International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 03 | Mar 2020 www.irjet.net p-ISSN: 2395-0072

ENLIGHTENING FARMERS ON CROP YIELD

IMRAAN. P[1], SRI PRATHYUSHA. L[2], SHAMIMA BHANU. S[3], VENKATA KAVYA. M[4]

[1,2,3,4]*B.Tech Students, Dept of CSE, Tirumala Engineering College, Narasaropet, Guntur, A.P., India*

Abstract - *Data Mining is emerging research field in crop yield analysis. Yield prediction is a very important issue in agricultural. Any farmer is interested in knowing how much yield he is about to expect. In the past, yield prediction was performed by considering farmer's experience on particular field and crop. The yield prediction is a major issue that remains to be solved based on available data. Data mining techniques are the better choice for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This research proposes and implements a system to predict crop yield from previous data. In this research we are using Random Forest Regression Algorithm This research helps the farmers in agriculture letting them know the prediction values of the production and recommends the user to select the crop for their better production in the future. It also shows them the visualized data in which it helps the farmer understand the downfalls of the previous crop which they held. It also recommends the user to sell their crop to the better industries in order make them profits. In this way the system guides them to the better production of the crops. The system lets the user or the farmer to know much more about their crop wanted to get high yield.*

Key Words: **Data Mining, Prediction, Recommender System, Data Visualization, Random Forest Regression.**

1. INTRODUCTION

India is an associate agricultural sub-continent, its finance structure preponderantly based on agricultural production growth plus allied agricultural business merchandise. It is the broadest economic sector and plays a vital role in overall development of the country. Thus, modernization of agriculture is extremely vital and therefore can lead the farmers of our country onto the Indian economical profit. Data Analytic (DA) is that the method of examining information sets so as to draw conclusions concerning the data they contain, progressively with the help of specialized systems and software system. The objective of this research is to allow the farmers to let know their crop yield and how to technologically use their fertilizers and soil benefits. Technologically, this terminology refers to the work we are doing, using the machine learning techniques to help farmers easy their jobs. This lets the farmers know their productivity issues, water usage, soil fertility, and fertilizer availability for their crop.

1.1 Crop Selection and Crop Yield Prediction

To maximize the crop yield, selection of the appropriate crop that will be sown plays a vital role. It depends on various factors like the type of soil and its composition, climate, geography of the region, crop yield, market prices etc. Techniques like Artificial neural networks, K-nearest neighbors and Decision Trees have carved a niche for themselves in the context of crop selection which is based on various factors. Crop selection based on the effect of natural calamities like famines has been done based on machine learning (Washington Okori, 2011). The use of artificial neural networks to choose the crops based on soil and climate has been shown by researchers (Obua, 2011). A plant nutrient management system has been proposed based on machine learning methods to meet the needs of soil, maintain its fertility levels, and hence improve the crop yield (Shivnath Ghosh, 2014). A crop selection method called CSM has been proposed which helps in crop selection based on its yield prediction and other factors (Kumar, 2009).

1.2 Weather Forecasting

Indian agriculture mainly relies on seasonal rains for irrigation. Therefore, an accurate forecast of weather can reduce the enormous toil faced by farmers in India including crop selection, watering and harvesting. As the farmers have poor access to the Internet as a result of digital-divide, they have to rely on the little information available regarding weather reports. Up-to-date as well as accurate weather information is still not available as the weather changes dynamically over time. Researchers have been working on improving the accuracy of weather predictions by using a variety of algorithms. Artificial Neural networks have been adopted extensively for this purpose. Likewise, weather prediction based on machine learning technique called Support Vector Machines had been proposed (M.Shashi, 2009). These algorithms have shown better results over the conventional algorithms.

2. Related Work

Agricultural management needs simple and accurate estimation techniques to predict rice yields in the planning process (Ji & Wan, 2007). The necessity of the present study were to: (Washington Okori, 2011) identify whether artificial neural network (ANN) models could effectively predict rice yield for typical climatic conditions of the mountainous region, (Miss.Snehal, 2014) evaluate ANN model performance relative to variations of developmental parameters and (Shivnath Ghosh, 2014) compare the

effectiveness of multiple linear regression models with ANN models.

