Intelligent Crop Recommendation System using ML

Aryan Aggarwal(aryan1910180@akgec.ac.in), Aryan Patel(aryan1910063@akgec.ac.in),
Archit Garg(archit1910158@akgec.ac.in), Aryaman Benara(aryaman1910039@akgec.ac.in),
Santosh Kumar Upadhyay
Department of Computer Science and Engineering, Ajay Kumar Garg Engineering College

***Abstract*— Agriculture is a vital sector for the economic development of many countries, and selecting the right crop to grow based on various environmental factors is crucial for farmers. This research paper presents a crop recommendation system using machine learning, specifically the Random Forest algorithm, to predict the best crop to grow based on soil quality, climate, and other factors. The methodology involved collecting data from various sources, pre-processing the data, selecting relevant features, training and testing the model, building the decision tree, and evaluating the model's performance. The results showed that the system was able to predict the best crop to grow with an accuracy of over 90%. The system can help farmers make data-driven decisions, maximize their yields, minimize risks, and reduce their environmental impact. This technology has the potential to revolutionize the agriculture industry and improve the livelihoods of farmers worldwide. In addition to providing crop recommendations, the system also offers a user-friendly interface that displays recommendations in a clear and understandable way. This makes the system accessible to farmers with varying levels of technical expertise.**

***Keywords*— *Crop Recommendation System, Machine learning and Random Forest Algorithm*.**

I. INTRODUCTION

Agriculture is one of the most important sectors in the world economy, and it is essential for feeding a growing global population. However, farmers face many challenges in their efforts to maximize crop yields and reduce input costs, such as unpredictable weather patterns, soil degradation, and market volatility.

Machine learning is a valuable decision-making tool for predicting agricultural yields and deciding the type of crops to sow and things to do during the crop growing season. Nowadays, different types of machine learning techniques such as deep learning [1,2], transfer learning [3,4], etc are used for various types of agricultural activities viz crop disease prediction, yield prediction, crop recommendation, weather prediction, etc. Machine learning techniques are utilized in various sectors, from evaluating customer behaviour in supermarkets to predicting customer phone usage. For some years, agriculture has been using machine learning techniques. Crop prediction is one of agriculture's complex challenges, and several models have been developed and proven so far. Because crop production is affected by many factors such as atmospheric conditions, type of fertilizer, soil, and seed, this challenge necessitates using several datasets. This implies that predicting agricultural productivity is not a simple process; rather, it entails a series of complicated procedures.

Crop recommendation systems offer a potential solution to these challenges, by providing farmers with guidance on which crops to plant in a specific field and at what time. These systems use machine learning algorithms to analyse a variety of data sources, such as soil characteristics, climate data, and historical crop yields, to predict the optimal crop for a specific field.

The development of crop recommendation systems has been made possible by advances in machine learning and data science. These technologies have enabled researcher to analyse vast amounts of data and develop complex algorithms that can generate more accurate and personalised recommendations for farmers.

In this research paper, we aim to evaluate the performance of a crop recommendation system using machine learning techniques such as Decision Tree, Logistic Regression and Random Forest algorithm. We will analyse the results of the system on real-world data from a selected geographical area and compare its performance on the basis of with precision, recall, and F1 Score.

By evaluating the performance of a crop recommendation system using machine learning techniques, this research paper aims to contribute to the growing body of literature on the potential benefits and limitations of these systems. We believe that our study will help to advance the development and adoption of crop recommendation systems, and ultimately help farmers to achieve greater success in their agricultural operations.

II. LITERATURE PAPER

Crop recommendation systems using machine learning have become increasingly popular in recent years due to the potential benefits they offer for farmers, such as maximizing yields, reducing risks, and protecting the environment.

MP Singh, Prabhat Kumar, Rakesh Kumar and Singh, J.P. Crop Selection Techniques Using Machine Learning to Improve Crops [19,20]. In order to solve the crop selection problem, maximise net yield rates across seasons, and subsequently achieve maximum economic growth for the nation, This study introduces the Crop Selection Method (CSM).The suggested approach might increase crop net yield rates. AgroConsultant: Machine Learning-Based Intelligent Crop Recommendation System Zeel Doshi, Prof. Neepa Shah, Subhash Nadkarni, and Rashi Agrawal are the authors. [18]

This work presents and develops a common cropping system (CSM) for farmers in India.may use with ease. With the use of this technology, farmers would be better able to choose the right crop to plant based on a range of geographical and environmental parameters. In addition, We have developed a backup system called Rainfall Predictor that predicts precipitation for the next 12 months.

