

Data Mining Technique to Predict Annual Yield for Major Crops

Rajshekhkar Borate¹ Rahul Ombale² Sagar Ahire³ Manoj Dhawade⁴ P.S. Kulkarni⁵

^{1,2,3,4,5}Department of Computer Engineering

^{1,2,3,4,5}NBN Sinhgad School of Engineering, Pune-411041

Abstract— The complexity of predicting the best crops is highly due to unavailability of proper knowledge discovery in crop knowledge base which affects the quality of prediction. However, Clustering is an important step in mining useful information. There are multiple clustering methods such as partition, hierarchical, model based grid-based. Constrained-based which make this task complicated due to problems related to optimization and noise. In this paper k-means clustering algorithm issued in solving the partition problem which led to select for performance evaluation and linear regression algorithm in order to get good quality of clusters for crop prediction. This project aimed to apply new data mining techniques on dataset to establish meaningful relationships can be found.

Key words: Data Mining, Crop Analysis, Yield Prediction, Clustering, K-Means, Linear Regression

I. INTRODUCTION

While the agriculture remains the main occupational sector, rural demand has witnessed a slowdown leading many to point over the simmering agrarian crisis. Agriculture growth rate in India has not crossed 3% in the last two years. However, West Asia has been a star performer in this region with 6% growth rate. It should be noted here that Half of India's labour is in agriculture. But, there is no suitable agriculture policy yet. Hence, according to experts, the present state of Indian agriculture is largely due to policy paralysis that's why there is necessity of making better policy for farmers in the sense of their farming. Ultimately the policy making is depend on annual yield of the different geographic area as well as particular climate .Hence we are using the predictive data mining algorithms and techniques to predict annual yield of the country

In our project research, we have considered various environmental and the areas of production are the factors for crop in India. Taking this into consideration we developed a database for various districts, for this we applies clustering techniques to partition regions, and then we apply suitable predictive algorithm to obtain crop yield predictions.

II. RELATED WORK

Bangladeshi student [1] propound Data mining techniques and algorithms to predict annual yield of crops and suggest planting different crops in different zones in Bangladesh. They considered the effects of environmental (weather), biotic (pH, soil salinity) and area of production as factors towards crop production in Bangladesh. Taking these factors into consideration as datasets for various districts, they applied clustering techniques to divide regions, and then they apply suitable classification techniques to obtain crop yield predictions.

In [2] Raorane A.A. and Kulkarni .V. Research paper they focus on the data mining as the tool for yield estimation. They said actually accurate information about the nature of historical yield of crop is important modeling

input, which are helpful to farmers & Government organization for decision making process in establishing proper policies.

In [3] Utkarsha P. Narkhede and K. P. Adhiya have research paper A Study of Clustering Techniques for Crop Prediction

In this review paper there is a comparative study of clustering algorithms. They said improved k-mean clustering algorithm is outstanding in solving the optimization problem.

In [4] M.C.S Geetha, study different data mining techniques in agriculture. Research paper aims of finding suitable data models which achieves a high precision and a high generality with respect to parameters namely rainfall, year and production. In research paper discusses applications of data mining techniques used in agriculture domain.

III. MOTIVATION

Agriculture plays a vital role in India's economy. Farm crisis in India is deepening, mainly due to back-to-back monsoon failures and falling crop prices. One indicator of the growing agrarian distress is farmer suicides. Reason behind this is agricultural organisation failed in policy making. So solution to this issue we are making a data mining tool that help to predict annual yield of crop depends on their environmental as well as crop zone area. So as the demand of food is increasing, the researchers, farmers, agricultural scientists and government are trying to put extra effort and techniques for more production. And as a result, the agricultural data increases day by day. As the volume of data increases, it requires involuntary way for these data to be extracted when needed.

Still today, a very few farmers are actually using the new methods, tools and technique of farming for better production. Data mining can be used for predicting the future trends of agricultural processes.

IV. SYSTEM ARCHITECTURE

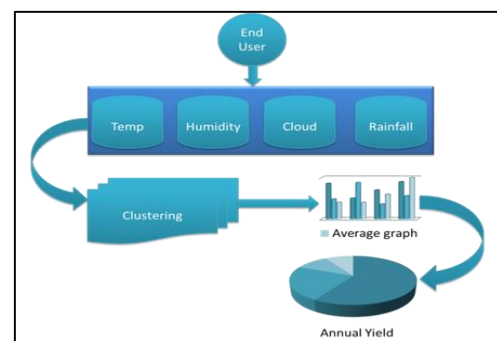


Fig. 1: System Architecture

V. METHODOLOGY

The method of our project is initially divided into two major parts: (1) Clustering (2) Classification.

A clustering of the selected districts. In our project, we have considered a total of 27 districts. In order to group the districts into distinct clusters, the assumption that we had to use was that the districts containing the similar values of relevant attributes should belong to the same cluster. According to this assumption, we categorized our selected attributes for the consideration of clustering the districts. For clustering we use K-means algorithm.

