Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier

Article *in* International Journal of Computer Science and Engineering · May 2020

DOI: 10.14445/23488387/IJCSE-V7I5P101

2 authors:

Archana Kumaravel
Sri Ramakrishna Engineering College
5 PUBLICATIONS **58** CITATIONS

SEE PROFILE

K G Saranya
PSG College of Technology
5 PUBLICATIONS **58** CITATIONS

SEE PROFILE

Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier

K. Archana[1], Dr.K.G.Saranya[2]
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore.

Abstract *-- Agriculture is the keystone of a developing country such as India. For the revenue, the majority of their population depends on agriculture. Machine Learning is an imminent field of informatics that can be applied quite efficiently to the agricultural sector. Crop yield prediction and forecasting is essential for agricultural stakeholders which can be acquired through machine learning techniques. When the farmers are not aware of the soil nutrition and soil composition that results in minimal crop yield. Thus the proposed system developed, which in turn focuses on the macronutrients (NPK), pH and electrical conductivity in the soil and temperature for providing the most appropriate crop suggestions. The proposed system constructs a collaborative system of crop rotation, crop yield prediction and forecasting and fertilizer recommendation. In this project a system is developed which incorporates the agricultural dataset wherein voting based ensemble classifier algorithm is applied to suggest the appropriate crops. Crop yield prediction and forecasting will increase the agricultural production. Periodical crop rotation will improve the soil fertility. This system supports farmer friendly fertilization decision making. The accuracy of this system was around 92%.*

Keywords *-- Nitrogen, Phosphorus, Potassium, soil nutrition, yield prediction, crop rotation, fertilize recommendation, Ensemble classifier, voting.*

I. INTRODUCTION

Agriculture plays an essential part in an economy's life. They are the backbone of our country's economy system. One of the key problems confronting farmers is selecting the right crop for cultivation. Selection of crops is determined by several factors such as temperature, soil composition, market prices etc.

Machine Learning is a technique that uses complex algorithms and a collection of predefined rules to operate intelligently. It uses past data to read the patterns and then perform the intended task according to the defined rules and algorithms based on the analysis it produces.

Machine Learning is an imminent field of informatics that can be applied quite efficiently to the agricultural sector. Machine Learning is everywhere throughout the entire growing and harvesting cycle

The major factors affecting crop yield are the soil type, land type and the macro-nutrients present in the soil. The purpose of this work is to categorize the soil samples according to the macro nutrients found therein and to predict the crops which can be grown in the soil.

The crop recommendation system which is developed incorporates the agricultural dataset. The Nitrogen (N), Phosphorus (P) and Potassium (K), Soil type, Soil texture, Land type, pH and Electrical conductivity of the soil are taken as input to recommend the crops.

The system for crop yield prediction, forecasting and fertilizer recommendation are all separate and distinct in the existing system. The proposed system constructs a collaborative system of crop yield prediction and forecasting, crop rotation and fertilizer recommendation.

In this project a system is developed in which Voting Based Ensemble Classifier is applied to recommend the appropriate crops

This system also proposes the required fertilizer to boost the nutrients contained in the soil and thus enhance the yield of the crop. Thus there arises a need for suggesting suitable crops and fertilizers using machine learning algorithm.

II. RELATED WORK

Mansi Shinde et al [1], Designed a system which furnish farmers, the expert advice to identify the appropriate fertilizer and crops. The farmers can use this model using smart phones based on android. This method enhances the production of the crop. This software also allows people to buy the recommended fertilizers from the shopping portal.

V. Sellam et al [2], Evaluated environmental parameters such as Cultivation Area, Annual Rainfall and Food Price Index which affects the crop yield. The crop yield is a dependent variable that depends on all of these environmental factors. From the results produced factors like Weather Conditions, Soil Parameters etc have great effect on the crop production when compared to Cultivation Area and Food Price Index.

