Comparative Analysis Of Naive Bayes, Decision Tree And XG-Boost For Crop Recommendation System

Ayushi Butani Dinesh Lakshmanan Nidhi Mehta Anil Hingmire Computer Engineering Dept. Computer Engineering Dept. Computer Engineering Dept. ComputerEngineering Dept. Vidyavardhini's College Of Vidyavardhini's College Of Vidyavardhini's College Of Vidyavardhini's College Of Engineering And Technology Engineering And Technology Engineering And Technology Engineering And Technology

Abstract

The need for increasing crop production is a significant challenge for farmers. This is due to a combination of factors, including climate change, soil degradation, and limited access to resources and technology. All of these factors can lead to decreased yields and, in turn, reduced income for farmers. To meet this challenge, farmers must develop new methods of production that are resilient, cost-effective, and sustainable. This will require a holistic approach, involving technological and agricultural advances, as well as improved access to resources, investments, and training. Crop cultivation anywhere in the world depends on the climate, seasons and various factors like soil properties. However, the process of increasing the production of crops depends on various factors, including mainly on temperature. Our proposed method utilises Machine Learning Procedures with the end goal that it proposes the appropriate crops dependent on the Temperature, humidity ,potassium, nitrogen, rainfall ph,phosphorus. This kind of framework, subsequently helps to reduce the losses and misfortunes looked by the farmers brought about by establishing the probability of bad harvests and furthermore it gives the information on the OCcasional characterization of yields what harvest is reasonable for which season. There are several machine learning algorithms available are used in this system., including KNN, Decision Tree, and To recommend the crop, Random Forest, Naive Bayes and XG-Boost are used.

Keywords—Naive Bayes, Decision Tree, XG-Boost, Machine Learning, Crop Recommendation System.

I INTRODUCTION

There is no doubt that the major source of income in rural India is agriculture and its allied industries. Moreover, agriculture contributes a lot to the country (GDP). The country is fortunate to have such a large agricultural sector. This may be one of the possible reasons for the higher suicide rate among marginal farmers in India. For farmers, this study offers a practical and easy-to-use yield forecasting system.

The proposed system provides connectivity to farmers through mobile or online applications. User location can be determined using GPS. The user enters a location and a soil type.

Machine learning algorithms help select the most profitable crop lists or forecasts. Some machine learning techniques are used to predict crop yields, including Support Vector Machines (SVM), Decision Trees (DT), XG Boost, Gaussian Naive Bayes, Random Forest (RF), and K Nearest Neighbors (KNN).

Among them, Random Forest showed the best results with 90% accuracy, while Naive Bayes showed 99%.

Agriculture in India is vital to the socio-economic structure of the country. For a country where nearly 58% of the population works in agriculture, the inability of farmers to use traditional, unscientific methods to choose the crops best suited to the soil is a serious problem. Farmers sometimes choose the wrong crops based on soil properties, planting season and location.

As a result, people commit suicide, stop farming and move to cities in search of work.

This research work proposes a mechanism to

help farmers solve this problem by choosing crops that take into account all factors, including planting season, soil and geographic location. Additionally, precision agriculture is gaining traction in emerging countries with an emphasis on site-specific crop management and integration using modern agricultural techniques.

India is a predominantly agricultural country where the economy and the daily life of the people are strongly affected by agriculture. To increase agricultural productivity, it is recommended to grow crops according to soil, weather conditions, humidity, rainfall and other factors. This not only benefits farmers, but the country as a whole, and lowers food prices.

This study demonstrates the use of machine learning techniques such as random forests and decision trees to predict which crops will perform best on a data set.

II. RELATED WORK

The most important actions to improve agriculture include the integration of technical knowledge and innovation to make agriculture more efficient and easier for farmers by using all ML techniques to predict ideal crops. This article discusses the benefits of various approaches including artificial neural networks, fuzzy networks, and data mining methods. Adding real-time datasets to all of this is another issue.

One of the first employees created a website dedicated to assessing the impact of weather conditions on agricultural productivity in designated areas of Madhya Pradesh. Select the region according to the crop area.

Based on these criteria, the top five districts with the largest agricultural area were selected. Crop selection for this study was based on the dominant crops in the selected areas. The crops selected were corn, soybeans, wheat and rice, and the yield of each was calculated for a continuous knowledge period of 20 years. For the selected crops, the models built were between 76% and 90% accurate, with an average accuracy of 82%. Another important study assesses soil quality and predicts crop yields along with appropriate fertiliser recommendations.

