Crop Recommendation System Using Machine Learning Classifiers

Md. Taukir Ahmed*, Md. Showrov Hossen†, Faria Rahman Annasha†,
Mst. Shikha Moni†, and Md. Al Amin Mridha†
*Dept of Electrical and Electronic Engineering, Varendra University, Rajshahi, Bangladesh
†Dept of Computer Science and Engineering, City University, Dhaka, Bangladesh
Email: taukirahmed.vu@gmail.com, sourovhossen96@gmail.com, rahmanfaria98@gmail.com,
shikha.moni.43.23@gmail.com, mmalamin551@gmail.com

***Abstract*— A major contributor to the nation's economic development and growth is agriculture. Farmer's poor crop selection is the main and most significant factor affecting agricultural productivity. A crop recommendation system that employs the technique of machine learning is to be created in order to increase crop productivity. By using the predictions of several machine learning models, the ensemble technique creates a model that can accurately select the best crop based on the kind and characteristics of the soil. Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Bagging, K Nearest Neighbor (KNN), Multilayer Perceptron, XGBoost, LightGBM, CatBoost, Naive Bayes, and SVM are the independent base learners that are employed in the ensemble model. With a respectable level of accuracy, each classifier produces a unique set of class labels. Last but not least, the Random Forest algorithm is quicker and more precise in this area. The suggested system makes use of a number of characteristics, such as soil composition and climatic data, to precisely forecast which crops will be best suited for a certain area. This technology has the potential to transform crop recommendation, improving crop yields, sustainability, and overall profitability for farmers of all sizes. By training and testing models with different configurations of machine learning algorithms, we have achieved near-perfect accuracy through exhaustive examination of a vast historical data set. Across all models, we routinely show accuracy above 95%, with the best accuracy of 99.2%.**

Keywords—crop recommendation, crop yields, machine learning, ensemble classifier, XGBoost, LightGBM, CatBoost, random forest

I. INTRODUCTION

Numerous input characteristics serve as the foundation for the agricultural crop recommendation system. This proposes a hybrid model for crop suggestions to south Indian states by considering several aspects, such as soil type, rainfall, groundwater level, temperature, fertilizers, pesticides, and season. The hybrid recommender model is developed using the classifier machine learning technique. The program will recommend the crop depending on the pertinent characteristics. A technology-based crop recommendation system for agriculture helps farmers increase crop yields by recommending a crop that is suitable for their region based on climatic and geographic parameters. The recommend model is found to be effective in recommending a suitable crop [1][9]. Updating crop yield production values has beneficial practical implications for directing agricultural output and informing farmers of changes in crop market prices. The goal of this research is to use the crop selection strategy to assist farmers and agriculture in resolving various problems. Crop recommendation systems can assess a variety of data, such as weather, soil, and market data. This information can be used to train machine learning algorithms that predict which crops will thrive in a given region [2]. Crop recommendation systems can also teach farmers the optimal practices for growing specific crops. The development of crop recommendation systems based on machine learning has the potential to increase agricultural sustainability and productivity. Additionally, crop suggestion systems can improve agriculture's climate change adaptation [3][8]. Furthermore, machine learning can assist with various other agricultural issues [4], including crop yield forecasting, disease and pest identification, crop yield maximization, water efficiency, fertilizer and pesticide consumption reduction, soil management, etc. Crops are essential to the world's population because they provide both food and fiber. By 2050, the World Resource Institute wants to figure out how to feed 10 billion people in a sustainable manner. Therefore, increasing the production of high-quality crops is essential [5]. The crops that are planted can have a significant impact on crop yields and profitability.

II. LITERATURE REVIEW

The development of a crop recommendation system that makes use of machine learning's ensemble technique might increase crop productivity. Based on the specific type and characteristics of the soil, Kulkarni et al. [1] created a model that uses the ensemble technique to integrate the predictions of several machine learning models to select the best crop with high accuracy. The ensemble model employs Random Forest, Naive Bayes, and Linear SVM as independent base learners. For each classifier, an acceptable accuracy set of class labels is provided. The majority vote process is used to merge the labels of each individual base learner class.

