RESEARCH ARTICLE	OPEN ACCESS

RANDOM FOREST ALGORITHM USE FOR CROP RECOMMENDATION

Pradip Mukundrao Paithane*[1]

[1] Head and Assistant Professor, AIDS department, VPKBIET Baramati, Pune, Maharashtra, India.

[1] http://orcid.org/0000-0002-4473-7544

Email: *paithanepradip@gmail.com

ARTICLE INFO	ABSTRACT

The proposed method seeks to assist Indian pleasant in selecting the optimum crop to produce based on the characteristics of the soil as well as external factors like temperature and rainfall by using an intelligent system called Crop Recommender. The Indian economy is significantly impacted by the agricultural sector. Whether publicly or covertly, the bulk of Indians are relying on agriculture for their living. As a result, it is undeniable that agriculture is significant to the country. The majority of Indian farmers believe that they should trust their intuition when deciding on a crop to grow in a particular season or they simply employ the methods they have been doing from the beginning of time. They are more at ease just adhering to conventional agricultural practices and standards than truly appreciating how crop yield is influenced by the present weather and soil conditions. The farmer can unintentionally lose money if he makes one bad decision, which would hurt both him and the surrounding agricultural industry. As the agriculture business is the foundation of the entire lateral system. Using the machine learning algorithm, this problem can be resolved. A crucial perspective for identifying a practical and workable solution to the crop production issue is machine learning (ML). Machine learning (ML) may predict a target or outcome from a set of predictors using supervised learning. A recommendation system is implemented using decision trees. The major goals of this system are to provide farmers with recommendations regarding the best crops to sow based on their soil and local rainfall patterns. We have employed the Random Forest Machine Learning technique to forecast the crop. Crop prediction is assessing the crop based on historical data from the past that includes elements like temperature, humidity, ph, and rainfall. It gives us a broad picture of the best crop that can be raised in light of the current field weather conditions. These predictions can be made by Random Forest, a machine learning technique. The highest level of accuracy, up to 90%, will be possible for crop predictions. The random forest algorithm achieved the accuracy about 99.03%.

I. INTRODUCTION

India has a lengthy agricultural history. In terms of farm output, India is currently ranked second worldwide. Nearly half of all jobs were in closely related industries to agriculture, such as forestry and fishing, and India's agricultural sector is no longer contributing significantly to GDP [1]. Predicting the best harvest is agriculture's principal source of income. Numerous factors, including meteorological, geographic, biological, and economic ones, have an impact on crop productivity [2]. Farmers find it challenging to choose when and what crops to plant due to shifting market pricing. The previous ten years. Farmers are confused about which crop to plant, when to start, and where to plant it because the weather is unpredictable. This may also be the cause of the farmer suicides. In this circumstance, the pace of crop output is steadily declining [3]. The problem can be resolved by giving the farmers access to a smart, user-friendly recommender system.

We offer a paradigm in this study that overcomes these difficulties. The recommended technique is unique in that it

teaches farmers how to choose the best crop for their soil system as well as the weather conditions in that place [4]. It suggests the best lucrative crop for a certain location. Crop selection is based on economic and environmental aspects, with the goal of reducing crop seed loss, efforts to take them, and components given to them such as water and fertilizers. Crop projections are made using a variety of variables such as "rainfall", "temperature", "area", "soil type", and so on. The method aids in determining the best time to apply fertilizer. The present crop production prediction system is hardware-based, costly to maintain, and complicated to utilize [5].

I.1 KEY CONTRIBUTIONS

1) Error rate and accuracy comparisons for crop prediction for specific regions using various machine learning approaches.
2) A simple web application that any user (including farmers) can use to access a user-friendly web application that recommends the most lucrative crop.
3) A GPS-based location identifier for retrieving rainfall and weather data estimates in a specific area.

