IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Farmer assistant and crop recommendation system

Vivek Rajput[1], Ajinkya Deshmukh[2], Utkarsh Patil[3], Harshal Ushire[4], Prof. Pramila M. Chawan[5]
[1,2,3,4] B. Tech Student, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[5]Associate Professor, Dept. of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

Abstract: In India, agriculture accounted for 18% of the GDP with about 42% of the workforce. The people involved in agriculture are high in percentage but their contribution to the economy is different. In today's world, people are getting more and more digitized. As a result, most people have smartphones and internet access. In rural areas people including farmers, shopkeepers and others are using these technologies to make their life easier. With this technological evolution, we have developed the farmer assistant to guide farmers such that they can improve their productivity and gain higher profits. We have developed a model that will suggest the appropriate crop to the farmer according to the conditions. The model will take into account all the soil quality parameters such as soil fertility which includes Nitrogen, potassium, Phosphorous value in soil. Also, the model will give the result according to the weather condition such as temperature, humidity, rainfall. This will give a huge benefit to the farmer. Also, there is a selling platform for the farmer where farmer can upload the advertisement on how much crop product he has and how much rate he expects. The buyer will then see the advertisement and he can contact the farmer for the product to finalize the deal.

***Key Words*:** Technology, Profit, Farmer, Cost, Machine Learning, Model, Ensemble Learning, Bagging, Boosting, Voting, Stacking

1. INTRODUCTION

In today's world, huge data is present in various sections, including agriculture. In 2022 the Indian government announced the scheme; every licensed shop is getting the status of a governmental shop. There will be many facilities like soil testing and guidance. But they are limited in number and everyone can't take the benefits from them. So, we came up with the idea of a farm assistant website. By using modern technology like machine learning, we have the ability to provide those services online. Machine learning (ML) helps us predict results from huge amounts of present data. Figure 1 represents the machine learning algorithm working, it pre-processes data and develops the ML model, whose output is to divide the soil according to similarity and predict which crop will be suitable in such type of soil**.** We have used machine learning models to predict the suitable crop for the soil. We have done the comparative analysis of the models to get the best model which will give accurate results.

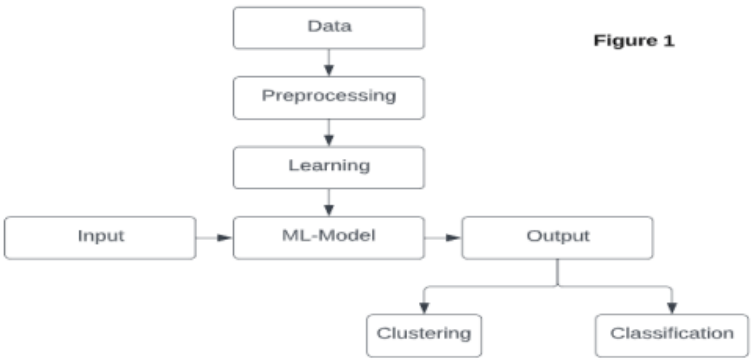


Figure 1

Fig -1: ML working

We have developed a Farming Assistance Web Service application which consists of a login, registration and E-Authentication System with OTP. There will be no password-based vulnerabilities in the system. It helps farmers in ensuring higher profitability by getting the exact crop that will be suitable for the soil and weather conditions. Also, there is a selling platform which will help in direct farmer to buyer communication. Farmer can publish the advertisement regarding how much quantity of crop product he has and what is the expected rate of the product. Buyer will see the advertisements and can contact farmer to tell what is the requirement and what should be the rate to finalize the deal. This will help farmer to stay active about the market rates to get maximum profit while selling the product.

The motivation behind this project was to create a useful model which will guide farmer on selection of crops that will be beneficial for him. This will help farmer in ensuring high profit and also this will reduce the increasing suicide rate of farmers.
Farmers are the most important component of our society and to protect and help him we have done this project.