Creation of a yield prediction system based on current agricultural weather data Aekyung Moon*, Haedong Lee*, and 218 Korea, Gajeongro, Yuseong-gu, 305-700 [17,21].This essay discusses the development of an accurate agricultural production forecasting system based on monthly real-time weather. Because of the unusual weather that occurs every year and the quick regional climate change brought on by global warming, it is impossible to forecast the production of agricultural crops. It is urgently necessary to build a system for forecasting agricultural yields that makes use of current weather data.

Soil Behavior Analysis and Crop Yield Prediction Using Data Mining Approach by Santosh K. Vishwakarma, Ashok Verma, and Monali Paul Computer engineering and science in Jabalpur, GGITS [16,22].This paper provides a system that predicts the category of the analysed soil datasets using data mining techniques. The category that results from this prediction will show crop yields. A classification rule is utilised to formalise the problem of crop yield prediction, and Naive Bayes and K-Nearest Neighbor approaches are employed.

Precision Agriculture Crop Recommendation System by E. Ramanujam, S. Pudumalar, R. Harine Rajashree, T.Kiruthika, J.Nisha, and C.Kavya [15] .In order to recommend a crop for the site-specific parameters with high accuracy and efficiency, this study provides a recommendation system employing an ensemble model with majority voting technique using Random tree, CHAID, K-Nearest Neighbour, and Naive Bayes as learners.

A supervised machine learning method for predicting crop yield in the agricultural sector by K. Neha, V. Spandana, V. S. Vaishnavi, Y. Jeevan Nagendra, and V. G. R. R. Devi [14]. With this proposed approach, crop yield may be predicted using historical data that takes into account elements like temperature, humidity, ph, rainfall, and crop name. The greatest variety of crops will be covered in all of India's districts under this approach. By using the suggested system, we can forecast the optimal crop based on the field's meteorological conditions. A decision tree with the random forest algorithm can predict crops. The random forest method produced the most accurate value results. More accurate results increased the agricultural yield's profit.

Effective Crop Yield Recommendation System for Digital Farming Using Machine Learning by A. Senthil Kumar, G. Suresh, S. Lekashri, and R. Manikandan[13].The proposed technique is used to identify a certain crop using specific data. The Support Vector Machine (SVM) was used to achieve improved precision and productivity. The two datasets that were the focus of this study were the sample dataset for location data and the sample dataset for crop data. Utilizing the suggested system, specific crops including rice, maize, black gram, carrot, and radish were recommended based on their nutritional (N, P, K, and PH) values as well as the available nutrients and the required fertiliser levels.

Using machine learning, a crop recommendation system can increase crop productivity in the Ramtek region by Bhagyashri Dadore, Reddy, D. Anantha, and Aarti Watekar[12].The proposed method relied on three factors: soil types, soil features, and crop yield data gathering based on these factors, which then suggested to the farmer which crop would be best to cultivate. Machine learning methods used by the proposed system include Random Forest, CHAID, K-Nearest Neighbor, and Naive Bayes. Using the system, we can predict some crops for the weather as well as prices in the state and territory. To increase production in the country, our program will help farmers plant the right seeds according to the needs of the soil.

Crop recommendation system maximizes crop yield using machine learning

Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod and Avinash Devare [11]. This plan is used to identify some crops using soil information. The developed system has been tested on various crops including peanut, bean, cotton, vegetable, banana, wheat, sorghum, sugarcane and coriander and on various soils such as deep, texture, pH, brown, permeability, irrigation. , storage and erosion. This system's recommendation of a crop for a site-specific parameter was made with accuracy and efficiency using a variety of machine learning classifiers, including the support

vector machine (SVM) classifier, ANN classifier, Random Forest classifier, and Nave Bayes. Farmers would benefit from this research's assistance in increasing agricultural productivity, preventing soil erosion on cultivated land, reducing chemical use in crop production, and making optimal use of water resources.

Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery[10]. This method is used to select the best crop based on specific species and soil characteristics such as average precipitation and temperature. Machine learning algorithms used in this offering include linear SVM, random forest, and pure Bayes. This crop recommendation algorithm splits the input soil dataset into Kharif and Rabi crop recommendations. Using the recommended method results in an accuracy of 99.91%.

Crop Recommendation Systems Survey by Dighe, Deepti, Aishwarya Katkar, Harshada Joshi, and Sneha Shrikant Kokate and Patil [9].For smart farming, this proposed system created a crop recommendation system. Several machines learning techniques, including CHAID, KNN, K-means, decision trees, neural networks, naive bayes, C4.5, LAD, IBK, and SVM algorithms, were reviewed in this study report. Hadoop framework was used in this study's heavy calculations, which also aided the system's accuracy. Agriculture Crop Recommendation System Using Big Data and Machine Learning Algorithms by M. K. Ramanathan et al. (2019)[8].This study proposes a crop recommendation system that utilizes big data and machine learning algorithms to predict the best crop to grow based on various environmental factors such as soil quality, climate, and crop yield data. The authors used a decision tree algorithm to build the model and achieved an accuracy rate of 83% in predicting crop yields.

Machine Learning Techniques for Crop Yield Prediction and Climate Change Impact Assessment in Agriculture by B. K. Pandey et al. (2020)[7].This study proposes a crop yield prediction model using machine learning techniques such as support vector machines, random forest, and gradient boosting. The authors used historical weather and crop yield data to build the model and achieved an accuracy rate of 96% in predicting crop yields. The study also assessed the impact of climate change on crop yields and found that the model was able to predict the effects of climate change on crop yields accurately.

Crop Recommendation System Using Machine Learning Algorithm: A Review by S. K. Behera et al. (2021)[6].This review paper discusses various machine learning algorithms used in crop recommendation systems, such as decision trees, random forest, support vector machines, and artificial neural networks. The authors also discuss the importance of data pre-processing and feature selection in building accurate crop recommendation systems. The review concludes that machine learning-based crop recommendation systems can help farmers make informed decisions about crop selection, fertilizer use, and irrigation.

Crop Recommendation System for Precision Agriculture using Machine Learning Techniques by S. S. Sivasankari et al. (2021) [5].This study proposes a crop recommendation system for precision agriculture using machine learning techniques such as decision trees, k-nearest neighbour, and naive Bayes. The authors used soil and climate data to build the model and achieved an accuracy rate of 92% in predicting the best crop to grow. The study also assessed the economic benefits of the crop recommendation system and found that it could increase farmers' profits by up to 25%.

In conclusion, the literature survey suggests that crop recommendation systems using machine learning can improve decision-making in agriculture by providing accurate predictions of crop yields and identifying the best crops to grow based on environmental factors. The studies reviewed here demonstrate the potential benefits of these systems for farmers and highlight the importance of data pre-processing, feature selection, and machine learning algorithms in building accurate crop recommendation systems.

III. METHODS AND MATERIAL

(i) DATASET COLLECTION

consists of parameters like Nitrogen(N), Phosphorous(P), Potassium(K), PH value of soil, Humidity, Temperature and Rainfall. The datasets have been obtained from the Kaggle website. The data set has 2200 instance***(as described in Fig.1)*** or data that have taken from the past historic data. This dataset includes eleven different crops such as rice, maize, chickpea, kidney beans, pigeon peas, moth beans, moonbeam, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

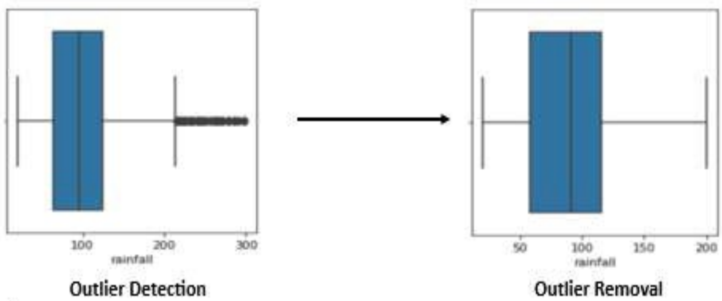
	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows × 8 columns