KNN clustering has been shown to perform better than SVM or regression. Using the latest regression techniques such as Enet, Lasso and Kernel Ridge algorithms, it can predict crop yield for a particular year.

Using stack regression improves the accuracy of the algorithm.

Using the Pandas analysis tool, filter the

Model inputs include Ph values and locations provided by the user. To predict the weather and temperature of the current location, an API is used. The system compares the results of supervised and unsupervised machine learning methods.

This is featured in Classifiers for crop yield prediction using Greedy. It has been proven that using features in decision tree classifiers can improve performance.

The proposed ensemble model has been shown to incorporate the effects of multiple models, often outperforming a single model.To predict agricultural yields, the random classification of forest sets uses several decision tree models. A ratio of 67% to 33% is used to divide the data into two groups, such as training data and test data, from which the mean and standard deviation are calculated. For the most accurate results, the method also groups related tumours has used many ML methods, and many efforts have been made in the field of agriculture. Increasing agricultural production and delivering it to consumers at the best price and with the best quality is the biggest problem in agriculture.

It has also been noted that at least 50% of farm food is wasted and never eaten. The proposed method provides suggestions for reducing wastage of agricultural products. A recent paper presents a crop yield prediction model using the KNN algorithm and clustering.

historical dataset to extract the Maharashtra dataset. Build a crop yield prediction model using a multi-layer perceptron neural network and tune biases, weights, and Adam optimizers to improve accuracy.

The proposed model predicts crop yield using an ANN with a three-layer neural network. A crop yield forecasting system is built using supervised learning methods.Associations between many historical traits have been established and have allowed the system to increase crop yields. Rainfall and temperature impact crop yields. To improve accuracy, recurrent neural network (RNN) and long-short-term memory (LSTM) algorithms are applied to this time-series data.

Using historical data, the ARMA (Autoregressive Moving Average), SARIMA (Seasonal Autoregressive Integrated Moving Average) and ARMAX (ARMA plus exogenous variables) methods are used to predict temperature and precipitation. In a crop yield forecasting system using fuzzy logic, the best model in the cluster is used.

Below is the table of comparative study of the IEEE papers which we have referred to regarding this recommendation system.

Sr. no	Authors	Focus of the paper	Key points in the coverage	Techniques used	Parameters/ Dataset	Research Gaps/ Limitations
1.	S. Pudumular J. Nisha , C. Kavya , R. H Rajashree	To develop a recommendation system through KNN and RF	Precision agriculture, recommendation system, ensemble model, knn	KNN , Random forest , Naive Bayes.	Depth , texture , ph,soil, permeability water holding	Less parameters and small data set
2.	Shilpa Pande, Dr.Prem , Anmol, B.R. Aishwarya, Karuna Rohilla, Kumar Shaurya	to predict crop yield for a user selected crop through RF and various machine learning alglorithms	crop yield prediction , ANN, KNN, Multivariate Linear Regression, Fertilizer	SVM, RF, KNN, MLR, ANN	soil, average rainfall, temperature	less parameters. analysis of previous datasets is not done.
3.	Nidhi Kulkarni, Dr. GN Srinivasan, Dr. BM Sagar, Dr. KN Cauvery	To classify the input soil dataset into recommendable crop type Kharif and Rabi.	ensemble, majority-voting Naive Bayes soil,crop recommendation	Ensemble learning , Linear SVM, naive bayes	average rainfall, surface temperature , sowing season, NPK content of the soil	small dataset
4.	Miftahul Jannat Mokarrama Mohamma d Shamsul Arefin	recommendation system named as RSF for farmers, which can recommend farmers most suitable crops to produce in different areas	Crop recommendation, recommender system, similarity calculation.	social pertinent trust walker (SPTW),C F-based inference method	Location Detection Module,Data Analysis , Storage Module,Crop Production Rate Database	less parameters and small data set

