International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211
Volume 10, Issue 3, March-2022, Impact Factor: 7.429, Available online at: www.ijaresm.com

Predictive and Preventive methods in Agriculture using Machine Learning

Subbhashit Mukherjee[1], Shivang Dholaria[2], Shree Ratn[3], Shreeharsh Gudibandi[4], Vidit Agarwal[5]

[1,2,3,4,5]Department of Computer Science And Engineering, BMS Institute Of Technology And Management. Karnataka, India

---*****************---

ABSTRACT

Artificial Intelligence & Machine learning being implemented in every possible field of application have drastically reduced human efforts and exponentially increased the outputs. One such field of Implementation is Agriculture. As Agriculture is one of the primary occupations in India, Applications of AI/ML proves to be a virtue to a large number of users. As agricultural supply chain is affected by middlemen, inconsistent pricing, the losses in the post-harvest sector have an adverse impact on Indian economy, farmers have been facing various difficulties due to insufficient knowledge of applicable technologies and use of resources such as water, fertilizers, irrigation methods. Cultivation of crops without knowing the soil fertility and the nutrients present the in soil of a particular region required for the crop is also one of the issues which may affect the quality and quantity of the yield produced. In this paper, we are providing solution for crop protection and yield. We propose a system where the user will be able to calculate the amount of fertilizers and pesticides, identify pest damage and diseases through image detection and get it diagnosed. The system will provide fertilizer and irrigation recommendations, predict the soil fertility, tips for cultivation with weather report and provide crop yield predictions. This System has the potential to help farmers calculate cost of fertilizers and pesticides required and saving the excess expenditure, get a better yield of crops

Keywords: Prediction System, Recommendation System, Plant Disease Detection System, XGBoost, LightGBM

I. INTRODUCTION

Highlight The practice of agriculture goes way back in history. It began independently in different parts of the globe in at least 11 separate centers of origin [3]. Over the course of history, agriculture evolved and now we use mechanized equipment and hybrid seed to get the maximum yield from the land. But still, some losses occur during the whole process of the harvest till getting the product to the end-user.Precision Farming is farm management system where in usage of different sensors are used and data collected from them is analysed to give farmers more control over their field and yield of the crop [4]. This approach is required to increase agriculture productivity, prevent soil degradation due to cultivating the same crop over and over again, reduce use to chemical application for crop production, irrigating water effectively, providing better socio-economic status of farmers, and many more.

Predictive analytics refers to the use of statistics and modelling techniques for making prediction on the future outcomes and performance using historical data. It's used to detect dangers and potential by looking for patterns in data [5] for agricultural purposes, it is employed to find the cuts in the cost, improve the efficiency of farming, for sustainable farming, optimize the supply chain management, and many more. This branch of analytics can be a game-changer for the agriculture business as it removes the guesswork for forecasting yields and helps in assessing risk.The term soil fertility refers to the capacity of soil to support the development of rural plants, to create favorable environmental conditions, and to produce sustained and consistent yields of good quality [6]. Fertility-promoting factors for soil are Soil depth to allow root growth and to retain water longer, adequate drainage for aeration for proper root growth, the pH level of soil, organic matter present on the topsoil for healthy structure, and moisture retention, and lastly the presence of microorganisms such as Azospirillum, Bacillus, Burkholderia that support plant growth. To produce the maximum yield of the crop from a given land, many factors need to be considered such as soil quality, air quality, weather pattern, etc. These factors together decide how much yield would be produced. Data from previous decades could be consolidated and be fed to a program to get results but they don't seem as accurate as they need to be. To approach this problem use of machine learning models such as statistical models, crop simulation models, etc., are made to run by giving datasets for previous weather patterns, climate change, soil fertility, etc. The output from the models would give very accurate data on the amount of fertilizer that plays an important role as a catalyst to speed up crop growth. As the saying goes "where there is water, there is life", water is an important natural resource through which the crops grow. Scarcity of water can lead to drought thus drying up the crop while excessive water can cause waterlogging, causing no aeration of soil thus destroying the crops from their roots. Through irrigation, water can be regulated thus providing the field with adequate water. But there still lies a major problem where which crop requires how much amount of water. There have been few irrigation methods developed to prevent water loss and maximize the yield. Surface irrigation, drip irrigation, sprinkler system, etc are some of the ways to irrigate water. The use of machine learning models can help farmers decide which method to use for what kind of crop. Different crops have different requirements of water, so the recommendation can suggest the irrigation methods that need to be implemented for increasing the yield.In this manuscript, we put forward an approach to tackle the issue of using the right amount of fertilizer for different crops on size of land, identifying the disease through object detection and suggesting proper diagnosis for the same, providing Irrigation Recommender for different crops, soil predictor to check the fertility of the land, cultivation tips with weather report and yield prediction for the crop harvest. A dataset containing parameters such as temperature, humidity, Moisture, Soil Type, Crop Type, etc., be used to create a model for fertilizer recommendation system which can tell the what fertilizer to use and by how much amount. The concept of object detection would be used with the help of OpenCV library to detect the diseased crop is affected with and identify the damage done and provide a diagnosis for the same. The as different crop requires a different amount of water and fertile, dataset having Temperature, Humidity, Moisture, etc., as the parameters would be used to create a model for creating fertilizer recommendation system and dataset having the parameters such as Crop Type, Crop Days, Soil Moisture, etc., would be used to create a model for irrigation recommendation system and in a proper format, the data will be presented to the farmer. Data from the soil would be used and compared with the model made using the available dataset from Kaggle which includes attributes such as pH, Nitrogen, phosphorous, etc., to determine the fertility and the status of the soil will be shown.

The rest of this paper is organized as follows. Section 2 methodology describes the research methodology and relevant methods used for the implementation of the recommendation system. Section 3 provides the conclusion of the paper by discussing our findings and comparing the algorithms used in making the recommendation models. Section 4 future step describes what additional work needs to be done upon the models to increase their accuracy for getting better predictions and results.