U.K.Diwan et al [3], Developed a climate based model to make a preliminary accurate crop yield

forecast in advance. The focus of this work was on the crop yield forecast model through the use of weather parameters and crop yield history. Temperature (maximum and minimum) and relative humidity were found to play a major role among all the weather factors in all the districts.

Rushika Ghadge et al [4], presented a tool to help farmers for monitoring soil quality based on data mining techniques. The system focuses on the soil quality inspection to predict crops that are highly suitable for the soil type. This method also recommends an appropriate fertilizer for crop yield optimisation. It recommends the crops by exploring soil nutrition contained in the agricultural land and suggest the crop that produces high productivity.

P.Priya et al [5], designed a framework that focuses on the prediction of crop yields based on the dataset of the Kharif and Rabi cropping seasons. This dataset is used to create the prediction model. The results obtained from this framework would be useful for farmers to forecast the yield before they are cultivated on agriculture land.

Vaneesbeer Singh et al [6], proposed a method to determine the yield class based on the macro and micro nutrients present in the soil, by using different machine learning algorithms. After analysis this system provides suggestions on the highly appropriate crops which results with the maximum profit.

Vrushal Milan Dolas et al [7] implemented an upgraded decision tree algorithm and the soil data set was incorporated by the classifier which is used in this application. The soil is with similar behaviour is grouped as one class, so that the farmers will be aware of the type of soil and can plant the crops appropriately.

R.Sujatha et al [8], proposed the fundamental concepts for interpreting phenology and suitable planting dates for distinct genotypes. This paper describes the idea of estimating the crop yield in advance and to make the right choice of the most appropriate crop to increase the value and gain of the agricultural field.

Supriya D M et al [9], Developed a system where data mining techniques were used to predict class of the analyzed soil datasets. Thus the predicted class will indicate high profitable crop yield. Classification algorithms were designed to label the unknown samples using the information provided by a sequence of classified samples.

S.Veenadhari et al [10], developed a web application to forecast the impact of climate variables on crop production and to find the most effective weather factor on crop production for selected plants in particular areas of Madhya Pradesh. The application provided a summary of the potential impact of different weather factors which are responsible for crop yield.

III. METHODOLOGY

A. Voting Based Ensemble Classifier

An ensemble classifier is a cluster of classifiers, where each classifier's distinct decisions are integrated to classify new examples. This technique provides a better predictive output in comparison to the prediction result of a single classifier model. The basic idea is to learn a set of classifiers, letting them vote.

a) Naïve Bayes Classifier

Naïve Bayes is a probabilistic classifier, which could be used for a wide range of classification tasks. The purpose of the probabilistic classifier is to evaluate the likelihood of the features occurring in each class, with features x1 to xn and classes c1 to ck and to return the most probable class. Probability P(c | x1… xn) for each class is calculated.

b) Random Forest Classifier

Random forest algorithm is a supervised algorithm for classification. Random forest is an ensemble method used for classification and other tasks. This works by constructing a multitude of decision trees at a time of training and by generating the class mode of individual trees classes. The higher the number of trees in the forest gives the higher precision results in random forest classifier.

c) CHAID Classifier

Chi square Automatic Interaction Detection (CHAID) is a tree classification method. CHAID partitions the data into mutually exclusive subsets which best describe the categorical dependent variable. CHAID is similar to decision tree but instead of information gain it uses gain ratio.

IV. SYSTEM ARCHITECHTURE

A. Data Collection

Agriculture (crop soil) data set is collected and checked for the presence of missing values. The data set is further analysed for thorough understanding and data clarity and the dataset is loaded to the crop recommendation system.

B. Checking for Skewness

Histogram chart is plotted for the data present in the crop soil dataset, to identify if the data is skewed. By analysing the histogram plot it is inferred that the data is normally distributed.

C. Pre - processing

Very few missing values were found in the agricultural dataset. The missing values were replaced with the mean of the corresponding attribute in the dataset.

