IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Prediction of Crop Yield Using Data Mining

[1] Nishiba Kabeer; [2] Dr.Loganathan.D; [3] Cowsalya.T

[1]PG Scholar, Department of Computer Science and Engineering,
Anna University, SVS College of Engineering,
Coimbatore, Tamilnadu, India. 642 109.

[2]Professor & Head, Department of Computer Science and Engineering,
SVS College of Engineering,
Coimbatore, Tamilnadu, India.642 109.

[3]Assistant Professor, Department of Computer Science and Engineering,
SVS College of Engineering,
Coimbatore, Tamilnadu, India.642 109.

Abstract - Agriculture is a major source of economy of the country. Agricultural crop production depends on various factors such as biology, climate, economy and geography. Several factors have different impacts on agriculture, which can be quantified using appropriate statistical methodologies .Predicting the crop yield well in advance prior to its harvest can help the farmers and Government organizations to make appropriate planning like storing, selling, fixing minimum support price, importing/exporting etc. As Prediction of crop deals with large set of database thus making this prediction system a perfect candidate for application of data mining. This work aims at finding suitable data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities. The main aim is to create a user friendly interface for farmers, which gives the prediction of production using Data Mining techniques like Regression and Clustering based on available data in all districts of Kerala.

Keywords *– Data mining, crop analysis, yield prediction, clustering, linear regression.*

.

1. Introduction

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. In this context, the farmers necessarily requires a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Agriculture is a task full of risks which is influenced by several factors like temperature, rainfall, sown area, past production record and other soil and climate related issues. Reliable information about these factors can be helpful to farmers as well as government in decision making.

a. It helps farmers in providing the historical crop yield record with a forecast reducing the risk management.
b. It helps government in making crop insurance policies as well as policies for supply chain operation.

Data Mining is widely applied to agricultural problems. Data Mining is used to analyse large data sets and establish useful classifications and patters in the data sets. The overall goal of the Data Mining process is to extract the information from a data set and transform it into understandable structure for further use. In this work the main aim is to create a user friendly interface for farmers, which gives the analysis of rice production based on available data. Different Data mining techniques were used to predict the crop yield for maximizing the crop productivity. In proposed work, data mining techniques like regression methods and clustering are used to predict the annual yield of major crops.

2. Literature Review

Ramesh and Vardhan [1] deal with the challenge of predicting the yield of various crops. One approach to this problem is to employ data mining techniques. In this paper, different types of data mining methods were applied and then evaluated on the datasets we prepared.

In [2], a software tool named Crop Advisor has been developed an user friendly web page for predicting the influence of climatic parameters on the crop yields. C4.5 algorithm is used to find out the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh. The selection of districts has been made based on the area under that particular crop. The prediction accuracy of the developed model varied from 76 to 90 percent for the selected crops and selected districts. Based on these observations the overall prediction accuracy of the developed model is 82.00 per

cent. With a high prediction accuracy the developed model can be used by the policy makers in arriving at a policy decision well in advance i.e., before the harvest of the crop.

In [3] its focus on application of data mining techniques to extract knowledge from the agricultural data to estimate crop yield for major cereal crops in major districts of Bangladesh. The dataset used in this research has been collected from BARI (Bangladesh Agricultural Research Institute). The dataset was pre-processed to select only the attributes which are important for the research: rainfall, maximum and minimum temperature, humidity, irrigated area for all districts; and cultivated area for every crop considered according to the districts.

One further environmental attribute: sunshine and two further biotic attributes soil salinity and soil pH were considered for the research. Clustering of the selected districts: In order to group the districts into distinct clusters, the assumption was that the districts containing the similar values of relevant attributes should belong to the same cluster. The predictions results were obtained according to the selected input attributes using appropriate classification and regression models For the purpose of use in learning models, two time periods of the dataset were considered.

3. Motivation

Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is enormous in Indian agriculture. The data when become information is highly useful for many purposes. Data Mining is widely applied to agricultural problems. Data Mining is used to analyze large data sets and establish useful classifications and patters in the data sets.