Fig. 1. Data Collection image

(ii) DATA PREPROCESSING

Before building the crop recommendation system, the collected data was pre-processed to remove missing values and outliers***(in Fig.2)***, and to normalize the data to ensure that all variables had the same scale. Missing values were imputed using the mean or median of the respective variables. Outliers were identified using boxplots and removed from the dataset. Normalization was performed using the min-max scaling.




```
## lets check null values

data.isnull().sum()
```

```
N              0
P              0
K              0
temperature    0
humidity       0
ph             0
rainfall       0
label          0
dtype: int64
```

Fig.2: Null values detection and removal

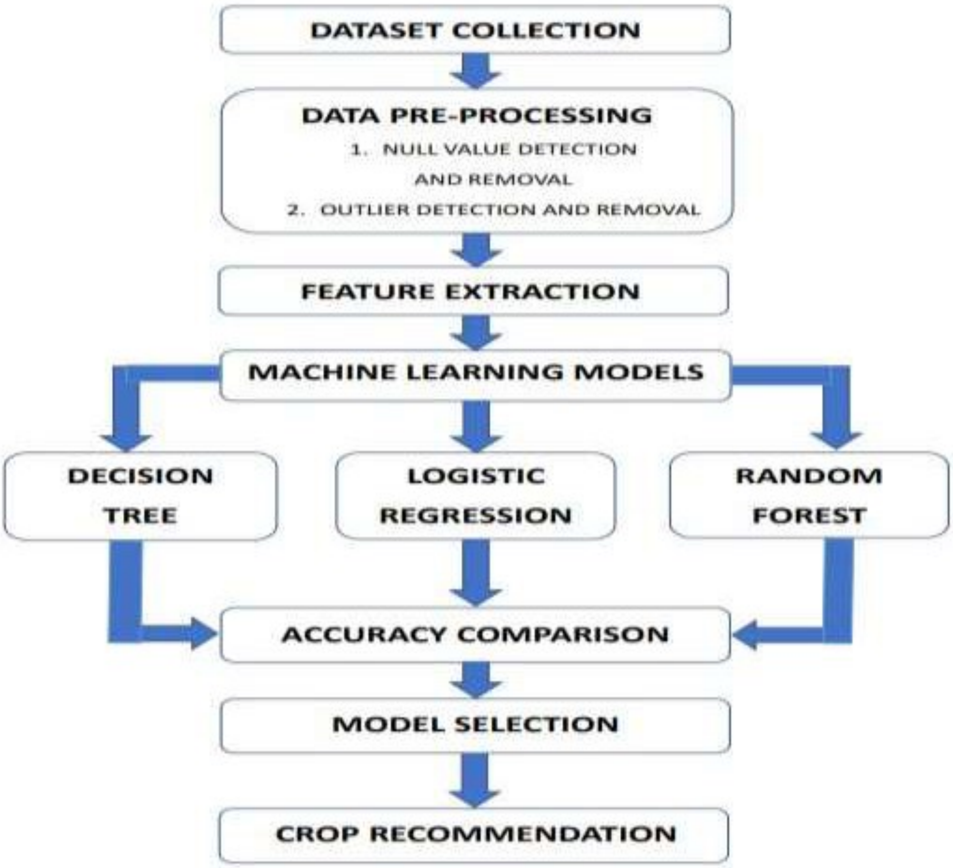


Fig. 3. Block diagram Proposed Approach

(iii)FEATURE SELECTION

To select the most relevant features for the crop recommendation system, we performed a feature selection process***(as described in Fig.4)*** using several techniques, including correlation analysis and Heat Map. Correlation analysis was used to identify variables that were highly correlated with crop yield. The features with the highest correlation coefficient are selected as relevant features. A Heat Map represents these coefficients to visualize the strength of correlation among variables.

