International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Fertilizer Recommendation and Crop prediction using Machine Learning Techniques

Mr. Ram Sharma[1], Ayush Singh[2], Rampal[3], Raj Kumar Chaurasiya[4], Ashish Kumar[5]

[1]Computer Science and Engineering Department Raj Kumar Goel Institute of Technology India ram.sharma@rkgit.edu.in
[2]Computer Science and Engineering Department Raj Kumar Goel Institute of Technology India ayushsingh6476@gmail.com
[3]Computer Science and Engineering Department Raj Kumar Goel Institute of Technology India rampalshekha1999@gmail.com
[4]Computer Science and Engineering Department Raj Kumar Goel Institute of Technology India raj445067@gmail.com
[5]Computer Science and Engineering Department Raj Kumar Goel Institute of Technology India ashishhkr14@gmail.com

ABSTRACT—

Precision agriculture is gaining momentum in recent years, and one of its main challenges is to recommend the appropriate fertilizer for a specific crop to achieve optimal yield. In this research paper, we propose a machine learning-based approach for crop and fertilizer recommendation. Our approach uses a dataset of soil and weather data along with crop yield data, which is preprocessed and used to train a model based on the Gradient Boosting Machine algorithm. The trained model can predict the optimal fertilizer type and amount for a given crop based on the current soil and weather conditions. Our proposed model achieved an accuracy of 86.5% in predicting the optimal fertilizer type and amount for various crops. We also compare our model's performance with traditional methods and demonstrate that our approach outperforms them in terms of accuracy and efficiency. Our approach can significantly reduce the cost of fertilizer application and help farmers achieve optimal crop yield, thus promoting sustainable agriculture practices.

Keywords—Machine learning, Gradient Boosting Machine, Soil and weather data, Random Forest.

I. INTRODUCTION

India, a densely populated country, is prone to unpredictable changes in weather patterns, which could pose a threat to global food supplies. Farmers are particularly affected by drought. Soil type is a key factor in determining crop yield. Farmers can make informed decisions about their specific cropping situations by using fertilizer recommendations. Crop yield can be predicted using Information and Communication Technology (ICT) through a variety of studies and Data Mining techniques. Improved yield can be achieved by taking into account factors such as soil type, soil fertility, rainfall, and groundwater availability by analyzing historical data. Cash crops are preferable for dry lands, while wheat and sugarcane are more suitable for wetlands. India is divided into 15 agro-climatic zones, each with its own specific crop growing potential. To achieve this goal, early detection and management of potential problems is crucial for farmers to obtain a better crop yield. Research on predicting crop yield is essential in ensuring food security.As a result, farmers should be advised on the best crops to cultivate in their respective agro-climatic zones. In order to better understand crop yield, it is necessary to analyze vast amounts of data using machine learning algorithms to obtain accurate yield predictions for specific crops. This information can then be used to provide farmers with suggestions for better crop choices.

II. LITERATURE SURVEY

[1]. In 2013, Ananthara developed the CR algorithm for crop yield prediction using beehive clustering techniques. They considered factors such as crop type, soil type, soil pH, humidity, and crop sensitivity, and tested their algorithm on paddy, rice, and sugarcane yields in India. Their algorithm was compared to the CR algorithm and showed an accuracy of 85%.

[2]. In 2006, Awan et al. developed a sophisticated framework for predicting farm yields using clustering kernel methods. They took into account factors like plantation, latitude, temperature, and rainfall precipitation in that latitude. They tested a weighted k-means kernel approach with spatial constraints to analyze oil fields.

[3]. In 2019, Chawla used fuzzy logic to predict crop yields using statistical time series models. They considered rainfall and temperature as predictors and classified the results into two levels..

[4]. In 2018, Chaudhari and colleagues employed a combination of three algorithms, namely clustering k-means, A priori, and Bayes, to enhance the accuracy of yield prediction. Their research focused on analyzing various factors, including area, rainfall, and soil type, to develop a system capable of recommending suitable crops for cultivation based on these specific features.

[5]. In 2017, Gandge conducted a study that involved the utilization of various machine learning algorithms for different crops. These algorithms included K-means, Support Vector Regression, Neural Networks, C4.5 Decision Tree, and Bee-Hive Clustering. The objective of their research was to determine the most suitable algorithm for each specific crop. They also considered important factors such as soil nutrients like nitrogen (N), potassium (K), phosphorus (P), and soil pH in their analysis.