5.	Priyadharshini A,Aayush Kumar,Swapneel Chakraborty	To develop a recommendation system using Machine learning	Profit analysis , Crop recommender, Crop Sustainability predictor .	CHAID, K-Nearest Neighbor, Naïve Bayes, and Random tree a	Data analysis,Data Preprocessing,soil characteristics,environmental characteristics	small dataset
6.	Samarth Arlikar, K.A. Shinde, Shubham Kothavade,	To predict, crop, yield based on climatic parameters,previous yield and soil attributes.	Machine learning and KNN	KNN	Contains columns such as district name, season name, crop name also predicts average cultivation and production obtained.	They can also consider economic aspect of farmers to recommend farmers most profitable crop
7.	Avinash kumar , Chittaranjan Pradhan	Crop recommendation system and pest control technique.	Svm , pest control techniques , Dt	DT , Logistic regression ,SVM .	Types of soil, Average temperature , ph , Rainfall	Developing this model with soil attributes and larger dataset.
8.	Zeal Doshi ,Subham Nadkarni ,Nipa Shah	Crop recommendation system using ML algorithms.	ML algorithms , soil parameters , Neural network	KNN , DT , RF ,Neural network	Rainfall , Temperature , Location , precipitation , ph , thickness	This model can be further improved to predict crop rotations to maximize the yield.

9.	J Madhuri M Indiramma	To develop a crop recommendation system using ANN.	Climate, Soil characters, ANN	ANN	Minimum Temperature ,Maximum Temperature, Sunshine hours,Soil texture,Slope ,Depth,Mean temperature, Soil drainage	It only works better on larger Dataset.
10.	Avinash Devare , Rohit kumar , pooja Shinde	To maximize crop yield using ML algorithm.	Precision agriculture , ANN , RF , DT	SVM , ANN , RF , Naive Bayes	Ph , soil nutrients , rainfall , temperature , location	The dataset can be large with many attributes

Fig. Comparative study of IEEE Papers

III. METHODOLOGY

Despite numerous new solutions being put out, there are still unresolved issues with developing a user-friendly website for crop advice. The proposed remedy seeks to address these drawbacks by creating a website that performs comparative analysis on the input crop and, based on the accuracy, recommends the crop. There are constraints. a user-friendly application that takes into account the factors that directly influence farming, such as rainfall, temperature, soil type, etc. The major goal is to increase the range of crops that can be cultivated throughout the season. By minimising the challenges farmers confront when selecting a crop and maximising productivity, the suggested system would ultimately assist to lower suicide rates.

With the supplied data sets, the proposed model forecasts crop yield. By raising yields and maximising resource use, integrating agriculture and ML will help the agriculture sector go further. The most important factors in predicting present performance are historical data. Historical data is gathered from a number of trustworthy sources, including data.gov.in and kaggle.com. Some databases with state and district details include the soil type as an attribute. The retrieved soil type column is combined with the primary data set. In a similar manner, the major data sets for the particular location are supplemented with temperature and average rainfall from a different dataset. The data sets have been prepared and cleansed. New mean values are used to replace the null values.

Before running the algorithms, the categorical attributes are translated into labels.

Fig 1:- System Architecture

Figure 1 is the system architecture of the proposed algorithm. Algorithms are selected on the basis of the accuracy of each algorithm. As in this model the data was passed through 3 various algorithms. And the algorithm having highest accuracy was selected.

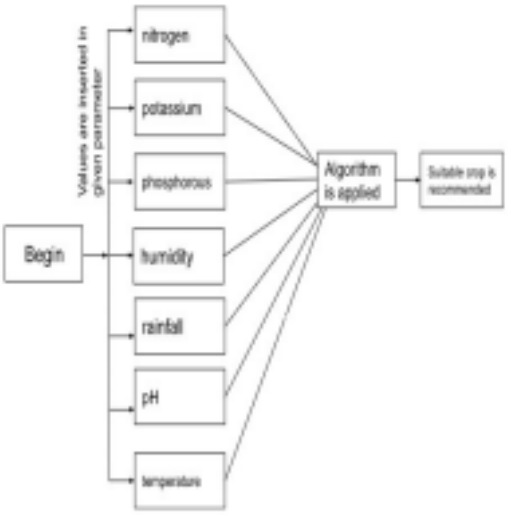


Fig 2:- Block Diagram

Figure 2 is the system architecture of the process of crop recommendation. in which the farmer has to input various soil parameters given, and once the values are uploaded, the proposed algorithm is applied on the parameters. On the basis of the results a suitable crop is recommended to the farmers.

The Algorithms which we have used for the implementation of this crop recommendation system are as follows:

i) Naïve Bayes

Classification problems are solved using the Naive Bayes supervised learning technique, which is based on the Bayes theorem.