D. Voting Based Ensemble Classifier

Naive Bayes Classifier, Random Forest Classifier and CHAID Classifier automatically gets executed internally once the user input is accepted by the Ensemble Classifier module and it suggests the suitable crop. Finally the voting will be performed for each unique crop suggested by all the three classifiers of the Ensemble classifier module. The crop with the highest vote will be suggested to the end user. Then for the recommended crop suitable fertilizer will be suggested and yield will be calculated for the forecasted crop as per the request of the end user.

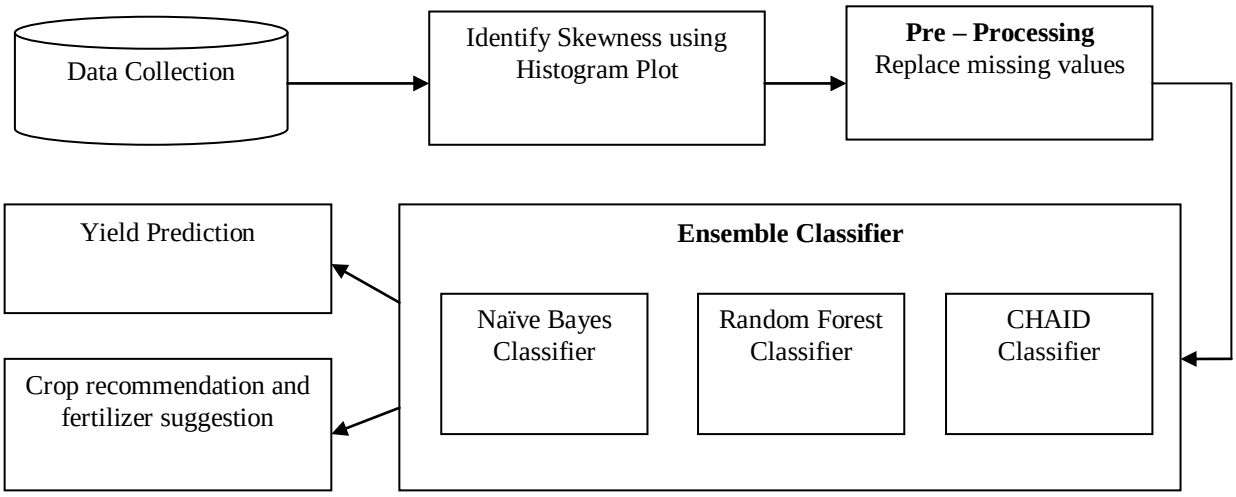

Figure 1 Proposed system architecture

E. Crop Recommendation and Fertilizer Suggestion

Naive Bayes Classifier, Random Forest Classifier and CHAID Classifier runs internally and as an output each classifier suggests crops separately for the user given input. Finally the crops with the highest vote are recommended to the user. The crops are suggested by considering the factors like nutrients that are already present in the soil (N,P,K) and the other soil parameters along with the temperature and hence the yield will be high for the forecasted crop. Additionally suggesting the fertilizer suitable for the recommended crop will tend to boost up the yield.

F. Yield Calculation

As the final step the yield calculation for the crop suggested by the Ensemble Classifier model can be determined by simply providing the area of the cultivation land in acres. As a result the yield prediction module returns the crop production in tonnes for the requested acres of land by the end user.

V. CROP ROTATION

Crop rotation module plays a vital role in crop recommendation system. Crop rotation must be considered as the important factor in order to maintain the soil fertility and to prevent from soil degradation. In this module crop cultivation season is provided as the input and the alternate crops other than the predicted crop that can also be grown in that season is provided as output by this module. Distance metric is used for the given NPK values in order to provide top three alternate crop suggestions.

VI. EXPERIMENTAL RESULT

A. Response Time Metric

Response time is used to calculate the processing time of the Voting based ensemble classifier and providing the crop suggestion to the user. Stopwatch is used to determine the elapsed time. It is used to calculate the execution time of the complete voting based ensemble classifier.