II. RELATED WORK

In order to find related work to our project, our team went through various websites, technical papers, articles, newspapers and documents from different sources. The needs of artificial intelligence in improving the present agriculture have been discussed in the paper [2] and paper [1] has discussed about improving the performance of agricultural knowledge and information systems (AKIS) using various alternatives for stimulating the development and to contribute to poverty reduction. After knowing the importance of artificial intelligence in agriculture we decided to make a web application to support the needs of farmers through our services. Our system aims to provide three systems: Predictive systems, Recommendation systems, and Detection systems.

A. Prediction System

We have two prediction systems in our web application: Soil Fertility predictor and Crop Yield Predictor.

Information related to crop yield and soil fertility is very useful for farmers. For loss reduction, It is very crucial to know the yield and fertility of the field. In past, the prediction was done on the basis of experience gained by the farmers who have worked for many years. Our prediction system also works in the similar fashion. It takes into consideration the data from previous years and uses it to predict the future crop yield and soil fertility. Weather and pesticides are the two important factors on which the crop yield mainly depends whereas soil fertility depends on soil nutrients. This prediction depends upon the quality of the information provided. Therefore, the proposed system predicts the yield and soil fertility and decreases the loss. An information combination strategy has been proposed by Mann, M. L., Warner, J. M., & Malik, A. S. et al. [18], for a considerable beneficiary of specially appointed imported nourishment to Ethiopia that consolidates both remotely detected information (RSD) and agrarian overview information. Janez Trontelj ml., Olga Chambers [19] has proposed a hypothesis that we can improve the accuracy of soil properties prediction using various machine learning algorithms.

B. Recommendation System

We are having three recommendation systems in our web application: Fertilizer recommendation system, Irrigation recommender system and Crop recommender system. The aim of the fertilizer recommendation is to estimate the nutrient content of soil and recommend a suitable fertilizer that can be used for higher productivity. Irrigation Recommendation helps in recommending if a land requires irrigation or not. Crop recommendation aims to provide the most suitable crop that can be grown in a certain field.In the paper [20], the author explains about "using fiber optics, temperature, humidity, temperature, and sunlight sensor to estimate the npk values of soil and other parameters and pass it to the microcontroller to remotely control the work in agriculture field such as fertilization". Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. et al. [21] Various Machine learning approaches have been discussed in this paper that are used in developing crop yield prediction systems for the Canadian Prairies. Their research was concentrated on determining the forecast yields of grain, canola, and spring wheat utilizing previously detected vegetation records.

C. Plant Disease Detection System

We will be providing a plant disease detection portal where one can upload a picture of that particular plant and our system will predict the disease for the same.Anand H. Kulkarni et al. [11], has proposed various image processing methods for early and accurate detection of plant diseases. Gabor filter has been used for feature extraction and ANN-based classifier has been used for classification which had an accuracy up to 91%. F. A fast approach for determining the parameters of co-occurrence matrix using supervised learning and maximum likelihood method has been proposed by Argenti, et al. [12] for fast classification. Detection of edges is done using homogenized techniques like Sobel and canny filter by P. Revathi et al. [13]. The classification uses the extracted edge features to identify the disease spots. A methodology has been proposed by Tushar H Jaware et al. [14], which involves the use of enhanced k-means clustering techniques to solve low-level image segmentations. The statistical texture features extraction has been extracted using spatial gray-level dependency matrices, by Sanjay B. Dhaygude et al. [15]. Hue Saturation Value (HSV) color space representation is obtained from RGB (red green blue) images and showed the H, S, and V components. Detection of leaf spot diseases in olives using fuzzy c-means classification and auto-cropping segmentation has been presented by Mokhled S. Al-Tarawneh [16]. Image enhancement is done by using median filter and RGB to Lab colorspace. S. Arivazhagan, et al. [17] has proposed methods for detection of unhealthy regions in the leaves of the plants and classification of the same with the usage of texture features. Algorithm proposed by them has been tested on ten species of plants which are: banana, jackfruit, beans, mango, lemon, tomato, potato and sapota. An accuracy of 94.74% has been achieved using the Support vector machine (SVM) classifier algorithm for the same.

III. METHODOLOGY

This section comprises various research methodology and relevant methods of application implementation broken down in how we collect data which are used for the training and validation, how these can predict the desired outcome. A web-based application approach is discussed, and suitable deep learning techniques are made use of – for plant disease detection while achieving high levels of accuracy.

A. Data Collection

1. RECOMMENDATION SYSTEM

There are many fertilizers that are available in the market but to use the most suitable one for farming, it is a critical decision to make .This recommendation system for fertilizers has parameters like Soil and Crop Type, temperature, humidity, Moisture and NPK (Nitrogen, Phosphorous, Potassium). Based on those parameters the Fertilizer name can be predicted by the machine learning algorithm.Currently our system can recommend these fertilizers: 10-26-26 , 14-35-14 , 17-17-17 , 20-20 , 28-28 , DAP and Urea [Dataset link]. The Irrigation Dataset [Dataset taken from Department of Computer Application, National Institute Of Technology, Raipur] can be used to train the Machine Learning model which will help to predict whether irrigation is required or not based on the certain values of 5 parameters: Crop Type, Crop Days, Soil Moisture, Temperature, Humidity. Based on these attributes the prediction system will suggest whether the crop requires irrigation or not.This system is made to ease the decision-making process of which crop to plant based on parameters like Nitrogen, phosphorous, and potassium content of the soil, temperature, humidity, and rainfall of the region, and pH scale measurements.Currently we have the data for crops of apple, banana, black gram, chickpea, coconut, coffee, cotton, grapes, jute, kidney beans, lentil, maize, mango, moth beans, mung beans, muskmelon, orange, papaya, pigeon peas, pomegranate, rice and watermelon for crop recommendation. Hence a suitable crop to plant and cultivate will be recommended based on input, and RMS do not have to be defined. [Dataset link].

