

PREDICTIVE ANALYTICS TILING METHODOLOGY BASED DATA MINING ON FARMING CROPS IN TAMILNADU

¹Muthaiah U, ²Dr.S.Chitra

¹Computer Science and Engineering, Sri Shanmugha College of Engineering and Technology, Salem,
²Er.Perumal Manimekalai College of Engineering, Krishnagiri

Abstract - The paper is to bring out the tiling based predictive analytics of data extraction over agriculture field. Data mining is not composed by efficient and effective algorithms and executed as standalone kernels. It is constituted by multifaceted application articulated in the non-trivial interaction among various components, running on large scale distributed environments. Our experimental setup initially focuses with retrieval of netting climatic data structure such that dynamic weather forecasting data as nodes and seasonable crops data as edges in order to categorize the fabulous or optimal data. This paper perform a detailed study of big data predictive analytics schema towards variant effect of manual data sampling in the field of farming which can be carried out with expected optimal output strategies. We have analyzed and compared many algorithms such as clustering, classification, regression analysis, sentiment analysis, association rule learning, machine learning, predictive theorem and graph technique. In near future we will work on challenging issues in various fields and provide solution using big data analytics methodology.

Keywords - clustering, regression, machine learning, association rule learning, sentiment analysis, graph theorem, nodes, edges, data retrieval, data predictive analytics.

I. Introduction

Yield prediction is a very import agricultural problem. Any farmer is interested in knowing how much yield he is about expect. 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. Rainfall conditions also influences the rice cultivation.

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. In the existing project, only rice production is analysed based on available data using Data Mining techniques. 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. 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 paper 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.

II. Literature Survey

From the research article [3], the researcher express that large amount of data which is collected and stored for analysis. Making appropriate use of these data often leads to considerable gains in efficiency and therefore economic advantages. There are several applications of Data Mining techniques in the field of agriculture. The researchers

implemented [4] K-Means algorithm to forecast the pollution in the atmosphere, the K Nearest Neighbour is applied [12] for simulating daily precipitations and other weather variables and different possible changes of the weather scenarios are analyzed [14] using Support Vector Machines.

Soil profile descriptions were proposed [15] by the researcher for classifying soils in combination with GPS based technologies. They were applied K-Means approach for the soil classification. In a similar approach, crop classifications using hyper spectral data was carried out [1] by adopting one of the data mining approach i.e. Support Vector Machines. One of the researcher used [9] an intensified fuzzy cluster analysis for classifying plants, soil and residue regions of interest from GPS based colour images.

In the agricultural science, clustering techniques are found in grading [5] apples before marketing. Weeds were detected [13] on precision agriculture. The researchers worked [8] on rainfall variability analysis and its impact on crop productivity. The effect of observed seasonal climatic conditions such as rainfall and temperature variability on crop yield prediction was considered [7] through an empirical crop model. Furthermore, there are two approaches to investigate the impact of climate change on crop production which include the crop suitability approach and the production function approach [6].

Researchers were found that the yields of winter wheat are reduced when temperatures rise, due to the consequent reduction of the growth phases of the plant [2] and also concluded that the complexity of a model was based on the

level of detailed analysis [10] or it was less detailed with only estimations of moisture content [11].

III. Overview Of Data

The data used for this paper are obtained for the years from 1955 to 2015 for Thanjavur district of Tamil Nadu in India. The preliminary data collection is carried out for all the districts of Tamil nadu in India. Each area in this collection is identified by the respective longitude and latitude of the region. The evaluation is considered for only Thanjavur district of Tamil Nadu in India. The data are taken in eight input variables. The variables are 'Year', 'Rainfall', 'Area of Sowing', 'Yield', 'Fertilizers' (Nitrogen, Phosphorous and Potassium) and 'Production'. The attribute 'Year' specifies the year in which the data are available in Hectares. 'Rainfall' attribute specifies the average rainfall in the specified year in Centimetres. 'Area of Sowing' attribute specifies the total area sowed in the specified year for that region in Hectares. 'Yield' specifies in Kilogram per hectare. 'Production' attribute specifies the production of crop in the specified year in Metric Tons. 'Fertilizers' specify in Tons in the specified year.

IV. Methodology

In this paper the statistical method namely Multiple Linear Regression technique and Data Mining method namely Density-based clustering technique were take up for the estimation of crop yield analysis.