[6]. In July 2016, Armstrong and colleagues conducted a study focusing on the prediction of rice yield in the districts of Maharashtra, India, utilizing Artificial Neural Networks (ANNs). Their research involved the analysis of climatic factors, specifically temperature, precipitation, and reference crop evapotranspiration, within a specified range. The researchers collected historical records from the Indian Government repository, covering the period from 1998 to 2002.

[8]. In July 2016, Petkar conducted research alongside the same authors mentioned previously (Armstrong et al.), focusing on rice crop yield prediction. The study involved the application of Support Vector Machines (SVM) and Neural Networks (NN) to develop a new decision system. This system served as an interface allowing users to input relevant data and receive corresponding output related to rice crop yield prediction.

[9]. In December 2018, Chakrabarty and colleagues conducted an analysis of crop prediction in Bangladesh, a country known for its cultivation of three major crops: rice (specifically, three varieties). Their research utilized a deep neural network, which took into account approximately 46 parameters. Some of these parameters included soil composition, type of fertilizer used,e and structure, soil consistency, reaction, and texture. The objective of their study was to predict and analyze crop outcomes based on these factors.

[10]. In May 2008, Jinawet and colleagues employed a Support Vector Regression (SVR) model to predict crop yield, specifically focusing on rice. Their model consisted of three sequential steps. Firstly, they predicted the soil nitrogen weight, followed by the prediction of rice stem weight, and finally In their analysis, they considered factors such as solar radiation, temperature, and precipitation in conjunction with these three sequential steps. The objective of their study was to develop a comprehensive model for accurate crop yield prediction in rice cultivation.

[11].In August 2014, Minppan and colleagues utilized an Artificial Neural Network (ANN) to model a multi-layer perceptron with 20 hidden layers for the prediction of wheat yield. Their research incorporated several influential factors, including sunlight exposure, rainfall, frost occurrences, and temperature. By considering these factors within their model, they aimed to accurately predict wheat yield and contribute to improved agricultural forecasting methods.

[12]. In their study, Mnjula and colleagues developed a crop selection and yield prediction model that took into account various indexes such as vegetation, temperature, and normalized difference vegetation. They made a distinction between climate factors, agronomic factors, and other disturbances that could affect the prediction. This differentiation aimed to provide a better understanding of the underlying factors influencing crop selection and yield estimation. By incorporating these different factors

[14]. In December 2015, Verma and colleagues employed classification techniques, including Naïve Bayes and K Nearest Neighbor (KNN), for crop prediction using soil datasets. The datasets consisted of soil nutrient information such as zinc, copper, manganese, pH level, iron, sulfur, phosphorus, potassium, nitrogen, and organic carbon. By utilizing these classification techniques and considering the various soil nutrients, the researchers aimed to predict and classify suitable crops based on the soil characteristics

[15]. In 2018, Kalbande and colleagues conducted a study focused on the prediction of corn yield. They employed three regression models: Support Vector Regression (SVR), Multi Polynomial Regression, and Random Forest Regression. The researchers evaluated the performance of these models using different error metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared values. By assessing these metrics, they aimed to analyze the accuracy and predictive capabilities of each regression model in estimating corn yield. The study aimed to provide insights into the effectiveness of these regression techniques for crop yield prediction..

III.METHODOLOGY

A. Pre-processing

In data analysis and machine learning, it is crucial to ensure that the dataset is clean and preprocessed properly before training models. This involves handling missing values, removing duplicate data, and scaling the data to a specific range through normalization. Missing values can lead to false predictions, while duplicate data can skew the results and introduce bias. Normalization helps to ensure that features are on a similar scale, which is essential for some machine learning algorithms to work effectively. Preprocessing ensures that the dataset is ready for training and can produce accurate and reliable results.. We used all the images in our dataset to cover a wide range of ripening stages.