2. PREDICTION

Since the crop growth is dependent on soil the fertility is measured by the different elements it contains and the pH scale. This data set includes pH, Nitrogen, phosphorous, zinc, iron, copper, manganese, sand, silt, clay, calcium carbonate to determine fertility [Dataset link]. This classifies fertility into two classes fertile and non-fertile. The dataset for yield prediction consists of attributes while making predictions for production: State_Name, District_Name, Crop_Year, Season, Crop and Area. Currently we have the data of for each state and a total of 124 crops [Dataset link].

3. PLANT DISEASE DETECTION

The data used in our web application is sourced from Kaggle and this dataset consists of about 87,000 high quality images of leaves of various plants which are divided into 13 different classes [Dataset link]. The total dataset is divided into the ratio of 80% training data and 20% validation data.Currently, the Dataset Contains 15 kinds of diseases for tomato, potato and pepper.

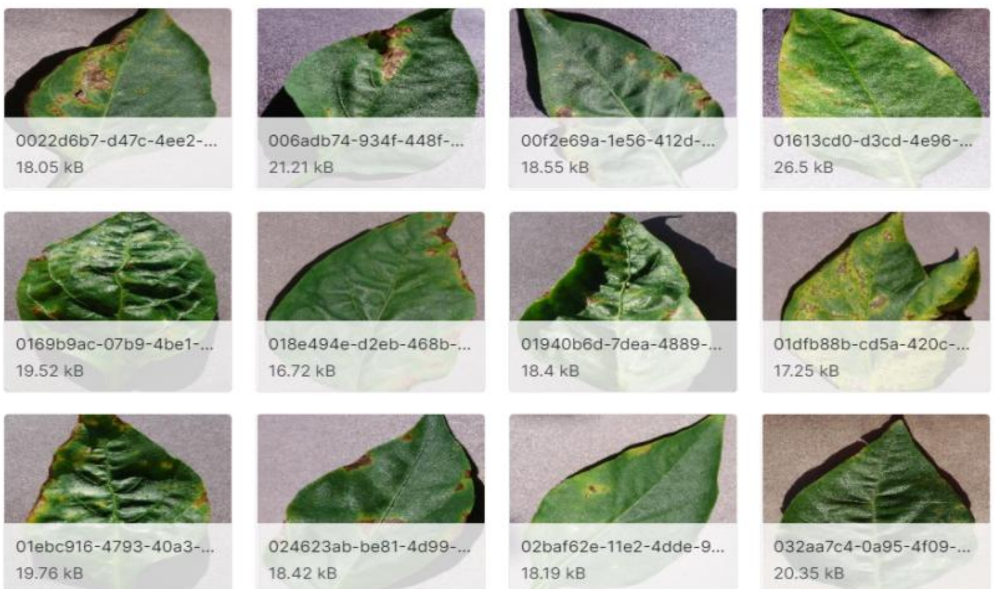

Fig 1.0 Plant Disease Dataset

B. Prediction System

We have implemented our prediction using various machine learning algorithms.

1. k-Nearest neighbor (KNN) Algorithm

k-Nearest Neighbor (KNN) algorithm is an algorithm which is used for classifying data point to its nearest cluster of data points. It can be also used for regression. However, it is mainly used for classifying the data and problems related to them. As discussed in paper [22], KNN bunches the elements into groups and orders the new information input based on likeness with recently prepared items. Data from the new input is assigned to the cluster of points that it shares the closest neighbors with.

2. Support Vector Machine Algorithm

The Support Vector Machine (SVM) algorithm is a kind of supervised machine learning technique. It is used to analyze data for classification and regression [23]. Using the training dataset, this algorithm aims to make a prediction model that is able to predict the target values of the test data using only the test data attributes as input. In SVM, the accuracy of the fit of the point is indicated by the training value, which refers to the residual of the output.

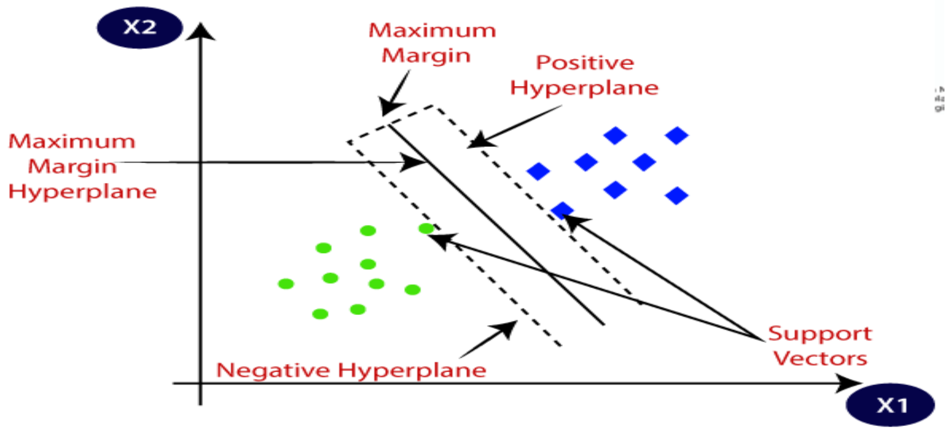

Fig 2.0 Working of SVM

3. Light GBM

Microsoft developed an open-source machine learning algorithm called as Light GBM. It employs a histogram-based technique to expedite the training process. It also minimizes the memory consumption. The leaf-wise technique is used to develop the trees and conduct the split by identifying a leaf with the biggest gain of variance. The LightGBM technique develops trees vertically whereas another algorithm grows trees horizontally indicating that later approach grows tree leaf-wise while LightGBM grows level-wise as stated in paper [25]. It will keep in growing by choosing the leaf having max delta loss.

