ESTIMATE THE CROP YIELD USING DATA ANALYTICS

SUBMITTED BY

MAHALAKSHMI V	(113219031083)
MAHALAKSHMI L	(113219031082)
KEERTHANA C	(113219031071)
RAMYA V S	(113219031121)

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

<h1 style="text-align:center"><u>LITERATURE SURVEY</u></h1>

<u>JOURNAL TITLE</u> 1

<p style="text-align:center">Analysis Of Crop Yield Prediction Using Data Mining Techniques</p>

<u>ABSTRACT</u>

Agrarian sector in India is facing rigorous problem to maximize the crop productivity. More than 60 percent of the crop still depends on monsoon rainfall. Recent developments in Information Technology for agriculture field has become an interesting research area to predict the crop yield. The problem of yield prediction is a major problem that remains to be solved based on available data. Data Mining techniques are the better choices for this purpose. Different Data Mining techniques are used and evaluated in agriculture for estimating the future year's crop production. This paper presents a brief analysis of crop yield prediction using Multiple Linear Regression (MLR) technique and Density based clustering technique for the selected region i.e. East Godavari district of Andhra Pradesh in India.

<u>ALGORITHM/METHODOLOGY</u>

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).

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.

BLOCK DIAGRAM
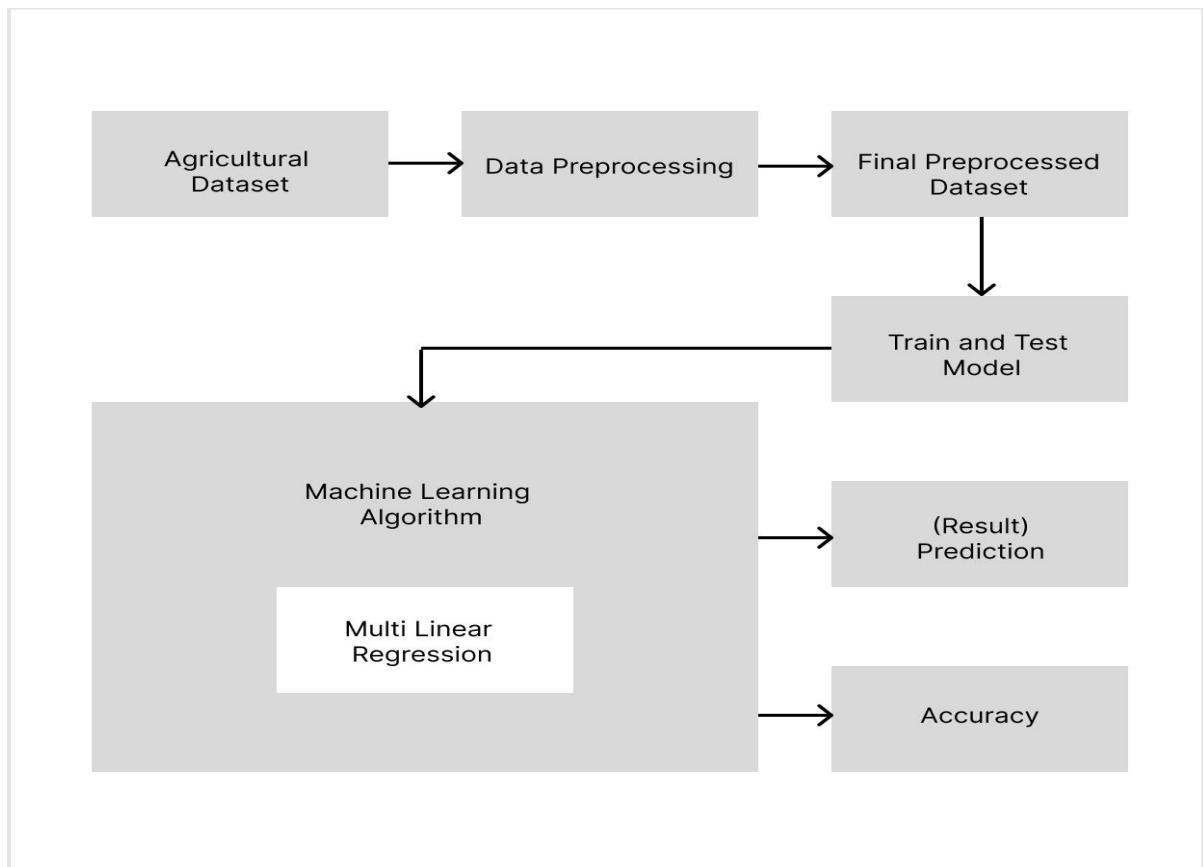
FEATURES

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.

DRAWBACKS

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.

JOURNAL TITLE 2

Design And Implementation Of Crop Yield Prediction Model In Agriculture

