Agriculture Data Analytics in Crop Yield Estimation using IBM Cognos

Lydia D. Isaac

lydia@sonatech.ac.in
Sona College of Technology

1 INTRODUCTION

1.1 Overview

Crop production in India is one of the important sources of income and India is one of the top countries to produce crops. As per this project we will be analyzing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India.

1.2 Purpose

The Cultivation process is usually planned by the farmers based on their previous experiences. Sometimes they end up in cultivating undesirable crops as their predictions are not precise. Application of Artificial Intelligence in this field can bring a boom in the income by changing the scenario through cultivating the most optimal crops. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. There are many factors influencing crop yield. Since analysing numerical data as it is would be complex, creating a dashboard that has the visualizations of the entire crop cultivation details would provide more insights on crop production.

2 LITERATURE SURVEY

2.1 Existing problem

There are different forecasting methodologies developed and evaluated by the researchers all over the world in the field of agriculture.

B M Sagar, Cauvery NK,[1] has done a review on various data analytics methods that could be used in Agriculture domain. Their work suggests that Agriculture is important for human survival because it serves the basic need. A well-known fact that the majority of population (≥55%) in India is into agriculture. Due to variations in climatic conditions, there exist bottlenecks for increasing the crop production in India. It has become challenging task to achieve desired targets in Agri based crop yield. Various factors are to be considered which have direct impact on the production, productivity of the crops. Crop yield prediction is one of the important factors in agriculture practices. Farmers need information regarding crop yield before sowing seeds in their fields to achieve enhanced crop yield. 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. The main challenge in using big data in agriculture is identification of effectiveness of big data analytics. Efforts are going on to understand how big data analytics can agriculture productivity. The present study gives insights on various data analytics methods applied to crop yield prediction and also signifies the important lacunae points' in the proposed area of research.

Priyanka G [2] has applied Big Data Analytics. Big data analytics is the process of collecting, organizing and analysing large sets of data to discover patterns and other useful information. Big Data is usually defined by 3V's. They are Volume, Variety and Velocity. The main aim of this paper is to analyze the given soil and its corresponding properties. This information can be used to determine the soil fertility which is essential for agricultural purposes. The dataset contains details about various soil types of the world and its composition. This dataset

is the World Harmonized Soil Database and it is taken from FAO. In this project there are two analyses. First analysis is to determine the draining capacity of the soil based on the various levels of sand, silt, clay and gravel given in the dataset. The second analysis is to take the organic carbon content and pH level values from the dataset and determine which crop can be cultivated from the given soil sample.

Jharna Majumdar, Sneha Naraseeyappa & Shilpa Ankalaki [3] focuses on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using data mining techniques like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the large amount of existing crop, soil and climatic data, and analysing new, non-experimental data optimizes the production and makes agriculture more resilient to climatic change.

Nabila Chergui, M-Tahar Kechadi, Michael McDonnell [4] has introduced the crop yield management process and its components, and then we focus on the crop yield monitoring. We then present a classification of data mining techniques applied for the crop yield monitoring tasks. This is followed by discussing each category of the classification throughout panoply of existing works and show their used techniques, then we provided a general discussion on the applicability of big data analytics into the field of digital agriculture.

Arjun S, Anish Joshi, Pooja Das H, Amutha R [5] recognizes the culminations of conventional cultivating practices and delivers how to build the yield of the agrarian things by utilizing present day PC innovations. Further, it additionally recognizes the basic figuring and analytic capacity of Big Data in preparing colossal volumes of value-based information continuously circumstances. The target of this paper is to display the revisions in the rural area and supports the dialogs on how government can cultivate developments in the enormous information examination to enhance the rustic agricultural framework.

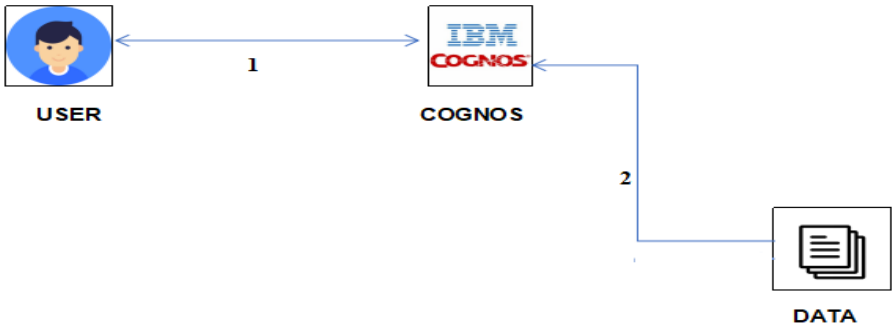
2.2 Proposed solution

Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions.