2.Literature Review:

There is a huge amount of data available on farming, and to make a decision from that data we have different machine-learning techniques. For example, Decision Tree, Gaussian Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Random Forest.

The system recommends the crop based on soil fertility. The system is based on the XGBoost algorithm. XGBoost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides a parallel tree boosting and is the leading machine-learning library for regression, classification, and ranking problems.

There are lots of systems based on other methods we have studied in the literature survey. Vrushali Bhuyar uses classification techniques to predict soil fertility by the decision tree method, naive Bayes, and random forest. They found out the decision tree is the best technique for classification. D Ramesh uses Multiple Linear Regression for predicting rice yield. Sandesh Ramesh, Department of Information Science and Engineering Nitte Meenakshi Institute of Technology, uses regression and neural networks for crop yield prediction.

S.S.Bhaskar made a study of soil classification of JRip, naïve bayes, and J48. They found J48 to be the best method. They also used regression techniques like linear regression and least square Median. They found that least median squares regression produces better results for prediction than the classical linear regression technique. Jay Gholap uses the J48 algorithm for predicting soil fertility class. Also, for performance tuning of the J48 algorithm, he uses attribute selection and boosting techniques. Suman clusters the data using the K-means Clustering on the soil dataset then linear regression is applied to classify the clusters.

Fatih Bal and Faith Kayaalp's “Review, of machine learning and deep learning models in agriculture” the identification of the plant species has been realised with ML and DL Methods depending on classification algorithms in smart agriculture applications based on artificial intelligence 126 citrus images obtained in different sizes and under various lighting conditions were trained with ML algorithms and a study was carried out to determine the green fruit.

“Machine Learning Applications for Precision Agriculture: A Comprehensive Review” by Abhinav Sharma, Arpit Jain, Prateek Gupta, (Student Member, IEEE), and Vinay Chowdary develop the manual spraying method for pesticides that leads to improper usage of resources and harms the environment. AI and IoT-enabled precision agriculture remove the randomness and assist new-age farmers to optimise every step of the farming process.

Gaitán provided a systematic study of the impact of extreme weather events, such as hail events, cold waves, and heat waves, and their impact on agricultural practices. The author reported floods, droughts, frost, hail, heatwaves, and pest outbreaks are impacted by climatic conditions.

Acar employed an extreme learning machine (ELM) based regression model for the prediction of soil surface humidity. The author selected two terrains having areas 4 KM2 and 16 KM2 located on the Dicle university campus for experimental analysis. The realtime field data was extracted using polarimetric Radarsat-2 data, which was pre-processed using the SNAP toolbox and features were added with the help of local measurements by separating the field into square grids. Once the pre-processing and feature extraction is done the data is passed to ELM based regression model to predict the soil surface humidity. The algorithm was tested with 5 different kernel functions and the prediction was validated using a leave-one-out cross-validation technique.

Y. Mekonnen developed a power-efficient WSN using an Arduino microcontroller and ZigBee module to monitor and control essential parameters that affect crop growth such as soil and weather conditions in Florida, USA. Pise and Upadhye explored Naive Bayes and SVM ML techniques for grading harvested mangoes based on their colour, size, features, quality, and maturity. Grading fruits increases the profit of the agriculture and food industries. A mango image dataset comprising three different colours red, green, and yellow is created and used for training and testing the ML algorithm. The proposed approach presents limited scope as it can detect defects in a particular surface area which can be overcome by creating a dataset of rotational view images.

Gradient Boosting is a boosting algorithm, in which each predictor corrects its predecessor’s error. XGBoost is an implementation of Gradient Boosted decision trees. In this algorithm, decision trees are created in sequential form. Weights play an essential role in XGBoost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. The significance of variables predicted wrong by the tree is increased and these variables are then fed to the second decision tree. These individual classifiers/predictors then ensemble to give a strong and more precise model. It can work on regression, classification, ranking, and user-defined prediction problems.