To help Indian farmers choose the optimum crop to grow based on the farm location, the season, soil properties, and meteorological conditions like temperature and rainfall, Doshi et al. [2] devised an intelligent system called AgroConsultant. Priyadharshini A et al. [3] suggested a method to help farmers choose crops by taking into account a number of variables, such as soil type, planting season, and geographic location. Additionally, developing nations are increasingly embracing precision agriculture, which stresses site-specific crop management, and sophisticated agricultural technologies.

Another study by Bandara et al. [4] includes a theoretical and conceptual platform for a recommendation system. In order to recommend a crop with high accuracy and efficiency for the selected land with site-specific parameters, this system uses integrated models to gather environmental factors using Arduino microcontrollers, machine learning techniques like support vector machines (SVM) and naïve bayes (multinomial), unsupervised machine learning algorithms like K-Means clustering, and natural language processing (sentiment analysis).
Rajak et al. [6] developed a recommendation system that uses the data collected from the soil testing lab to perform an ensemble model using a majority voting technique, employing support vector machines (SVM) and artificial neural networks (ANN) as learners, in order to recommend a crop for a site-specific parameter with a high degree of efficiency and accuracy. Reddy and associates [7] produced This study examines many machines learning (ML) methods, including SVM, ANN, RANDOM TREE, and NB classifier, that are used to estimate crop productivity. It also offers a thorough evaluation of the methods' accuracy.

In [10], the harvests were predicted using a neural network, which had an accuracy of 89.88%. Despite the fact that crop rotation has not been thoroughly investigated in this study, this report forecasts suitable crops. The study has suggested a way to help farmers choose crops by considering all the variables, including soil type, sowing season, and geographic location.

Similarly, the temperature, moisture content, and humidity of the soil were identified as key variables in a Naive Bayes algorithm for crop prediction [11]. This study uses machine learning (ML), one of the most cutting-edge technologies in crop prediction, to give new farmers a method that helps them plant healthy crops. Furthermore, a supervised learning method called Naive Bayes provides direction on how to proceed. The forecast accuracy of these models needs to be increased. To assess the different soil properties and recommend the crop for production, another machine learning technique was proposed [12]. Even though they used k-nearest neighbor (KNN) algorithms, they only used the soil's properties to make predictions.

The Random Forest (RF) method was used to forecast crop yields in the agriculture industry [13]. The best crop production model is produced by the RF approach by considering as few models as possible. The results imply that crop output forecasting is advantageous to the agriculture sector. A winter wheat prediction model was developed by estimating the parameters of the soil using online soil spectroscopy and a prototype sensor [14].

The model comprised counter-propagation artificial neural networks with a self-organizing map, XY-fused networks, and supervised Kohonen networks. Researching soil-related factors alone will not be enough to maximize crop output, despite the method's useful data. Crop forecast depends on several factors, which makes feature selection crucial. Using soil properties and environmental data, including rainfall, season, texture, and temperature, a comparative study of several feature selection techniques was conducted to forecast crops using different classifiers [15].

III. METHODOLOGY

The combination of agricultural data and machine learning will transform farmers' knowledge and approach optimization. With the increasing amount of data coming from sources like weather stations, satellites, sensors, and agricultural equipment, machine learning algorithms can now analyze massive volumes of data and extract useful information. Intricate relationships, patterns, and predictive models that were previously hidden by the data can be uncovered using these algorithms. By combining machine learning techniques with agricultural data, farmers can make data-driven decisions on anything from yield prediction and pest control to crop selection and irrigation management. This integration enables sustainable practices, resource management, and increased efficiency, ultimately increasing agricultural profitability and output. The proposed methodology flowchart that followed by experiments outlined in Fig. 1.

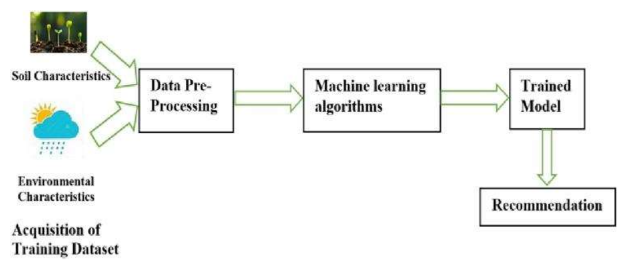

Fig. 1. Machine Learning based crop recommendation system.