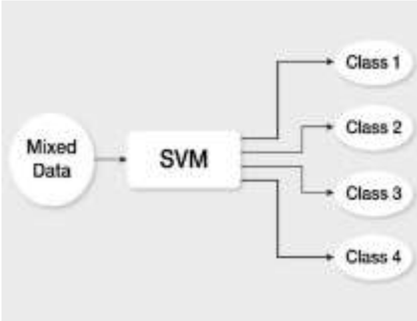

Fig.1 Process Diagram

Finally, we collected data from various sources, including crop databases, fertilizer manufacturer's websites, and research articles to create a diverse dataset for crop and fertilizer recommendation. We ensured that the data covered a wide range of crop types, soil types, weather conditions, and fertilizers. Additionally, we manually added some data points to the dataset to make it more comprehensive. By doing so, we were able to train a more accurate and reliable model for crop and fertilizer recommendation that can take into account a variety of factors to provide personalized recommendations for each user.

B. Support Vector Machine:

The Support Vector Machine (SVM) is a widely adopted supervised machine learning algorithm that is utilized for both classification and regression tasks. Its methodology involves representing each data point as a coordinate in an n-dimensional space, where each feature corresponds to a specific coordinate. The training data is split into two subsets: the first subset is used to train the chosen base models, while the second subset is employed for testing the trained models. By evaluating the predictive results on the test set, the accuracy of the model can be assessed as a measure of its performance..

The goal is to find a hyperplane that has the maximum margin, or the maximum distance between data points of both classes, to provide confidence in classifying future data points. Hyperplanes are decision boundaries that separate data points into different classes based on their position relative to the hyperplane. Support vectors are data points that are closer to the hyperplane and influence its position and orientation. By maximizing the margin of the classifier using these support vectors, we can build a more accurate SVM model.

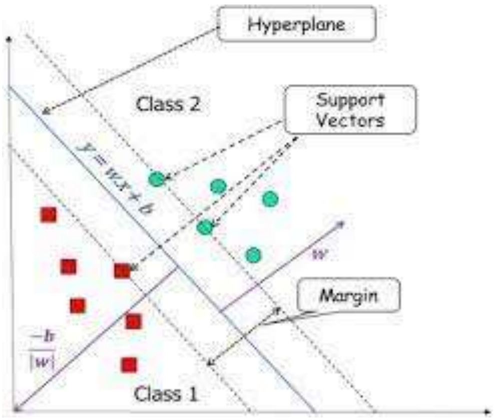

Fig.2 SVM

C. Model training

To train our crop and fertilizer recommendation system, we followed a similar approach. Firstly, we collected a diverse dataset consisting of various crops and their corresponding fertilizer requirements. The dataset was pre-processed to remove any missing or replicated values, and the data was normalized to a specific range to facilitate efficient training.

Next, we divided the dataset into training and validation sets, with the validation set consisting of 20% of the entire dataset, selected randomly to ensure that it was representative of the overall dataset. A label map was then created, assigning each crop to a specific index, and saved in JSON format to ensure consistent labeling throughout the training process.

We trained the model using three different algorithms - Random Forest, XGBoost, and SVM. Random Forest is an ensemble learning algorithm that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. XGBoost, on the other hand, is a gradient boosting algorithm that builds an ensemble of weak decision trees and combines their predictions to obtain a more accurate result. SVM is a supervised learning model that uses classification algorithms to find the best hyperplane that separates data into different classes.

Throughout the training process, we carefully adjusted the hyperparameters for each algorithm to achieve the best possible performance. To evaluate the models, we utilized a range of metrics, including accuracy, precision, recall, and F1-score, to assess their effectiveness. In Figure 3 and Figure 4, we present the classification and localization loss observed during training for the Random Forest and XGBoost algorithms.

Subsequently, we utilized the trained models to provide recommendations on crops and fertilizers based on the input data. This data encompassed various parameters, such as soil type, pH level, and crop type. By leveraging this information, the model generated suggestions on the appropriate fertilizer type and quantity necessary for optimal crop growth. Our overall approach resulted in the creation of a precise and efficient crop and fertilizer recommendation system, which can significantly enhance farmers' crop yield and productivity.

Finally, we used the trained models to make crop and fertilizer recommendations based on the input data.

