Title: Crop Recommendation System

Abstract - In today's farming world, technology plays a big role in helping farmers grow crops better. One important tool is the crop recommendation system. It's like a smart assistant for farmers, giving them advice on which crops to grow based on things like the weather and soil.This report dives into how these systems work, what they're good at, and where they could do better. We'll explore different types of systems, from simple ones with basic rules to fancy ones using advanced computer smarts. We'll also talk about what makes these systems work well and what challenges they face, like getting the right data and making sure they're easy for farmers to use.But we don't stop there. We suggest ways to make these systems even better. For example, by combining different kinds of data, like pictures from satellites and info about the soil, we can make the advice more accurate. We also suggest making the systems easier to understand and use, so farmers can get the most out of them.By looking at what's already out there, finding ways to improve, and making suggestions for the future, this report aims to help farmers grow crops smarter and feed more people.

I. Introduction

Crop recommendation systems leverage machine learning techniques to analyze agricultural and environmental data and suggest the most suitable crops for cultivation in a given area. These systems consider factors such as soil properties, climate conditions, historical crop performance, and market trends to make informed recommendations. By providing tailored suggestions to farmers, crop recommendation systems aim to optimize crop yields, maximize profitability, and promote sustainable agricultural practices.In modern agriculture, the utilization of data-driven technologies has revolutionized traditional farming practices. One such advancement is the development of crop recommendation systems, which leverage various parameters such as nutrient levels (N, P, K), environmental factors (temperature, humidity, rainfall), and soil characteristics (pH) to provide tailored guidance to farmers regarding suitable crops to cultivate in their specific regions.The primary objective of this project is to evaluate the effectiveness of different machine learning models in predicting crop recommendations based on a combination of input features including nutrient levels, environmental conditions, and soil attributes. The models under consideration include Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbours, Naive Bayes, and Support Vector Machine (SVM).The dataset comprises several key attributes including N (Nitrogen), P (Phosphorus), K (Potassium), temperature, humidity, pH levels, rainfall, and a label indicating the type of crop. The label data consists of a variety of crops commonly grown across different regions, such as rice, maize, chickpea, and others.Throughout this project, we aim to assess the performance of each machine learning model in accurately predicting the appropriate crop based on the provided input features. By comparing the strengths and weaknesses of these models, we seek to identify the most effective approach for crop recommendation in agricultural contexts. This evaluation holds significant implications for optimizing crop yield, resource utilization, and sustainability in farming practices.

II. Literature Review

Research paper utilizes various machine learning algorithms for crop recommendation, including Decision Tree, Naive Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), and XGBoost .These algorithms are applied in different phases of the project, such as dataset collection, pre-processing, feature extraction, and the recommendation system. Each algorithm plays a specific role in analyzing the soil parameters collected by sensors and providing suggestions for suitable crop growth based on the data analyzed through these algorithms [1].

This paper discusses the use of Agro Consultant to assist Indian farmers in making decisions about crop selection based on factors like sowing season, geographical location, soil parameters, temperature, and rainfall .It mentions a system that utilizes data mining techniques to predict crop yield by analyzing soil datasets, using classification rules like Naive Bayes and K-Nearest Neighbor methods .Machine Learning is highlighted as a crucial tool for supporting crop yield prediction, aiding in decisions on which crops to grow during specific seasons. Various classifiers like KNN, Naive Bayes, SVM, Logistic Regression, and Random Forest are used for crop yield prediction, with the paper specifically focusing on the KNN Algorithm. Factors considered in the prediction include crop production and the area of crop yield produced [2].

This research paper Agriculture's vital role in developing economies like India is highlighted, emphasizing the need for sustainability and food security.The paper proposes an innovative Agriculture system integrating IoT and ML technologies for crop prediction based on Metrological Agriculture theory.By utilizing IoT for real-time data collection and ML for predictive analysis, the system aims to enable smart farming practices.The integration of IoT and ML technologies is expected to enhance overall yield and product quality in agriculture.This system contributes to the modernization and efficiency of agricultural practices, assisting farmers in making informed decisions for optimal crop management[3].