Given a framework that gathers the necessary data, the decision making to be performed requires knowledge extraction from these data. Towards this direction, we can use IBM Cognos, a web based integrated business intelligence suite by IBM. It is an analytical tool that integrates reporting, modelling, analysis, dashboards, stories and event management so that effective business decisions can be made.

3 THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

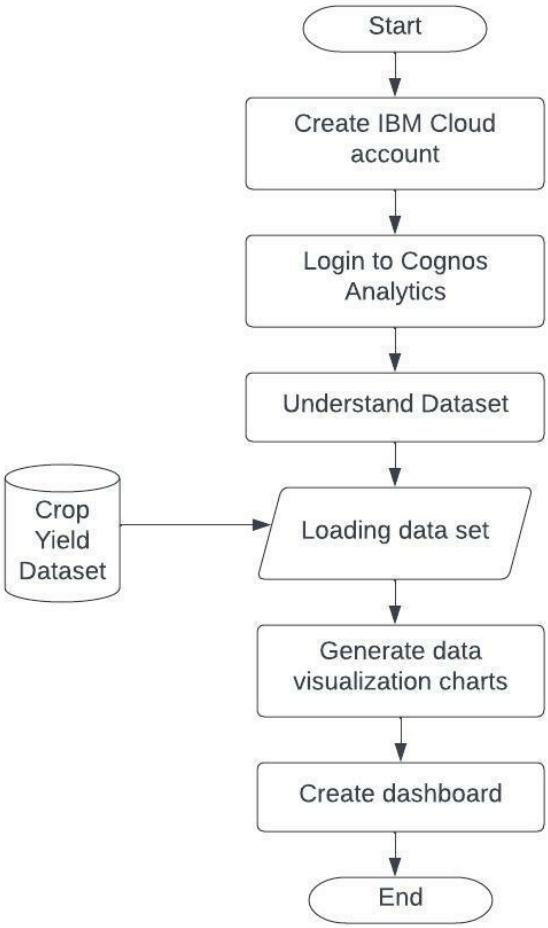
IBM Cloud
IBM Congnos

4 EXPERIMENTAL INVESTIGATIONS

- This dataset provides a huge amount of information on crop production in India ranging from several years.
- It has 2,46,092 data points (rows) and 6 features (columns) describing each crop production related details.
- Features in the dataset include:
    - State Name - All the Indian State names.
    - District Name -Different District names.
    - Crop Year- contains the crop years.
    - Season – Different seasons for crop production.
    - Area- Total number of areas covered.
    - Production- production of crops.

5 FLOWCHART

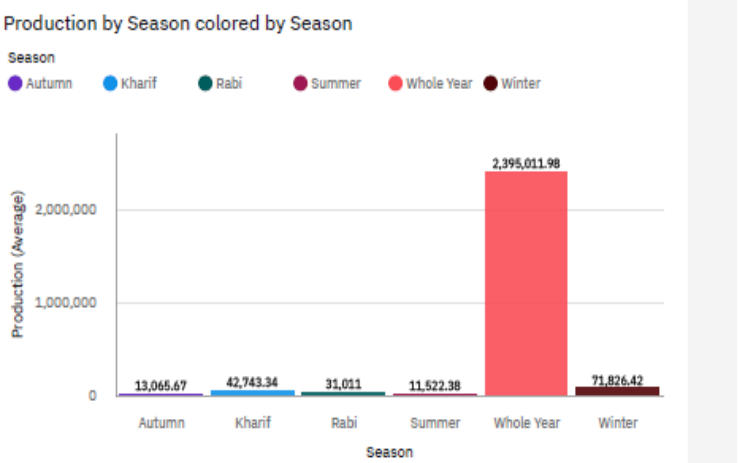
The flow chart given below describes the tasks to be done:



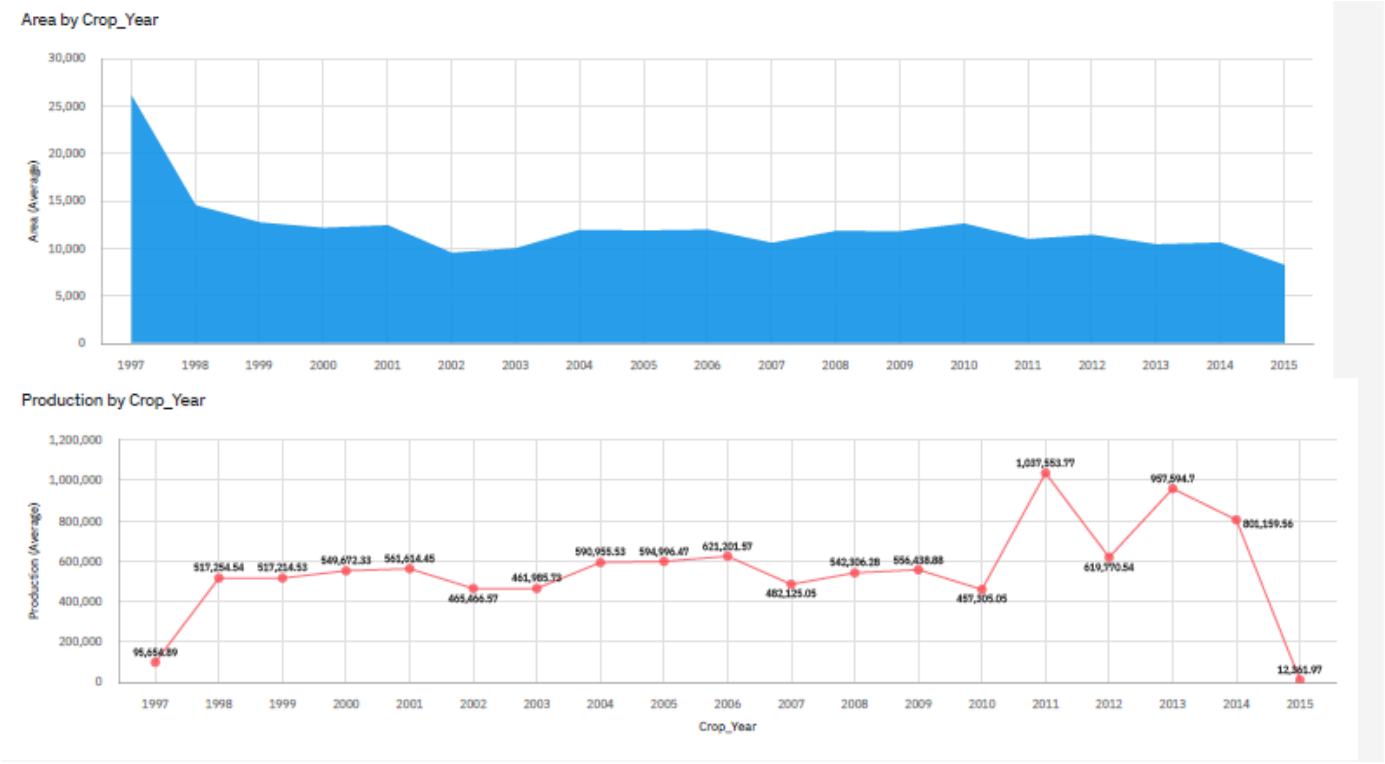
6 RESULT

The visualizations done:

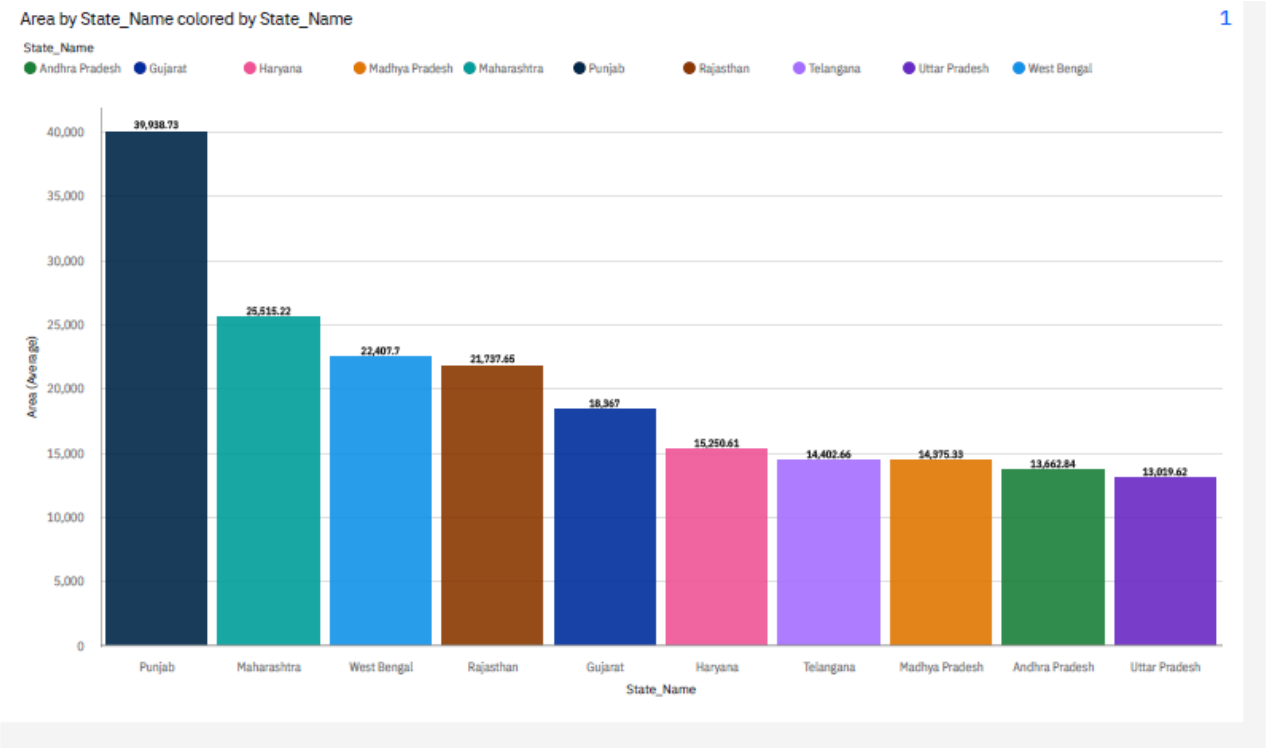
a) Seasons with Average Productions



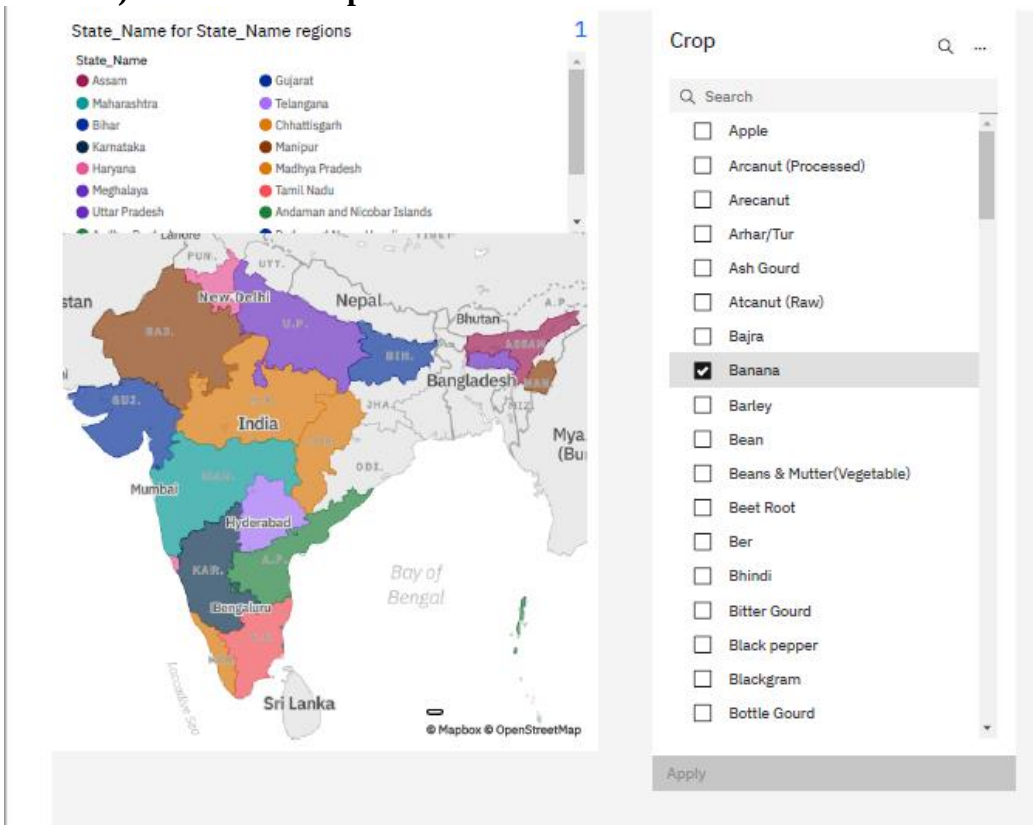
b) Years usage of Area and Production