B. Accuracy Metric

Accuracy is used to measure the degree to which our forecast result is precise. Accuracy is computed by the formula,

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made}$$

TABLE I

Performance Evaluation Table				
Nitrogen	**Phosphorus**	**Potassium**	**Crop Name**	**Accuracy**
86	6.30	210	Sunflower	92.02 %
82	4.78	303	Onion	93.54 %
79	0.10	317	Guava	89.06 %
60	5.60	155	Sugarcane	93.89 %
76	3.70	456	Grapes	91.31 %
93	4.50	204	Banana	94.38 %
71	4.30	235	Cholam	94.52 %
42	4.90	240	Greengram	91.17 %
79	4.00	322	Mango	92.69 %
65	4.20	190	Paddy	88.78 %
71	6.30	225	Gingelly	93.07 %
79	4.00	266	Ragi	92.33 %
75	6.80	227	Cotton	94.90 %
76	4.00	357	Bajra	92.16 %
78	3.60	345	Tomato	93.14 %
58	4.20	154	Groundnut	90.16 %
80	6.78	226	Chilly	93.79 %
73	5.60	140	Blackgram	92.02 %
81	6.30	224	Redgram	87.32 %

VII. CONCLUSION

By classifying the soil based on the soil type, land type and macro nutrients Nitrogen (N), Phosphorus (P) and Potassium (K) present in the soil, along with Temperature, pH and Electrical Conductivity of the soil the highly appropriate crop along with the suitable fertilizer to enrich the soil and boost up the productivity is suggested to the agricultural stakeholder. The yield prediction is also provided for the crop requested by the farmer. Alternate crops that can be grown for that particular cropping season and also suitable for the NPK ratio contained in the soil is taken into account for providing suggestion to the farmer. Alternate crop rotation module provides the additional three different crops that can be grown for that soil conditions. The proposed crop recommendation system provides 92% of accuracy.

REFERENCES

[1] Mansi Shinde, Kimaya Ekbote, Sonali Ghorpade, "*Crop Recommendation and Fertilizer Purchase System*", IJCSIT International Journal of Computer Science and Information Technologies, Volume 7, Issue 2, 2016.

[2] V. Sellam, E. Poovammal, "*Prediction of Crop Yield using Regression Analysis*", IJST Indian Journal of Science and Technology, Volume 9, Issue 38, October 2016.

[3] U.K. Diwan, H.V. Puranik, G.K. Das, J.L. Chaudhary, "*Yield Prediction of Wheat at Pre-Harvest Stage Using Regression Based Statistical Model for 8 District of Chhattisgarh, India"*, IJCMAS International Journal of Current Microbiology and Applied Sciences, Volume 7, Issue 1,2018.

[4] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya.RL, "*Prediction of Crop Yield using Machine Learning"*, IRJET International Research Journal of Engineering and Technology, Volume 5, Issue 2, February 2018.

[5] P.Priya, U.Muthaiah, M.Balamurugan, "*Predicting yield of the crop using machine learning algorithm*", IJESRT International Journal of Engineering Sciences & Research Technology, Volume 7, Issue 4, April 2018.

[6] Vaneesbeer Singh, Vinod Sharma, Abid Sarwar, "*Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach*", IJARCS International Journal of Advanced Research in Computer Science, Vol. 8, June 2017.

[7] Vrushal Milan Dolas, Prof. Uday Joshi, "*A Novel Approach for Classification of Soil and Crop Prediction*", IJCSMC International Journal of Computer Science and Mobile Computing, Volume 7, Issue 3, March 2018.

[8] R.Sujatha, Dr.P.Isakki, "*A Study on Crop Yield Forecasting Using Classification Techniques*",2016 IEEE

[9] Supriya D M, "*Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach*", IJIRCCE International Journal of Innovative Research in Computer and Communication Engineering, Volume 5, Issue 5, May 2017.

[10] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh , "*Machine learning approach for forecasting crop yield based on climatic parameters*", International Conference on Computer Communication and Informatics (ICCCI - 2014).