Weather forecasting has become extremely difficult as a result of global warming and increased pollution. We use our traditional ways for crop selection because we have been farming for so long. We determine which crop to take based solely on our assumptions, without employing any sophisticated methodologies. These conventional systems rely solely on global weather, but because forecasting weather is difficult, the results can be disastrous for farmers. That is why, when making decisions such as which crop to plant, there should be a smart system that will tell us which crop will produce the best results depending on our soil as well as weather-based observations such as temperature, rainfall, and ph. The system will employ an efficient algorithm to make the best decision about the main crop. There are also dynamic parameters for soil type and weather conditions [6]. Because of its accuracy, robustness, interpretability, scalability, and ability to manage missing data, the Random Forest algorithm is an excellent foundation for a crop recommendation system. Because of these characteristics, it is a popular candidate for machine learning-based crop recommendation applications [7].The purpose is to anticipate the most suited crop(s) to be grown on a specific farm or agricultural location given a set of input factors such as soil type, climate conditions, crop traits, and historical yield data. The Random Forest algorithm is used by the recommendation system to create these predictions [8]. The system's primary goal is to provide crop recommendations based on input factors such as soil type, climate conditions, crop traits, and historical yield data. The system's goal is to recommend the best crop(s) to grow in a certain agricultural region or farm. The system is strongly reliant on the availability and accuracy of historical crop production data, soil data, climate data, and other pertinent aspects. The accuracy and dependability of suggestions might be impacted by limited or incomplete data [9].

The user enters the region and soil type as input. Machine learning algorithms can be used to determine the most profitable crop list or to estimate crop yield for a crop chosen by the user. Machine Learning algorithms such as "Support Vector Machine (SVM)", "Artificial Neural Network (ANN)", "Random Forest (RF)", "Multivariate Linear Regression (MLR)", and "K-Nearest Neighbor (KNN)" are used to forecast crop productivity. The unpredictable nature of the environment makes it difficult for farmers to decide which crop to grow, when to plant it, and where to begin. Due to changes in seasonal weather patterns and important resources like "soil", "water", and "air", the use of various fertilizers is also unclear. Crop yields in this situation are steadily declining. As a result of study, a ground-breaking system for crop suggestion that addresses farmers' challenges has been developed. The fundamental goal of our suggested approach is to aid farmers in maximizing agricultural productivity and choosing the most profitable crops suitable for their individual regions [10].

The most significant promise of block-chain for the agricultural sector is that it will do away with the need for third parties to guarantee trust in buyer-seller relationships or other source-destination links. Blockchain technology enables peer-to-peer transactions, which do away with the need for middlemen. Peer-to-peer transactions are made possible by blockchain, which also makes it possible to create "smart contracts" that carry out the terms of any agreement when specific conditions are met. When something of value is exchanged, whether it is real commodities, services, or money, the transaction can be documented, providing a very long history of the product or exchange from its origin to its destination. Blockchain technology could be quite handy in this situation. Putting all data linked to agricultural happenings on a blockchain allows for the creation of a dependable and transparent system. Farmers also have rapid access to information on a variety of areas, such as seed quality, weather and environment, payments, soil moisture, demand, and sale price [11].

Stable agricultural growth in India has raised questions. Using data on paddy yield, area, and production from the years 1970–1971 to 2011–2012, an analysis of 41 years is conducted to better understand the problem of instability in India's rice production. The research revealed that while the acreage, output, and yield of rice had positive compound annual growth rates over all of India, they had been steadily declining over time. There has been an increase in precariousness at the national level in India's regions, production, and rice yield over the past ten years (2000-01 to 2011-12). The rise in instability may have been caused by a decline in the usage of fertilizer, seeds, and other agricultural inputs as well as a low ratio of irrigated land to total cropland. The wholesale price of paddy has fluctuated considerably between states during the reform, from 1990–1991 to 2016–17, whereas the price of paddy harvested on farms has been less erratic. Although there has been a lot of research on agricultural sector instability, this paper intends to explicitly look into the topic of instability in India's rice output. Over 10% of India's entire agricultural production value is made up of paddy rice, with China being the world's top producer and India coming in second. Over 16 states' worth of farmers harvest rice, this is a basic crop for about 60% of the Indian population [12].

A key viewpoint for securing a real-world and practical solution to the crop yield problem is artificial intelligence (AI). By using directed learning, machine learning (ML) may predict an objective or result from a set of indicators. A good function must be created by a group of variables that will map the input variable to the intended output in order to get the desired results. Crop yield prediction includes predicting a crop's yield based on historical information such as temperature, humidity, pH, rainfall, and the crop's name. It provides information about the best crop that may be expected to be grown in a field [13]. These predictions can be made using the machine learning method Random Forest. The crop prediction will be as precise as feasible. The ideal crop yield model is found using the random forest approach by considering the fewest number of models. Predicting

crop yield is extremely useful in agriculture [14]. The suggested approach functions as an informed and sophisticated tool for farmers, taking into account a number of important elements like soil quality, weather forecasts, and yield. The method improves precision, allowing farmers to maximize crop yield and eventually boost earnings. The use of accurate data is necessary to achieve increased precision. The suggested system analyses all available data using data mining techniques and delivers accurate harvest yield projections. With the aid of this forecast, farmers are better equipped to understand their unique needs and make wise decisions [15].