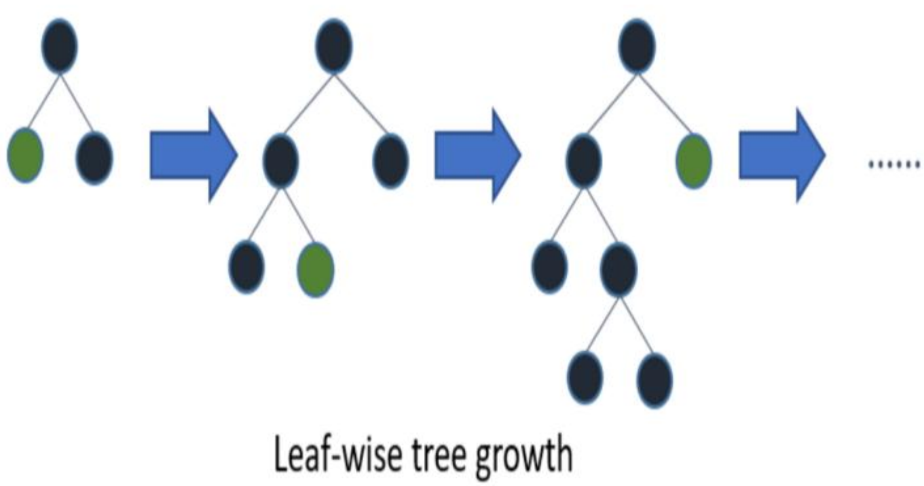

Fig 3.0 Leaf-wise growth of Trees in Light GBM

Soil Fertility Predictor

The word soil fertility relates to the potential of soil to support the growth of rural plants, to establish suitable environmental circumstances, and to generate continuous and consistent harvests of excellent quality. Soil fertility can be improved using two types of fertilizers: organic and inorganic fertilizers. In soil fertility prediction, we take various attributes of the soil such as N, P, K, Ph, EC, OC, OM, etc. to predict if the soil is fertile or not. Using the mentioned machine learning algorithms,we determine the optimal parameters for predicting the soil fertility and equation is formed for the same. Our model is simple and provides exceptionally good predictions. For training our model, we have used the input data set which is divided into 2 parts. The input data set makes up 80% of the data in the data set, while the validation data set makes up the remaining 20%. After completion of the training of model on input dataset, the trained model is fed with test data for making predictions. The accuracy of the different models are shown below.

Table 1.0 Comparison of Error Values for The Applied Models

Algorithm	RMSE	Accuracy
KNN	999.26	82%
SVM	512.78	85%
LightGBM	78.20	94%

The analyses of all three algorithms are shown in the above given table. We see that the k-nearest neighbor algorithm has given 82% of accuracy, which is the least among the three algorithms. Support vector machine gave 85% of accuracy. On comparing it with the k nearest neighbor algorithm, it is kind of acceptable. LightGBM outperformed both of them and got the best accuracy of 94%, so we used the same for soil fertility prediction.

Yield Predictor

Crop yield gives a very important piece of information to farmers. In the past, experienced farmers used to predict the yield based upon their past experiences. Our proposed framework likewise works along these lines. Information from the previous years is taken of the yield and the same is used to predict the future yield. Weather and pesticides are important factors for predicting crop yield. Our proposed system provides suitable forms for uploading the input data. Using the mentioned machine learning algorithms, we determine the optimal parameters for predicting the soil fertility and equation is formed for the same. Our model is simple and provides exceptionally good predictions. For training our model, we have used the input data set which is divided into 2 parts. The input data set makes up 80% of the data in the data set, while the validation data set makes up the remaining 20%. After completion of the training of model on input dataset, the trained model is fed with test data for making predictions. The accuracy of the different models are shown below.

Table 2.0 Comparison of Error Values for The Applied Models

Algorithm	RMSE	Accuracy
KNN	8881.69	72%
SVM	1028.23	78%
LightGBM	90.20	91%

The analyses of all three algorithms are shown in the above given table. We see that the k-nearest neighbor algorithm has given 82% of accuracy, which is the least among the three algorithms. Support vector machine gave 85% of accuracy. On comparing it with the k nearest neighbor algorithm, it is kind of acceptable. LightGBM outperformed both of them and got the best accuracy of 94%, so we used the same for soil fertility prediction.

C. Recommendation System

We have implemented our recommendation systems using various machine learning algorithms.

1. **Random Forest Algorithm**

Random Forest is a supervised learning machine learning algorithm. This algorithm uses precision techniques to create forest and makes it random. It builds a forest of an ensemble of decision trees, as discussed in paper [27]. In Random forest algorithm, multiple decision trees are created and later merged into a single tree to obtain accurate prediction as in paper as in paper [24]. As the number of trees in the forest increases, the accuracy of prediction increases as well.

2. Gradient Boosting Algorithm

Gradient boosting algorithm is a supervised machine learning algorithm. Bias and Variance Error are two types of errors in machine learning algorithms. "Gradient boosting is one of those boosting algorithms that are used to minimize bias error of the machine learning model", as discussed in paper [28].Regression tasks as well as classification tasks can be achieved using Gradient boosting algorithm. Cost function when gradient boosting algorithm is used as a regressor is Mean Squared Error (MSE) and cost function is log loss when it is used as a classifier .

3. XGBOOST

"XGBoost is an ensemble Machine Learning algorithm that uses a gradient boosting framework" as discussed in paper [9]. Decision tree-based algorithms are considered as the best for making predictions when we have to deal with small or medium size structured dataset.

