



ANALYSIS OF CROP YIELD PREDICTION USING FUZZY CLUSTERING TECHNIQUES

Mrs. K.A.Poornima

Assistant Professor in Computer Science,
Gobi Arts and Science College (Autonomous),
Gobichettipalayam, Tamil Nadu, India

Dr.G.Dheepa

Assistant Professor of Computer Science,
P.K.R.Arts College for Women (Autonomous),
Gobichettipalayam. Tamil Nadu, India

Abstract: Nowadays the most important field in the real world is agriculture and it is the main occupation and backbone of our Indian economy. Agriculture data analysis is one of the latest drift research fields in data mining. Crop yield prediction is vital as it can support decision makers in agriculture zone. Data mining have modern techniques and algorithms for finding best yield prediction. This paper presents a brief comparative study on different views that deals with various performances used to figure out the different crop yield with less error rate. Fuzzy C-Means(FCM), Fuzzy logic (FL), Adaptive Neuro Fuzzy Inference System(ANFIS), Multiple Linear Regression(MLR), Linear Discriminant Analysis (LDA) are used to survey out high accuracy and less error rate prediction capabilities.

Keywords: Data Mining, Fuzzy Clustering, Agriculture, Different Techniques.

I. INTRODUCTION

Data Mining is the progression of analyzing, extracting and predicting the meaningful evidence from enormous data to extract some pattern. The process of Data Mining includes first selection of data followed by pre-processing of data and then transforming the data to get patterns which can then be used to predict useful insights. Pre-processing includes verdict outliers and detecting missing values whereas transformation finds the correlation between objects.

Applying the data mining techniques to the agricultural yield primarily depends on environmental factors such as rainfall, temperature, Soil. Clustering algorithms are used widely in different areas of research. Clustering algorithms arise due to need to find data groups that share similar characteristics in a given dataset, currently there are several fuzzy clustering algorithms such as Fuzzy C-Means (FCM) [1] [2], Possibilistic C-Means (PCM) [3] [2], Fuzzy Possibilistic C-Means (FPCM) [4], and Possibilistic Fuzzy Means (PFCM) [5] [2] which is widely used technique for crop yield prediction in data mining. Each data of data sets should be clearly decided to one certain class by the traditional clustering algorithm. The conception of fuzzy clustering is more appropriate to the essence of matter. At present, the fuzzy c-means clustering algorithm is one of the most popular clustering algorithms.

In this paper the main aim is to create a user friendly interface for farmers, which gives the analysis of crop production based on available data. Different Data mining techniques were used to predict the crop yield for maximizing the crop productivity.

This paper organized as follows: In Section II explains about the related work already done by previous researchers from this particular agriculture domain. In Section III describes different classification techniques used for crop yield prediction. This section explains within the classification

analytical model that suggestion of crop with relevancy environmental condition. In Section IV, it describes what are all the various performances and methods used for this proposed work. In Section V conclusion describes about different methods and techniques are given best method of an approach to the problem of crop yield prediction

II. LITERATURE WORK

The research study on various fuzzy clustering algorithms to find out the performance of the algorithms.

Anjana Gosain, Sonika Dahiya[1] researchers particularly focused on FCM, PCM, PFCM, FCM- σ , DOFCM, T2FCM, KT2FCM, IFCM, KIFCM, IFCM- σ , KIFCM- σ , NC, CFCM fuzzy clustering algorithms [1]. They used standard data points with present of outliers and noise for better performance analysis. Researchers used standard dataset for implementation. From various fuzzy clustering algorithms, DOFCM yields the best performance for their dataset.

A.Mucherino, P.Papajorgj and P. M. Pardalos surveyed about different data mining techniques and how they can be useful in agriculture sector [2].

A predictive model that provides a cultivation plan for farmers to get high yield of paddy crops using data mining techniques is proposed by **Anita Arumgam** [3]. The research focused to develop predictive model for paddy yield prediction using data mining techniques such as K-means clustering, Decision tree induction and so on. Researcher collected the real time dataset from the farmers under the area of Thamirabarani river basin. The proposed predictive model constructed using WEKA tool. Data pre-processing focused in four ways such as data cleaning, attribute selection, transformation and integration. K-means was taken for clustering and then classification techniques are used to build the model. Finally, the researcher concludes that the random forest classifier achieved high accuracy with 97.5% than other classifier.

