Survey Paper on Crop Suggestion and Yield Prediction using Machine Learning

Pratik Jaiswal[#1], Neha Korade[#2], Vaidehi Rasane[#3], Pavan Soni[#4], S.A.J. Patel[#5]

[#]*Information Technology, Sinhgad Institute of Technology and Science, Narhe*

[1]*pratikjaiswal.1234@gmail.com*

[2]*nehakorade5@gmail.com*

[3]*rasanevaidehi@gmail.com*

[4]*sonipavan002@gmail.com*

[5] *sapatel_sits@sinhgad.edu*

Abstract

India being an agricultural country, its economy predominantly depends on agriculture yield growth and agro-industry products. The major problem in the crop productivity is that farmers do not choose the right crop for cultivation by taking into consideration various factors like the location of their farm, type of soil, climatic conditions, etc. Hence in order to increase the crop productivity, a crop recommendation system is to be developed which will help the farmer to choose the right crop. Similarly, the yield of the crop is also important. Yield prediction is an important issue in the agriculture. Every farmer is interested in knowing the amount of yield he is about to expect. Hence along with suggestion of right crop it is also important to build a system which will predict the yield of the suggested crop.

Keywords— Crop suggestion, yield prediction, machine learning, agriculture, data analysis

I. INTRODUCTION

Agricultural researchers insist on the need for an efficient mechanism to predict and improve the crop growth and its yield. India being an agricultural country, its economy predominantly depends on agricultural yield growth and agro-industry products. Yield prediction is an important issue in agriculture. Any farmer is interested in knowing how much yield he is about to expect. For this the agricultural data is analysed and mined to get useful results using technologies like data analytics and machine learning.

The Data Analysis is the process of inspecting, cleaning, modelling data with the goal of discovering useful information and conclusions. It is a process of analysing, extracting and predicting the meaningful information from huge data to extract some pattern.

Most farmers rely on their long-term experiences in the field on particular crops to expect a higher yield in the next harvesting period but still they don't get worth price of the crops. It is mostly happening due to improper irrigation or inappropriate crops selection or also sometimes the crop yield is less than that of expected.

The idea is to identify the suitability of crops and yield based on various factors that affect the production can increase the quality and the yield of crops, thereby increase the economic growth and attain profitability. The outcome of crop yield primarily depends on parameters such as State, District, Area, Soil type, Year, Season, Crop name, Production.

II. RELATED WORK

Literature survey is the most important step in any kind of research. It describes how the proposed research is related to prior research. Before we start developing, we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers. In this section, we briefly review the related work on Crop Suggestion and Crop Yield Prediction and their different techniques.

A. In paper [1] authors have mainly focussed on analysing the soil dataset that comprise of physical and chemical characteristic of soil and also various climatic condition to build a crop recommendation system using ensemble modelling and majority voting technique, to improve the crop productivity.

B. Paper [2] determines two factor that are important in determining crop yield that are soil type and soil fertility.
The model build recommends required fertilizers ratio on the basis of atmospheric and soil parameters and predict precise and accurate crop yield.

C. Yield prediction is important agriculture issue as earlier farmer use to determine the yield on the basis of there previous experience. Due to varying climatic condition its difficult for farmers to understand which crop to be grown to get better yield. In this paper [3] authors have analysed the dataset that comprise of attributes like land(acres), rainfall, production, season, crop name, soil type and used smart technologies to predict the crop yield with maximum accuracy.

D. Soil is one the key factor in determining which crop is to be grown in particular soil and season. In this paper [4] the model classifies various kind of soil series and suggest suitable crop to be grown to improve productivity

III. FREQUENTLY USED ALGORITHMS

A. Random Forest Algorithm:

It is a method that operates by constructing multiple decision trees during the training phase. The decision of majority of the trees is chosen by random forest as final decision. The main advantages of random forest are that it provides better accuracy, easy to understand and parallelize and requires training time.