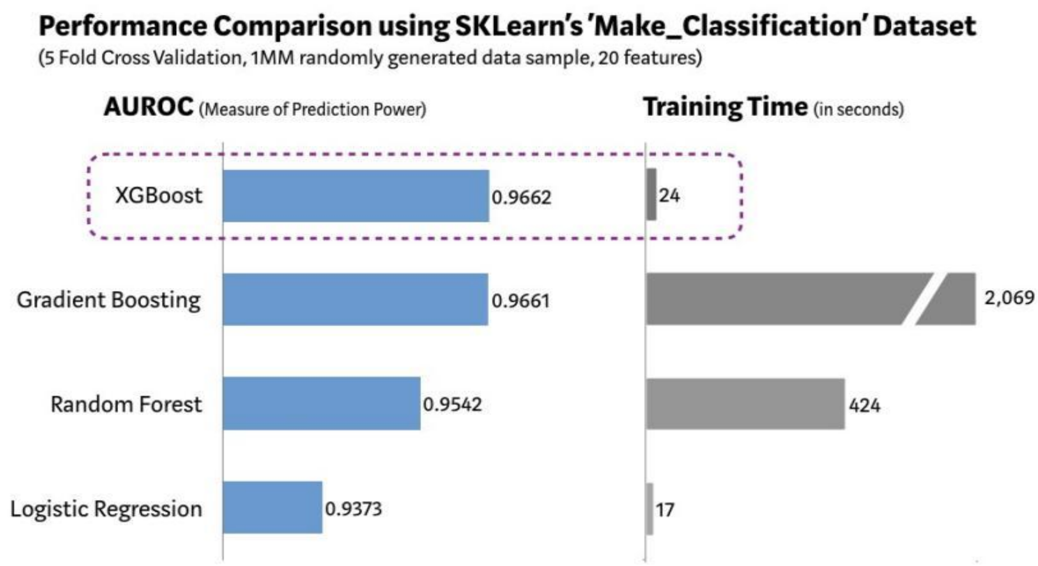

Fig 4.0 Performance comparison of XGBoost with other ML models

"As shown in the chart above, prediction performance and processing time of the XGBoost model is the best in comparison to other algorithms", as discussed in the article [26]. It also shows the system optimization and algorithmic enhancements that XGBoost provides, which helps it in outperforming various other machine learning algorithms.

Fertilizer Recommendation System

The aim of our proposed fertilizer recommendation system is to estimate the nutrient contents of the soil and recommend a suitable fertilizer that van be used for increasing soil fertility. Under application of fertilizer results in low yield due to insufficient nutrients present in the soil for the crop. Over usage of fertilizer results in soil pollution. The food products from the polluted soil will be food poisoning and health issues for the consumer. In Fertilizer Recommendation System, we take various attributes like Temperature, Humidity, Moisture, Soil Type, Crop Type, Nitrogen, Potassium, and Phosphorous. Using the mentioned machine learning algorithms, we determine the optimal parameters for predicting the fertilizer needed and equation is formed for the same. Our model is simple and provides exceptionally good predictions. We trained our model using the input data set, which is further divided into training(80% of input data) and validation dataset (20% of input data). After completion of the training of model on input dataset, test data is given as input to the trained model to make predictions. The accuracy of the different models are shown below.

Table 3.0 Comparison of Error Values for The Applied Models

Algorithm	RMSE	Accuracy
Random Forest	112.64	90%
Gradient boosting	812.90	82%
XGBoost	86.4	94%

We analyze that the gradient boosting algorithm has produced 82% accuracy as shown in the bable which is the least among the three algorithms. Random Forest Algorithm produced 90% accuracy which is comparatively better when compared with gradient boosting algorithm. While XGBoost outperformed both of them and got the best accuracy of 94%, so we used the same for the fertilizer recommendation system.

Crop Recommendation System

Environmental factors needed to be identified to compare and give recommendations. The dataset is cleaned and then processed for better crop recommendation. A machine learning model is used to predict the best crop type for a particular land as the environmental conditions may differ from region to region. Random Forest, Gradient Boosting, and XGBoost algorithms are used to select the best crop type. This is done by analyzing factors like humidity, temperature, soil moisture, pH level, and sunlight.Our recommendation system recommends 4 types of crops by using the above-mentioned machine learning algorithms. We trained our model using the input data set, which is further divided into training (80% of input data) and validation dataset (20% of input data). After completion of the training of model on input dataset, test data set is given as input to the trained model to make predictions. The accuracy of the different models is shown below.

Table 4.0 Comparison of Error Values for The Applied Models

Algorithm	RMSE	Accuracy
Random Forest	92.84	95%
Gradient boosting	312.90	85%
XGBoost	56.4	98%

We analyze that the gradient boosting algorithm has produced 85% accuracy as shown in the table which is the least among the three algorithms. Random Forest Algorithm produced 95% of accuracy which is comparatively better than gradient boosting algorithm. While XGBoost outperformed both of them and got the best accuracy of 98%, so we used the same for the crop recommendation system.

Irrigation Recommendation System

Various sensors available are used to gather all the important attributes for our recommendation system and these values are collected already and the dataset is publicly available in Kaggle. The collected dataset is used as input to our ML-based model for recommendation. The Machine Learning unit is the heart of our irrigation recommendation system. The parameters that are considered for our recommendation system are atmospheric pressure, precipitation, solar radiation, and wind speed. We provide input to our real time deployed model using multi-step forms. This classification model categorizes the samples to be predicted into two classes: irrigation is required (Yes) or irrigation is not required (No). The results are showed in the website after the form submission.

Using the mentioned machine learning algorithms,we determine the optimal parameters for predicting if the land requires irrigation or not and equation is formed for the same. Our model is simple and provides exceptionally good predictions. We trained our model using the input data set, which is further divided into training (80% of input data) and validation dataset (20% of input data). After completion of the training of model on input dataset, test data is given as input to the trained model to make predictions. The accuracy of the different models are shown below.

Table 5.0 Comparison of Error Values for The Applied Models

Algorithm	RMSE	Accuracy
Random Forest	52.64	96%
Gradient boosting	178.93	92%
XGBoost	39.45	98%

We analyze that the gradient boosting algorithm has produced 92% accuracy as shown in the table which is the least among the three algorithms. Random Forest Algorithm produced 96% of accuracy which is comparatively better than gradient boosting algorithm. While XGBoost outperformed both of them and got the best accuracy of 98%, so we used the same for the irrigation recommendation system.