A. Multiple Linear Regression

A regression model that involves more than one predictor variable is called Multiple Regression Model. Multiple Linear Regression (MLR) is the method, used to model the linear relationship between a dependent variable and one or more independent variables. The dependent variable is sometimes termed as predictant and independent variables are called predictors.

Multiple Linear Regression (MLR) technique is based on least squares and probably the most widely used method in climatology for developing models to reconstruct climate variables from tree ring services. This crop yield prediction model is presented with the use of Multiple Linear Regression (MLR) technique where the predictant is the Production and there are seven predictors namely Year, Rainfall, Area of Sowing, Yield and Fertilizers (Nitrogen, Phosphorous and Potassium)

B. Density Based Clustering Technique

The primary idea of Density-based clustering techniques is that, for each point of a cluster, the neighborhood of a given unit distance contains at least a minimum number of points. In other words the density in the neighborhood should reach some threshold. However, this idea is based

on the assumption that the clusters are in the spherical or regular shapes.

These methods group the objects according to specific density objective functions. Density is usually defined as the number of objects in a particular neighborhood of data objects. In these approaches, a given cluster continues to grow as long as the number of objects in the neighborhood which exceeds some parameter. This is considered to be different from the idea in partitioning algorithms that use iterative relocation of points that give a certain number of clusters.

V. Result And Discussion

In this paper an effort is made in order to know the region specific crop yield analysis and it is processed by implementing both Multiple Linear Regression technique and Density-based clustering technique. These models were experimented in respect of all the districts of Tamil Nadu, but the process of evaluation is carried out with only Thanjavur district of Tamil Nadu in India.

The exact value along with the corresponding estimated value using Multiple Linear Regression technique for 40 years interval of sample data about Thanjavur is shown in the Table-1.

The estimated results using Multiple Linear Regression technique which are ranging between -14% and +13% for 40 years interval.

Table-1: Exact production and estimated values using Multiple Linear Regression technique.

Observation Year	Production (Exact)	40 Years Interval	
		Production (Estimation)	Percentage Of Difference
2001	683423	592461	13
2002	579850	566050	2
2003	551115	579433	-5
2004	762453	722638	5
2005	743614	742752	0
2006	348727	399062	-14
2007	547716	551541	-1
2008	715472	691069	3
2009	716609	697227	3
2010	616567	633494	-3
2011	683423	592461	13
2012	579850	566050	2
2013	551115	579433	-5
2014	762453	722638	5
2015	743614	742752	0

The estimation of the crop yield prediction using Density-based clustering technique for 6-clusters approximation of

sample data about Thanjavur district is shown in the Table-2. The estimated results using Density-based clustering technique which are ranging between -13% and +8% for 6-clusters approximation.

Table-2: Exact production and Estimated values using Density-based clustering technique

Observation Year	Production (Exact)	6 Clusters	
		Production (Estimation)	Percentage Of Difference
2006	683423	666011	3
2007	579850	651103	-12
2008	551115	566972	-3
2009	762453	703914	8
2010	743614	737897	1
2011	348727	392770	-13
2012	547716	534709	2
2013	715472	791589	-11
2014	716609	676321	6
2015	616567	695574	-13

The comparison between exact production along with the corresponding estimated value using Multiple Linear Regression technique for 40 years interval and Density-based clustering technique for the 6-clusters approximation about Thanjavur District is shown in the following Table-3 and Figure-1.

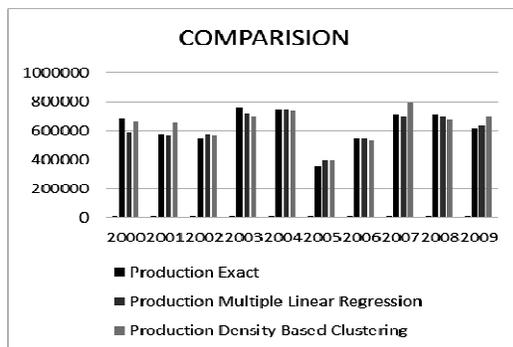


Fig-1: Comparison between Multiple Linear Regression technique and Density-based clustering technique

Observation Year	Production (Exact)	Production(Estimation)	
		Multiple Linear Regression technique	Density-based clustering technique
2006	683423	592461	666011
2007	579850	566050	651103
2008	551115	579433	566972
2009	762453	722638	703914
2010	743614	742752	737897
2011	348727	399062	392770
2012	547716	551541	534709
2013	715472	691069	791589
2014	716609	697227	676321
2015	616567	633494	695574

Table-3: Comparison between Exact production and estimated values using Multiple Linear Regression technique and Density-based clustering technique.

