CROP YIELD PREDICTION SYSTEM USING MACHINE LEARNING

M. Kaleeswari and Mr. R. Raja guru

Final year PG student and assistant professor,

Department of Information Technology

G. Venkataswamy Naidu college, kovilpatti, Tamil Nadu, India

ABSTRACT

Agriculture is one of the major and the least paid occupation in India. Machine learning can bring a boom in the agriculture field by changing the income scenario through growing the optimum crop. This paper focuses on predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc.

Keywords: crop yield prediction; simple RNN; random forest; machine learning.

I. INTRODUCTION

The history of agriculture in India dates back to the Indus Valley Civilization Era. India ranks second in this sector. Agriculture and allied sectors like forestry and fisheries account for 15.4 percent of the GDP (gross domestic product) with about 31 percent of the workforce. India ranks first globally with the highest net cropped area followed by US and China. Agriculture is demographically the broadest economic sector and plays a significant role in the overall socio-economic fabric of India. Due to the revolution in industrialization, the economic contribution of agriculture to India's GDP is steadily declining with the country's broad-based economic growth. The problem that the Indian Agriculture sector is facing is the integration of technology to bring the desired outputs. With the advent of new technologies and over use of non-renewable energy resources patterns of rainfall and temperature are disturbed. The inconsistent trends developed from the side effects of global warming make it cumbersome for the farmers to clearly predict the temperature and rainfall patterns thus affecting their crop yield productivity. In order to perform accurate prediction and handle inconsistent trends in temperature and rainfall various machine learning algorithms like RNN, LSTM, etc can be applied to get a pattern. It will complement the agricultural growth in India and all together augment the ease of living for farmers. In past, many researchers have applied machine learning

techniques to enhance agricultural growth of the country.

II. LITERATURE REVIEW

Balamurugan[1] have implemented crop yield pre- diction using only random forest classifier. Various features like rainfall, temperature and season were taken into account to predict the crop yield. Other machine learning algorithms were not applied to the datasets. With the absence of other algorithms, comparison and quantification were missing thus unable to provide the apt algorithm.

Mishra[2], has theoretically described various machine learning techniques that can be applied in various forecasting areas. However, their work fails to implement any algorithms and thus cannot pro- vide a clear insight into the practicality of the proposed work.

Manjulaset[3] research aimed to propose and implement a rule-based system to predict the crop yield production from the collection of past data by applying association rule mining on agriculture data from 2000 to 2012. The dataset used in this research is limited to the southern district of India thus limiting its scope for pan India implementation.

Dahikar and Rode[4] in their research provided the datasets of different features and applied Artificial Neural Networks for crop yield prediction. However, Dahikar fails to give a practical implementation of his proposed work. This paper focuses on the practical application of machine learning algorithms and its quantification. The work presented here also takes into account the inconsistent data from rainfall and temperature datasets to get a consistent trend. Crop yield prediction is determined by considering all the features in contrast with the usual trend of determining the prediction considering one feature at a time.

III. METHODOLOGY

A. *Factors affecting Crop Yield*

There are number of factors that influence the yield of any crop prediction. These are basically the features that help in predicting the production of any crop over the year. Some of the major factors are as follows:

1. Temperature
2. Rainfall
3. Area
4. Season

B. Data Pre-Processing

Data Preprocessing is a method that is used to convert the raw data into a clean data set. The data are gathered from different sources, it is collected in raw format which is not feasible for the analysis. By applying different techniques like replacing missing values and null values, we can transform data into an understandable format. The final step on data preprocessing is the splitting of training and testing data. The data usually tend to be split unequally because training the model usually requires as much data-points as possible. The training dataset is the initial dataset used to train ML algorithms to learn and produce right predictions (Here 80% of dataset is taken as training dataset).

C. Comparison And Selection Of Machine Learning Algorithm

Before deciding on an algorithm to use, first we need to evaluate and compare, then choose the best one that fits this specific dataset. Machine Learning is the best technique which gives a better practical solution to crop yield problem. There are a lot of machine learning algorithms used for predicting the crop yield. In this paper we include the following machine learning algorithms for selection and accuracy comparison:

- *Logistic Regression*:- Logistic regression is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes. When logistic regression algorithm applied on our dataset it provides an accuracy of 87.8%.