Agriculture is crucial for sustaining life and economic growth, blending traditional practices with modern technologies .Previously, crop selection was based on practical experience, but now growers can use a recommendation system powered by Machine Learning Algorithms to choose the most suitable crop based on soil characteristics . The recommendation system utilizes Neural Networks to make accurate crop suggestions without human intervention, as the nested layers within the network can learn from data independently [4].

The research paper titled ”Crop Recommendation System” presents an integrated model for automating crop recommendations based on environmental factors. It addresses the lack of a comprehensive system that suggests the most suitable crops for a given area. The proposed system utilizes Arduino microcontrollers, machine learning techniques, and natural language processing to collect and process environmental data with high accuracy and efficiency. The literature survey explores the benefits of digital agriculture and emphasizes the importance of considering multiple factors for optimal crop growth. The system also includes a feedback mechanism to track plant growth and provide necessary precautions. Overall, the paper aims to improve agricultural practices and decision-making through advanced technology[5].

This paper "Crop Recommender System Using Machine Learning Approach" by Pandey et al. presents a comprehensive solution to address low crop yields in rural India, a significant contributor to the nation's GDP. By leveraging machine learning algorithms like Random Forest, the system achieves high accuracy in predicting crop yields, enhancing profitability for farmers. The proposed mobile application integrates features such as GPS-based location identification and optimal fertilizer timing, catering to the specific needs of agriculturalists. Through a user-friendly interface, the system addresses challenges in crop selection, empowering farmers to make informed decisions amidst fluctuating market dynamics and climatic uncertainties. This innovative approach marks a significant advancement in agricultural technology, offering practical solutions to improve farmers' livelihoods and mitigate socio-economic challenges in rural communities[6].

This research paper explores various techniques in crop recommendation, including ensemble models and data mining approaches, setting the context for their proposed SVM-based system. Their contributions encompass classification, model training, crop prediction, and evaluation methods, demonstrating a holistic approach to addressing the issue of inaccurate crop selection.Ultimately, the study aims to equip farmers with a reliable recommendation system, integrating soil analysis and machine learning algorithms. This approach empowers farmers to make informed decisions, potentially enhancing profits and overall agricultural productivity[7].

This research paper "Crop Prediction using Machine Learning Approaches" addresses the pressing challenges faced by Indian farmers, including soil degradation and reduced crop yields due to repetitive cultivation practices. By leveraging machine learning algorithms, the proposed system offers a solution to suggest optimal crop choices based on soil composition, weather conditions, and nutrient requirements. The use of Support Vector Machine (SVM) for rainfall prediction and Decision Tree for crop recommendation showcases the versatility of machine learning in agricultural applications. The authors underscore the significance of embracing technological advancements in agriculture to enhance profitability and environmental sustainability. Overall, the document contributes to the growing body of research on utilizing machine learning for improving crop yield and promoting sustainable farming practices[8].

This paper explores leveraging machine learning in agriculture for precise crop prediction, emphasizing its critical role in addressing climate variability and global food demands. By integrating various algorithms, including decision trees and linear regression, it aims to optimize crop yield forecasts, aiding farmers in adapting to changing environmental conditions and enhancing decision-making processes. Through its innovative approach, the paper underscores the potential of machine learning to revolutionize agricultural practices, offering insights into crop selection and management strategies crucial for sustainable farming in diverse regions[9].

This paper "Intelligent Crop Recommendation System using Machine Learning" contributes to addressing the challenges faced by Indian farmers by proposing a sophisticated system that leverages machine learning techniques to enhance crop selection processes. Emphasizing the significance of agriculture in India's socio-economic landscape, the paper underscores the complexities farmers encounter in selecting suitable crops based on diverse factors. By integrating machine learning algorithms such as Decision Tree Learning and random forest, the system aims to provide accurate and automated recommendations tailored to specific soil, environmental, and economic conditions. The study aligns with existing research highlighting the potential of machine learning and data mining in improving crop recommendation accuracy and efficiency, offering insights into the integration of advanced technologies to empower farmers and enhance agricultural practices[10].

	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

Figure 1: dataset

III. Methodology

A. Data Collection

The dataset is collected sourced from Kaggle link [https://www.kaggle.com/datasets/atharvaingle /crop-recommendation-dataset] a popular platform for sharing datasets . The dataset contains information about various crops and their corresponding attributes, including N, P, K levels, temperature, humidity, pH, rainfall, and the label indicating the type of crop.Data set contains 2200 instences and 8 columns. The dataset was downloaded and stored for further analysis.