It primarily uses a huge training set for text categorization. The Naive Bayes Classifier, one of the simplest and most effective classification algorithms, facilitates the creation of quick machine learning models that are capable of making precise predictions. It provides predictions based on the likelihood that an object will occur because it is a probabilistic classifier. A few applications for Naive Bayes algorithms include spam filtration, sentiment analysis, and article classification.

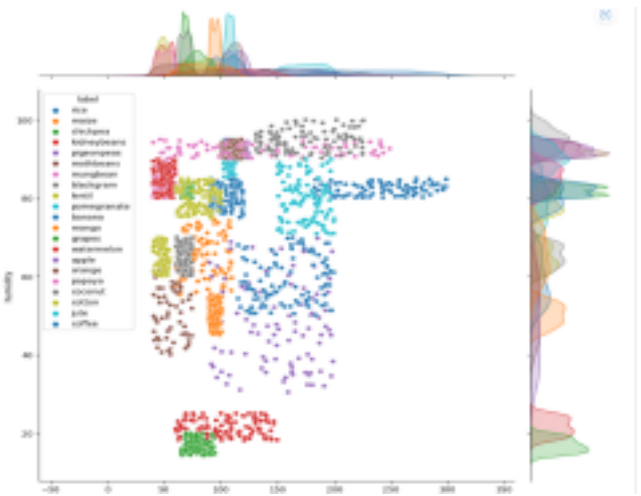
ii) Decision Tree

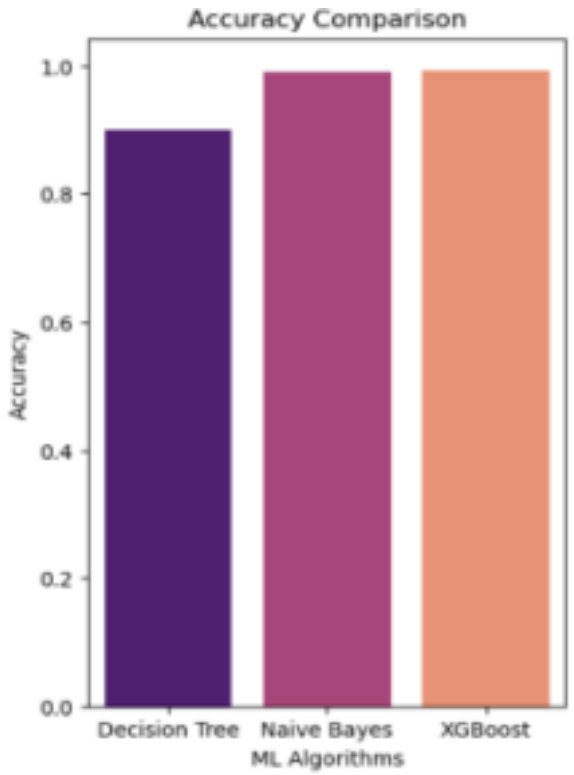
Decision Tree is a member of the supervised learning algorithm family. Unlike other supervised learning techniques, the decision tree technique can handle classification and regression problems. A Decision Tree is used to create a training model that may be used to predict the class or value of the target variable by teaching simple choice rules generated from prior data. In decision trees, we start at the root when predicting the class label of a record. We compare the values of the attribute on the record with those of the root attribute. We jump to the node based on the comparison and follow the branch connected to that value.

vi) XGBoost

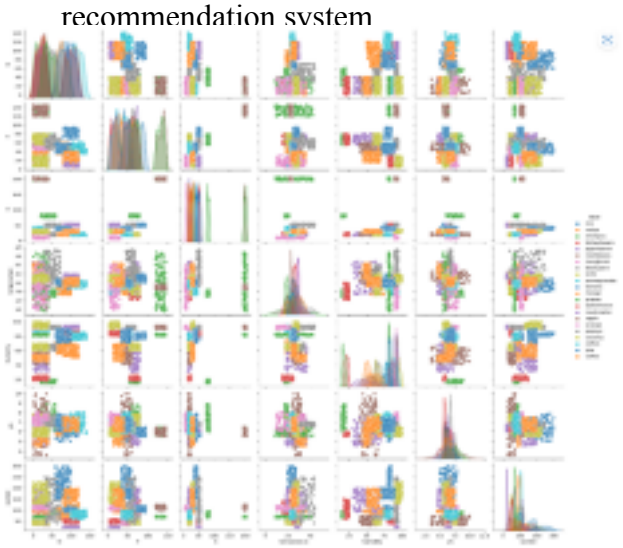
A distributed gradient boosting toolkit called XGBoost has been enhanced for speed and scalability in the training of machine learning models. Using the ensemble learning method, a number of weak models' predictions are pooled to produce a stronger forecast. Because it can handle massive datasets and excel at the forefront of many machine learning tasks, including classification and regression, Extreme Gradient Boosting, or XGBoost, is one of the most well-known and widely used machine learning algorithms.