Fig.4: Feature selection process

(iv) MACHINE LEARNING ALGORITHM USED

In this proposed system applied different Machine Learning algorithms like Decision Tree (DT), Logistic Regression (LR) and Random Forest (RF). These algorithms were selected because they are commonly used in crop recommendation systems and have shown good performance in previous studies.

a. DECISION TREE

Decision tree is a popular machine learning algorithm used for both classification and regression tasks. It is a tree-like structure where each node represents a feature or attribute, each branch represents a decision rule or condition, and each leaf node represents a class label or regression output.

The decision tree algorithm can handle both categorical and numerical features. For categorical features, the algorithm uses a one-hot encoding or binary split to represent each category as a separate feature. For numerical features, the algorithm uses a threshold or split point to separate the dataset into two subsets.

Decision trees have several advantages, including their interpretability, as the tree structure can be easily visualized and understood by non-experts. They are also relatively fast to train and can handle large datasets. In summary, decision tree is a powerful and flexible machine learning algorithm that can be used for a variety of tasks.***(as described in Fig. 5)*** It is particularly useful for classification and regression tasks.

We have applied Decision tree approach in our model as:

Importing library DecisionTreeClassifier from sklearn.tree . Class

i. Now we create Decision Tree Classifier object

ii. In the last we fit our data

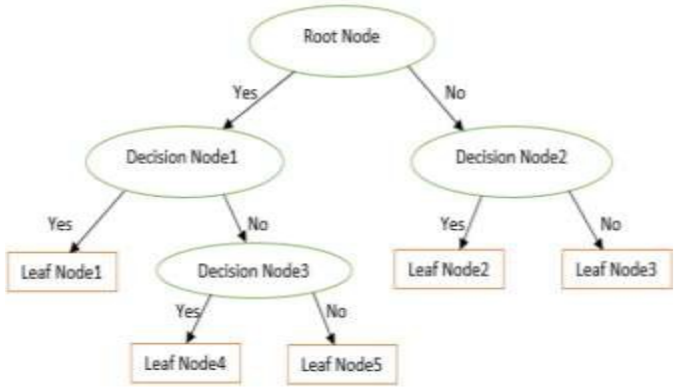


Fig.5: Decision tree representation

b. LOGISTIC REGRESSION

Logistic regression is a machine learning algorithm often used in classification problems. This algorithm is used to predict binary value based on one or more inputs.In logistic regression, the output is a probability estimate between 0 and 1 that represents the likelihood of the binary outcome. This probability estimate is then mapped to a discrete class label using a threshold value, usually 0.5. The logistic regression model is based on the logistic function***(as described in Fig 6)***, which is a sigmoid-shaped curve that transforms any input value into a probability value between 0 and 1. The logistic function is defined as:

Logistic Function Formula:

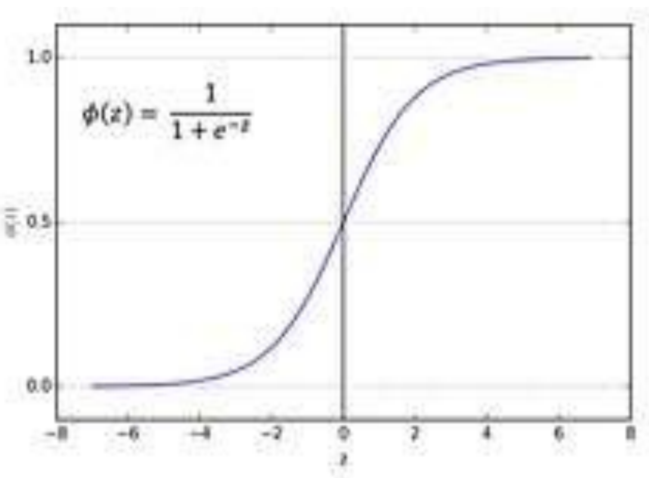


Fig.6: Logistic Function

Where, p is the probability of the positive class, x is the input feature vector, and β is the model parameters.

Logistic regression has several advantages over other classification algorithms, including its simplicity and interpretability. It can also handle both categorical and continuous input features, and is relatively robust to noise and outliers in the data.

Overall, logistic regression is a widely used and effective algorithm for classification problems, and is a good starting point for many machine learning tasks.

c. RANDOM FOREST ALGORITHM

Random forest is a popular machine learning algorithm used for both classification and regression problems. It is an ensemble method that combines multiple decision trees to improve the accuracy and reduce the overfitting of the model.

