Novelty Journals

ISSN 2394-7314

International Journal of Novel Research in Computer Science and Software Engineering
Vol. 3, Issue 1, pp: (34-37), Month: January-April 2016, Available at: www.noveltyjournals.com

Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India

[1]Rajshekhar Borate, [2]Rahul Ombale, [3]Sagar Ahire, [4]Manoj Dhawade,
[5]Mrs. Prof. R. P. Karande

[1, 2,3,4,5] Department of Computer Engineering, NBN Sinhgad School of Engineering, Pune-411041

Abstract: **In India there are different agriculture crops production and those crops depends on the various kind of factors such as biology, economy and also the geographical factors. And this several factors have the huge different impact on crops, which can be quantified using appropriate statistical methodologies. Applying such methodologies and techniques on historical yield of different crops, it is possible to obtain information or knowledge which can be helpful to farmers and government organizations for making better decisions and for make better policies which help to increased production. In this paper, our focus is on application of data mining techniques which is use to extract knowledge from the agricultural data to estimate better crop yield for major crops in major districts of India.**

Keywords: **data mining; crop analysis; yield prediction; clustering; K-means; K-NN; linear regression; neural net.**

I. INTRODUCTION

India is a country of rural economy and it is predominately agriculture oriented. In India more than 82 % are farmers belongs to small and large marginal farmers. Crop yield prediction is a very important area of research till which is use to ensure the food security all over the world. India's economy is almost depends on the crop yield. For Indian farmers they should need to know exactly which crop have to plant at which environment. Different district in India have the different climates at various time and so it is very important to consider the environmental factors of these different areas. This will help to select the best crop for planting in the different district.

All the crops are depend on the various factors like temperature, rain, sun light, humidity, moisture and carbon dioxide (CO_2) concentration to produce grains and other crop products that are so essential to our nutrition and health. However the amount of rainfall is vary between one districts to other and also vary in one year to another. Crop management is about managing climate risk so as to have financially viable and sustainable agricultural systems.

The most important part of farming is a pesticides. Without pesticides crops would die significantly due to insects and other pests which leads to sudden drop in the crop yield. And also the too much pesticides may affect the crop, while too little may not useful for crop. So the amount of pesticides required by crop is very important parameter.

In our project research, we have considered the various environmental, biotic(soil salinity) 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 divide regions, and then we apply suitable classification techniques to obtain crop yield predictions.

II. RELATED WORK

Bangladeshi student [1] purposes Data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts 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 [4] David H White and S Mark Howden, Research paper they focus on the climates determinants of crop productivity. They considered how the climate envelopes of different crops based on light, temperature and moisture influence the distribution of cropping and other land uses around the world. They also discuss how these and other climatic variables influence the growth and yield of crops. Adaption strategies are also discussed that helps a lot to assist the crop producers to cope with the rising global temperatures and carbon dioxide (CO2) levels, along with the often reduced rainfall, soil moisture and water availability.

In [5] 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 is the backbone of the Indian nation. In spite of the fact that large areas in India have been brought under irrigation, only one-third of the cropped part is irrigated. The productivity of agriculture is very low as compare to land which farmers has

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.

Indian farmer must have a good understanding of the soil type, the biotic factors governing the soil, environmental factors and also a thorough knowledge about the traditional agricultural practices to gain maximum crop yield.

IV. DATA SET

The dataset used in this project has been collected from Mahatma Phule Krushi Vidyapith Rahuri.

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 every crop considered according to the districts.

We also considered the two biotic attributes – soil salinity and soil pH for our project.

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 20 districts of India. 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 as follows:

1) Cluster Type-1 is based on the following attributes: Rainfall, minimum temperature, maximum temperature, humidity and sunshine. These are the environmental or climatic attributes considered for our research. The degree of similarity of the collection of these attributes should indicated distinct clusters for the selected districts.

2) Cluster Type-2 is based on the following attributes: soil pH and soil salinity. As discussed earlier, these biotic factors contribute largely towards the prediction of the crops.

3) Cluster Type-3 is based on irrigated area. Clustering is based on the area attributes for each district was considered because we can obtain the separate clusters based on distinct ranges of areas that were irrigated for each district.

4) Cluster Type-4 is based on the individual crop yields of rice, potato and wheat. This type of clustering was considered in order to classify the districts into separate clusters with similar crop yields and after analysis of the results, to see whether they exhibit a pattern related to effects from the selected attributes.

K-Means Clustering: The K-means clustering algorithm is used to produce non-hierarchical groups of similar points in the data based on the centroid. For our project, k-means clustering was used upon the selected districts according to the categorized types mentioned previously.

Rapid Miner Studio software is useful to implement k-means clustering. Clustering results were separately written to Excel files for each cluster type (1 to 4) for the convenience of result showcasing and analysis.

B. Prediction of crop yields using classification techniques: In our project, we determined prediction results for yields of selected crops for the selected districts in India. The predictions results were obtained according to the selected input attributes using appropriate classification and regression models in Rapid Miner.

The following classification/regression models were used to obtain the crop yield prediction results:

a) Linear Regression: 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 changing variables known as independent variables. If independent variable contains multiple input attributes like in our research (rainfall, sunshine hours, humidity, pH etc.), then it is termed as multiple linear regressions. Linear regression provides a model for the relationship between a scalar variable and one or more explanatory variables. This is done by fitting a linear equation to the observed data.

b) k-NN: The k-nearest neighbour algorithm compares a given test example with training examples which are similar. Each example denotes a point in dimensional space. Thus, all of the training examples are saved in an n-dimensional pattern space. K is a positive integer, usually small. For our purpose, the basic k-NN algorithm was applied. It first finds the k examples from the training set that are closest to the unknown example. Then it takes the most common occurring classification for the k examples.

c) Neural Net: An artificial neural network (ANN) is a mathematical model inspired by the structure and functional aspects of biological neural networks for instance in our brains. In most cases an ANN is an adaptive system that modifies its structure based on external or internal information that flows through the network during the learning phase. The basic neural network model consists of three layers: the input layer, the hidden layer and an output layer.

VI. 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 to post process the data and give the best three possible crops in order of preference to choose from for farming across all the major agricultural districts. If there are not any feasible choices the program simply outputs ‘NONE’. These recommendations are based on a combination of annual yield of that crop species per hectare area of a district.

VII. FUTURE WORKS

In our project we considered the 5 environmental variables, 2 biotic variables and also 2 area related variables to determine crop yield in the different districts. In the near future, geospatial analysis will be added to our data processing model to improve accuracy and also implement a better geographical data.

VIII. CONCLUSION

In our project we found that the accurate prediction of different specified crop yields across different districts will help to farmers of India. From this Indian farmers will plant different crops in different districts.

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] D. Diepeveen and L. Armstrong, "Identifying key crop performance traits using data mining" World Conference on Agriculture, Information and IT, 2008.

[3] Mohammad Motiur Rahman, Naheena Haq and Rashedur M Rahman "Comparative Study of Forecasting Models on Clustered Region of Bangladesh to Predict Rice Yield", 17th. IEEE International Conference on Computer and Information Technology (ICCIT), Dhaka, 2014.

[4] Soils,Plant growth and crop production- Vol.I-Climate and its Effects on Crop Productivity and Management-S Mark Howden,Devid H White.

[5] A Survey on Data Mining Techniques in Agriculture.M.C.S.Geetha Assistant Professor, Dept. of Computer Applications, Kumaraguru College of Technology, Coimbatore, India.