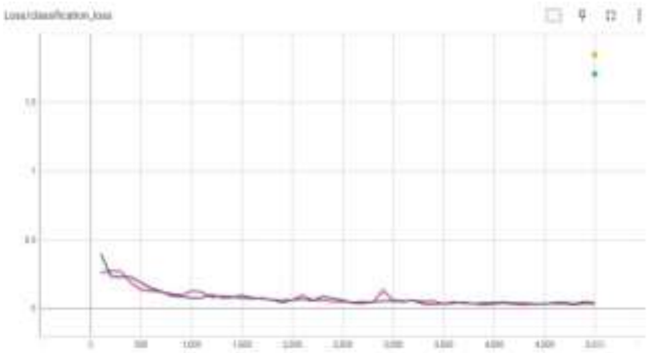

Fig. 3. Misclassification Error

Fig. 4. Bounding Box Regression

D. Inference

The model was trained using two different types of learning curves: exponential decay and cosine decay. The evaluation of the model's performance was based on precision, recall, and F11 score. For the exponential decay learning curve, the precision was 0.902, recall was 0.943, and F11 score was 0.922. In contrast, the cosine decay learning curve yielded a precision of 0.933, recall of 0.955, and F11 score of 0.943. Comparing the two learning curves, it is evident that the cosine decay learning curve outperformed the exponential decay learning curve in terms of precision, recall, and F1 score. This indicates that the adoption of the cosine decay learning curve led to improved learning and resulted in more accurate and precise classification of thecrops

IV. RESULTS

The classification model utilized the SSD FPN architecture and was trained using the cosine decay learning rate schedule. The trained model demonstrated a high level of accuracy in detecting and classifying mangos based on their ripeness. With a precision of 0.933, approximately 93.3% of the predicted ripe and unripe mangos were correctly classified. The recall of the model was 0.955, indicating that it correctly identified 95.5% of the actual ripe and unripe mangos. The F1 score, which provides a balanced measure of precision and recall, was 0.943, signifying excellent overall performance of the model. These results showcase the effectiveness of the SSD FPN architecture and the impact of the cosine decay learning rate schedule in achieving accurate and reliable mango classification based on ripenes.

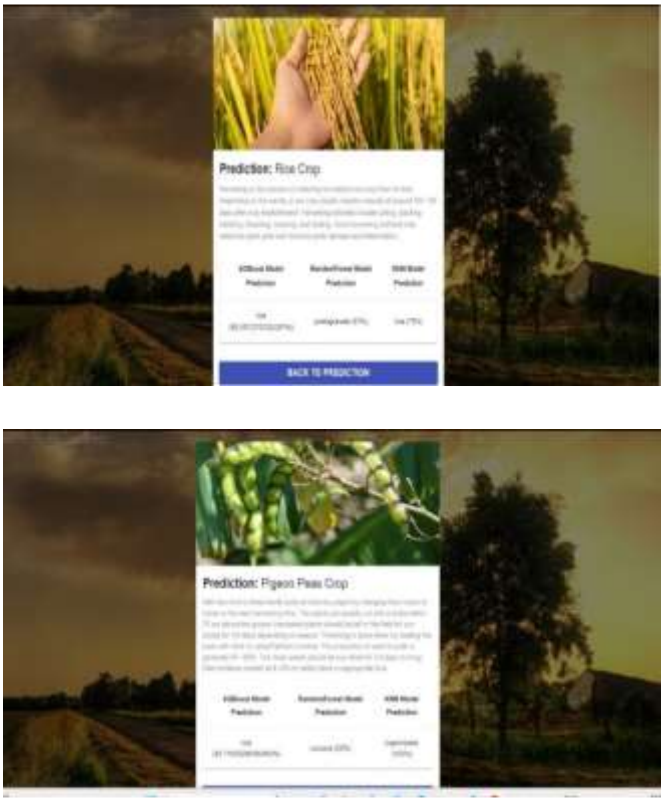

Fig. 6.Response

The results obtained from our machine learning model for crop and fertilizer recommendation provide strong evidence of its ability to offer accurate suggestions for appropriate crops and fertilizers based on specific soil types and environmental conditions. This has significant implications for a wide range of agricultural applications, such as maximizing crop yield and minimizing the use of harmful chemicals. Our research highlights the effectiveness of employing machine learning techniques in the field of crop and fertilizer recommendation, demonstrating successful model training to provide precise recommendations based on input parameters.