For this experiment, we used the Crop Recommendation dataset. The open-source Kaggle dataset served as the source of our dataset. This dataset consists of 8 columns and 2200 rows. The dataset should contain variables such as soil properties (pH, nutrient levels), weather conditions (temperature, precipitation, humidity), historical crop yields, pest occurrences, and geographic information.

The dataset includes the soil-specific parameters that were gathered using Kaggle. Comparable internet resources for general agriculture knowledge were also looked at. Rice, maize, chickpeas, kidney beans, pigeon peas, moth beans, mung beans, black grams, lentils, pomegranates, bananas, mangoes, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee are among the crops that are taken into account by our model. This is how the dataset is examined. This shows how many of each crop there are in the training dataset. Rainfall, pH, temperature, humidity, nitrogen (N), potassium (K), and phosphorus (P) were all taken into account. The above-mentioned soil properties have a significant impact on a crop's capacity to draw water and nutrients from the soil. For crops to flourish as much as they can, the soil must provide a suitable environment. The earth is the anchor for the roots.

Nitrogen has an important role in plant leaf growth. Root growth, fruit and flower development, and other processes are largely attributed to phosphorus. One nutrient that supports the plant's overall health is potassium. One important element in the growth and development of plants is temperature. Plant development and, ultimately, crop productivity are influenced by the light, carbon dioxide, water, nutrients, and humidity in the air.

Humidity has an indirect effect on photosynthesis, pollination, leaf growth, disease incidence, and ultimately economic production, while having a direct effect on a plant's water relations. The availability of soil nutrients is significantly impacted by the soil's acidity or alkalinity (Ph). PH can have an impact on the activity of the current bacteria as well as the quantity of exchangeable aluminum in the soil. Rainfall also affects how quickly a crop grows from seed and when it will be ready for harvest. Both the time it takes for seeds to germinate and the period between planting and harvest can be shortened by promoting faster plant growth through the use of appropriate irrigation and rainfall. To make categorical variables (such crop or soil type) easier to employ in machine learning algorithms, convert them into numerical representations using methods like label encoding or one-hot encoding. To convert these categorical data into numerical values, we employed the label encoding technique. Utilize statistical methods like correlation analysis to identify significant relationships between variables and the target variable (crop yield or suitability). The correlation heatmap of the characteristics in the CRS dataset is shown in Fig. 2.

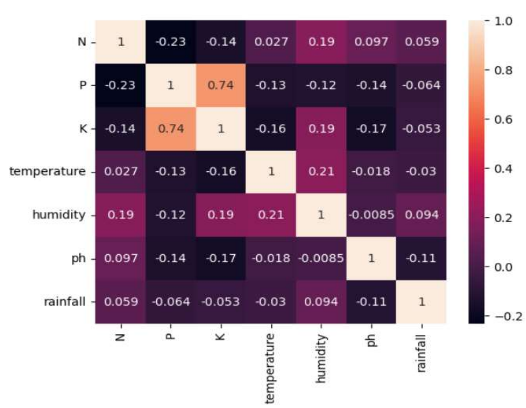

Fig. 2. Correlation heat map of the features of CRS dataset

In Fig.3, showed the Nitrogen, Phosphorous, Potassium values comparison among the different crops.

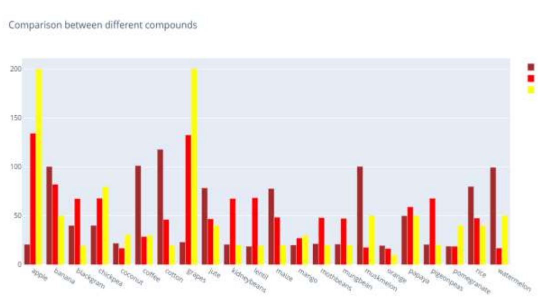

Fig. 3. Correlation of Nitrogen, Phosphorous, Potassium values among crops

In Fig.4, showed the Nitrogen, Phosphorous, Potassium ratio for rice, cotton, jute, maize and lentil.