4. Data Set

The data available in this work is obtained for the years from 2002 to 2012 in 14 districts of Kerala in India (Fig 1). The data is taken in six input variables. They are Year, Season, Crop, Area of Sowing, Production and Rainfall .Year attribute specifies the year in which the data available .Season specifies the name of the Season. Crop specifies the predominate crops in the each district. Area of sowing attribute specifies the total area sowed in the specified year. Production attribute specifies the production of crop in the specified year in Tons. Rainfall attribute specifies the Rainfall in the specified year in Centimeters.

The preliminary data collection is carried out for all the districts of Kerala in India. Each area in this collection is identified by the respective longitude and latitude of the region. The information gathering process is done with three government units like Indian Meteorological Department, Statistical Institution and Agricultural department. In this research the estimation of the crop yield is analyzed with respect to four parameters namely Year, Rainfall, Area of Sowing and Production.

	state_name	district_name	crop_year	season	crop	area	production	rainfall
1	Kerala	ALAPPUZHA	2002	Autumn	Rice	3721	1911.5	2475
2	Kerala	ALAPPUZHA	2002	Kharif	Sesamum	74	1911.6	2475
3	Kerala	ALAPPUZHA	2002	Summer	Rice	8765	1901	2475
4	Kerala	ALAPPUZHA	2002	Whole Year	Arecanut	2441	1910	2475
5	Kerala	ALAPPUZHA	2002	Whole Year	Banana	470	1911	2475
6	Kerala	ALAPPUZHA	2002	Whole Year	Bhindi	16	1912	2475
7	Kerala	ALAPPUZHA	2002	Whole Year	Bitter Gourd	86	1913	2475
8	Kerala	ALAPPUZHA	2002	Whole Year	Black pepper	1940	1914	2475
9	Kerala	ALAPPUZHA	2002	Whole Year	Brinjal	55	1915	2475
10	Kerala	ALAPPUZHA	2002	Whole Year	Cashewnut	4313	1916	2475
11	Kerala	ALAPPUZHA	2002	Whole Year	Cashewnut Raw	4313	1917	2475
12	Kerala	ALAPPUZHA	2002	Whole Year	Coconut	55407	1918	2475
13	Kerala	ALAPPUZHA	2002	Whole Year	Drum Stick	825	1919	2475
14	Kerala	ALAPPUZHA	2002	Whole Year	Dry ginger	125	1915	2475
15	Kerala	ALAPPUZHA	2002	Whole Year	Jack Fruit	3015	1915	2475
16	Kerala	ALAPPUZHA	2002	Whole Year	Mango	5708	1915	2475
17	Kerala	ALAPPUZHA	2002	Whole Year	Other Fresh Fruits	883	1915	2475
18	Kerala	ALAPPUZHA	2002	Whole Year	other oilseeds	135	1915	2475
19	Kerala	ALAPPUZHA	2002	Whole Year	Other Vegetables	1977	1915	2475
20	Kerala	ALAPPUZHA	2002	Whole Year	Papaya	1099	1915	2475
21	Kerala	ALAPPUZHA	2002	Whole Year	Pineapple	105	1915	2475
22	Kerala	ALAPPUZHA	2002	Whole Year	Rubber	3825	1915	2475

Fig. 1 Data Set

5. Methodology

The method in this paper is initially divided into two major parts: Clustering and Regression.

5.1 Prediction of crop yields using Regression techniques

In this paper, to determine the prediction results for yields of selected crops for the selected districts in Kerala. The predictions results were obtained according to the selected input attributes using appropriate classification and regression models.

The below regression model is used to obtain the crop yield prediction results: Simple 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 (regular attributes).Simple linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and

dependent(y) variable. The line can be modeled based on the linear equation Eq (1)

$$y = mx + c. \quad (1)$$

Based on the given data points, we try to plot a line that models the points the best. The motive of the linear regression algorithm is to find the best values for m and c.The attributes considered in this work are rainfall which is the predictor and production which is the target.

