

An Overview of Data Mining Techniques for Maize Yield Prediction

M.Tamilselvi¹, B.Jaison²

¹Department of CSE, Roever Engineering College, Perambalur, India

²Department of CSE, R.M.K Engineering College, Chennai, India

Abstract

Agriculture is the main source for the people in India which ensures major source of economy for the country. Major harvest generation depends on suitable climatic conditions. More than 60 percent of crop depends on rainfall. However various natural disasters such as flood or lack of rainfall spoil the entire crop yield production. This leads to enormous financial loss for the farmers which in turn affect the crop production. Predicting the crop yield in advance through various attributes like rainfall, temperature, type of soil, soil nutrient, and seed rate will help the farmers and government organizations to make prior arrangement regarding sales, storage, Import/Export Details etc. Through data mining we extort the knowledge from the massive size of data. Different Data Mining Techniques can be utilized for estimating the future crop production. This paper deals with Maize crop yield prediction through Apriority Algorithm for association rule Mining and SVM classifier for dimensionality reduction. Furthermore performance of the proposed technique is improved with AdaBoost algorithm. The success of any crop yield prediction system heavily relies on how exactly the features have been extracted and how properly classifiers have been employed.

Keywords: crop yield prediction, Maize, data mining, Apriori algorithm, SVM Classifier, Ada boost algorithm.

1. Introduction

Maize is a standout amongst the most versatile rising product bit more extensive adaptability under fluctuated agro-climatic conditions. Maize has the most astounding characteristic yield potential among the oats and subsequently otherwise called ruler of oats since it. It is developed on almost 150 m ha in around 160 nations having more extensive assorted variety of soil, climate, biodiversity and the board rehearses that contributes 36 % (782 m t) in the worldwide grain creation. The biggest maker of maize contributes almost 35 % in USA. The USA has the most astounding profitability (> 9.6 t ha⁻¹) which is twofold than the worldwide normal (4.92 t ha⁻¹). Though, the normal efficiency in India is 2.43 t ha⁻¹. Maize is the third most critical nourishment crops after rice and wheat.

The maize is developed during the time in all conditions of the nation for different purposes including grain, grub, green cobs, sweet corn, child corn, pop corn in peri-urban zones. It tends to be developed effectively in assortment of soils stretching out from loamy sand to mud topsoil. Soils with great natural issue content having high water holding limit with impartial pH are viewed as useful for higher efficiency.

Yield forecast is an imperative agricultural issue. Each agriculturist is occupied with knowing, how much yield he is about anticipate. Before, yield forecast was performed by thinking about agriculturist's past involvement on a specific harvest. The volume of

information is tremendous in Indian horticulture. The information when moved toward becoming data is exceptionally valuable for some reasons. There are various approaches to increment and enhance the harvest yield and the nature of the crops.

Data innovations have the capacity to keep the hazard related with the farming and furthermore foresee the product yield all the more absolutely preceding harvest.



Fig 1: Sample Maize crop

Yield prediction requires different kinds of data gathered from different sources like meteorological data, agri - meteorological, soil (pH, N, P, K) data, remotely sensed data, agricultural statistics etc[]. To handle such a huge data the best option we have is Data Mining.

Data Mining is extensively useful to agricultural problems also helpful for predicting the crop yield fabrication. Data Mining is used to analyze large data sets and create useful classifications and patterns in the data sets. The overall goal of the Data Mining process is to mine the information from a data set and transform it into understandable structure for further use. This research focuses on evolution of a prediction model which may be used to predict crop yield fabrication.