B. Data Preprocessing:

After collecting datasets from kaggle. Dataset must be prepossessing before training to the model. The data prepossessing can be done by various stages, begins with reading the collected dataset the process continues to data cleaning. In data cleaning the datasets contain some redundant attributes, those attributes are not considering for crop prediction. So, we have to drop unwanted attributes and datasets containing some missing values we need to drop these missing values or fill with unwanted Nan values in order to get better accuracy. Then The define the target for a model. After data cleaning the dataset will be split into training and test set by using sklearn library.

Data cleaning Dataset was inspected for any missing values (null values). Fortunately, there were no null values present in the dataset, eliminating the need for imputation or removal of missing data.Additionally, the dataset was checked for duplicate entries, ensuring that each data point is unique and eliminating redundancy. So the collected data the collected dataset contains no duplicate values.

Correlation Correlation analysis measures the strength and direction of the relationship between two variables. It provides insights into how changes in one variable may affect the other. A correlation coefficient close to +1 indicates

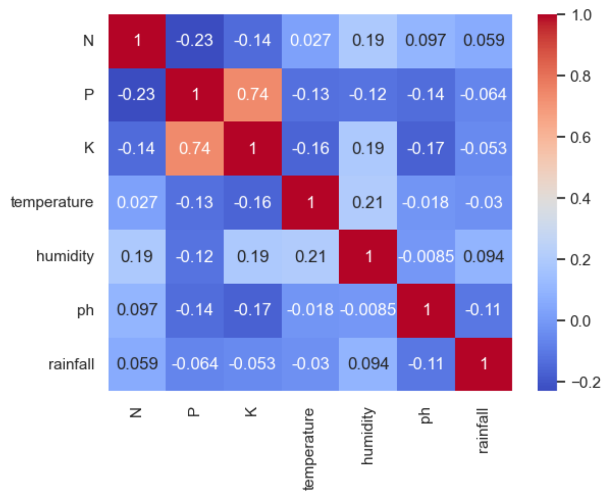

Figure 2: Correlation

a strong positive relationship, while a coefficient near -1 suggests a strong negative relationship. Understanding correlation aids in uncovering patterns and making informed decisions in various fields, from finance to healthcare and beyond.

C. Data Normalization: Normalization is an important step in preprocessing the data to ensure that all features are on a similar scale. This prevents certain features from dominating the model training process due to their larger magnitude. In this project, the features such as N, P, K levels, temperature, humidity, pH, and rainfall were normalized using techniques such as Min-Max scaling or Z-score normalization. Min-Max scaling resales the data to a specific range (typically between 0 and 1), while Z-score normalization standardizes the data to have a mean of 0 and a standard deviation of 1.Normalizing the features helps in improving the performance and convergence of machine learning models, especially those sensitive to the scale of input features, such as K-Nearest Neighbours and Support Vector Machine.