3.Proposed System:

3.1 Problem statement

"To develop a Farming Assistance Website that will help farmers in ensuring high profitability by predicting suitable crops according to soil fertility ."

3.2 Dataset Description

The dataset we used here is crop dataset Crop Recommendation Dataset from kaggle. This dataset was built by augmenting datasets of rainfall, climate and fertiliser data available for India. The dataset contains the 2200 entries of rows. The columns are:

- N - ratio of Nitrogen content in soil
- P - ratio of Phosphorus content in soil
- K - ratio of Potassium content in soil
- temperature - temperature in degree Celsius
- humidity - relative humidity in %
- ph - ph value of the soil
- rainfall - rainfall in mm
- label- crop

3.3 Data Preprocessing

The data preprocessing is the gathering the required data from model development. Here we transform our data which were categorised into the number, mainly the label column we transform into the numerical value. Feature selection Create new features that can improve the accuracy of the crop recommendation model. This includes selecting only seven columns from our dataset which are essential in model development.Split the dataset into training, validation, and testing sets. This is done to evaluate the performance of the crop recommendation model on new, unseen data. We split the data 70 for training and 30 for testing. We train each model on a randomly selected training set.

We compare the same model on different splitting of training and testing dataset. We choose the best one from those splits.

4.Architecture and Implementation

We employed various algorithms, including Decision Tree, Logistic Regression, KNN, XGBoost, Gaussian Naive Bayes, and Neural Network, in our analysis. Among these algorithms, Decision Tree yielded the lowest accuracy, achieving a score of approximately 86%. aIn contrast, XGBoost produced the best results, with an accuracy of 99.54%. By analysing the accuracy of models, we chose the XGBoost model as best. Following table represents the accuracy of each algorithm:

Algorithms	Accuracy Score
Decision Tree (DT)	85.909
Gaussian Naive Bayes (GNB)	99.24
Logistic Regression (LR)	96.212
Random Forest (RF)	98.939
XGBoost (XGB)	99.545
KNN	98.48
Multilayer Perceptron (MLP)	94.54

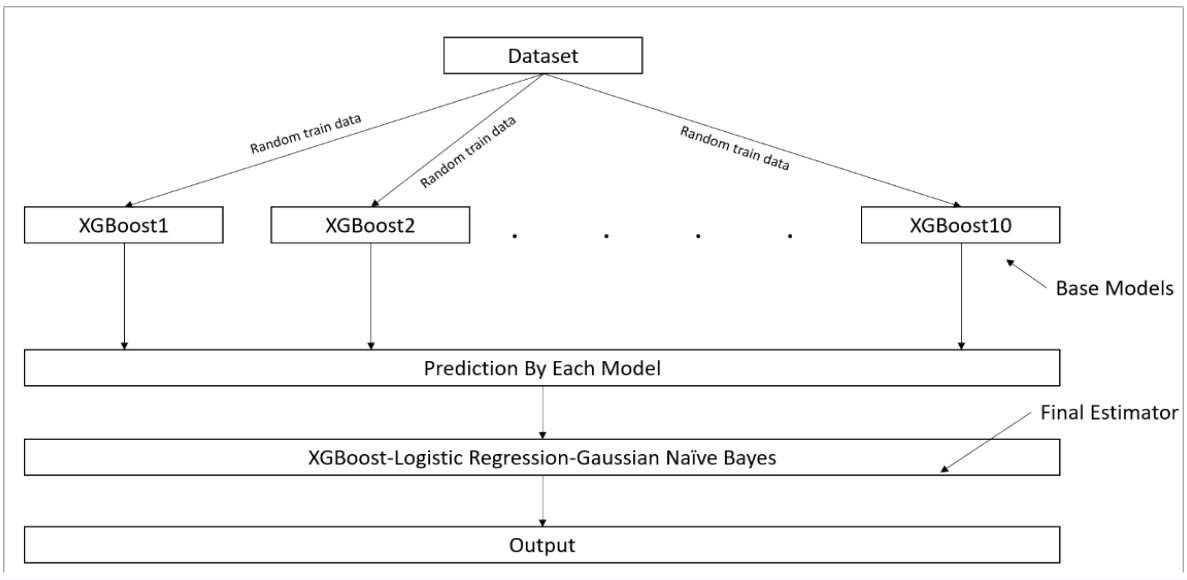
We combined each algorithm with XGBoost. We used the voting classifier for the combination. We make the pairs of two-two, three-three models, where in each pair XGBoost combined with other models. Following table represent the accuracy of combined models:

Algorithms	Accuracy Score
XGBoost-Decision Tree	98.939
XGBoost-RandomForest	99.39
XGBoost-GaussianNB	99.09
XGBoost-Logistic Regression	97.42
XGBoost-KNN	98.78

XGB-RF-DT	99.545
XGB-RF-LR	96.212
XGB-RF-GNB	99.54
XGB-DT-LR	96.212
XGB-DT-GNB	99.54
XGB-LR-GNB	96.121

Then we prepared the model using the bagging method of ensemble learning. We took the XGBoost and prepared 500 submodules of XGBoost with randomly selected training dataset. And provided this to the bagging classifier, which gave us the same accuracy as given by the single XGBoost model, which was 99.54.

We choose the stacking for model building, here we use all the seven algorithms (Table 1) as a base model. We use logistic regression as a final estimator which combines the output of each individual base model. The accuracy of the stacking model was 99.54, which was the same as XGBoost. Because of this we choose the XGBoost as a base model for building the stacking model. We first trained the ten individual models of XGBoost using the different training dataset. Used these models as base models for the stacking model building. Following is the architecture of model.



We used each algorithm from table 1 and table 2 as a final estimator in stacking. Following is the accuracy table given by each algorithm as a final estimator:

Algorithms	Accuracy
Decision Tree	52.73
Gaussian Naive Bayes	95.0
Logistic Regression	99.65
RF	99.24
XGBoost	99.24
KNN	99.39
MLP	99.65
Stack_7	99.24
XGBoost-DecisionTree	99.09
XGBoost-RandomForest	99.24
XGBoost-GaussianNB	98.94
XGBoost-LogisticRegression	99.55
XGBoost-KNN	99.24
XGB-RF-DT	99.24
XGB-RF-LR	99.68
XGB-RF-GNB	99.24
XGB-DT-LR	99.68
XGB-DT-GNB	99.24
XGB-LR-GNB	99.70

From table 3, we can see the combination of XGBoost, Linear Regression and Gaussian Naive Bayes as a final estimator provide the best accuracy score, which is 99.70. So after the analysis of various models we got the best accuracy through the stacking of XGBoost.

5.Conclusion:

We implemented various algorithms on the crop recommendation dataset. In a single algorithm XGBoost provided the best accuracy. In ensemble learning we merged two, and three models respectively. In merging none of the models provided us with more accuracy than the XGBoost. Bagging of XGBoost provided less accuracy than XGBoost. Stacking on XGBoost, where we used ten XGBoost models as base learner and a combination of XGBoost, Linear Regression and Gaussian Naive Bayes to combine the result. This approach provided the best accuracy which was 99.7% more than XGBoost single model. We took this model as our final model for prediction after comparing it with various models.

REFERENCES:

[1] . C. F. Gaitán, ‘‘Machine learning applications for agricultural impacts under extreme events,’’ in Climate Extremes and Their Implications for Impact and Risk Assessment. Amsterdam, The Netherlands: Elsevier, 2020, pp. 119–138.

[2]. Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, and S. Bhansali, ‘‘Review—Machine learning techniques in wireless sensor network based precision agriculture,’’ J. Electrochem. Soc., vol. 167, no. 3, Jan. 2020, Art. no. 037522.