I.2 RESEARCH CONTRIBUTION

1. In this research work, in detail comparison has been carried out on various machine learning algorithm.
2. The Random Forest algorithm provides higher result as compare to other machine learning approaches
3. The proposed random forest algorithm works on the basis of various variable parameters like rainfall, temperature, area, soil type, and various soil parameters.
4. The proposed random forest algorithm, predict the crop on the basis of parameter used in dataset.
5. The accuracy of random forest algorithm is about 99.09%, and which is higher than Decision tree with accuracy 90.00%, Support Vector Machine with accuracy 97.90% and Logistic Regression with accuracy 95.22%.

II. METHODOLOGY

For farmers, crop production is a crucial piece of information. Knowing the yield that lowers loss is really helpful. Farmers with experience used to predict the yield. The way the suggested system functions is likewise similar. It makes use of the historical data to predict the future yield. Crop productivity is significantly impacted by both weather and pesticide use. It is required for the accuracy of the data used to make this prediction. As a result, the proposed technique anticipates yield and minimizes losses.

Given data sets from the chosen region, the suggested model forecasts the crop. Integrating ML and agriculture will lead to significant industry improvements.

For forecasting current performance, past performance data is crucial. Historical data is compiled from a variety of trustworthy sources, including "data.gov.in," "kaggle.com," and "indianwaterportal.com." Other databases including information on states and districts include soil type as an attribute. The primary data set is combined with the soil type column that was retrieved. Similar to this, average temperature and rainfall from a different dataset are added to the main data sets for the specific place. The data sets have been organized and purified. The null values are swapped out for the mean values. The attributes of the category are converted into labels before the algorithms are processed.
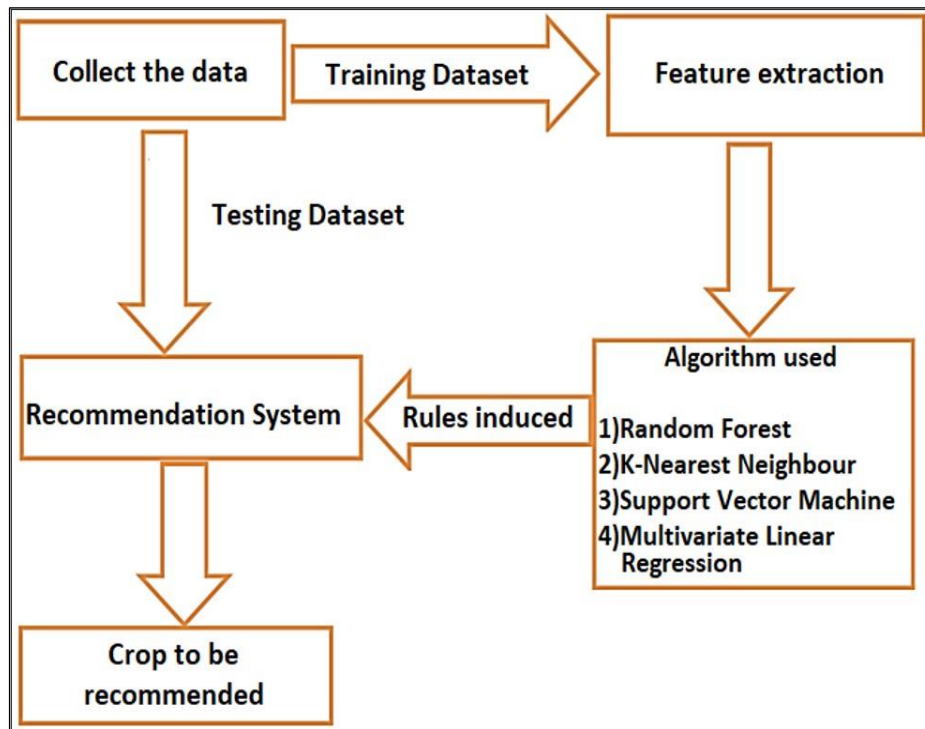

Figure 1: System Architecture.
Source: Author, (2023).

Figure 1 depicts the architecture of the created crop recommender system. The main applications of the crop recommender system are:

The first step is to gather all the data (in the form of a dataset) from all the locations. Since we are employing a supervised machine learning technique, training will follow. Following that, there will be feature extraction, in which the raw data will be transformed into a numerical feature to produce an output with a higher yield and greater efficiency.