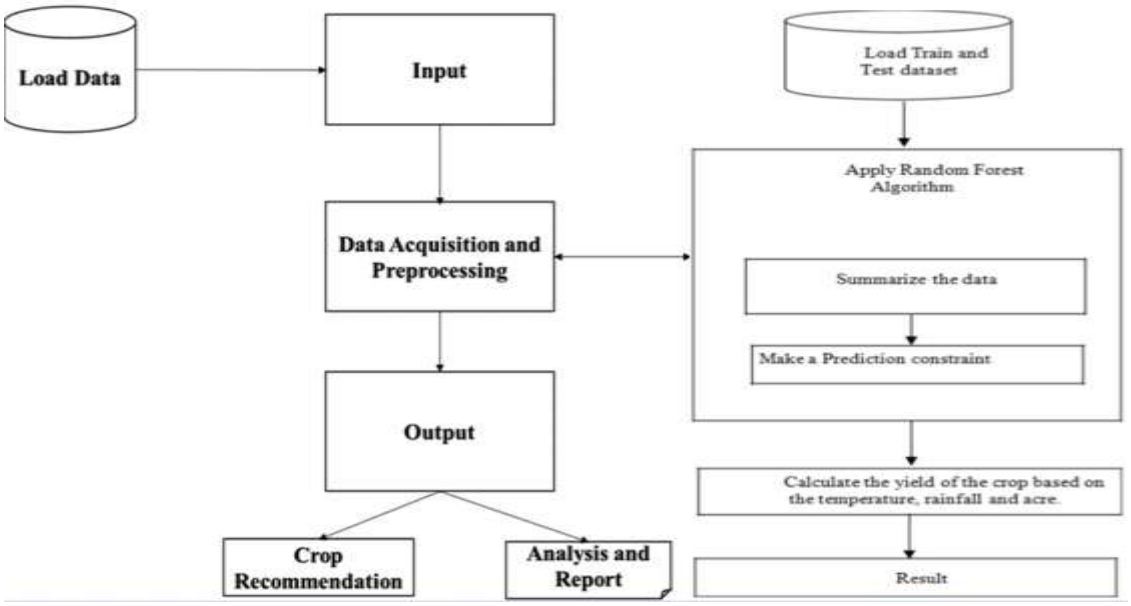
Generalized Regression Neural Networks (GRNN) method is used for forecasting of agricultural crop production (Chaochong, 2008). They found GRNN to be a good technique for prediction grain production in rural areas. It was reported that GRNN model is suitable for non-linear, multi-objectives and multivariate forecasting.

Time series analysis is a method to analyze time on parametric, series data to extract meaningful statistics and other characteristics of the data. Time series forecasting is a model to predict future values based on previously observed values. New concept of crop yield under average climate conditions was described and it is used in time series techniques on the past yield data to set up a forecasting model.

3. CROP YIELD PREDICTION

Data Mining is widely applied to agricultural issues. 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.

This paper analyzes the crop yield production based on available data. The Data mining technique was used to predict the crop yield for maximizing the crop productivity. Figure 1 shows the flow of proposed crop yield prediction.



3.1 Data Set Gathering:

A data set is a collection of data. In other words, a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set. Gathering data is the most important step in solving any supervised machine learning problem. Our text classifier can only be as good as the dataset it is built from.

3.2 Data Set Overview

3.3 Data Pre-Processing

Data pre-processing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data pre-processing prepares raw data for further processing.

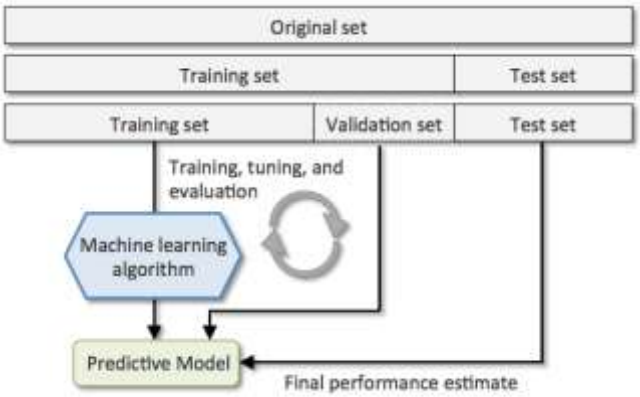
3.4 Data Transformation (Normalization)

Data Normalization is necessary for Machine Learning models. Normalization is a technique often applied as part of data preparation for machine learning. The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values.