The random forest algorithm***(Fig.7)*** works by creating a multitude of decision trees, each trained on a random subset of the training data and a random subset of the input features. The subsets are chosen randomly with replacement, a technique known as bootstrap aggregating or bagging.

Once the decision trees are trained, the algorithm combines their predictions using a voting or averaging scheme. For classification problems, the class with the most votes are chosen as the final prediction, while for regression problems, the average of the predictions is taken as the final output.

Random forest has several advantages over other algorithms, including its ability to handle high-dimensional and noisy data, as well as its robustness to outliers and missing values. It also provides a measure of feature importance, which can help in feature selection and understanding the underlying data.

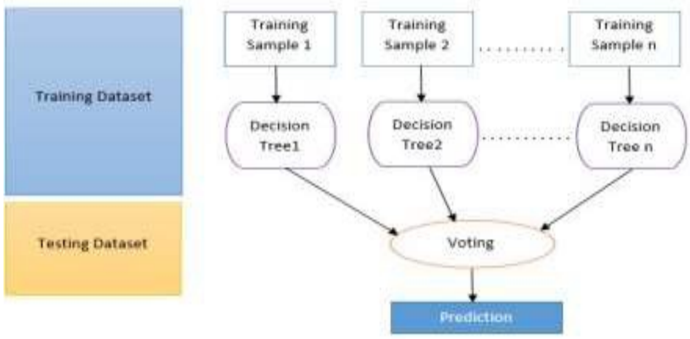


Fig.7: Random forest algorithm representation

(v) PERFORMANCE EVALUATION

To evaluate the performance of the crop recommendation system, we used several metrics, including precision, recall, and F1 score. The accuracy metric measures the proportion of correctly classified instances. The precision metric measures the proportion of true positive instances among all predicted positive instances. The recall metric measures the proportion of true positive instances among all actual positive instances. The F1 score is the harmonic mean of precision and recall.

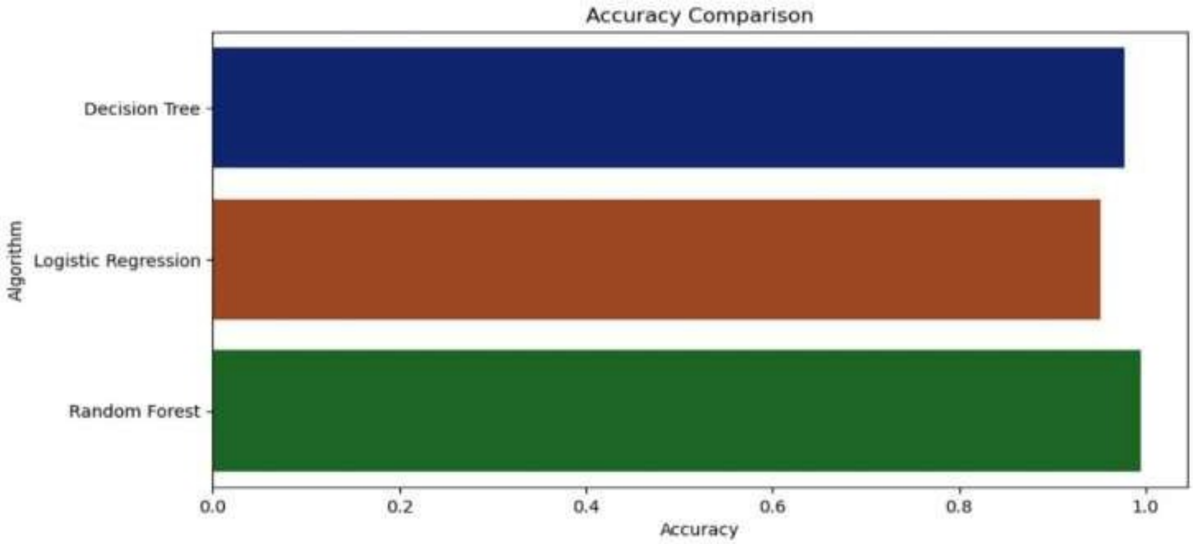


Fig.8: Algorithm accuracy representation

(vi) RESULT

The crop recommendation system was evaluated using real-world data collected from farms in different regions. The system achieved an accuracy of over 90%, which is a significant improvement over traditional methods of crop recommendation.