After that, we only choose the "Random Forest Algorithm" from among the available methods because it produces a greater result. The rules generated by our algorithm are then represented in a figure, and they illustrate how our system actually operates by selecting and forecasting the crop, which is our ultimate objective [16].

II.1 Algorithm: Random Forest

Steps:

1. *Choose random samples from a given data or training set.*

2. *Make a decision tree for each piece of training data.*

3. *Then data is trained and tested based on dataset.*

4. *After training the data, real time weather data is fetched into the system.*

5. *Then based on crop data, algorithm will calculate its final output based on decision tree.*

6. *Choose the prediction result with the most votes as the final prediction result.*

7. *Find each decision tree's predictions for the new data points and assign them to the category that receives the most votes.*
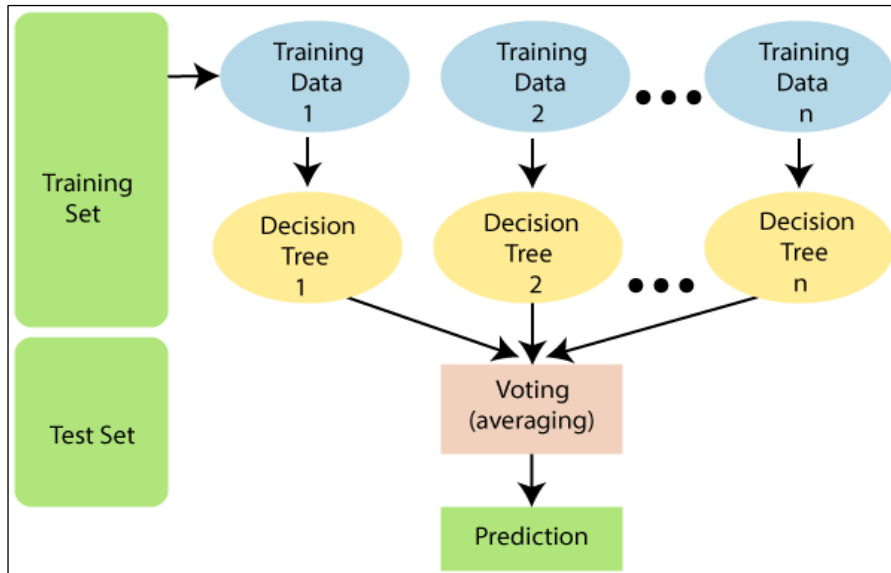

Figure 2: Working of Random Forest Algorithm.
Source: Author, (2023).

To increase the accuracy of the input dataset, the Random Forest classifier applies a number of decision trees to various subsets of the input dataset and averages the results. The core of the trees is ensemble learning, a method for combining several classifiers to handle challenging problems and improve model performance. The random tree uses the variation from each decision tree instead of relying solely on one, and it predicts the outcome based on votes for prediction maturity.

The random forest algorithm builds a forest out of a number of decision trees, adding randomness as the trees get bigger. The strategy enhances the model and adds more diversity by searching for the best characteristics among the random subset of features while splitting a node.

For visualizing through-out the system, there is a login system in which user first need to sign-up by his/her credentials (like name, username, mobile number and password). After that the user can login into system and need to click on 'Get the Crop' option.

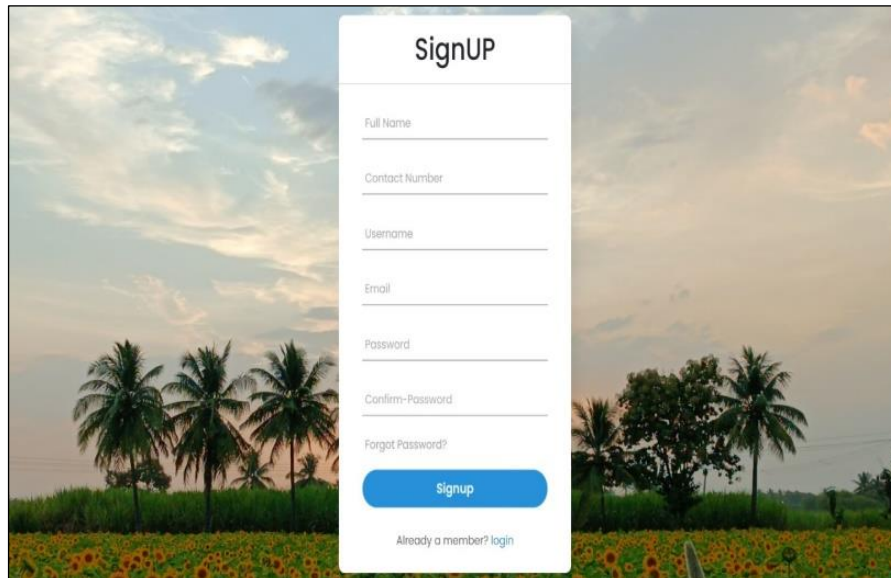Figure 3: Login Page of System.
Source: Author, (2023).

