International Research Journal of Computer Science (IRJCS)
Issue 08, Volume 09 (August 2022)
ISSN: 2393-9842
https://www.irjcs.com/archives

AGRICULTURE ANALYSIS USING ENSEMBLE LEARNING TECHNIQUES

Chithra N
Dept.of Information Science and Engineering,
Vemana Institute of Technology, Bangalore, India
Lakshmi V
Dept.of Information Science and Engineering,
Vemana Institute of Technology, Bangalore, India
Monisha R
Dept.of Information Science and Engineering,
Vemana Institute of Technology, Bangalore, India
Muthyanjali A
Dept.of Information Science and Engineering,
Vemana Institute of Technology, Bangalore, India
Sridevi K N
Assistance professor
Dept. of Information Science and Engineering,
Vemana Institute of Technology, Bangalore, India

Publication History
Manuscript Reference No: IRJCS/RS/Vol.09/Issue08/SPAUCS10081
Research Article | Open Access
Peer-review: Double-blind Peer-reviewed
Article ID: IRJCS/RS/Vol.09/Issue08/SPAUCS10081
Volume 2022 | Article ID SPAUCS10081 http://www.irjcs.com/volumes/Vol09/iss-08/02.SPAUCS10081.pdf
Received: 25, July 2022 | Accepted: 04, August 2022 | Published: 12, August 2022
doi: https://doi.org/10.26562/irjcs.2022.v0908.002
Citation: Chithra,Lakshmi,Monisha,Muthyanjali,Sridevi(2022).Agriculture Analysis using Ensemble Learning Techniques. International Research Journal of Computer Science (IRJCS), Volume IX, Issue VIII, 2022, Pgs162-166
BibTeX key:
Academic Editor-Chief: Dr.A.Arul Lawrence Selvakumar

Abstract: Agriculture is an important occupation in India. Agriculture contributes to approximately 17% of India's Gross Domestic Product and it's still the most popular occupation amongst 70% of India's population. Due to the various climatic factors the agriculture productivities in India are continuously decreasing over a decade. Farmers are still struggling when it comes to picking the right decision for better crop production. Thus, to accelerate the crop production, different technologies are being proposed worldwide. Data mining and machine learning in agriculture is a new approach for predicting crop yield production. The reasons for continuously decrease in production over a decade was studied mostly using regression analysis. The emerging technologies can change the situation of farmers and help in decision making. For this we are using Machine learning algorithms called k-means algorithm, AdaBoost and Random-forest algorithm. So, these algorithms take in different parameters to give us best recommendations which not only leads to better yields but also minimum use of resources and capital. In this model, the algorithm classifies the dataset into 4 classes namely, 0-poor yield, 1- average yield,2-good yield, 3-very good yield. This model predicts the expected accurate yield by considering the parameters like, state name, crop, production of crop, and season.
Keywords: Data mining, Machine learning, K-means algorithm, Ada Boost algorithm, Random-forest algorithm.

I. INTRODUCTION

The Indian economy is reported to be mostly based on agriculture. In India, this profession supports close to 70% of the population. Every day, the population is growing quickly. Food thus becomes more in demand. The agriculture sector faces numerous issues. The majority of these issues have quite unpredictable natures. These variables ultimately cause the depth of the farmers' troubles to expand. This is a significant factor in suicides. Farmers, agricultural scientists, and researchers are developing a variety of ways and strategies to forecast greater yield output with the least amount of resources and capital, so that it simultaneously generates a lot of profit for the farmers. The concept of precision agriculture is not widely utilised in India, and farmers continue to employ the conventional methods of cultivating crops. This increases the danger of failure and doesn't produce a greater yield. The art of cultivating the soil, growing crops, and raising livestock is known as agriculture.

It involves preparing plant and animal items for human consumption and distributing them to marketplaces. Most of the food and textiles in the world are produced by agriculture. Agriculture generates. Wood for building and paper goods. In the modern world, data mining is a relatively recent method of estimating crop production in the field of agriculture. Farmers can make informed judgments thanks to data mining tools' effective analysis. They use the knowledge in these decisions to save costs and increase productivity. The following are steps in the data mining process: maintaining the data in multidimensional databases after it has been extracted, transformed, and loaded. Data mining gives analysts employing application software access to data. Graphs make the analysed data simple to understand. Emerging technologies are developing a variety of approaches and procedures to improve the position of farmers, who still struggle to choose the best crop for a higher yield. As greater accuracy may be attained, several algorithms including K means clustering, the Ada Boost algorithm, and the random forest algorithm are applied.