[3]. Mohamed Hamdy Eldefrawy, Khaled Alghathbar, Muhammad Khurram Khan, "OTP-Based Two-Factor Authentication Center of Excellence in Information Assurance (CoEIA), King Saud University, Saudi Arabia, Information Systems Department, College of Computer and Information Sciences.

[4]. Jude Immaculate H*, Evanzalin Ebenanjar P, Sivaranjani K, and Sebastian Terence J, “Applications of Machine Learning Algorithms in Agriculture ”, Volume 82, page number:9312, November 2019.

[5]. Faith Bal and Faith Kayaalp,” Review of machine learning and deep learning models in agriculture ”, International Advanced Researches and Engineering Journal, Volume 05, Issue 02, pages: 309-323, 2021.

[6]. Majwega Jackson, Ggaliwango Marvin, Amitabha Chakrabarty, “ Robust Ensemble Machine Learning For Precision Agriculture ”, IEEE 2022.

[7]. Suman, Bharat Bhushan Naib "Soil Classification and Fertilizer Recommendation using WEKA" IJCSMS International Journal of Computer Science & Management Studies, Vol. 13, Issue 05, July 2013

[8] Jay Gholap " Performance Tuning of J48 Algorithms for Soil Fertility" 2012. Asian Journal of Computer Science and Information Technology 2: 8(2012) 251–252

[9]. ABHINAV SHARMA, ARPIT JAIN, PRATEEK GUPTA, And VINAY CHOWDARY," Machine Learning Applications for Precision Agriculture: A Comprehensive Review" IEEE 2021.

[10]. Crop Yield Prediction Using Regression and Neural Networks by Sandesh Ramesh, Anirudh Hebbar, Varun Yadav, Thulasiram Gunta, Balachandra A Department of Information Science and Engineering Nitte Meenakshi Institute of Technology Bangalore-56006

[11] S.S.Baskar L.Arockiam S.Charles "Applying Data Mining Techniques on Soil Fertility Prediction" International Journal of Computer Applications Technology and Research Volume 2–Issue 6, 660-662, 2013 .

BIOGRAPHIES:

Vivek R. Rajput
B Tech Dept. of Computer Engineering- VJTI, Mumbai

Ajinkya V. Deshmukh
B Tech Dept. of Computer Engineering- VJTI, Mumbai

Utkarsh R. Patil
B Tech Dept. of Computer Engineering- VJTI, Mumbai

Harshal S. Ushire
B Tech Dept. of Information Technology- VJTI, Mumbai

Prof. Pramila M. Chawan, is working as an Associate Professor in the Computer Engineering Department of VJTI, Mumbai. She has done her B.E. (Computer Engineering) and M.E. (Computer Engineering) from VJTI College of Engineering, Mumbai University. She has 28 years of teaching experience and has guided 85+ M. Tech. projects and 130+ B. Tech. projects. She has published 143 papers in International Journals and 20 papers in National/International Conferences/ Symposiums. She has worked as an Organizing Committee member for 25 International Conferences and 5 ICTE/MHRD-sponsored Workshops/ STTPs/ FDPs.She has participated in 16 National/International Conferences. Worked as Consulting Editor on – JEECER, JETR, JETMS, Technology Today, JAM&AER Engg. Today, The Tech. World Editor – Journals of ADR Reviewer -IJEF, Inters science She has worked as NBA Coordinator of the Computer Engineering Department of VJTI for 5 years. She had written a proposal under TEQIP-I in June 2004 for 'Creating Central Computing Facility at VJTI'. Rs. Eight Crore was sanctioned by the World Bank under TEQIP-I on this proposal. Central Computing Facility was set up at VJTI through this fund which has played a key role in improving the teaching-learning process at VJTI. warded by SIESRP with Innovative & Dedicated Educationalist Award Specialization: Computer Engineering & I.T. in 2020 AD Scientific Index Ranking (World Scientist and University Ranking 2022) – 2nd Rank- Best Scientist, VJTI Computer Science domain 1138th Rank- Best Scientist, Computer Science, India