Figure 4: Sign Up Page of System.
Source; Author: (2023).

After that a new page will open which asks your location and basic elements of your soil (like Nitrogen, Phosphorus and Potassium value). Then some element like temperature are fetched from real time website, and based upon that model gives recommendation for the crop.
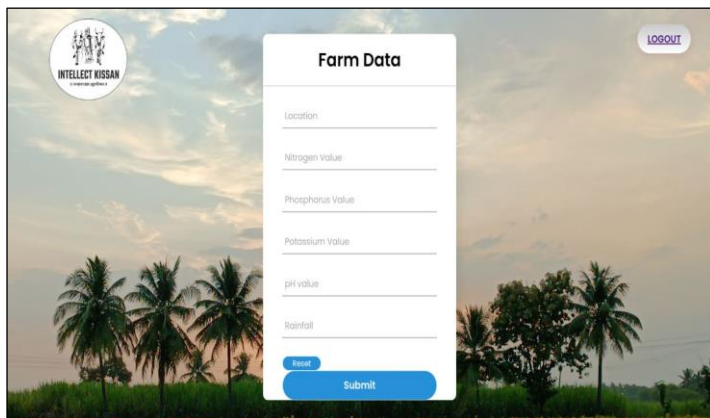

Figure 5: Crop Data Filling Form.
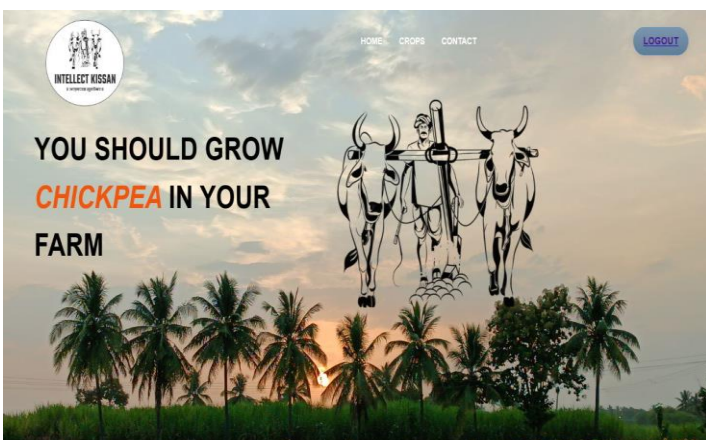Source: Author, (2023).

Figure 6: Final Output as a Recommended Crop.
Source: Author, (2023).

Final output we get based on all of our values is the recommended crop that will get maximum yield or suitable for respective climate.

III. RESULTS AND DISCUSSIONS

Result Analysis for Crop Recommendation system vs. Traditional Approach

Several elements should be taken into account when comparing the result analysis between a Crop Recommendation System employing machine learning and a Traditional Approach. Here are some things to take into account when comparing the outcomes of the two methods:

III.1 CROP RECOMMENDATION SYSTEM USING MACHINE LEARNING

Accuracy: Machine learning models can leverage large amounts of data and complex algorithms to make predictions. The accuracy of a Crop Recommendation System using machine learning can be evaluated based on how well it predicts suitable crops for specific conditions compared to actual crop yields in the given region.

Personalization: Machine learning models can take into account individual factors such as soil type, weather patterns, historical crop yields, and other relevant data points to provide personalized recommendations. The ability to provide tailored suggestions based on specific requirements can be a significant advantage.

Scalability: Machine learning models can handle large datasets and scale well, making them suitable for analyzing vast amounts of historical data and incorporating new data points as they become available. This scalability allows the system to continually improve its recommendations over time.

Adaptability: Machine learning models can adapt to changing conditions and learn from new data, enabling them to adjust recommendations based on evolving factors like climate change or updated agricultural practices. This adaptability can lead to more accurate and relevant crop recommendations.

III.2 TRADITIONAL APPROACH

Expert Knowledge: Traditional approaches often rely on expert knowledge and experience in agriculture. Crop recommendations are made based on established guidelines, local knowledge, and expertise in agricultural practices. The accuracy of recommendations depends on the proficiency and experience of the experts involved.

