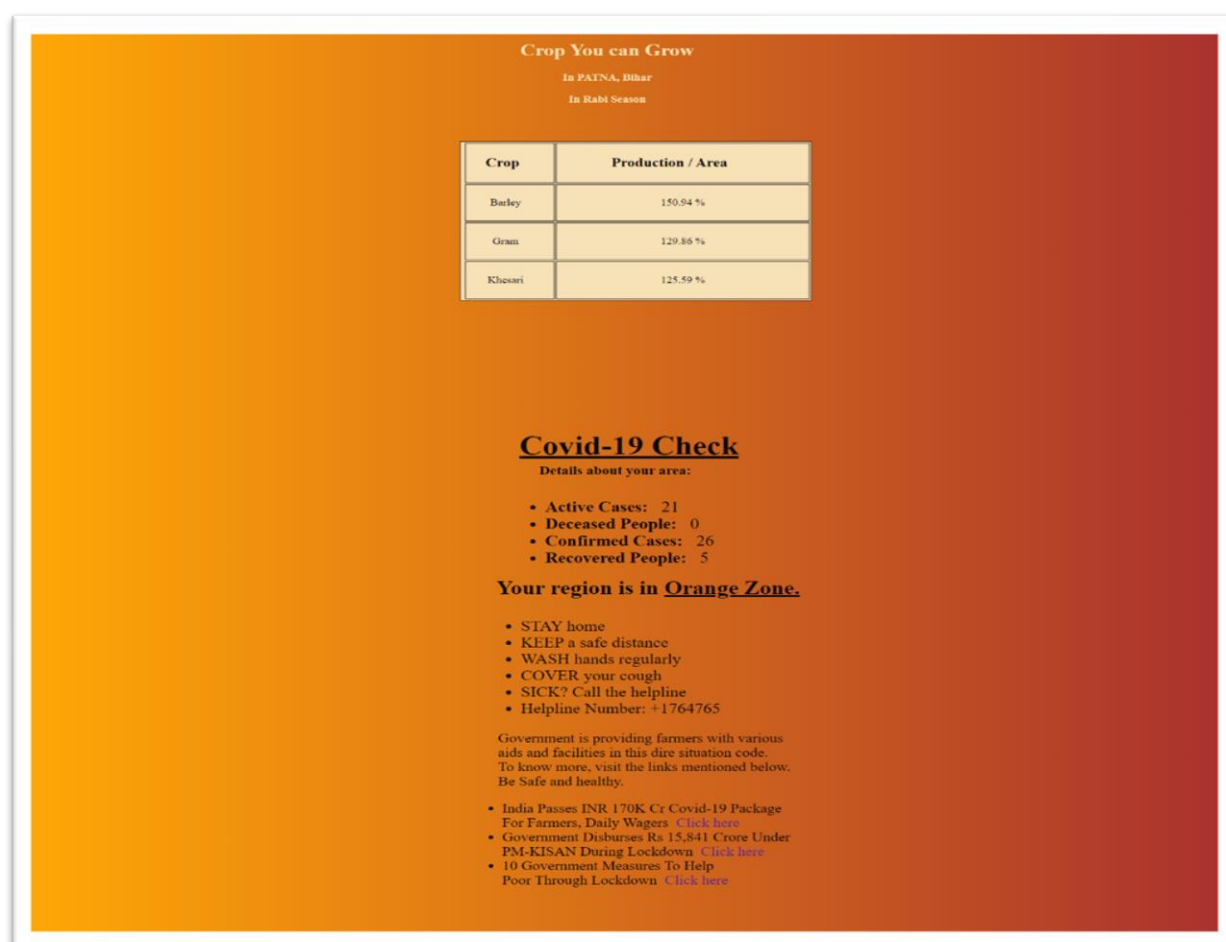

Figure 5: A screenshot from our project

3. Future Prospectives

The technology in our time is changing and improving fast. Life has come to smart phones and tablets from desktops and laptops and everyone own smart phones. As of now this application is only available for farmers as a web application but the android application of this project will also be made by which some more problems of farmers will get reduced. Android application will be based on location-based services, by which farmers will have to make less efforts, they won't need to type or select anything, they will just open up the application and on the basis of their location and choices, results will get displayed. Moreover, during an emergency they can be notified by push notifications and does not need upon the consumer to open the app to get informed and they can also be notified about the weather conditions of their regions periodically in a harvesting Season.

In future, on this project ML lib will be applied which is an apache spark's scalable machine learning library by using this library many algorithms like regression, clustering, collaborative and classification will get implemented on this application by which it will give better results and will predict more results.

References:

[1] Ch. Chandra Shekhar, J. Uday Kumar, B. Kishore Kumar, Ch. Shekhar, Effective use of Big Data Analytics in Crop planning to increase Agriculture Production in India.

[2] Open data ecosystem as per National Data Sharing and Accessibility Policy (NDSAP)

initiated Open Government Data (OGD) Platform, https://data.gov.in/catalogs/sector/Agriculture9212.

[3] M. Moorthy, R. Baby and S. Senthamaraiselvi, "An Analysis for Big Data and its Technologies", International Journal of Computer Science Engineering and Technology (IJCSET), vol. 4, no. 12, (2014) December, pp. 412-418.

[4] A. Pal, K. Jain, P. Agrawal and S. Agrawal, "A Performance Analysis of MapReduce Task with Large Number of Files Dataset in Big Data Using Hadoop", 4th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, doi: 10.1109/CSNT.2014.124, (2014), pp. 587-591.

[5] K. Grolinger, M. Hayes, M. Hayesm, A. L'Heureux and D. S. Allison, "Challenges for MapReduce in Big Data", 2014 IEEE World Congress on Services, June 27-July 2, IEEE Computer Society Washington, DC, USA © 2014, pp. 182-189.

[6] G. Nasrin Fathima, et al., "Agriculture Crop Pattern Using Data Mining Techniques",IJARCSSE, Volume 4, Issue 5, May 2014, ISSN: 2277 128X, pp. 781-786.

[7] Joseph O. Chan, "An Architecture for Big Data Analytics", IIMA, Volume 13 Issue 2, 2013, pp. 1-14

[8] Steve Sonka, "Big Data and the Ag Sector: More than Lots of Numbers", IFAMA, Volume 17 Issue 1, 2014, pp. 1-20

[9] IEEE International Conference by CSI on Big Data on IT in Business, Industry and Government (CSIBIG), March 8-9, 2014, ISBN: 978-1-4799-3063-0

[10] Anirban Mukherjee, et. al., "Shared Disk Big Data Analytics with Apache Hadoop", 978-1-4673-2371-0/12© 2012 IEEE

[11] Yang, et. al., "Data Clustering on Taiwan Crop Sales under Hadoop Platform", Proceedings of the Institute of Industrial Engineers Asian Conference 2013

[12] https://covid-19-india-data-by-zt.p.rapidapi.com/GetIndiaDistrictWiseDataForState