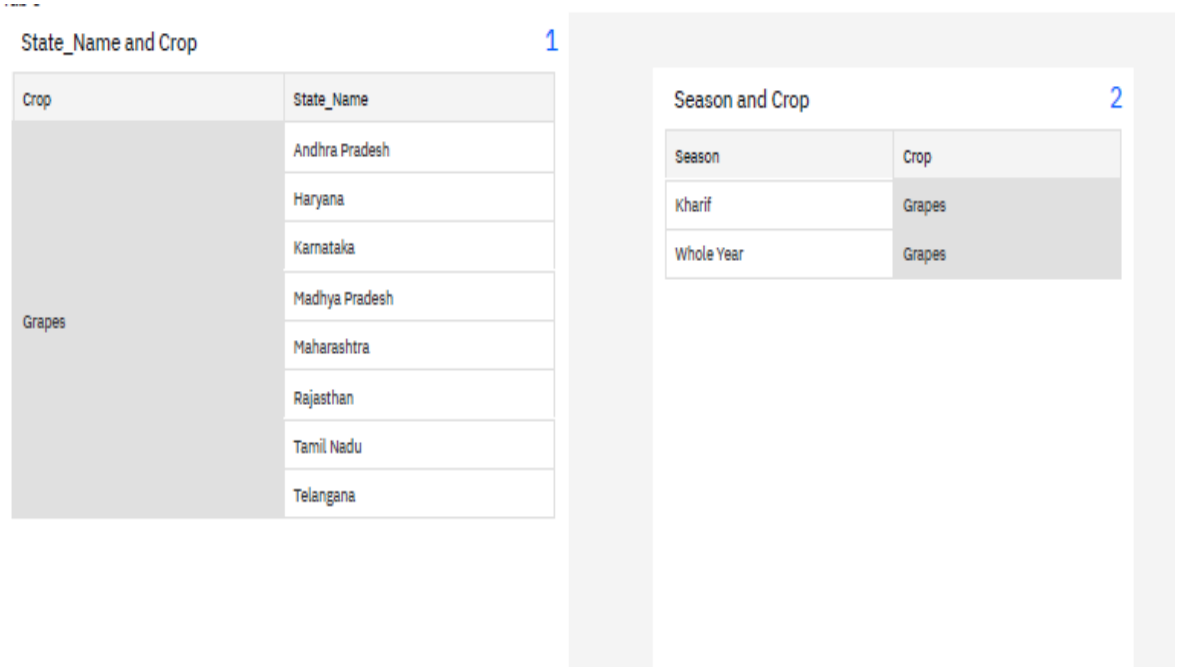
c) Top 10 states with Most area



d) State with Crop Production



e) States with the Crop Production along with season(Text table)