VI. Conclusion

Initially the statistical model Multiple Linear Regression technique is applied on existing data. The results so obtained were verified and analyzed using the Data Mining technique namely Density-based clustering technique.

In this procedure the results of two methods were compared according to the specific region i.e. Thanjavur district of Tamil Nadu in India. Similar process was adopted for all the districts of Tamil Nadu to improve and authenticate the validity of yield prediction which are useful for the farmers of Tamil Nadu for the prediction of a specific crop.

In the subsequent work a comparison of the crop yield prediction can be made with the entire set of existing available data and will be dedicated to suitable approaches for improving the efficiency of the proposed technique.

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.

References

- [1] Camps-Valls G, Gomez-Chova L, Calpe-Maravilla J, Soria-Olivas E, Martin-Guerrero J D, Moreno J, "Support Vector Machines for Crop Classification using Hyper Spectral Data", Lect Notes Comp Sci 2652, 2003, pages : 134-141.
- [2] G R Batts, "Effects Of CO2 And Temperature on Growth and Yield of Crops of Winter Wheat over Four Seasons", European Journal of Agronomy, vol. 7, 1997, pages : 43-52.
- [3] G Ruß, "Data Mining of Agricultural Yield Data : A Comparison of Regression Models", Conference Proceedings, Advances in Data Mining – Applications and Theoretical Aspects, P Perner (Ed.), Lecture Notes in Artificial Intelligence 6171, Berlin, Heidelberg, Springer, 2009, pages : 24-37.
- [4] Jorquera H, Perez R, Cipriano A, Acuna G, "Short Term Forecasting of Air Pollution Episodes", In: Zannetti P (eds) Environmental modeling , WIT Press, UK, 2001.
- [5] Leemans V, M F Destain, "A Real Time Grading Method of Apples Based on Features Extracted from Defects", J. Jood Eng., 2004, pages : 83-89.
- [6] M J Foulkes, "Raising Yield Potential of Wheat", Journal of Experimental Botany, vol. 62, 2011, pages : 469-486.
- [7] M Trnka, "Projections of Uncertainties in Climate Change Scenarios into Expected Winter Wheat Yields",

- Theoretical and Applied Climatology, vol. 77, 2004, pages : 229-249.
- [8] Mehta D R, Kalola A D, Saradava D A, Yusufzai A S, "Rainfall Variability Analysis and Its Impact on Crop Productivity - A Case Study", Indian Journal of Agricultural Research, Volume 36, Issue 1, 2002, pages : 29-33.
- [9] Meyer G E, Neto J C, Jones D D, Hindman T W, "Intensified Fuzzy Clusters for Classifying Plant, Soil and Residue Regions of Interest from Color Images". Computer Electronics Agric Vol. 42, 2004, pages : 161-180.
- [10] R J Brooks, "Simplifying Sirius : Sensitivity Analysis and Development of A Meta-Model for Wheat Yield Prediction", European Journal of Agronomy, vol. 14, 2001, pages : 43-60.
- [11] R V Martin, "Seasonal Maize Forecasting for South Africa and Zimbabwe Derived From an Agroclimatological Model", Journal of Applicable Meteorology, vol. 39, 2000, pages : 1473-1479.
- [12] Rajagopalan B, Lall U, "A K-Nearest Neighbor Simulator for Daily Precipitation and Other Weather Variables", Wat Res Res 35(10), 1999, pages : 3089-3101.
- [13] Tellaeche A, X P Burgos Artizzu, G Pajares, A Ribeiro, "A Vision-Based Classifier for Weeds Detection in Precision Agriculture through the Bayesian and Fuzzy K-Means Paradigms", Adv.Soft. Comp., 2008, pages : 72-79.
- [14] Tripathi S, Srinivas V, Nanjundiah R S, "Downscaling of Precipitation for Climate Change Scenarios: A Support Vector Machine Approach", J Hydrol, 2006, pages : 621-640.
- [15] Verheyen K, Adriaens D, Hermy M, Deckers S, "High-resolution continuous soil classification using morphological soil profile descriptions", Geoderma Vol.101, 2001, pages : 31-48.