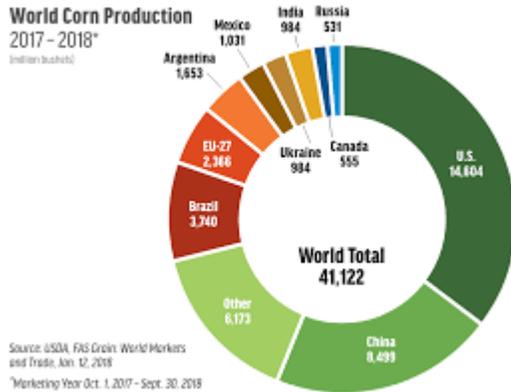


Fig 2: Maize yield in various countries

Data mining methods are mainly divided in two groups:

1. Classification
2. Clustering.

Classification techniques are considered to classify unidentified samples using information provided by a set of confidential samples. This set is usually referred to as the training set it is used to train the classification technique to perform its classification. In case a training set is not there, then there is no knowledge about the data to categorize. In such cases clustering technique can be used to associate set of unknown samples into clusters []. The proposed method use data mining technique to predict the crop yield production based on the Apriori algorithm.

The rest of the paper is organized as follows, Section II overviews the existing techniques. Section III presents the proposed methodology and Section IV concludes.

2. Related Work

From the research article [1], the researcher express that large amount of data which is collected and stored for analysis. Based on this data considerable gains in efficiency and therefore economic advantages.

In [2] R. Medar and V. Rajpurohit (2014), presents the various crop yield prediction methods using data mining techniques. Farming system is very difficult since it deals with huge amount data position which comes from a number of factors. Typical Data Mining strategies, extremely late utilizations of information mining procedures in farming field. Information mining innovation has gotten an extraordinary advance with the quick improvement of software engineering, computerized reasoning.

Information Mining is a developing examination field in agribusiness edit yield investigation. Information Mining is the way toward recognizing the concealed examples from extensive measure of information. The issue of yield forecast can be fathomed by utilizing information mining methods. In this paper a few information mining strategies were embraced keeping in mind the

end goal to gauge trim yield examination with existing information and their utilization in information mining.

Maize crop forecasting [3] has been done using multilayered feed forward network of ANN. They considered maize crop yield data as response variable and total human labor, farm power, fertilizer consumption, and pesticide consumption as predictors.

Crop prediction model framework was developed and concluded that climate related variables were not the main determinants of corn yield, rather yield was greatly affected by planting practices, particularly by the application right amount of fertilization [4]

In the research paper by David H White and S Mark Howden[5], they focus on the climates determinants of crop productivity. They considered how the climate envelopes different crops based on temperature, moisture and light 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.

3. Proposed Methodology

Data mining procedure is separated into seven methods[]:

- Data cleaning
- Data integration
- Data selection
- Data transformation
- Data mining
- Pattern estimation
- Knowledge display

To come to gather rapidly corn requires warm temperatures, rich soil, and even, standard watering. Corn is wind-pollinated so planting in squares or different lines to guarantee fertilization is critical. Early, Plant Effectively

- Practice Seasonal Soil Rotation
- Know The Yield Potential
- Always Scout Your Fields
- Ensure Proper Water Drainage
- Utilize Fertilizers

1. Input

Climatic parameters like temperature, humidity, rainfall and some agronomical parameters like soil, nutrient contents like N, P, K, and pesticides etc. The values of these variables have been taken as input.

2. Preprocessing

Huge dataset is required for the successful completion of the yield prediction. The data which is attained from various resources are sometime in raw form. It may contain some incomplete, redundant, inconsistent data. Therefore in this step such redundant data should be filtered. Information ought to be standardized utilizing Min-Max standardization method.

The system which gives straight change on unique scope of information is called Min-Mix Normalization. The procedure which keeps relationship among unique information is called Min-Mix Normalization. Min-Max standardization is a basic method where the strategy can explicitly fit the information in a pre-characterized limit.

3. Feature Extraction (Attribute Selection)

This step aims at identifying and using most relevant attribute from the dataset. Through this process irrelevant and redundant information is removed for the application of classifiers. Apriori algorithm is a traditional algorithm in data mining. It is used for mining regular datasets and relevant association rules. It is devised to function on a record containing a lot of communication, for example, objects brought by customers in a store.