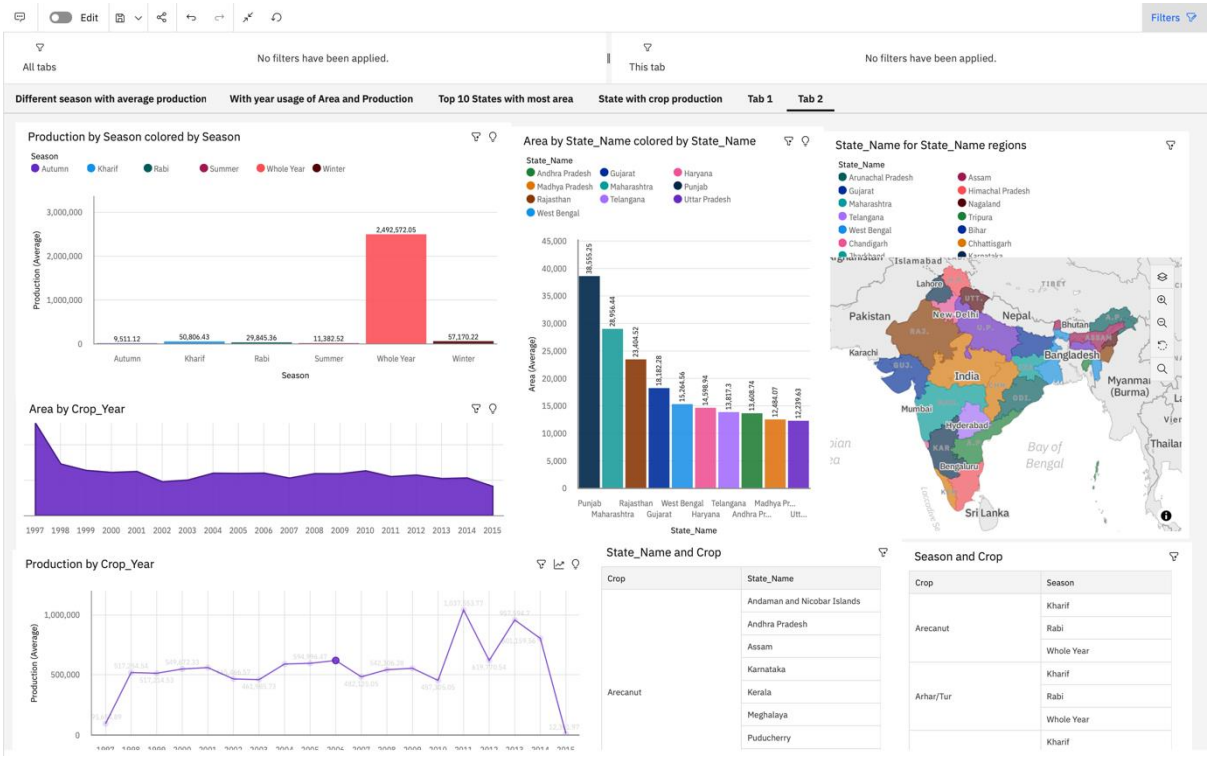
State_Name and Crop 1

Crop	State_Name
Grapes	Andhra Pradesh
	Haryana
	Karnataka
	Madhya Pradesh
	Maharashtra
	Rajasthan
	Tamil Nadu
	Telangana

Season and Crop 2

Season	Crop
Kharif	Grapes
Whole Year	Grapes

f) Final Dashboard created:



7 ADVANTAGES & DISADVANTAGES

- Cognos Active Reports are transportable. They can be e-mailed
- Cognos Active Reports are great for offline consumption. They can be executed while away from the office.
- Cognos Active Reports work with iPads
- Cognos Active Reports can be flashy.

8 APPLICATIONS

This can be used by farmers and business decision makers to analyze and investigate data sets and summarize their main characteristics. Also can be used to see what data can reveal beyond the formal modeling or hypothesis testing task and provides a better understanding of data set variables and the relationships between them. It can also help determine if the statistical techniques you are considering for data analysis are appropriate.

9 CONCLUSION

The proposed work deals with data visualization using IBM Cognos Analytics platform. Crop yield estimation dataset was used to perform the task. Creating a dashboard and by generating individual visualizations, we will get most of the insights of Crop production in

India. The dashboard created using this platform acts as a single point where each and every changes in the data could be analysed.

10 FUTURE SCOPE

The proposed work involves visualization of the dataset. By applying various Machine Learning models we can use the analytics report to predict the yield in the future.

11 BIBILOGRAPHY

1. https://www.researchgate.net/publication/329467349_Agriculture_Data_Analytics_in_Crop_Yield_Estimation_A_Critical_Review
2. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3423354
3. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0077-4
4. https://ieeexplore.ieee.org/document/9151851
5. https://www.ijert.org/research/big-data-analytics-for-agricultural-development-in-india-IJERTCONV4IS22002.pdf
6. https://www.ibm.com/academic/badges/improving-healthcare-the-role-of-cloud-computing
7. https://www.ibm.com/academic/badges/improving-healthcare-the-role-of-data-analytics-and-data-science
8. https://cognitiveclass.ai/courses/python-for-data-science
9. https://cognitiveclass.ai/courses/data-analysis-python
10. https://cognitiveclass.ai/courses/data-visualization-python
11. https://smartbridge.teachable.com/courses/enrolled/1830742

APPENDIX

A. Source Code

Only visualizations are done hence source code is not applicable for this proposed work.