## ABSTRACT

Agriculture is the best utility region especially inside the developing worldwide areas like India. Usage of records age in agriculture can substitute the circumstance of decision making and Farmers can yield in higher manner. About portion of the number of inhabitants in India relies upon on farming for its occupation however its commitment towards the GDP of India is just 14 percent. One suitable explanation behind this is the deficiency of adequate decision making by farmers on yield prediction. There isn't any framework in location to suggest farmer what plants to grow. The proposed machine learning approach aims at predicting the best yielded crop for a particular region by analyzing various atmospheric factors like rainfall, temperature, humidity etc., and land factors like soil pH, soil type including past records of crops grown. Finally our system is expected to predict the best yield based on dataset we have collected.

ALGORITHM/METHODOLOGY

Machine learning mainly consists of three learning methods, namely supervised learning, reinforcement learning and unsupervised learning methods of training a model. Supervised learning is a learning method which maps known input resulted into output which maps from input to output. But in case of unsupervised leaning we would not know targeted output in this learning we should train the model in order to get desired output.

BLOCK DIAGRAM
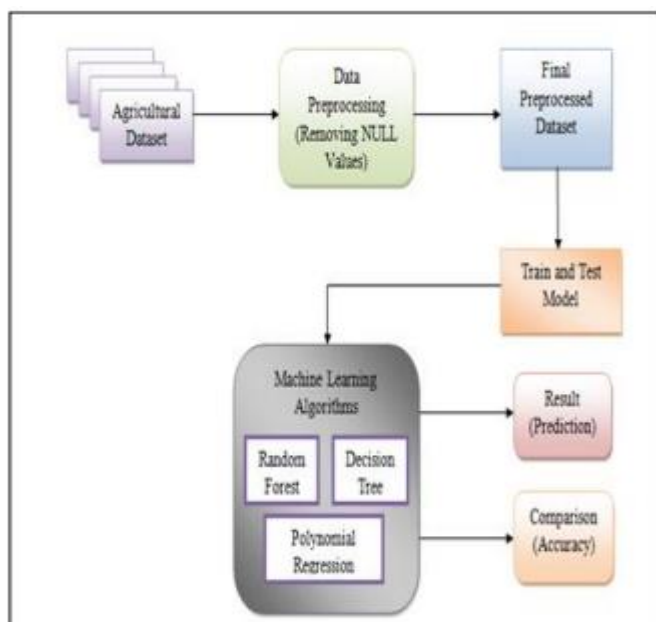
FEATURES

This project is undertaken using machine learning and evaluates the performance by using Random forest, Polynomial Regression and Decision Tree algorithms. In our proposed model among all the three algorithm Random forest gives the better yield prediction as compared to other algorithms. Along with random forest, Polynomial Regression, Decision Tree model classify the output that shows improvements in dataset. So we analyzed that proposed model has got more efficiency than the existing model for finding crop yield.