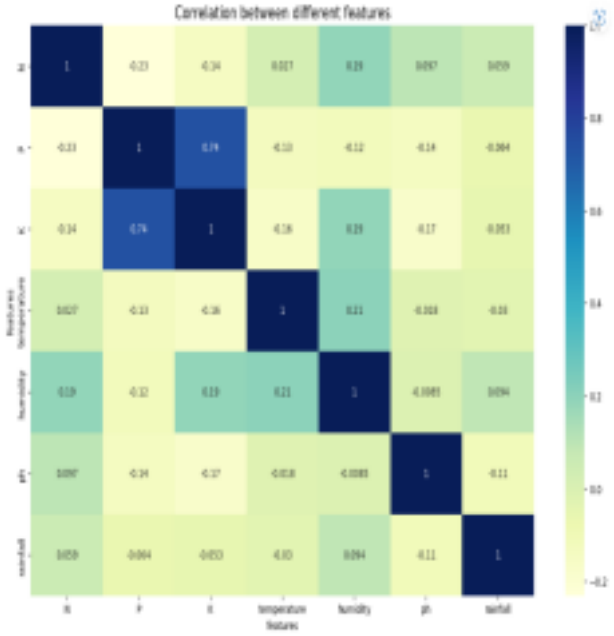
IV. RESULT ANALYSIS





Above is the Comparative analysis of the accuracy of the algorithms we have performed in this recommendation system





V. CONCLUSION

The proposed method helps farmers choose the best crops by providing information that typical farmers do not monitor, thus reducing the likelihood of crop failure and increasing yields. It also protects them from losses. Millions of farmers across the country will in the future be able to access advice on planting crops through a combination of web interface and mobile app.

Future work will focus on updating the dataset from time to time to produce accurate predictions, and the process could be automated. Another feature that needs to be should be analysed.

Thus, the work proposed makes it possible to help farmers make precise choices of crops to plant. This creates an exponential increase in crop productivity which, in turn, boosts the country's economy.

implemented is providing the right type of fertiliser for a given crop and location.

To carry out this thorough study of available fertilisers and their relationship to soil and climate, must be done. Available statistical data

VI. REFERENCES

[1] Umamaheswari S, Sreeram S, Kritika N, Prasanth DJ, “BIoT: Blockchain-based IoT for Agriculture”, 11th International Conference on Advanced Computing (ICoAC), 2019 Dec 18 (pp. 324-327). IEEE.

[2] Jain A. “Analysis of growth and instability in the area, production, yield, and price of rice in India”, Journal of Social Change and Development, 2018;2:46-66

[3] Manjula E, Djodiltachoumy S, “A model for prediction of crop yield” International Journal of Computational Intelligence and Informatics, 2017 Mar;6(4):2349-6363.

[4] Sagar BM, Cauvery NK., “Agriculture Data Analytics in Crop Yield Estimation: A Critical Review”, Indonesian Journal of Electrical Engineering and Computer Science, 2018 Dec;12(3):1087-93.

[5] Wolfert S, Ge L, Verdouw C, Bogaardt MJ, “Big data in smart farming– a review. Agricultural Systems”, 2017 May 1;153:69-80.

[6] Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Godfray HC, Herrero M, Howitt RE, Janssen S, Keating BA, “Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural systems", 2017 July

[7] Johnson LK, Bloom JD, Dunning RD, Gunter CC, Boyette MD, Creamer NG, “Farmer harvest decisions and vegetable loss in primary production. Agricultural Systems”, 2019 Nov 1;176:102672.

[8] Kumar R, Singh MP, Kumar P, Singh JP, “Crop Selection Method to maximise crop yield rate using a machine learning technique”, (ICSTM), 2015 May 6 (pp. 138-145). IEEE.

[9] Sriram Rakshith.K, Dr.Deepak.G, Rajesh M, Sudarshan K S, Vasanth S, Harish Kumar N, “A Survey on Crop Prediction using Machine Learning Approach”, In International Journal for Research in Applied Science & Engineering Technology (IJRASET), April 2019, pp(3231- 3234)