D. Plant Disease Detection System

1. Adam

We have used adam as an optimizer in our plant disease detection model. This algorithm computes the exponentially weighted average of the gradients that is used to get the point of minima at a faster pace, as discussed in paper [8]. Its working algorithm with equations is explained below.

Require: α: Stepsize
Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates
Require: $f(\theta)$: Stochastic objective function with parameters θ
Require: θ_0: Initial parameter vector
$m_0 \leftarrow 0$ (Initialize 1st moment vector)
$v_0 \leftarrow 0$ (Initialize 2nd moment vector)
$t \leftarrow 0$ (Initialize timestep)
while θ_t not converged **do**
$t \leftarrow t + 1$
$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)
$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)
$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)
$\widehat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)
$\widehat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)
$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \widehat{m}_t / (\sqrt{\widehat{v}_t} + \epsilon)$ (Update parameters)
end while
return θ_t (Resulting parameters)

Fig 5.0 Pseudocode for Adam Optimizer

2. Loss Function

The loss function that we have used for our plant disease detection system model is binary cross-entropy. We have first converted the multi-class problem in binary class problem using label binarizer. The binary cross-entropy function computes the cross-entropy loss between actual labels and predicted labels. It then calculates the score which penalizes the probabilities on the basis of distance from the expected values. That means, if the predictions are closer to the actual values, then the loss will be minimum and if the predictions are far away from the actual values, the loss value will be the maximum. The mathematical annotations can be seen here:

$$logloss = -\frac{1}{N}\sum_{i}^{N}\sum_{j}^{M} y_{ij}\log(p_{ij})$$

- N is the number of rows
- M is the number of classes

..eq.(1)

3. ReLU (Rectified Linear Unit)

ReLU activation function has been used for training and validation in our model. “ReLU stands for rectified linear unit and is a very popular activation function” as discussed in paper [10]. It is defined by the formula, Y = max (0, X). Visually, it looks like the following: ReLU is a widely used activation function in neural networks in the recent times, especially in CNNs. Graph of ReLU looks as follows:

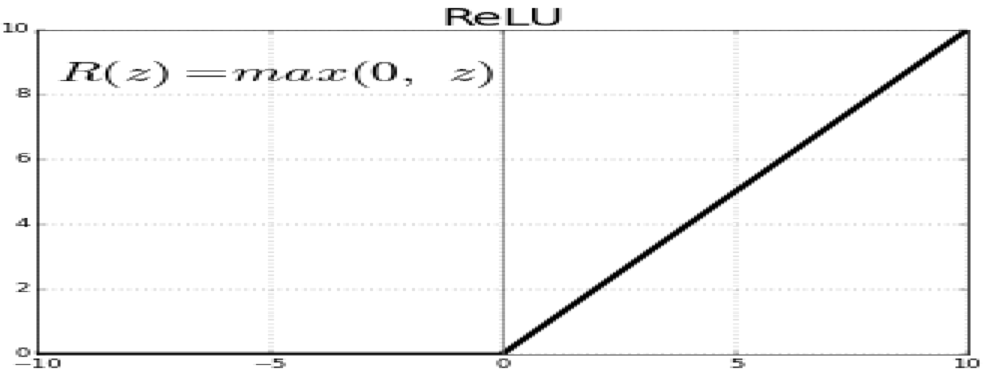

Fig 6.0 Pseudocode for Adam Optimizer

For the objective our plant disease detection system to be fulfilled, a detection system was made using object detection techniques. Object detection is vast area of research in the field of computer vision and image processing that helps in locating objects of interest as per the requirements. We initialized learning rate for our model as 0.001. We converted shape of the images after pre-processing to (256,256,3) >S

For our proposed system, we have used convolutional neural network for image classification for disease detection. "Convolutional neural network is widely used in image classification tasks" as discussed in paper [7]. We have used Keras for developing our deep learning model for plant disease detection. Keras is beginner-friendly and easy to use for beginners. There is also a wide variety of pre-trained models available in Keras like ResNet and VGG. But for our system, we have developed a custom model for plant disease detection. We start our model by initializing an input layer. Convolution layer is used that performs different kinds of operations on the input image, so that features like edges or curves are detected. We have used a filter of (3,3) for the same. After this, we apply the pooling layer that extracts only the key features with a pool size of (2,2) and thereby helps in dimensional reduction of the data, so number of computations decreases. We used ReLU as our activation function and used Binary Crossentropy as our loss function having the formula as given in eq.(1) Finally, we pass the data into the fully connected network layers, which try to combine the features obtained from the previous layers and a classification is done using the output layer.

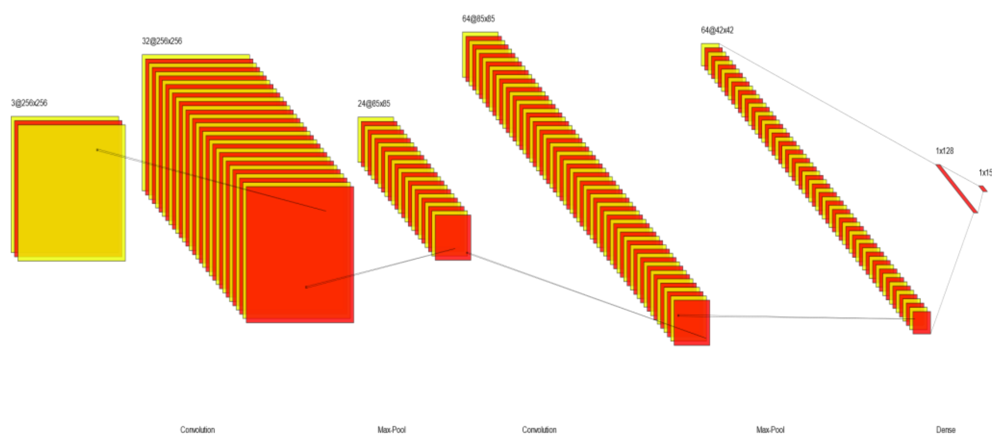

Fig 7.0 Our Custom Image Classification Model

We used the above-specified activation function, loss function, and optimizer to train our predictive model. We have provided a suitable image input form in our system for predicting if the plant is suffering from a disease or not. If the plant is suffering from a disease, it will showcase the name of the particular disease.