DRAWBACKS

we analyzed that proposed model has got more efficiency than the existing model for finding crop yield. The implementation of above system would help in better cultivation of the agricultural practices of our country. Further it can be used to reduce the loss faced by the farmers and improve the crop yield to get better capital in agriculture. The model can be improved by integrating this with other departments like horticulture, sericulture, and others towards the agricultural development of our country.

JOURNAL TITLE 3

Crop yield forecasting using data mining

ABSTRACT

India is a heavily reliant on agriculture. Organic, economic, and seasonal factors all influence agricultural yield. Estimating agricultural production is a difficult task for our country, particularly given the current population situation. Crop production assumptions made far in advance can help farmers make the necessary planning for things like storing and marketing. Crop production prediction involves a huge amount of data, making it a per fect candidate for data mining methods. Data mining is method of accumulating previously unseen anticipated information from vast database. Data mining assists in the analysis of future patterns and character, enabling com panies to make informed decisions. For a specific region, this research provides a fast inspection of agricultural yield forecast using the Random Forest approach.

ALGORITHM/METHODOLOGY

The Important Classification Algorithm Random Forest is applied to the data collection provided from the Official Government website. To ensure accuracy, the datasets are examined. Random Forest is a supervised learning technique for classifying a dicting datasets. It will choose a collection of features at random from the dataset's attributes and construct a set of decision trees by locating the root nodes and splitting the attributes. Following the creation of the forest, the best decision is made based on the highest number of votes among the projected targets as the classifier's final prediction. Crop yield prediction

systems provide for better planning and decision-making to increase production. The proposed system involves a prediction module based on data mining classification algorithm namely Random Forest used to forecast the yield of major crops based on historical data.

BLOCK DIAGRAM
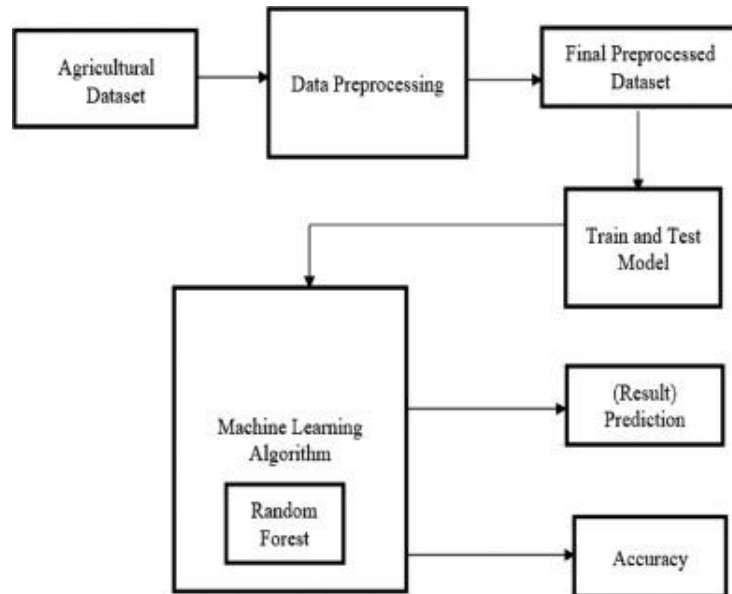
FEATURES

The algorithms for predicting crop yield based on temperature, season, and location. A Yield prediction for a specific district can be made by combining Precipitation, Temperature, and other parameters such as season and location. When all the factors are considered, Random Forest emerges as the greatest classifier. The dataset which is in use with more features increases the accuracy rate. Random forest is the superior prediction algorithm when compared to other technologies that are multiple linear regression and decision trees

DRAWBACKS

Our dataset contains a lot more variables, resulting in more accurate predictions. The introduction of this project which are helpful to the farmers to reduce their losses and increase crop yields to increase their resources in agriculture. This will not only help farmers choose the best crop to cultivate in the future season, but it will also help bridge the technological and agricultural divide. Limitation of our project is, Yield is predicted for 100acres and implemented for 30 districts. The Future work of our project is to overcome our limitations.