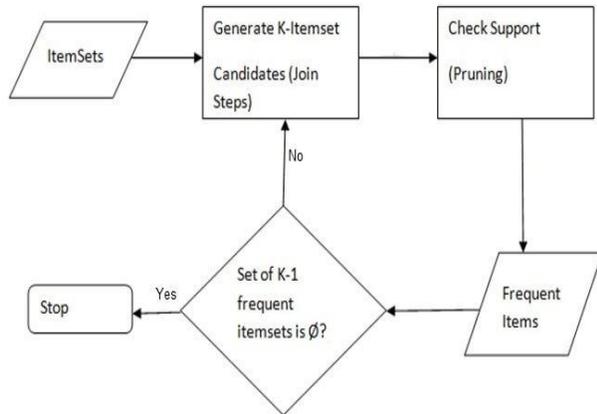


Fig 3: Steps in Apriori algorithm

AdaBoost is regularly alluded to as the best out-of-the-crate classifier. At the point when utilized with choice tree learning, data accumulated at each phase of the AdaBoost calculation about the relative 'hardness' of each preparation test is bolstered into the tree developing calculation to such an extent that later trees will in general spotlight on harder-to-arrange models. Ada-help , retrains the calculation iteratively by picking the preparation set dependent on exactness of past preparing. The weight-age of each prepared classifier at any cycle relies upon the precision accomplished.

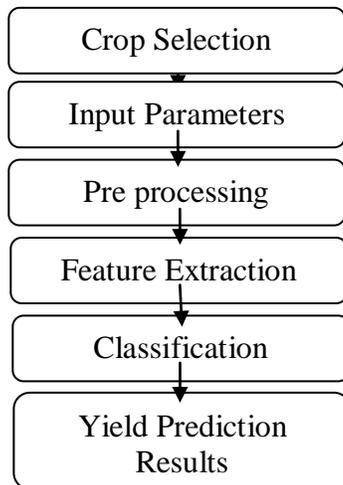


Fig 4: Proposed System

A brief overview about the proposed system

1. Input data: Parameters such as information regarding soil (Nitrogen (N), Phosphorus(P),Potassium(K) content), Micronutrients

present in soil, Moisture in soil etc. are collected over some period of time.

2. Preprocessing: Data which is collected should be preprocessed to remove redundant data and to avoid inconsistency using Min-Max normalization technique.

3. Attribute Selection: Important Features are extracted by Apriori algorithm.

4. Classification Algorithm: An efficient Adaboost classifier is built.

5. Result: Prediction or recommendation can be provided to the farmers based on the results obtained.

4. Conclusion

It can be observed that, the proposed system provides prediction about the maize crop yield using data mining which helps the farmers to support for increasing financial income. Classification and Rule mining techniques are used to predict the maize crop yield. Attributes such as rainfall, humidity, soil nutrient etc., are considered as parameters to determine the yield. Apriori algorithm is used as rule mining technique for extracting the frequent datasets. Additionally Adaboost algorithm is used for improving the performance of classification.

References

- [1] 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.
- [2] R. Medar, V. Rajpurohit, "A survey on data mining techniques for crop yield prediction", International Journal of Advance Research in Computer Science and Management Studies, vol. 2, no. 9, pp. 59-64, 2014.
- [3] Prajneshu, R. K. (2008). Artificial Neural Network Methodology for Modelling and Forecasting. Agricultural Economics Research Review, 21, 5-10.
- [4] Rossana MC, L. D. (2013). A Prediction Model Framework for Crop Yield Prediction. Asia Pacific Industrial Engineering and Management Society Conference Proceedings Cebu, Phillipines, 185.
- [5] White, David, Howden, S. Mark, Climate Change: Significance for Agriculture and Forestry, Springer, May 1994.
- [6] Ramesh A. Medar, Vijay S. Rajpurohit, "A survey on Data Mining Techniques for crop yield prediction", International Journal of Advance Research in Computer Science and Management Studies. Volume 2 , Issue 9, Sept 2014.
- [7] Jeysenthil.KMS, Manikandan.T, Murali.E "Third Generation Agricultural Support System Development Using Data Mining" International Journal of Innovative Research in Science, Engineering and Technology Vol. 3, Issue 3, March 2014.