SCOPE

Agriculture plays an important role in increasing the country economic. The product from agriculture based industries such as jute, cloth, tinned food, etc. contributed by agriculture sector produces the huge income. The emerging technologies are coming up with various methods and technologies towards changing the situation of farmers and help in the decision making in terms of agricultural field in a better way.

CLASSIFICATION TECHNOLOGIES AND CLUSTERING METHODOLOGY

CLASSIFICATION TECHNOLOGIES:

Random forest is an algorithm that supports both the classification and regression problems.
Step1: We import library called Random Forest Classifier.
Step2: By creating a variable rfc, we test the data based on the training data i.e., x_train.
Step3: This process will identify the points and then divides the dataset into several decision tress.
Step4: By taking the average of the decision trees, it calculates the accuracy score and confusion matrix results.
Step5: This continues till the given data is processed and the predictions give the accuracy as an output.

AdaBoost algorithm, short for Adaptive Boosting. The AdaBoost can be given as follows:
Step1: Firstly, we import the Ada boost classifier library.
Step2: We fit the x_train and y_train subsets to the model then we predict the test results.
Step3: This process will continue until the complete training data fits without any error.
Step4: We calculate the accuracy score, predicted classification report of the model and the confusion matrix.
Step 5: These predictions give us the results of the accuracy of the model.

CLUSTERING METHODOLOGY

K-Means Clustering is an unsupervised learning algorithm. This groups the unlabelled dataset into labeled dataset by forming different clusters. The working of the k-means algorithm is explained in the below steps:

Step1: Select the number K randomly which is predefined to decide the number of clusters.
Step2: Assign each data point to their closest centroid. Doing this will form the predefined clusters that is K clusters.
Step3: Calculate the variance. Now place a new centroid of each cluster.
Step4: Repeat the above steps, which mean reassign each data point to the new closest centroid of each cluster.
Step5: go to step3 in case any reassignment occurs else the model is ready.

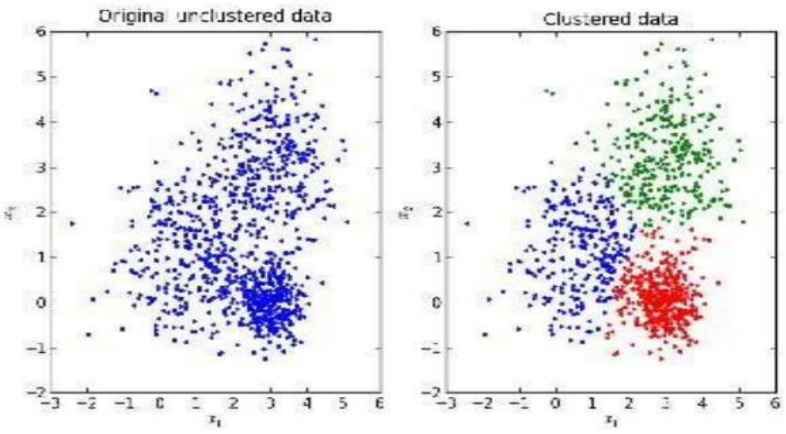

Fig: Unclustered data and clustered data

IV. PROPOSED METHODOLOGY

In this model, the agricultural dataset undergoes few processes such as clustering and data pre-processing where the erroneous and null values are removed. After successful completion of this process the data is further subjected to splitting of the data where the dataset is split into two: training data and test data.

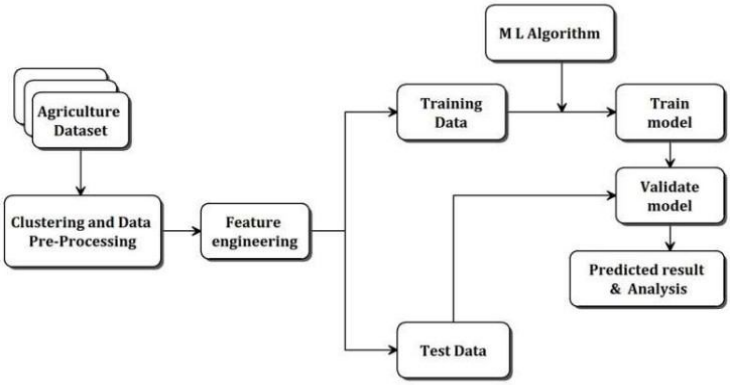

Fig: System Architecture

The approach of the proposed system is it provides the efficient and accurate result of the crop analysis by using the popular and modern technologies in data mining and machine learning like Random Forest algorithm and AdaBoost algorithm. The parameters include state name, season name and the crop. In this model, the algorithm classifies the data set into 5 classes namely 0 - poor yield, 1 - average yield, 2 - good yield, 3 - very good yield. Implementation of system means the process of converting a new or revised system design into operational one. The implementation requires the following: Data acquisition and data pre-processing, Feature selection and data preparation, Model construction and model training, Model validation and result analysis.