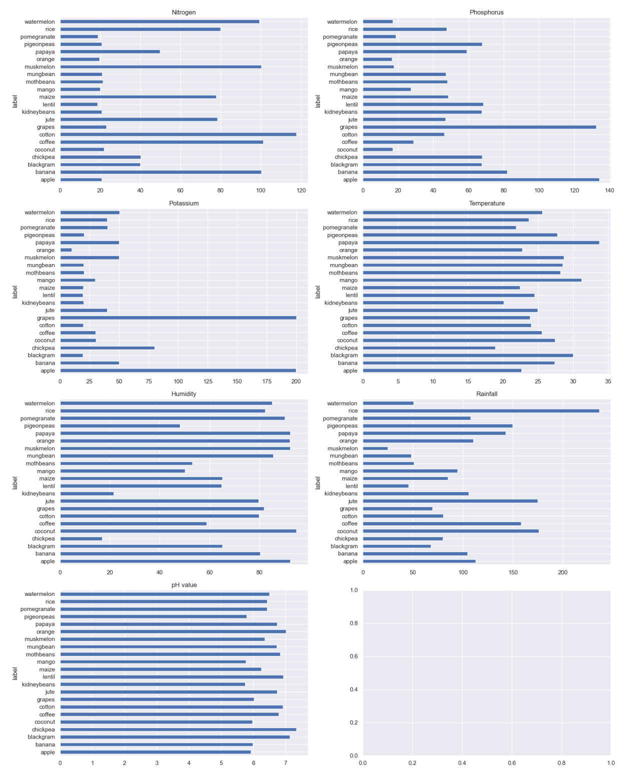

Figure 3: Features visualization

D. Data visualization Above figure 3 is all needs according to the our feature need 'Nirogen', 'Phosphorous', 'Potassium', 'temperature', 'humidity', 'phosphores', 'rainfall',of the very important steps in machine learning is data visualization. In this process, data is checked for missing values, outliers, null values etc. It also helps to understand the data by using the various graphical tools such as bar charts, scat-

```
train_X, test_X, train_y, test_y = train_test_split(
    features, labels,
    test_size=0.2,
    random_state=42,
    stratify=labels
)
```

Figure 4: splitting data

ter plots, line plots etc. we used scatter plots to know the relationship between the two variables.

1.1 Data splitting

The above Fingure.4 represent data splitting step that can be performed after data prepossessing and before model training.split the dataset into training and testing sets to evaluate the performance of machine learning models. This step involves randomly dividing the dataset into two subsets: one for training the model and the other for evaluating its performance. I have set 80 percent of the data is used for training, while the remaining 20 is used in testing.The training set is used to train the model, while the testing set is used to assess its performance on unseen data. This helps in estimating how well the model generalizes to new, unseen instances.

1.2 Model Selection: we have Explore and evaluate various machine learning algorithms suitable for crop recommendation, including but not limited to decision trees, random forests, support vector machines, and Logistic Regressin. we have selected the most appropriate algorithm(s) based on performance metrics such as accuracy, interpret ability, and scalability.

Model Training and Validation: Split the dataset into training and validation sets and train the selected machine learning model(s) using appropriate techniques such as cross-validationto optimize model performance and generalization ability.

2.1 Model Performance and Evaluation metrics

1. Classification: The most common evaluation metrics for classification models are accuracy, precision, recall, and F1-score. These metrics are computed using a confusion matrix, which is a table that summarizes the predictions of a classification model.

2. Accuracy: The proportion of correct predictions over the total number of predictions. It is computed as (TP + TN) / (TP + TN + FP + FN) where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

3. Precision The proportion of true positive predictions over the total number of positive predictions. It is computed as TP / (TP + FP).

4. F1-score: A harmonic mean of precision and recall. It is computed as 2 * (precision * recall) / (precision + recall).

5. Object detection: The most common evaluation metrics for object detection models are mean average precision (mAP) and intersection over union (IoU).

6. map: A measure of the average precision of the model across all classes. It is computed

as the area under the precision-recall curve, which is a graph that shows the trade-off between precision and recall for different threshold values.

I. Logistic regression When the goal variable is categorical, a supervised learning approach called logistic regression is applied. The target variable in Logistic Regression is categorical, therefore we must restrict the range of predicted values. Multiclass logistic regression is also known as multinomial logistic regression. We used this algorithm to predict the crop . This model provide us 0.94 percent accuracy on testing data.

II. Decision Tree Model The Decision Tree model for crop recommendation employs a sophisticated algorithm to analyze various agricultural factors and determine the most suitable crops for a given region. By utilizing either Gini impurity or Entropy as splitting criteria, the model efficiently partitions the feature space to create decision nodes. This process enables it to accurately classify crops based on parameters such as soil type, climate, and nutrient levels. With accuracy rates of 97.955 percent using Gini and 98 percent with Entropy, the model serves as a reliable tool for optimizing crop selection and maximizing agricultural productivity.

III. Random Forest A random forest is a supervised machine learning system that uses decision tree algorithms to construct it. This algorithm determines the outcome based on decision tree predictions. It predicts by averaging the output of various trees. The precision of the result improves as the number of trees grows. It reduces dataset over fitting and improves precision. We applied Random forest classifier for crop name prediction and Random forest regressor yield prediction which give us 99.54 percent accuracy.