IV. CONCLUSION

It is necessary to consider the climatic factors and soil factors for better crop recommendation, irrigation recommendation, and fertilizer recommendation. Usually, the failures in agriculture are caused due to the poor irrigation recommendation, poor choice of crops, poor selection of fertilizers, or faulty prediction of weather. Our proposed prediction and recommendation systems can help in avoiding such things.In this paper, we are using various classification techniques for the prediction of crop yield and soil fertility. Three classifier algorithms namely KNN, SVM, and LightGBM were used for both cases. We are also using various supervised machine learning algorithms for our fertilizer recommendation system, a crop recommendation system, and an irrigation recommendation system. Three classifier algorithms namely random forest algorithm, gradient boosting algorithm, and XGBoost were used for all three cases. The outcome of these techniques is compared based on error values and accuracy. The result of the experiment showed that for the yield Prediction system, lightGBM gets the highest accuracy value of 91%, the lowest accuracy was for KNN 72% And the highest error value 8881.69 was for KNN. For the soil fertility prediction system, lightGBM gets the highest accuracy value of 94%, the lowest accuracy was for KNN 82% And the highest error value 999.26 was for KNN. For the fertilizer recommendation system, XGBoost gets the highest accuracy value of 94%, the lowest accuracy was for gradient boosting algorithm 82% And the highest error value 812.90 was for gradient boosting algorithm. For the crop recommendation system, XGBoost gets the highest accuracy value is 98%, the lowest accuracy was for gradient boosting algorithm 85% And the highest error value 312.90 was for gradient boosting algorithm. For the irrigation recommendation system, XGBoost gets the highest accuracy value 98%, the lowest accuracy was 92% for the gradient boosting algorithm and the highest error value 178.93 was for the gradient boosting algorithm. We have used mean F1-score for evaluating the performance of our plant disease detection system. It is a better in comparison to accuracy-score as it takes both false negatives and positives into consideration while making predictions. When we trained our model, it was clear that our model was able to correctly classify the input data, with a minimum F1-score of 96%.

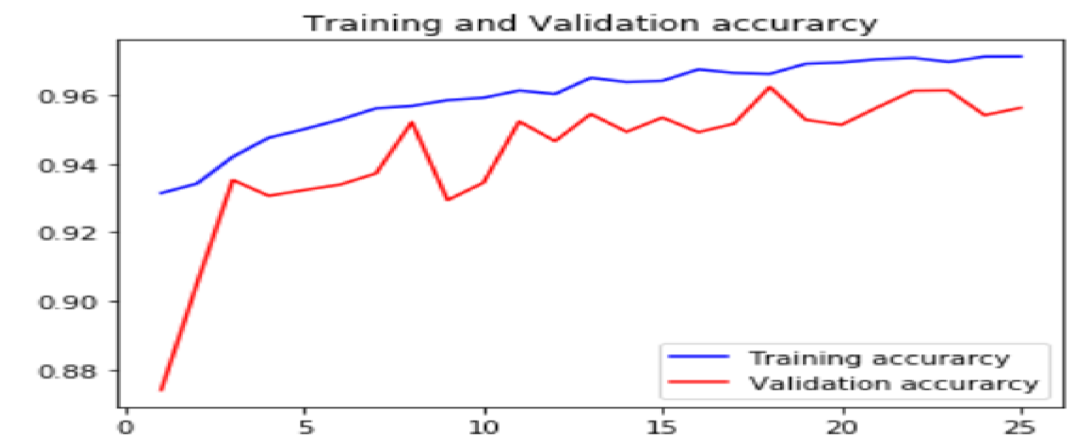

Fig 8.0 Plot of training and testing accuracy charts of our plant disease detection model

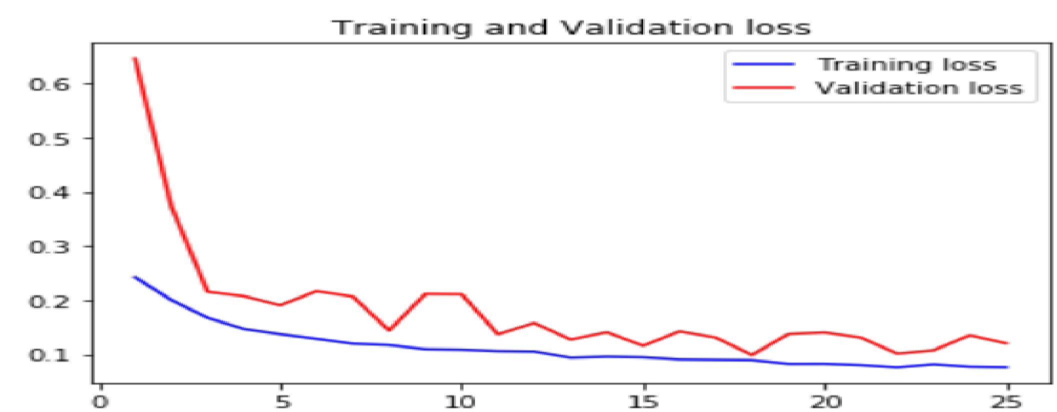

Fig 9.0 Plot of training and testing accuracy charts of our plant disease detection model

Our system will help in reducing the struggle and hardships faced by the farmers. Analysis of the important soil and location properties is done and based on that we are dealing with our proposed prediction and recommendation systems. This will act as a simple solution to help the farmers with the necessary information to obtain great yield and therefore maximize their productions and therefore will reduce their difficulties.