The data flow model of the proposed system is given below:

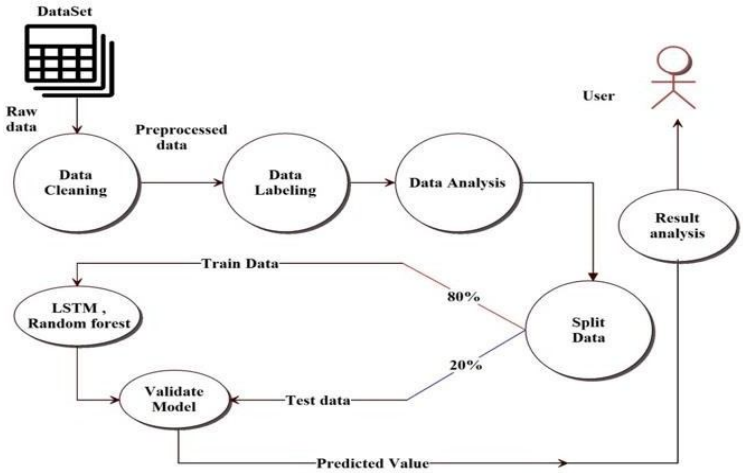

Fig: Data flow diagram of agriculture analysis

The sequence diagram of the design is given below:

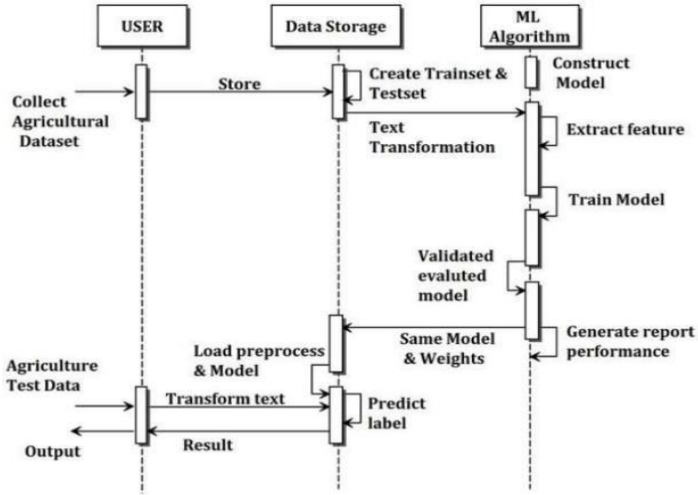

Fig: sequence diagram of agriculture analysis

V. RESULTS

After extracting and loading the dataset, the data is pre-processed wherein the null values, duplicated values and outliers' values are eradicated from the dataset. The seasons will be assigned with numerical constraints and the same will be reflected in the dataset output and depending on each season's crop count the pie chart is plotted.

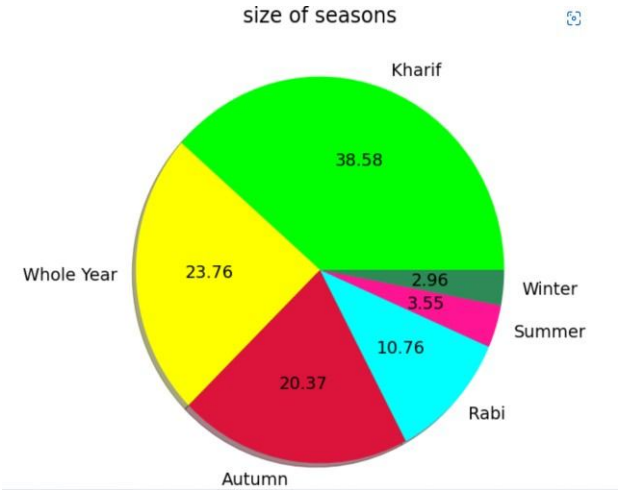

Fig: pie chart of seasons

The below figure plots the horizontal bar graph for the comparison of the crop size. This depends on the value count of the crops.