5.2 Clustering of the selected districts based on attribute Rainfall

In this paper, it was considered a total of 14 districts of Kerala. 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, It was categorized our selected attributes for the consideration of clustering the districts.

DBSCAN is a clustering method that is used in machine learning to separate clusters of high density from clusters of low density. Given that DBSCAN is a density based clustering algorithm, it does a great job of seeking areas in the data that have a high density of observations, versus areas of the data that are not very dense with observations.

DBSCAN works as such:

- Divides the dataset into *n* dimensions
- For each point in the dataset, DBSCAN forms an *n* dimensional shape around that data point, and then counts how many data points fall within that shape.
- DBSCAN counts this shape as a cluster. DBSCAN iteratively expands the cluster, by going through each individual point within the cluster, and counting the number of other data points nearby.

6. Implementation

The crop yield prediction includes repeatedly all essential parameters that are needed for the well yield of crop. This improves the classification outcomes of the crop yield. All the essential parameters are thought-about as inputs. In common, one in all the issues faced with in the prediction method is that almost all of the required parameters that are essential to consider for the exact prediction are not consider.

Crop prediction is that the art of predicting crop yields and manufacture before the yield really takes place. Before harvest prediction was done by considering the farmer's knowledge on a selected field and crop. This work presents a system (Fig 2) that uses data processing strategies so as to predict the analyzed datasets. The anticipated sort can specify the yielding of crops.

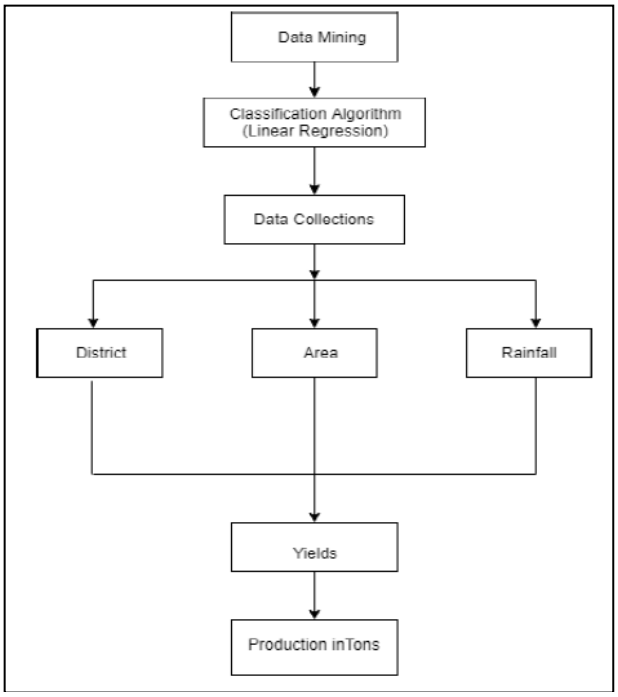

Fig.2 Linear Regression Model for Crop Yield Prediction

Architecture is a system that unites its parts or components into a coherent and purposeful complete. The crop information base consists of farm data like crop varieties, crop year, area and rainfall. The knowledge-based additionally contains of zones furthermore district information, ecological parameter like extreme and lowest temperature value and average precipitation.

The crop yield prediction model that includes associate input module that is in charge of taking input from the farmer. The input module includes land area, average rainfall, crop year and district. The feature selection model is in charge offset. Selection of associate attribute from crop particulars.

The crop yield prediction model used to predict the yield. Once feature selection, the data go to classification rule for grouping similar contents crop parameters used to predict crop growth can be predicted. Then prediction rules are going to be applied to the output of classifying crop particulars in terms of crop name, season and total yield details.

7. Results

The aim of the work was to construct a user friendly website that will help the farmers and other policy makers to predict the crop yield based on the data set parameters. A website was thus developed. The farmer or the user of the application will enter the details such as name of district, average rainfall and area of field in hectare. After entering the details user will press 'Predict'. The prediction will be made using the regression method and the result of the prediction of the crop yield will be send to the user and it will be displayed in tonnes.