3.5 Training and Testing data set

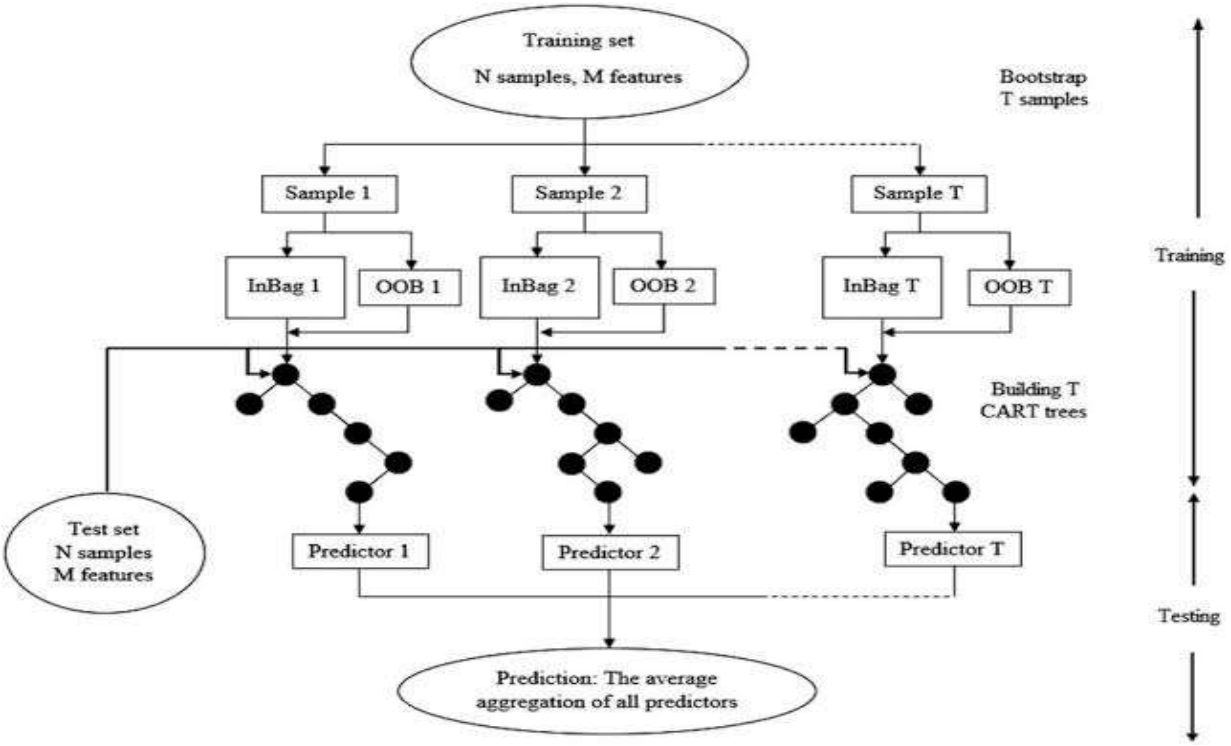
The training data set is the one used to train an algorithm to understand how to apply concepts to learn and produce results. It includes both input data and the expected output. Training sets make up the majority of the total data, around 60 %. In testing, the models are fit to parameters in a process that is known as adjusting weights.

The test data set is used to evaluate how well your algorithm was trained with the training data set. In AI projects, we can't use the training data set in the testing stage because the algorithm will already know in advance the expected output which is not our goal. Testing sets represent 20% of the data. The test set is ensured to be the input data grouped together with verified correct outputs, generally by human verification.



3.6 Random Forest Regression

The random forest is a classification algorithm consisting of many decisions trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree.