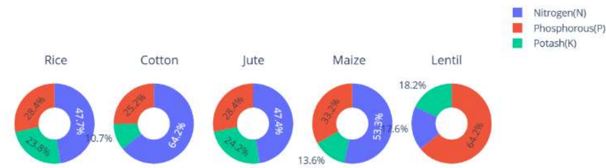

Fig. 4. Correlation of Nitrogen, Phosphorous, Potassium ratio among rice, cotton, jute, maize and lentil

In Fig.5, showed the Nitrogen, Phosphorous, Potassium ratio for apple, banana, grapes, orange, mango, coconut, papaya, pomegranate watermelon and muskmelon.

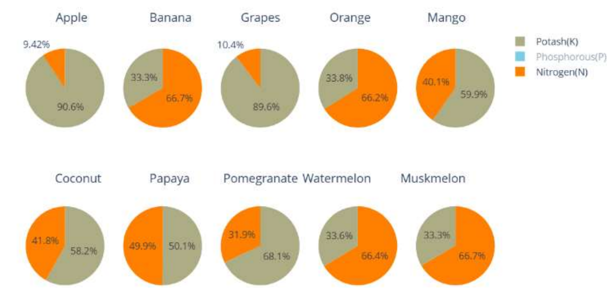

Fig. 5. Correlation of Nitrogen, Phosphorous, Potassium ratio among fruits

In Fig.6, showed the correlation of rainfall, temperature and humidity values among different crops to represent the impact of rainfall, temperature and humidity on crops production.

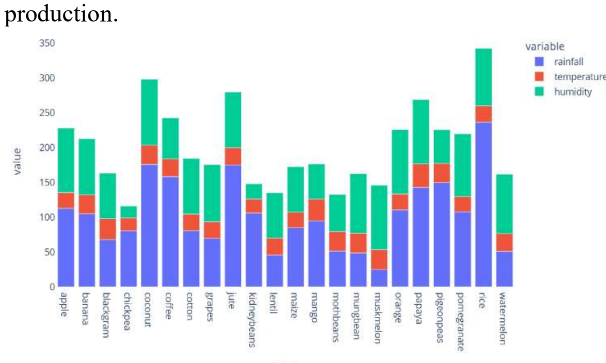

Fig. 6. Correlation of rainfall, temperature and humidity values among crops

IV. RESULTS AND DISCUSSION

This section presents the classifier's performance evaluation. The most often used performance metrics for any classification task are accuracy, precision, and recall. Using the following formula, we determined the accuracy, precision, and recall values to gauge the performance of the machine learning classifiers.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

where, TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative.

Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Gradient Boosting, AdaBoost, Bagging, Support Vector Machine (SVM), K Nearest Neighbor (KNN), Naïve Bayes, Multilayer Perceptron, XGBoost, LightGBM, and CatBoost are some of the machine learning classifiers that we used for the crop classification. To choose the optimal model for the CRS dataset, it is essential to compare the accuracy of various classifiers. Table I displays the classification reports of various classifiers used for crop classification or recommendation.

TABLE I. COMPARISON OF THE CLASSIFICATION REPORTS

Classifiers	Precision	Recall	Accuracy (%)
Logistic Regression (LR)	0.947730	0.946247	94.69
Decision Tree (DT)	0.988045	0.988159	98.78
Random Forest (RF)	0.994172	0.991883	99.24
Gradient Boosting	0.984426	0.983381	98.33
AdaBoost	0.156670	0.227273	21.06
Bagging	0.990263	0.987299	98.78
Support Vector Machine (SVM)	0.96774	0.964396	96.36
K-Nearest Neighbor (KNN)	0.979634	0.977677	97.72
Naïve Bayes	0.863305	0.866754	85.75
Multilayer Perceptron	0.960283	0.956402	95.75
XGBoost	0.983368	0.979940	98.03
LightGBM	0.978906	0.976360	97.57
CatBoost	0.991498	0.989031	98.93

Table I shows that the Random Forest classifier outperforms the other classifiers and offers a considerably higher accuracy of 99.24%. Decision Tree, Gradient Boosting, Bagging, XGBoost, and CatBoost are some more ensemble classifiers that also offer 98% accuracy. The Python Jupyter Notebook environment was used for the entire experiment.