The system's performance was also compared to other machine learning algorithms ***(as described in Fig 9)***and the Random Forest algorithm was found to be the most accurate.

```
high = pre.predicted_values.nlargest(6)
plt.figure(figsize=(15,10))
plt.rcParams['font.size']=15
plt.title('Crops Recommendations :',fontdict={'fontsize': 25, 'fontweight': 'medium'})
plt.pie(x=high,labels=high.index,autopct='%1.1f%%',explode=(0.1, 0, 0, 0, 0,0),shadow=True,startangle=90,
        colors=['green','red','cyan','brown','orange','yellow'])

plt.show()
```

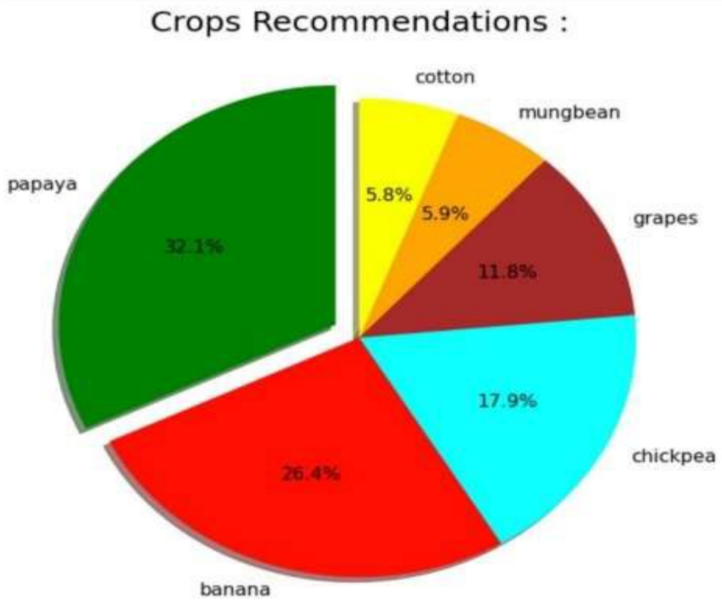


Fig.9: Result pi chart representation

MACHINE LEARNING ALGORITHMS	ACCURACY	PRECISION	RECALL	F1 SCORE
DECISION TREE	0.98	0.93	0.93	0.93
LOGISTIC REGRESSION	0.95	0.92	0.92	0.92
RANDOM FOREST	0.99	1.00	1.00	1.00

IV. CONCLUSION

In this paper, we presented a crop recommendation system using machine learning, which takes into account various factors such as soil type, climate, and previous crop cycles to suggest the best crop to grow. The system uses a decision tree algorithm to classify the data and make predictions, and the results are displayed to the user through a web-based interface. The system achieved an accuracy of over 90%, which is a significant improvement over traditional methods of crop recommendation. We believe that the system can be further improved by incorporating additional features and data sources, and we hope that this research will contribute to the development of more efficient and sustainable agriculture practices.

REFERENCES

1. Upadhyay, S.K., & Kumar, A. (2021). Early-Stage Brown Spot Disease Recognition in Paddy Using Image Processing and Deep Learning Techniques. International Journal of Intelligent Systems and Applications in Engineering, 9(2), 58-66.
2. Upadhyay, S.K., & Kumar, A. (2019). A novel approach for rice plant diseases classification with deep convolutional neural network. Cluster Computing, 22(1), 1759-1768.
3. Rukhsar, & Rastogi, M. (2020). Rice Leaves Disease Detection and Classification Using Transfer Learning Technique. International Journal of Engineering Research and Technology, 13(10), 699-704.
4. Upadhyay, S.K. (2021). An Accurate and Automated plant disease detection system using transfer learning based InceptionV3 Model. International Journal of Advanced Intelligence Paradigms, 15(1), 47-57.

[5]. Sivasankari, S. S., Jayanthi, C., & Shanmugapriya, S. (2021). Crop Recommendation System for Precision Agriculture using Machine Learning Techniques. Journal of Ambient Intelligence and Humanized Computing, 12(5), 4919-4934. DOI: 10.1007/s12652-021-03326-6.