Simplified Models: Traditional approaches may use simplified models or rules of thumb based on historical practices and observations. These models may not account for as many variables or adapt as effectively to changing conditions compared to machine learning models.

Limited Data: Traditional approaches may rely on limited historical data or general knowledge about crop suitability in certain regions. They may not be able to leverage the vast amount of available data that machine learning models can analyze.

Time and Cost: Traditional approaches may require significant time and resources to gather expert opinions, conduct surveys, or analyze historical data manually. The efficiency and cost-effectiveness of traditional approaches may vary depending on the expertise available.

When analyzing the results, it is essential to compare the accuracy, efficiency, scalability, and adaptability of both approaches. Machine learning-based systems can leverage large datasets, personalize recommendations, and adapt to changing conditions, potentially leading to more accurate and dynamic crop recommendations. On the other hand, traditional approaches may rely on expert knowledge and local expertise but may lack the scalability and adaptability of machine learning models. The specific context, available resources, and the accuracy of the results must be carefully evaluated to determine which approach is more suitable for a particular crop recommendation system.

In the below figure 7,8,9 and 10 shows the detail comparison of proposed random forest algorithm with other machine learning algorithm.
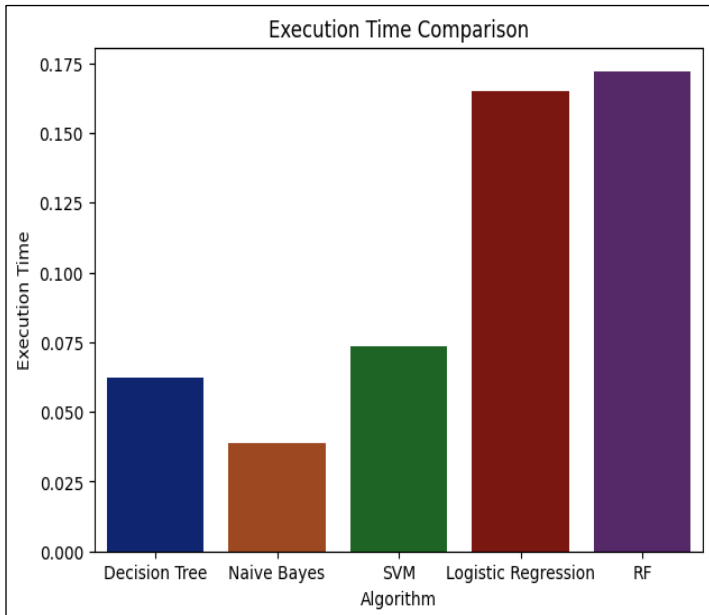

Figure 7: Comparison using Time Evaluation Parameter.
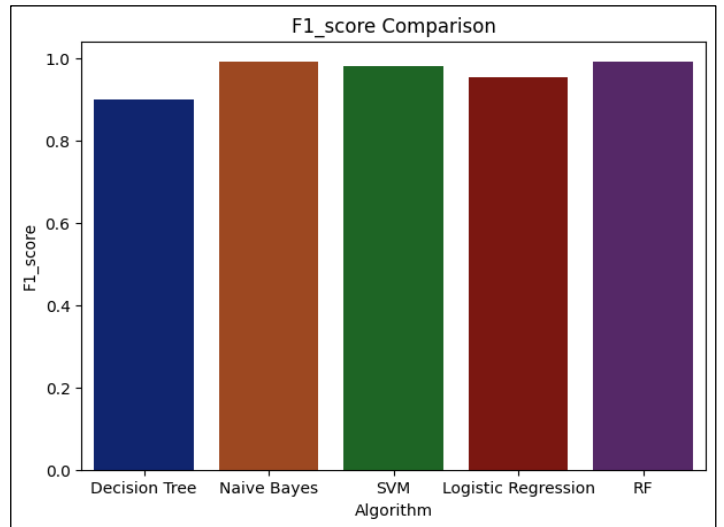Source: Author, (2023).