V. CONCLUSION

Crop recommendation models that predict which crops would flourish using machine learning algorithms have been introduced by this research. The approach is very flexible enough to accommodate new data and various nations or regions. The results of the study have several positive implications for the agriculture industry. In the first place, growers can use the method to choose crops more wisely. Second, the approach can be used by governments to develop policies that support the agricultural sector. Third, businesses can utilize the technology to create new products and services that help the agriculture industry; fourth, it will help keep agricultural product prices steady. The challenges facing agriculture and some exciting future prospects were then discussed in detail. All things considered the field of agriculture has greatly benefited from this study. Farmers, governments, and businesses can all benefit from the method's accuracy, scalability, and ease of use. Several machine learning classifiers were employed in this study to select crops, and it was shown that Random Forest classifiers outperformed the others in terms of accuracy, with an accuracy of 99.24%. An end-to-end system for end users (farmers or agribusiness owners) can be constructed on top of this work as a mobile application, which is what we plan to create in the future.

REFERENCES

[1] Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). Improving crop productivity through a crop recommendation system using ensembling technique. In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) (pp. 114-119). IEEE.

[2] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE.

[3] Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021, April). Intelligent crop recommendation system using machine learning. In 2021 5th international conference on computing methodologies and communication (ICCMC) (pp. 843-848). IEEE.

[4] Bandara, P., Weerasooriya, T., Ruchirawya, T., Nanayakkara, W., Dimantha, M., & Pabasara, M. (2020). Crop recommendation system. International Journal of Computer Applications, 975, 8887.

[5] Rajak, R. K., Pawar, A., Pendke, M., Shinde, P., Rathod, S., & Devare, A. (2017). Crop recommendation system to maximize crop yield using machine learning technique. International Research Journal of Engineering and Technology, 4(12), 950-953.

[6] Dahiphale, D., Shinde, P., Patil, K., & Dahiphale, V. (2023). Smart Farming: Crop Recommendation using Machine Learning with Challenges and Future Ideas.

[7] Rushika G., Juilee K, Pooja M, Sachee N, and Priya R.L.(2018). Prediction of Crop Yield using Machine Learning, Issue 02 IRJET (pg 2337-2339).

[8] Ruchita T, Shreya B, Prasanna D, and Anagha C (2017). Crop Yield Prediction using Big Data Analytics, Volume 6, Issue 11, IJCMS.

[9] Monali P, Santosh K, Vishwakarma, and Ashok V (2015). Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach, ICCICN 2015.

[10] Mrs. N. Hemageetha, and Dr. G.M. Nasira (2016). Analysis of Soil Condition based on pH value using Classification Techniques, IOSR-JCE, Issue 6, (pp.50-54).

[11] E. Manjula, and S. Djodiltachoumy (2017). Data Mining Technique to Analyze Soil Nutrients based on Hybrid Classification. IJARCS.

[12] Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare (2017). Crop Recommendation System to maximize Crop yield using Machine Learning. Issue 12 IRJET.

[13] S. Veenadhari, Dr. Bharat Misra, Dr. CD singh (2014). Machine Learning Approach for Forecasting Crop Yield based on Climatic Parameters. ICCCI-2014

[14] Sadia A, Abu Talha K, Mahrin Mahia, Wasit A, and Rashedur M.R.(2018). Analysis of Soil Properties and Climatic Data To Predict Crop Yields and Cluster Different Agricultural Regions of Bangladesh, IEEE ICIS 2018 (pp.80-85).

[15] Arun K, Navin K, and Vishal V (2018). Efficient Crop Yield Prediction using Machine Learning, (pp.3151-3159), IRJET.