- *Random Forest*:- Random Forest has the ability to analyze crop growth related to the current climatic conditions and biophysical change. Random forest algorithm creates decision trees on different data samples and then predict the data from each subset and then by voting gives better solution for the system. Random Forest uses the bagging method to train the data which increases the accuracy of the result. For our data, RF provides an accuracy of 92.81%.

D. Random Forest Model For Crop Prediction
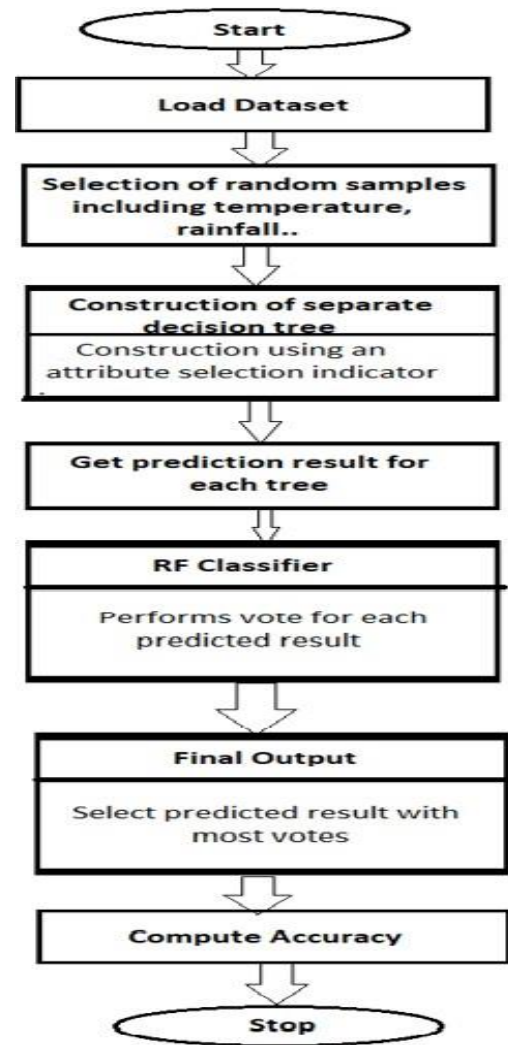
Flow chart for Random Forest Model

Random forests are the aggregation of tree predictors in such a way that each tree depends on the values of a random subset sampled independently and with the same distribution for all trees in the forest. Random Forest used the bagging method to trained the data which increases the accuracy of the result. For getting high accuracy we used the Random Forest algorithm which gives accuracy which predicate

by model and actual outcome of predication in the dataset. The predicted accuracy of the model is analyzed 91.34%. Fig.2 shows the flowchart of random forest model for crop yield prediction.

IV. CONCLUSION

Data mining is the latest research area of agriculture. This is reasonably a fresh research field and it is projected to grow in the future. Nowadays, farmers are struggling to produce the yield because of unpredictable climatic changes and range of rainfall. The experiments conducted to analyze a small number of traits contained within the dataset to decide their effectiveness when compared with standard statistical techniques. Feature selection is very important to classify the crop data. In this research work, decision tree concept was used to build a model in finding a solution of which crop is better for cultivation. RF algorithm is used to classify the crops and a solution to increase their yield by analyzing the climatic changes. This work helps to predict the best crop for analyzing the temperature and rainfall. This will help farmers to increase the yield and income level. The obtained accuracy of RF algorithm is 92.83 % and it is better than other algorithms.

V.REFERENCES

[1] Aruvansh Nigam, Saksham Garg, Archit Agrawal "Crop Yield Prediction using ML Algorithms ",2019.

[2] Leo Brieman, "Random Forests", 2001

[3]Priya, P., Muthaiah, U., Balamurugan, M."Predicting Yield of the Crop Using Machine Learning Algorithm",2015

[4] Mishra, S., Mishra, D., Santra, G. H.,"Applications of machine learning techniques in agricultural crop production",2016

[5] Dr.Y Jeevan Kumar,"Supervised Learning Approach for Crop Production",2020

[6] Ramesh Medar,Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2019

[7] Ranjini B Guruprasad, Kumar Saurav, Sukanya Randhawa,"Machine Learning Methodologies for Paddy Yield Estimation in India: A CASE STUDY", 2019

[8] Sangeeta, Shruthi G, "Design and Implementation of Crop Yield Prediction Model in Agriculture", 2020.