S. No	Author	Different types of Crop	Method	Best Method
1.	Rajendrachoudhary and Dr.Aakashsaxena (2018)	Wheat Production	Fuzzy logic, Adaptive Neuro Fuzzy Inference system and MLR	ANFIS Model give Better Performance compare to FL and MLR.
2.	P.Surya,I.Laurance Aroquiaraj (2019)	Any Particular crop	K-means, K-medoids and Artificial bee colony with weighted Based fuzzy clustering.	Artificial bee colony with weighted Based fuzzy clustering
3.	D.Ramesh, B. Vishnu Vardhan.	Prediction of any crop yield	Multiple Linear Regression and Density based Clustering Techniques	Density based Clustering is best.
4.	RabinaDayal, ArunKumar Yadav (2018)	Wheat	Fuzzy c-means and LDA, ANFIS.	LDA Show envisions then ANFIS.
5.	M.Mohanadevi V.Vinodhini (2018)	Sugarcane	Linear Regression model, REP And Random Forest	Linear Regression give Correct Correlation.
6.	P.Surya ,I.Laurance Aroquiaraj (2019)	Any Particular crop	K-means, K-Medoids and Modified Fuzzy Clustering with genetic algorithms	Compare to other cluster algorithm Modified Fuzzy Clustering with genetic algorithm better one.

Rajendrachoudhary and Dr.Aakashsaxena, “Hybrid algorithm of Neuron Fuzzy Inference System and LDA for Crop Yield Prediction” [4]. In this paper analysed wheat production using Fuzzy logic, Adaptive Neuro Fuzzy Inference system and MLR. ANFIS Model gives better performance compare to FL and MLR.

P.Surya, Dr.I.Laurance Aroquiaraj, “Performance Analysis of K-Means and K-Medoid Clustering Algorithms Using Agriculture Dataset”. In this paper analysed Crop Prediction Analysis in North western Zone of Tamilnadu using Artificial Bee Colony with Weighted based Fuzzy Clustering [5].

D Ramesh, B VishnuVardhan, researchers build a prediction for crop yield using multiple linear regression technique and density based clustering technique for Andhra Pradesh in India. The result is achieved by Density based clustering technique [6].

III. CLASSIFICATION TECHNIQUES

A. Fuzzy C-Means:

The FCM algorithm is derivative by Dunn is a clustering algorithm, and further enriched by Bezdek. This algorithm is useful when the no. of cluster is pre-derived. This method tries to set every data points to one clusters. The FCM is not quite the same as other is that does not characterize the total participation of an information point to given cluster whereas ascertain the information point (level of membership) that indicate the group. FCM doesn't evaluate the absolute membership function but is extremely fast to desired cluster related to accuracy. In Automatic FCM (AFCM) calculation, few clusters contain the statistical patterns, having unordinary participation esteems. The membership values of a measurement to accumulate means the examination among the given statistical models of the cluster. At certain arrangement of n insights designs patterns (i.e. $X = k \times n \times \dots \times \dots \times \dots \times 1$), it is utilized to minimization of the succeeding target work $F(M,N)$ AFCM by an iterative procedure.

B. Linear Regression (LR):

A Linear regression methodology that's used to analyse a response variable that alterations with the value of the interference variable. A way of predicting the value of a response variable from a given value of the explanatory variable is also referred to as prediction. The least-square fit, that is capable of fitting each linear additionally as polynomial relations, is that the most typically used linear regression. The method of applying model estimate to values outside the domain of the first knowledge is thought as extrapolation. A linear regression model is computed to analyze the yield.

Linear Regression model is used for crop yield prediction. To develop the Linear Regression models for crop yield prediction, Linear Regression analysis is majorly used for prediction functions because it provides predicted entity as a function of depended entities.

It is a statistical measure that can be used to determine the strength of the association between one dependent variable and a sequence of other changing variables known as independent variables (regular attributes).

In independent variable contains several input attributes like in our analysis (rainfall, sunshine hours, humidity, pH etc), then it's termed as multiple linear regressions. Linear regression provides a model for the relationship between a scalar variable and one or more explanatory variables.

C. Fuzzy Clustering:

The other name of the Fuzzy clustering is known as soft clustering techniques.

Fuzzy Clustering techniques can be applied to data that are quantitative that is numerical, qualitative that is categorical and or a mixture of both data (numerical and categorical). In this method, it allows simultaneously all the objects to belong to several clusters at the time. Compare to the hard clustering, fuzzy clustering is most natural.