B. Naïve Bayes Algorithm:

It is a classification algorithm for binary and multiclass classification problems. It works on the principal of conditional probability as given by Bayes Theorem. It calculates the probability of occurrence of an event based on prior knowledge of condition that may be related to the event. This algorithm is easy to understand, can make prediction in real time, can handle both continuous and discrete data and is fast and highly scalable.

C. Linear SVM Algorithm:

Linear SVM is the current machine learning algorithm that is the quickest to solve the multi class classification problems. It is a supervised learning algorithm that processes the data and creates a line or a hyperplane which separates the data into classes. It works well with extremely large datasets and provides output with high accuracy.

D. Majority Voting Technique:

Majority voting technique is method of combining the class labels of independent classifiers and the result label is the one with maximum votes. This technique is also called as plurality vote.

E. Back-Propagation Algorithm:

This algorithm is used for large datasets which have no proper relationships between the attributes of the dataset to form a network model by training the dataset and predict the output. This algorithm mainly consists of three layers in network model. They are input layer, hidden layer and output layer. It is the method of updating the weights of the network in order to reduce the error in prediction. For this a random weight is chosen initially and an output is generated. This output is compared to the actual output and the error is calculated. If the error is positive then the weight is increased. If error is negative then the weight is decreased and if zero then that is the answer. After the adjusting the weight throughout the network the output generation and error calculation step is repeated until the error becomes zero or is the least. The training period of this algorithm is longer than the prediction period.

F. K-Means Clustering:

This algorithm is used to create clusters of the pre-processed data. It is an unsupervised, non-deterministic, numerical, iterative method of clustering. In this each cluster is represented by the mean value of object in the cluster. Data objects are similar to one another in the same cluster and dissimilar to objects in another cluster. In this algorithm, we firstly select k data values as initial cluster enter, then calculate distance between each data value and each cluster enter and assign it to the nearest cluster, update the averages of all clusters, repeat this process until the criteria is not matched.

G. Apriori Algorithm:

It is an algorithm used for mining frequent item sets for Boolean association rules. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as *candidate generation*), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found. It is used to count frequently occurring features which help to predict crop yield for a specific location.

IV. CONCLUSIONS

Survey of these paper helped in incorporating various mechanism to build efficient crop suggestion and crop yield prediction system with maximum profit. In order to suggest crop and predict the yield analysis of data take place and then classified.

Paper [1] in this classification is done using various data mining algorithms and technique. Using this paper, we analyzed classification rules and Linear SVM is appropriate algorithm for our dataset and use in our project.

Paper [2,4] These papers proposed system in order to improve the production of crop and crop yield using various algorithms Random Forest, Back-propagation, K means clustering. Using this paper, we analyzed Random forest algorithm works efficiently with small and large dataset and with high accuracy on comparing with other algorithms.

Paper [3] This paper discusses various algorithms to predict the crop yield in order to achieve maximum accuracy. This paper provides the insight of technique decision trees and Naïve Bayes and

its efficiency with various dataset. On analysis Naïve Bayes works efficiently with our data set to achieve precise and accurate yield.

Paper [1] this paper implemented the Ensembling model and Majority voting technique in order to achieve maximum efficiency. On analysis, we understood using models individually are less efficient and also provide less accuracy than used in Ensemble model and generate output using majority voting technique.

REFERENCES

[1] Nidhi H Kulkarni, Dr. G N Srinivasan, Dr. B M Sagar, Dr. N K Cauvery, "Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique", 3rd IEEE International Conference on Computational Systems and Technology for Sustainable Solutions, 2018.

[2] S Bhanumathi, M Vineeth, N Rohit, "Crop Yield Prediction and Efficient Use of Fertilizer", International Conference on Communication and Signal Processing, April 4-6, 2019, India.

[3] Ms. Shreya V. Bhosale, Ms. Ruchita A. Thombare, Mrs. Prasanna G. Dhemey, Ms. Anagha N. Chaudhary, "Crop Yield Prediction Using Data Analytics & Hybrid Approach", Fourth International Conference On Computing Communication Control And Automation, 2018.

[4] Neha Rale, RaxitKumar Solanki, Doina Bein, "Prediction of Crop Cultivation", 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.