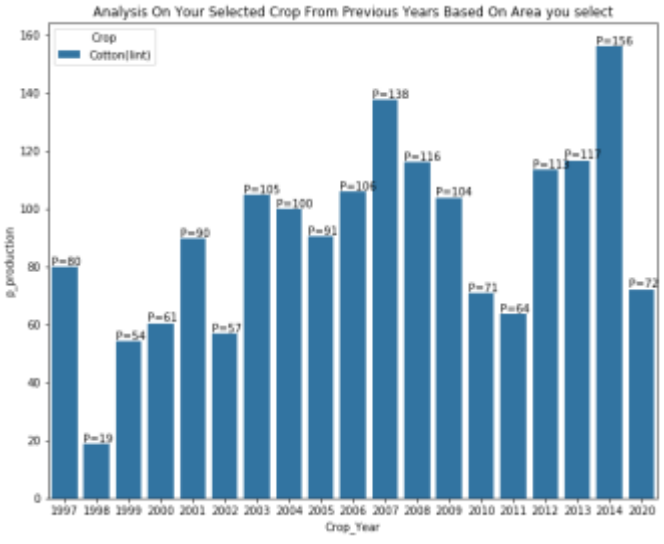


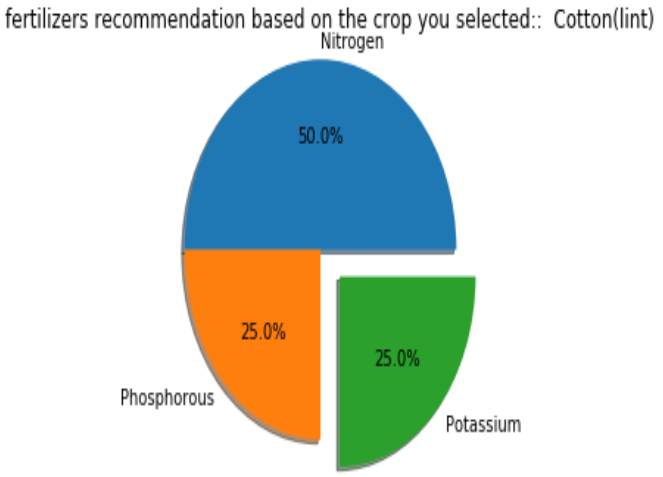
3.7 Prediction Model for Recommendation System

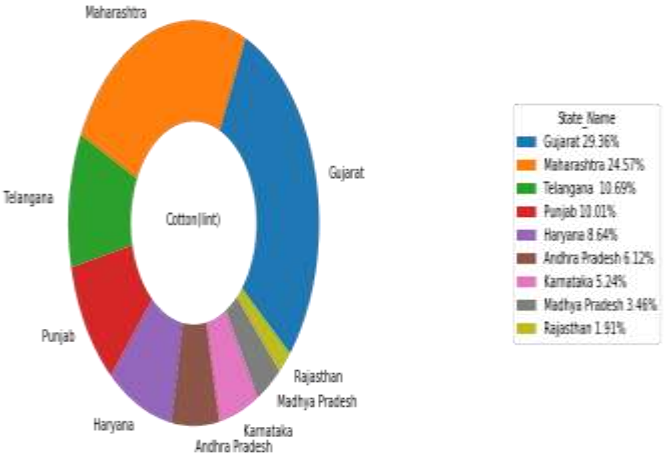
Predictive modelling is a process that uses data mining and probability to forecast outcomes. Each model is made up of a number of predictors, which are variables that are likely to influence future results. Once data has been collected for relevant predictors, a statistical model is formulated. Using Machine learning techniques, a recommender system makes prediction based on users' historical behaviors. Specifically, it's to predict user preference for a set of items based on past experience.

3.8 Visualization

Data Visualization is the presentation of data in graphical format. It helps people understand the significance of data by summarizing and presenting a huge amount of data in a simple and easy-to-understand format and helps communicate information clearly and effectively.







4. CONCLUSION

Crop yield prediction is still remaining as a challenging issue for farmers. The aim of this research is to propose and implement a regression based system to predict the crop yield production from the collection of past data. This has been achieved by applying as random forest regression mining on agriculture data from 2000 to 2014.

REFERENCES

[1]. Boryan, Claire. (2012) "A Review of Four Area Sample Designs", Paper prepared for course GGS796 at George Mason University; Available from Research Division, National Agricultural Statistics Service, USDA, Fairfax, Virginia.

[2]. the state of agricultural commodity markets, Agricultural Trade, Climate Change and Food security. http://www.fao.org/3/I9542EN/i9542en.pdf

[3]. Data flow diagrams for the crop-based techniques, https://www.researchgate.net/figure/Dataflow-diagram-level-0-agricultural-management-system-centre_fig2_322941382

[4].https://www.google.com/search?q=indian+farmers+suiciding+2017&rlz=1C1CHWL_enIN864IN864&oq=indian+farmers+sui&aqs=chrome.1.69i57j0l4.12066j0j7&sourceid=chrome&ie= UTF-8

[5]. Pocket book of agricultural statistics by Government of India, Ministry of Agricultural and Farmers Welfare, http://agricoop.nic.in/sites/default/files/pocketbook_0.pdf

[6].https://www.economicsdiscussion.net/india/agriculture-india/development-of-agriculturalarunder-5-year-plans-in-india/21674

[7]. Richard Kidd Perrin, Analysis and prediction of crop yields for agricultural policy purposes, Part of the Agricultural and Resource Economics Commons, and the Agricultural Economics Commons.

[8].http://cidportal.jrc.ec.europa.eu/thematicportals/agri4cast