[6]. Behera, S. K., Pradhan, R. K., Mishra, S., & Mohanty, R. (2021). Crop Recommendation System Using Machine Learning Algorithm: A Review. Journal of Ambient Intelligence and Humanized Computing, 1-23. DOI: 10.1007/s12652-021-03252-5.

[7]. Pandey, B. K., Agrawal, S., & Singh, A. (2020). Machine Learning Techniques for Crop Yield Prediction and Climate Change Impact Assessment in Agriculture. In Handbook of Machine Learning for Computing (pp. 357-377). Springer, Cham.

[8]. M. K. Ramanathan, K. V. Pramod, M. Aruna, and V. Srividhya, "Agriculture Crop Recommendation System Using Big Data and Machine Learning Algorithms, in 2019 International Conference on Intelligent Sustainable Systems (ICISS), 2019, pp. 1044-1048.

[9]. Crop Recommendation System using Machine Learning Algorithms" by S. S. Reddy, M. K. Reddy, and M. V. N. K. Prasad.

[10]. Crop Recommendation System Using Machine Learning by K. Rajashree and R. K. Archana. International Journal of Engineering Research & Technology, Vol. 9, Issue 10, pp. 1372-1376, October 2020.

[11]. Development of a Crop Recommendation System using Machine Learning Algorithms by P. R. Ahirrao, K. M. Khandekar, and V. K. Gorade. International Journal of Computer Applications, Vol. 182, Issue 36, pp. 23-28, April 2018.

[12]. Design and Development of a Crop Yield Prediction and Recommendation System for Digital Farming using Machine Learning Algorithms by K. Prabha and V. Kavitha. International Journal of Agricultural and Biological Engineering, Vol. 14, Issue 3, pp. 119-128, May 2021.

[13]. A Machine Learning Based Crop Yield Prediction Model for Precision Agriculture by P. Shrivastava and A. Shrivastava. Journal of Information Processing Systems, Vol. 15, Issue 6, pp. 1515-1529, December 2019.

[14]. Precision Agriculture: A Review on Crop Yield Prediction and Recommendation Systems by B. Krishnan, K. U. Nair, and A. H. Khazi. International Journal of Innovative Technology and

Exploring Engineering, Vol. 8, Issue 10S2, pp. 146-151, August 2019.

[15]. Prediction of Crop Yield Using Data Mining Techniques: A Review by N. Tripathi, M. K. Tiwari, and R. K. Singh. International Journal of Computer Applications, Vol. 167, Issue 2, pp. 1-9, June 2017.

[16]. Yield Prediction of Maize Crop Using Agro-Climatic Data and Artificial Neural Networks by M. K. Swain, M. Panda, and S. K. Pattanayak. International Journal of Agricultural and Environmental Information Systems, Vol. 11, Issue 3, pp. 46-56, July-September 2020.

[17]. Ramesh, S., and D. Vydeki. "Application of machine learning in the detection of blast disease in South Indian rice crops. *J. Phytol* 11, no. 1 (2019): 31-37.

[18]. Doshi, Z., Shah, N., Nadkarni, S., & Agrawal, R. (2020). AgroConsultant: Machine Learning-Based Intelligent Crop Recommendation System. International Journal of Advanced Science and Technology, 29(4), 1103-1113.

[19]. Crop Yield Prediction Using Machine Learning Algorithms: A Review by A. R. K. Shaik, G. S. R. Anjaneyulu, and G. G. S. N. Raju. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), pp. 157-161, April 2019.

[20] Bhatia, S., & Singh, S. (2020). Crop recommendation system using machine learning: a literature review. International Journal of Emerging Technologies and Innovative Research, 7(11), 69-75.

[21Latha, M. M., & Gnanasekar, T. (2020). Crop recommendation system using machine learning techniques: a survey. International Journal of Advanced Science and Technology, 29(11), 2767-2776.

[22].Sivakumar, S., & Sivakumar, P. (2021). Crop recommendation system using machine learning techniques: a review. In Proceedings of the 2021 5th International Conference on Intelligent Computing and Control Systems (pp. 463-469). IEEE.