Figure 8: Comparison using F1_Score Evaluation Parameter.
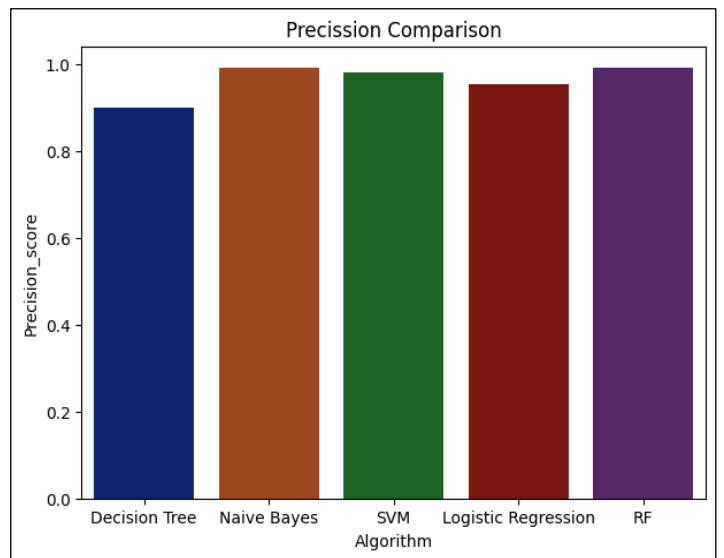Source: Author, (2023).

Figure 9: Comparison of Random Forest with State-of-Art using Precision.
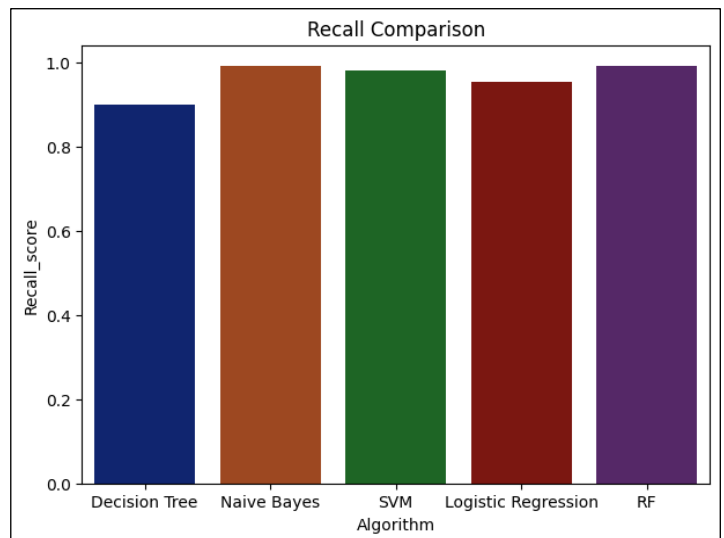Source: Author, (2023).

Figure 10: Comparison of random Forest with State-of-Art using Recall.
Source: Author, (2023).

Table 1: Comparison of random forest algorithm using accuracy parameter.

Sr. No	Algorithm	Accuracy
1	**Random Forest**	**99.09%**
2	Decision Tree	90.00%
3	Naïve Bayes	98.99%
4	Support Vector Machine	97.64%
5	Logistic Regression	95.22%

Source: Author, (2023).

Table 2: COMPARISON OF RANDOM FOREST ALGORITHM USING ACCURACY PARAMETER.

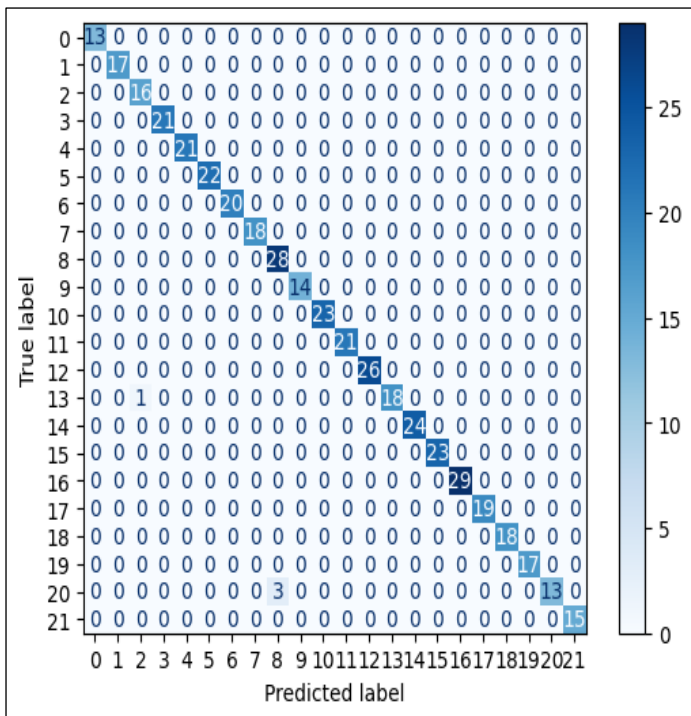| Sr. No	Algorithm	Accuracy
1	Random Forest	0.170
2	Decision Tree	0.045
3	**Naïve Bayes**	0.035
4	Support Vector Machine	0.075
5	Logistic Regression	0.169

Source: Author, (2023).