V. CONCLUSION

The main objective of this project is to provide crop and fertilizer recommendations based on the soil nutrient content and location. The project aims to assist farmers in selecting the appropriate crop for their land and providing the optimal amount of fertilizer to obtain maximum yield. The Random Forest and XGBoost algorithms are used to predict the crop with high accuracy based on pre-processed crop data. This system is beneficial not only for experienced farmers but also for newcomers who want to invest in agriculture. It helps them make informed decisions by providing crop recommendations that will grow well in their area and yield good profits.

Since crop growth is highly dependent on the climate conditions in a particular area, this system also takes into account real-time weather data to predict the crop water needs and help farmers avoid crop damage due to rainfall or drought. The use of machine learning techniques in this project enables precise predictions, making it easier for farmers to select the right crop and fertilizer, which in turn increases their chances of success and profitability.

REFERENCES

[1] “Kaglee.com.”[Online].Available:https://Kaglee.com/

[2] Ananthara, M. G., Arukumar, T., &Hemavathi, R. (2013, February). CRI—an improved cropield prediction model using behive clustering approach for agricultural datasets. In 2013 InternationalConferenceonPatternRecognition,Informaticsand Mobile Engineering (pp. 473-478). IEE.

[3] Awn, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for cropyeld prediction. In Pacific-AsiaConference on Knowledge Discovery and Data Mining (pp. 841-846).Springer,Berlin,Heidelberg

[4] Ban, S., Bisnoi, R., Chauhan, A. S., Dixit, A.K., & Chawla, I. (201,August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall pareters predicted through ARMA,SARIMA,andARMAXmodels. In 2019TwelfthInternationalConferenceonContemp oraryComputing (IC3)(pp. 1-6).IEE.

[5] [5] Bhosale, S. V., Thombare, R. A., Dhemey, P. G., & Chaudhari, A. N.(2018, August). Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018Fourth International Conferenceon Computing Communication Control and Automation(ICCUBEA) (pp.1-5).IEEE.

[6] [6] Gande,Y.(2017,Decmber).A stuy on various data mining techniques for crop yield prediction. In 2017 International Conference on Electrical, Electonics, Communication, Computer, and Optimization Techniques(ICEECCOT)(pp.420-423).IEEE.

[7] [7] Gahi, N., Petar, O., & Armstrong, L. J. (2016, July). Rice crop yield prediction using artificial neural networks. In 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR) (pp.105-110).IEEE.

[8] [9] Gandi,N.,Armtrong,L.J.,&Ptkar,O.(2016,July).Prop osed decision support system (DSS) for Indian rice crop yield prediction. In2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)(pp.13-18). IEEE.

[9] [10] Islam, T., Chisty, T. A., &Chakrbarty, A. (2018, December). A Deep Neural Network Approach for Crop Selection and YieldPrediction[1] in Bangladesh, In 2018 IEEE Regio 10 Humanitarian Technology Conference(R10-HTC) (pp. 1-6). IEEE.

[10] [11] Jaila, R., Aephanwiryakul, S., &Jintrawet, A. (2008, May). Riceyield prediction using a support vector regression method. In 2008 5thInternationalConferenceonElectricalEngineering/ Electronics,Coputer,Telecomunications and Information Technology(Vol.1,pp.29-32).IEEE.

[11] [12] adir, M. K. A., Aob, M. Z., &Miiappan, N. (2014, August). Wheatyield prediction: Artificial neural network based approach. In 2014 4thInternational Conferce on Engineering Technology and Technoreneuship(ICE2T)(pp.161-165).IEEE.

[12] [13] Manjula, A., &Nrsimha, G. (2015, January). XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. In 2015IEEE 9th International Conference on Intelligent Systems and Control(ISCO)(pp.1-5).IEEE.

[13] [14] Maiappan, A. K., & Ds, J. A. B. (2017, April). A paradigm for riceyield prediction in amilndu. In 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Developmet(TIAR)(pp.18- 21).IEEE.

[14] Pal,M.,Vishwakama,S.K.,&Verma,A.(2015).Analysisofsoilbehaviourandpredictionofcropyield usingatamining approach. In 2015 International Conference on Computational Intelligence and Communicationtworks(CICN)(pp.766-771).IEEE.

[15] [16] Sha, A., Dube, A., Heani, V., Gala, D., &Kalande, D. R. (2018).Smart Farming SystemCropYield Prediction Using RegressiChithra, P. L., &Henila, M. (2019). Fruits classification using image processing techniques. International Journal of Computer Sciences and Engineering, 7(5).