A. K-means:

The k-means algorithm captures the insight that each point in a cluster should be near to the center of that cluster. It works like this: first we choose k, the number of clusters we want to find in the data. Then, the centers of those k clusters, called centroids, are initialized in some fashion

We recalculate each centroid's location as the mean (center) of all the points assigned to its cluster. We then iterate these steps until the centroids stop moving or equivalently until the points stop switching clusters.

B. Linear Regression:

Regression is function that relates a data item to a real-valued prediction variable. In this case regression help to predict annual yield. It is a statistical measure that can be used to determine the strength of the relationship between one dependent variable and a series of other.

VI. DATA SET

The source of the dataset used in this project is Bangladesh agriculture research council. From the dataset, we pre-processed and selected only the attributes which are important for our project rainfall, temperature humidity etc. And also cultivated area for crop considered according to the districts.

VII. RECOMMENDED SYSTEM

After getting all the result graph charts and tables, then we have write a program which takes into account the necessary tables from our results as well as it will show the annual yield for user entered area also but that is approximately yield which compare the environmental conditions with pre-existed dataset and show result graph. These recommendations are based on a combination of annual yield of that crop species per hectare area of a district.

VIII. TEST CASE AND RESULT

A. Test Case 1:

Module Name: Import Dataset		
Action	Input	Expected Output
Select dataset	Upload Action	Dataset Upload Successfully
Result : Success		

Table 1: Test case 1

B. Test Case 2:

Module Name: Clustering		
Action	Input	Expected Output
Select pre-processing button	Cluster making	Dataset pre-processed Successfully
Result : Success		

Table 2: Test case 2

1) Average Result:

After applying K-means algorithm on the dataset then cluster will be generated and data is pre-processed. So these districts consider only at time of prediction

This graph shows the average of temperature, cloud, humidity, rainfall for pre-processed districts.

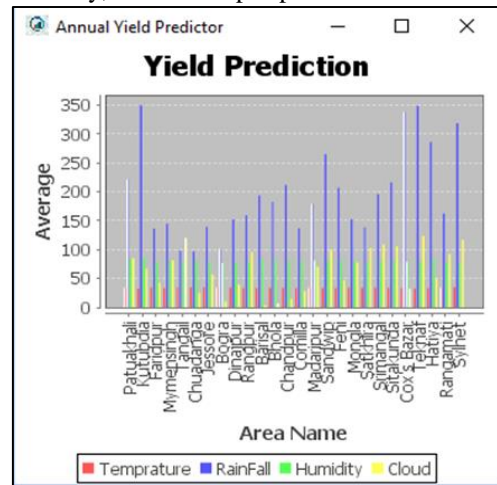


Fig. 2: Yield prediction

2) Yield Prediction:

Current year data can be taken as a input from user and that particular data relate with the appropriate cluster through regression technique annual yield predicted according to the particular area given by the user.

This graph shows the historical average yield prediction of the pre-processed districts.

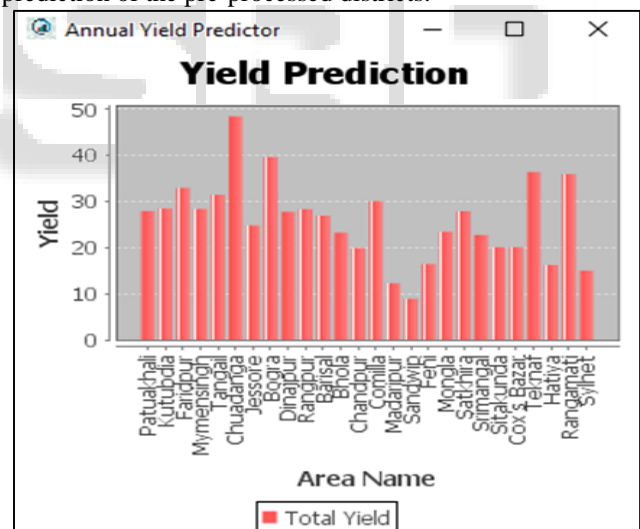


Fig. 3: Yield prediction

IX. CONCLUSION

Agriculture is the most influencing and significant application area particularly in the developing countries like India. Use of information technology in agriculture can change the situation of decision making and farmers can yield in better way and agricultural organisation get idea about yield and they make better policies for famers

Data mining plays a crucial role for decision making on several issues related to agriculture field. This projects integrates the work of various authors in one place so it is useful to get information of current scenario of data mining techniques and applications in context to agriculture field and predictive data mining.

REFERENCES

- [1] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, Rashedur M Rahman “Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in Bangladesh” IEEE paper Issue 1, June 2015
- [2] Raorane A.A., Kulkarni R.V. “Data Mining: An effective tool for yield estimation in the agricultural sector”
- [3] Utkarsha P. Narkhede, K. P. Adhiya “A Study of Clustering Techniques for Crop Prediction - A Survey”
- [4] A Survey on Data Mining Techniques in Agriculture. M.C.S.Geetha Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.