Fig.3 Yield Prediction Page

User also has the option to see the cluster graph (Fig.4) based on Rainfall and Production.

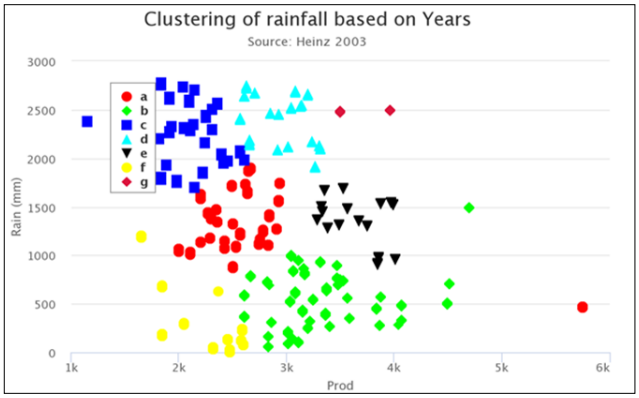

Fig.4 Cluster Graph

8. Conclusions and Future Enhancement

The work demonstrated the potential use of data mining techniques in predicting the crop yield based on the input parameters average rainfall and area of field. The developed webpage is user friendly and the accuracy of predictions are above 90 per cent. The districts selected in the study indicating higher accuracy of prediction. The user friendly web page developed for predicting crop yield can be used by any user by providing average rainfall and area of that place. The process was adopted for all the districts of Kerala to improve and authenticate the validity of yield prediction which are useful for the farmers of Kerala for the prediction of a specific crop.

The future work aimed at the analysis of the entire set of data and will be devoted to suitable strategies for improving the efficiency of the proposed algorithm. Use of such kind of approach to forecasting is not restricted to agriculture alone. The clustering and regression is one of the capable tool in field of data mining which can be used in several different ways.

The clustering can also be implement in the concept of soil type clustering so that soils having similar kind of features can be used for similar kind of crops. The concept can be further merged with the market data to predict the price of crop, as well as to predict the fertilizers consumption. This is not limited to agriculture; the concept can be deployed in weather forecasting also. All these combined together can be a very good work in the field of precision agriculture.

References

[1] D Ramesh, B Vishnu Vardhan. "Data Mining Techniques and Applications to Agricultural Yield Data". International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 9, September 2013.

[2] S.Veenadhari, Dr. Bharat Misra, and Dr. CD Singh, "Machine learning approach for forecasting crop yield based on climatic parameters", International Conference on Computer Communication and Informatics (ICCCI - 2014), Jan 2014, Coimbatore.

[3] A.T.M Shakil Ahamed, Navid Tanzeem Mahmood, Nazmul Hossain, Mohammad Tanzir Kabir, Kallal Das, Faridur Rahman, and 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 2015.

[4] D Ramesh, and B Vishnu Vardhan, "Analysis of Crop Yield Prediction Using Data Mining Techniques", IJRET: International Journal of Research in Engineering and Technology, Jan 2015.

[5] Ye Nong; Data Mining: Theories, Algorithms, and Examples, CRC Press, 2013.
[6] http://books.irri.org/0471097608_content.pdf
[7] http://docs.rapidminer.com/studio
[8] http://www.barcapps.gov.bd/dbs/index.php
[9] http://www.faostat.fao.org/site/339/default.aspx
[10] http://www.assignmentpoint.com/science/zoology/agrisector of bangladesh.html

Authors –

Ms.Nishiba Kabeer is currently pursuing M.E, CSE at SVS college of Engineering affiliated to Anna University. Her area of interests are Data Mining, Machine Learning, and Predictive Analysis.

Dr.D.Loganathan is a Professor and Head of Computer Science and Engineering department in SVS College of Engineering, Coimbatore, Tamilnadu. He has published several research articles in various international journals and presented several research papers in various international and national conferences.

Ms.T.Cowsalya working as a Assistant Professor in the Computer Science and Engineering department at SVS College of Engineering, Coimbatore, Tamilnadu.