Figure 11: Confusion matrix of random forest algorithm.
Source: Author, (2023).

IV. CONCLUSIONS

This research highlighted the limitations of current methods and their applicability for crop recommendation. The proposed approach then connects the farmers with a functional crop recommender system through a web application. The web application gives users a number of options from which to choose a crop. Farmers that use the built-in suggestion technology can predict crop output. A user can research possible crops using the built-in recommender system to make better decisions. Machine learning algorithm (Random Forest) is deployed on the Keggle datasets that are provided, together with the rainfall Jeta and real meteorological data, and its prediction accuracy is evaluated. A useful technique for giving farmers and stakeholders data-driven advice on the best crops for particular environmental conditions is a crop recommendation system employing Random Forest. The system offers recommendations that can help optimize agricultural practices and maximize yields by utilizing historical crop yield data and analyzing the correlations between input features and crop performance. The random forest algorithm's accuracy is higher than that of Naïve bias, SVM, Decision Tree, and Logistic Regression, but its execution time is longer than that of Decision Tree.

In future, to reduce execution time of random forest algorithm is next step of this research

V. AUTHOR'S CONTRIBUTION

Conceptualization: Dr.Pradip Mukundrao Paithane.
Methodology: Dr.Pradip Mukundrao Paithane.
Investigation: Dr.Pradip Mukundrao Paithane.
Discussion of results: Dr.Pradip Mukundrao Paithane.
Writing – Original Draft: Dr.Pradip Mukundrao Paithane.
Writing – Review and Editing: Dr.Pradip Mukundrao Paithane.
Resources: Dr.Pradip Mukundrao Paithane.
Supervision: Dr.Pradip Mukundrao Paithane.
Approval of the final text: Dr.Pradip Mukundrao Paithane.

VI. REFERENCES

[1] Shilpa Mangesh Pande, Dr. Prem Kumar Ramesh, Anmol, B.R Aishwarya, karuna Rohilla, Kumar Shaurya. "Crop recommender system using machine learning approach", Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021) IEEE Aydin, O.F., Gokasar, I. and Kalan, O., 2020. Matching algorithm for improving ride-sharing by incorporating route splits and social factors. PloS one, 15(3), p.e0229674.

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

[3] Jain A. "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development, 2018.

[4] Paithane, Pradip Mukundrao. "Yoga Posture Detection Using Machine Learning." Artificial Intelligence in Information and Communication Technologies, Healthcare and Education: A Roadmap Ahead (2022): 27.

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

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

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

[8] S. Pavani, Augusta Sophy Beulet P., "Heuristic Prediction of Crop Yield Using Machine Learning Technique", International Journal of Engineering and Advanced Technology (IJEAT), December 2019, pp(135-138).

[9] Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE, 2019.

[10] Paithane, Pradip Mukundrao, and S. N. Kakarwal. "Automatic Pancreas Segmentation using A Novel Modified Semantic Deep Learning Bottom-Up Approach." International Journal of Intelligent Systems and Applications in Engineering 10.1 (2022): 98-104.

[11] Paithane, Pradip M., S. N. Kakarwal, and D. V. Kurmude. "Automatic Seeded Region Growing with Level Set Technique Used for Segmentation of Pancreas." Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020) 12. Springer International Publishing, 2021.

[12] Wagh, Sarita Jibhau, Pradip M. Paithane, and S. N. Patil. "Applications of Fuzzy Logic in Assessment of Groundwater Quality Index from Jafrabad Taluka of Marathawada Region of Maharashtra State: A GIS Based Approach." International Conference on Hybrid Intelligent Systems. Cham: Springer International Publishing, 2021.

[13] Kakarwal, Sangeeta, and Pradip Paithane. "Automatic pancreas segmentation using ResNet-18 deep learning approach." System research and information technologies 2 (2022): 104-116.

[14] Paithane, Pradip, Sarita Jibhau Wagh, and Sangeeta Kakarwal. "Optimization of route distance using k-NN algorithm for on-demand food delivery." System research and information technologies 1 (2023): 85-101.

[15] Paithane, Pradip M., and S. N. Kakarwal. "Automatic determination number of cluster for multi kernel NMKFCM algorithm on image segmentation." Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2. Springer International Publishing, 2020.