REFERENCES

[1] Berdegué, Julio & Escobar, Germán. (2001). Agricultural knowledge and information systems and poverty reduction.
[2] Eli-Chukwu, Ngozi & OGWUGWAM, EZEAGWU. (2019). Applications of Artificial Intelligence in Agriculture: A Review. Engineering, Technology and Applied Science Research. 9. 4377-4383. 10.48084/etasr.2756.
[3] devlet, Abdulgani. (2020). Agriculture History and Policy. International Journal of Science Letters. 2. 39-51. 10.38058/ijsl.685412.
[4] Pandey, Himanshu & Singh, Devendra & Das, Ratan & Pandey, Devendra. (2022). Precision Farming and Its Application. 10.1007/978-981-16-6124-2_2.
[5] Kumar, Vaibhav & L., M.. (2018). Predictive Analytics: A Review of Trends and Techniques. International Journal of Computer Applications. 182. 31-37. 10.5120/ijca2018917434.
[6] Sathyanarayana, Eetela & Bharghavi, J & Saranya, Suddakanti & Manchala, Santhosh & Mandla, Rajashekhar & Veeranna, Jatoth & Rajashekar, Banda & Reddy, Dr T Prabhakar & Thallapally, Saideep & Ingle, Krishnananda & Chaitanya, Alamuru & Duta, Asik & Umarajashekhar, Alavala & Sunita, Kumari & Jatav, Hanuman. (2022). Overview of Soil Fertility from Past To Present.
[7] Yingying Wang,Yibin Li,Yong , 'Facial Expression Recognition Based on Random Forest and Convolutional Neural Network'
[8] Alom, Mohammed. (2021). Adam Optimization Algorithm.
[9] Duyen Thi Do,Nguyen Quoc Khanh Le,'Using extreme gradient boosting to identify origin of replication in Saccharomyces cerevisiae via hybrid features'
[10] Victor Jacobson,J. Jenny Li,Kevin, 'Visualizing Neural Networks for Pattern Recognition'
[11] Anand H. Kulkarni, Ashwin Patil R. K., Applying image processing technique to detect plant diseases, International Journal of Modern Engineering Research, vol.2, Issue.5, pp: 3661-3664, 2012.
[12] F. Argenti,L. Alparone,G. Benelli ," Fast algorithms for texture analysis using co-occurrence matrices" Radar and Signal Processing,IEE Proceedings , vol. 137, Issue 6, pp:443-448 , No. 6, December 1990.
[13] P. Revathi, M. Hemalatha, Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques, IEEEInternational Conference on Emerging Trends in Science, Engineering and Technology, pp-169-173, Tiruchirappalli, Tamilnadu, India, 2012.
[14] Tushar H. Jaware, Ravindra D. Badgujar and Prashant G. Patil, Crop disease detection using image segmentation, National Conference on Advances in Communication and Computing, World Journal of Science and Technology, pp:190-194, Dhule, Maharashtra, India, 2012.
[15] Prof.Sanjay B. Dhaygude, Mr.Nitin P. Kumbhar, Agricultural plant Leaf Disease Detection Using Image Processing, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , S & S Publication vol. 2, Issue 1, pp: 599-602, 2013.
[16] Mokhled S. Al-Tarawneh An Empirical Investigation of Olive Leave Spot Disease Using Auto-Cropping Segmentation and Fuzzy C-Means Classification, World Applied Sciences Journal , vol.23, no.9, pp:1207-1211,2013.
[17] S. Arivazhagan , R. Newlin Shebiah , S. Ananthi, S. Vishnu Varthini , Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, Commission Internationale du Genie Rural(CIGR) journal, vol. 15, no.1,pp:211-217, March 2013.
[18] Mann, M. L., Warner, J. M., & Malik, A. S. (2019). Predicting high-magnitude, low-frequency crop losses using machine learning: an application to cereal crops in Ethiopia. Climatic Change, 154(1- 2), 211-227.
[19] Janez Trontelj ml.,Olga Chambers,Machine Learning Strategy for SoilNutrients Prediction UsingSpectroscopic Method. Sensors 2021,21, 4208. https://doi.org/10.3390/s21124208.
[20] L. C. Gavade, "Detection of N, P, K using Fiber Optic Sensor and PIC Controller," International Journal of Engineering Science,vol. 13787, 2017.
[21] Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., & Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. Agricultural and forest meteorology, 218, 74-84.
[22] K. Taunk, S. De, S. Verma and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.

[23] S. Ghosh, A. Dasgupta and A. Swetapadma, "A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification," 2019 International Conference on Intelligent Sustainable Systems (ICISS), 2019, pp. 24-28, doi: 10.1109/ISS1.2019.8908018.

[24] Afshin Jamshidi,Jean-Pierre Pelletier,Johanne Martel-Pelletier ,'Machine-learning-based patient-specific prediction models for knee osteoarthritis'

[25] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 3146–3154. [Online]. Available: http://papers.nips.cc/paper/6907-lightgbm-a-highlyefficient-gradient-boosting-decision-tree.pdf

[26] Vishal Morde , "XGBoost Algorithm: Long May She Reign!"

[27] Shipway, N.J., Huthwaite, P., Lowe, M.J.S. et al. Performance Based Modifications of Random Forest to Perform Automated Defect Detection for Fluorescent Penetrant Inspection. J Nondestruct Eval 38, 37 (2019). https://doi.org/10.1007/s10921-019-0574-9

[28] N. Aziz, E. A. P. Akhir, I. A. Aziz, J. Jaafar, M. H. Hasan and A. N. C. Abas, "A Study on Gradient Boosting Algorithms for Development of AI Monitoring and Prediction Systems," 2020 International Conference on Computational Intelligence (ICCI), 2020, pp. 11-16, doi: 10.1109/ICCI51257.2020.9247843.