Iv. Naïve Bayes algorithm: The Naïve Bayes algorithm for crop recommendation leverages probabilistic principles to classify crops based on a set of input features. By assuming independence between features, Naïve Bayes efficiently computes the probability of each crop belonging to a particular class. With an impressive accuracy of 99.545 percent, this algorithm excels in handling large datasets and diverse agricultural variables. Its simplicity and speed make it a practical choice for crop recommendation systems, offering farmers reliable insights for optimizing yields and resource utilization.

2.2 Model Comparison

Comparing machine learning algorithms The chosen machine learning algorithms are applied to the prepared data to solve the problem. The training dataset is applied for each of the algorithms one by one and trained. Using each of these algorithms, crop yield prediction and prediction of crop name is done. For crop name prediction classification algorithms are used and for crop yield prediction regression algorithms are used. Then, using testing data evaluation of each model is done. It is important to compare the performance of multiple different machine learning algorithms consistently. Each model will have different performance characteristics.We have used five algorithms to train and to build the model after data pre-processing step. After comparing the accuracy of the five ML algorithms the two model which have the

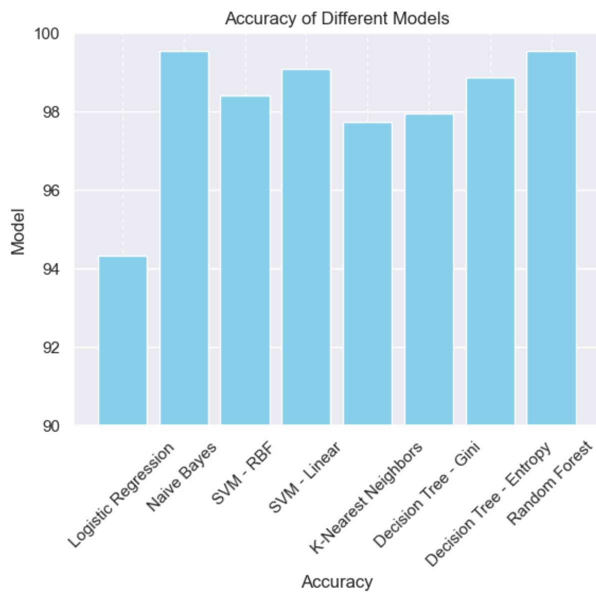

Figure 5: model accuracy

high accuracy will be used on testing data to recommend suitable crop.From the comparison and description of accuracies we can see below that both Random Forest 99.5 and Naïve Bayes algorithms have the highest accuracy of 99.56. So that we have used both Random Forest and Naïve Bayes algorithms to recommend the crop on test data.

v.confusion Matrix of Random Forest Achieving a remarkable accuracy of 99.5, our Random Forest model showcases exceptional performance in classifying crops for recommendation. This accuracy is further reinforced by the confusion matrix, which likely demonstrates strong diagonal dominance, indicating precise predictions across various crop categories. Such high accuracy underscores the reliability and effectiveness of the model in guiding farmers to-

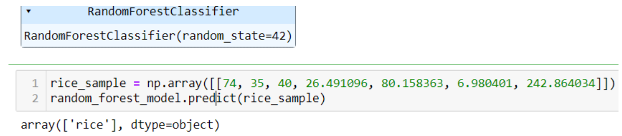

Figure 6: Ramdom Forest prediction

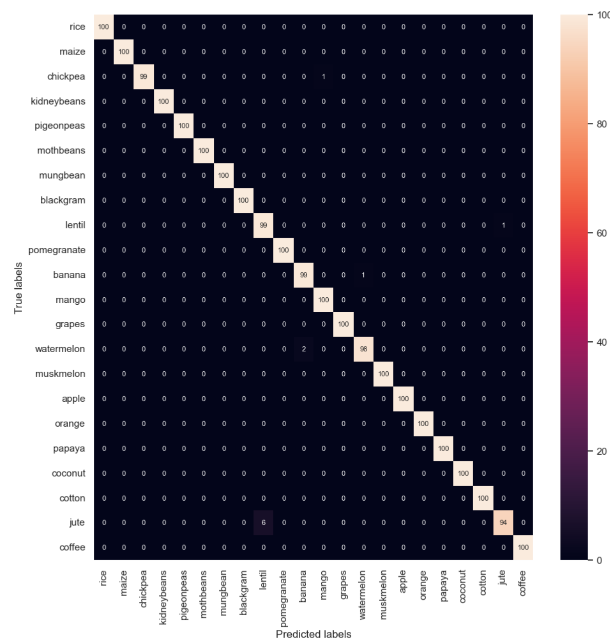

Figure 7: RF confusion matrix

wards optimal crop selections, potentially revolutionizing agricultural decision-making processes with its remarkable precision.