D. Random Forest (RF):

The Random Forest models were trained to predict crop yield using numerous biophysical variables as predictors. Environmental variables integrated soil, water, climate, and photoperiod and fertilization data. The equivalent data were used for training MLR models for benchmarking purposes. The RF algorithm as such set aside partial data for its own internal validation, referred to as out-of-bag (OOB) data.

However, to ensure a reasonable and conservative comparisons between MLR and RF, we tend to used only a random half of every dataset (i.e., potato, wheat, silage maize, maize grain) for training ('training dataset') both RF and MLR models. The other half that was not intended for training was then used because the 'test dataset' to validate and evaluate performances between the RF and MLR models.

E. Adaptive Neuro Fuzzy Inference System (ANFIS):

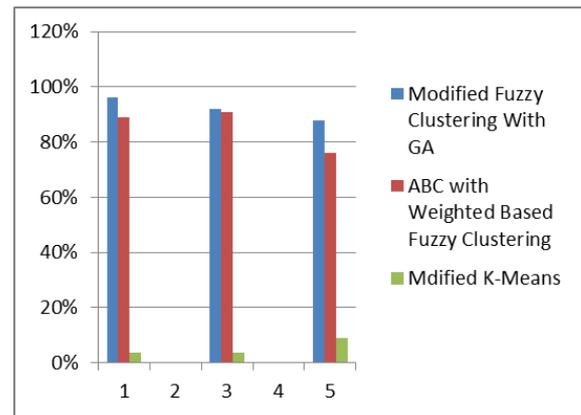
An adaptive neuro-fuzzy inference system is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system.

ANFIS is considered to be a universal estimator. For using the ANFIS in a more efficient and optimal way, one can use the best parameters obtained by genetic algorithm.

We compared ANFIS crop prediction model, ANFIS crop prediction model and MLR and FL based on RMSE values. When compared to all 3 models, ANFIS is give better accuracy then MLR and FL with lower RMSE value. This work analyses how yield of a particular crop is determined by few attributes. In this paper, the models of Fuzzy logic (FL), Adaptive Neuro Fuzzy Inference System (ANFIS) and Multiple Linear Regression (MLR) are used for predicting the yield of wheat by considering biomass, extractable soil water(esw), Radiation and rain as input parameters. The outcome of the prediction models will assist agriculture agencies in providing farmers with valuable information as to which factors contribute to high wheat yield. We compare all these models based on RMSE values. Results show that the ANFIS model performs better than MLR and FL models with a lower RMSE value.

F. Performance Analysis

Figure: 1 Analysis Graph



IV. RESULT AND DISCUSSION

This paper is an opening for further research in predicting the crop yield using data mining algorithms. In agriculture, there are a growing number of applications of data mining techniques are available and an increasing number of data that are presently available from several resources. This paper presents in which different methods and techniques are given best method of an approach to the problem of crop yield prediction

While comparing with other algorithms Modified Fuzzy Clustering with Genetic Algorithms gives best accuracy and precision value.

V. REFERENCES

- [1]. Anjana Gosain, Sonika Dahiya, "Performance Analysis of Various Fuzzy Clustering Algorithms: A Review", 7th International Conference on Communication, Computing and Virtualization 2016.
- [2].Mucherino A, Papajorgji P, and Pardalos PM "A survey of data mining techniques applied to agriculture". Oper Res. 2009, 9 (2): 121-140.
- [3].Anita Arumugam, "A predictive modeling approach for improving paddy crop productivity using data mining technique", Turkish Journal of Electrical Engineering & Computer Sciences, 2017.
- [4].Rajendrachoudhary and Dr.Aakashsaxena, "Hybrid algorithm of Neuron Fuzzy Inference System and LDA for Crop Yield Prediction", " International Journal of Research in Electronics and Computer engineering, 2018.
- [5].P.Surya, Dr. I.Laurence Aroquiaraj, "Performance Analysis of K-Means And K-Medoid Clustering Algorithms Using Agriculture Dataset", Journal of Emerging Technologies and Innovative Research (JETIR), Volume 6, Issue 1, January 2019.
- [6].D.Ramesh ,B Vishnu Vardhan, "Analysis of Crop Yield Prediction using Data Mining Techniques", Ijret: International journal of research in Engineering and Technology, eISSN: 2319-1163,pISSN:2321-7308,Volume:04 Issue:01,Jan-2015.
- [7].B M Sagar, Cauvery N K, "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review",Indonesian Journal of Electrical Engineering and Computer Science Vol.12, No.3, December 2018.