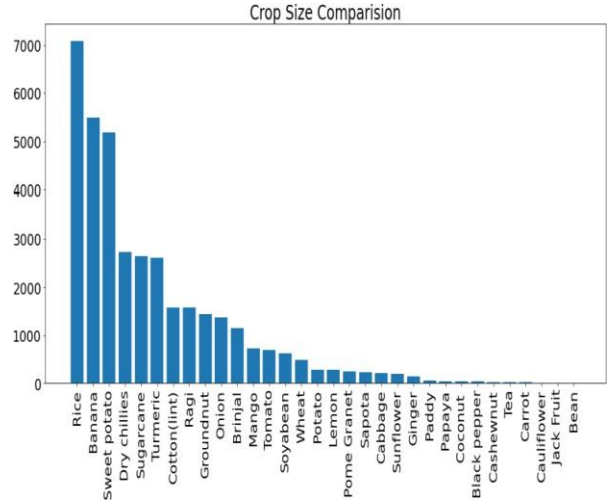

Fig: Bar chart of crops

The agricultural dataset will be assigned with the following labels: 0-Poor, 1-Average, 2-Good and 3- Very good. For each assigned label the count of the crops is computed.

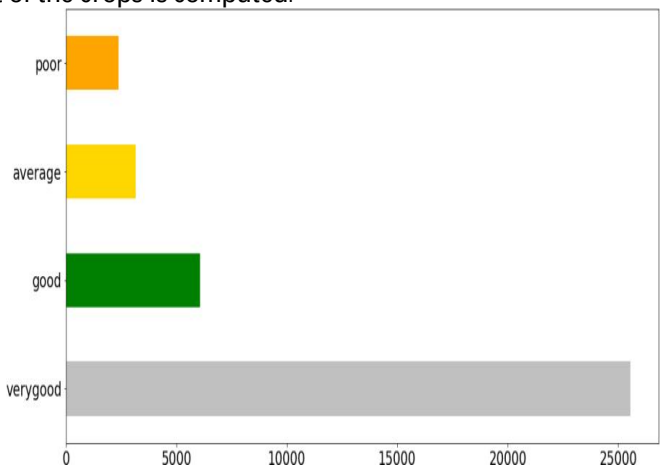

Fig: Labelled class graph for entire agricultural dataset records

Based upon the comparison of the result analysis of both the models. The below figure resembles the model accuracy plot of Random Forest and AdaBoost classifier model.

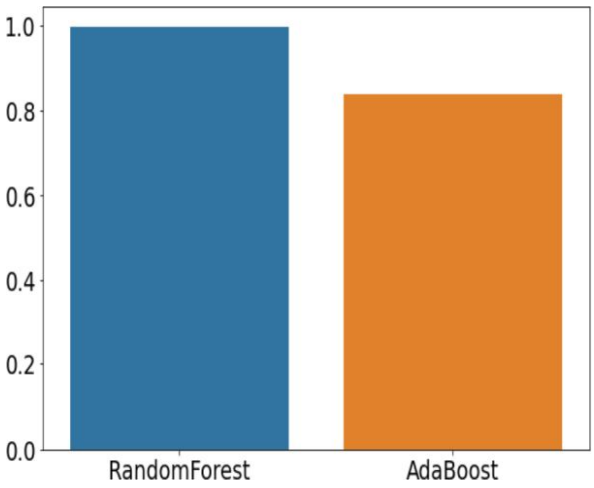

Fig: Model Accuracy plot

VI. CONCLUSION

This project was proposed to overcome the problems faced by farmers in picking up the right crop for better yield. The proposed system is based on Agricultural analysis which results in providing the highest accuracy in terms of agriculture for predicting the crop yield using various Machine learning algorithms. The system can be trained on a bigger dataset comprising of wide range of previously recorded data. The accuracy of the models and their relevance can be improved significantly with availability of agricultural datasets with very large number of parameters. Random Forest algorithm and Ada boost algorithm is more suitable for numerical processing especially in agriculture Analysis classification. We conclude that the experimental result what we will be getting from developed model provides more accurate results. This model can be enhanced in future by creating a web/mobile application with an user friendly interface, this will make the technology more accessible to the farmers and subsequently leads to more profitable yields. To get rid of these situations' farmers, agricultural scientists and researchers are trying various ways to obtain better crop yield. Upon extensive study and research, recommendations for improvement and enhancement of the agricultural analysis program this system can be enhanced further with various techniques.

REFERENCES

1. Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2(1).
2. Impact of Machine Learning Techniques in Precision Agriculture Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper.
3. Agriculture Analysis Using Data Mining And Machine Learning Techniques SABEERA D, Dr. D ARADHANA.
4. Crop Yield Prediction using Machine Learning Algorithm D Jayanarayana Reddy, Dr M Rudra Kumar.
5. Agriculture analysis using Data Mining and Machine Learning techniques: Vanitha CN, Archana N, Sowmya R.
6. Predictive Agriculture Using Data Analysis and Machine Learning: Aastha Doshi, Anuradha Chopade, Sonali Bodekar.