”””””””””””””””””””””’

Results

With impressive accuracy scores across various models,our crop recommendation system demonstrates robust performance in predicting suitable crops for cultivation. The high accuracies, especially evident in Random Forest and

Model	Accuracy Result
Logistic Regression	95.86
Decition Tree	98.68
KNN	98.13
Random Forest	99.5
Naive Bayes Theorem	99.5
SVM	97.68

Figure 8: Model Result

Naive Bayes models, underscore the system's reliability in guiding agricultural decisions. Leveraging advanced algorithms like Decision Trees and SVM, it ensures precise recommendations, potentially optimizing yields and resource utilization for farmers. Such a proficient system holds promise for enhancing agricultural productivity and sustainability, contributing significantly to the agricultural landscape.

Conclusion

The proposed project aims to develop a crop recommendation system using machine learning algorithms to assist farmers in making informed decisions about crop selection. By leveraging agricultural and environmental data, this system seeks to provide personalized recommendations tailored to specific farming conditions and constraints. Successful implementation of this project could help farmers optimize yields, improve profitability, and contribute to sustainable agricultural practices, ultimately enhancing food security and rural livelihoods. A crop recommendation and yield prediction system has been developed successfully using Supervised Machine Learning Approach. The analysis began with data prepossessing and cleaning, followed by exploratory analysis using an agricultural dataset. After that, we used the dataset to train multiple machine learning models, and made various evaluation processes to find the best algorithm. As a result of evaluation, we use Decision tree classifier for crop recommendation and Random Forest regress for yield prediction as they give best accuracy. Our trained algorithm can predict crops based on the specified characteristics, as well as crop yield rate. As this system will cover the widest range of crops, farmers will be able to learn about crops that have never been farmed before and will be able to see a list of all possible crops, which will aid them in deciding which crop to plant. We intent to help people who are planning to invest in farming without any prior knowledge about farming and how much they can make a profit out of it. It is also to help people who are new to farming and try to make their way in learning the practice that has been followed for generations. Thus, our approach can assist farmers in Tamil Nadu, particularly newcomers, in deciding which crop to produce by predicting crop and yield based on local climatic circumstances.

Reference

[1] Gosai, D., Raval, C., Nayak, R., Jayswal, H., Patel, A. (2021). Crop recommendation system using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(3), 558-569.

[2] Ikram, A., Aslam, W., Aziz, R. H. H., Noor, F., Mallah, G. A., Ikram, S., ... Ullah, I. (2022). Crop Yield Maximization Using an IoT-Based Smart Decision. Journal of Sensors, 2022, 1-15.

[3] Sundari, V., Anusree, M., Swetha, U. (2022). Crop recommendation and yield prediction using machine learning algorithms. World Journal of Advanced Research and Reviews, 14(3), 452-459.

[4] Pande, S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., SHAURYA, K. (2021, April). Crop recommender system using machine learning approach. In 2021 5th international conference on computing methodologies and communication (ICCMC) (pp. 1066-1071). IEEE.

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

[6] Pande, S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., SHAURYA, K. (2021, April). Crop recommender system using machine learning approach. In 2021 5th international conference on computing methodologies and communication (ICCMC) (pp. 1066-1071)IEEE.

[7] Modi, D., Sutagundar, A. V., Yalavigi, V., Aravatagimath, A. (2021, October). Crop recommendation using machine learning algorithm. In 2021 5th International Conference on Information Systems and Computer Networks (ISCON) (pp. 1-5). IEEE.

[8] Nischitha, K., Vishwakarma, D., Ashwini, M. N., Manjuraju, M. R. (2020). Crop prediction using machine learning approaches. International Journal of Engineering Research Technology (IJERT), 9(08), 23-26.

[9] Patil, P., Panpatil, V., Kokate, S. (2020). Crop prediction system using machine learning algorithms. Int. Res. J. Eng. Technol.(IRJET), 7(02).

[10] riyadharshini, 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.