Crop and Fertilizer Management: A Machine Learning Approach for Optimized Yield by Season

Lav A. Chaudhari (202101135)
Mentor: Prof. Srimanta Mandal

***Abstract*—Machine learning techniques are increasingly being harnessed to address the complexities of modern agriculture. In regions like Gujarat, Rajasthan, and Uttar Pradesh, where farming is a significant source of employment, the impact of environmental changes on agricultural productivity is profound—factors such as humidity, rainfall, sunlight, soil type, and temperature influence crop outcomes. Recognizing the crucial role of accurate information in agricultural decision-making, our research integrates insights from multiple sources, including agricultural metrics and environmental factors.**

By leveraging machine learning algorithms such as Decision trees, random forests, k-nearest neighbors, Logistic Regression, SVM, and Naïve Bayes, our model offers personalized recommendations for crop selection and fertilizer use based on localized conditions. These Models identify similarities between farm features to enhance recommendations such as crop and fertilizer. Through a holistic approach that considers soil characteristics, climate, crop history, and fertilizer usage patterns, our model provides farmers with actionable insights to optimize yield while minimizing resource use.

Central to our proposal is developing a platform that empowers farmers with easy-to-use tools for crop selection and planting guidance, like which fertilizer they should use. By harnessing the power of machine learning, our approach not only supports modern agricultural practices but also promotes sustainability and enhances the overall profitability of agriculture in diverse climatic regions such as India.

***Index Terms*—Crop & Fertilizer recommendation, Nitrogen-Phosphorus-Potassium (NPK), environmental factors, Machine Learning (ML), K-Nearest Neighbour (KNN), SVM, Logistic Regression, Trees.**

I. Introduction

INDIA, often revered as the land of agriculture, stands proudly among the top three global producers of numerous agricultural commodities. With agriculture being the cornerstone of its economy and livelihood for millions, Indians dedicate much of their time to cultivating the land. However, amidst a competitive environment and the ever-changing climate, agriculture in India faces both challenges and opportunities. The agricultural landscape of India is deeply intertwined with environmental factors such as temperature variations, soil composition, humidity levels, rainfall patterns, and the availability of essential nutrients like nitrogen, potassium, and phosphorus. These factors and distinct seasonal variations delineate the cropping patterns across the nation.

By harnessing the power of machine learning algorithms and leveraging vast datasets encompassing environmental parameters and geographical nuances, this project endeavors to empower farmers with predictive insights. By analyzing crucial factors such as temperature, rainfall, humidity, soil type, nutrient levels, and geographical location, the system strives to optimize crop selection for maximizing productivity and profitability.

Using the right fertilizers is one of the main things important for growing healthy crops and getting good yields. So, our project is expanding to include giving advice on which fertilizers to use. This advice is crucial because it affects how fertile the soil is, how healthy the crops are, and how much produce you get. But figuring out the best fertilizers to use can be tricky. It depends on what kind of soil you have, what crops you're growing, the weather, and what nutrients your crops need at different times.

We're simplifying this process with fancy computer programs and data analysis techniques. We aim to make fertilizer advice more accurate and personalized for each farmer. We'll do this by looking at lots of data about soil, nutrients, the types of crops, and the environment. Then, we'll give each farmer a specific plan for which fertilizers to use to grow their crops better while also being good for the environment.

II. Literature Survey

1) Doshi et al. (2019): "Smart Farming using IoT, a solution for optimally monitoring farming conditions."
Doshi and colleagues propose a solution for intelligent farming utilizing Internet of Things (IoT) technology to monitor farming conditions in real time. Farmers can collect data on various parameters such as soil moisture, temperature, and humidity by integrating sensors and IoT devices. This data-driven approach enables farmers to make informed decisions regarding irrigation, fertilization, and other crop management practices, leading to optimized yields and resource efficiency.
2) Thilakarathne et al. (2022): "A cloud-enabled crop recommendation platform for machine learning-driven precision farming."
Thilakarathne et al. presents a cloud-enabled platform based on machine-learning algorithms for crop recommendations. By analyzing data on soil characteristics, climate conditions, and crop requirements, the platform generates personalized recommendations for farmers, guiding them in crop selection, fertilization, and irrigation management. This machine learning-driven approach facilitates precision farming, enhancing crop yields and sustainability.
3) Kumar et al. (2015): "Crop Selection Method to maximize crop yield rate using machine learning technique."
Kumar and co-authors propose a machine learning-based method for crop selection to maximize yield rates. By

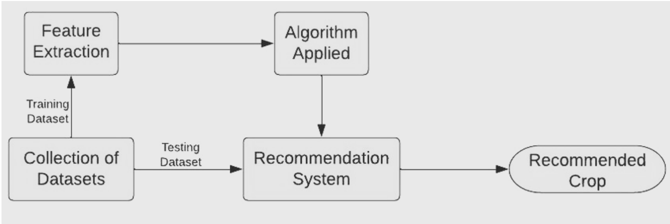

Fig. 1. General block diagram of the application process.

	A	B	C	D	E	F	G	H
1	N	P	K	temperature	humidity	ph	rainfall	label
2	90	42	43	20.87974371	82.00274	6.502985	202.9355	rice
3	85	58	41	21.77046169	80.31964	7.038096	226.6555	rice
4	60	55	44	23.00445915	82.32076	7.840207	263.9642	rice
5	74	35	40	26.49109635	80.15836	6.980401	242.864	rice
6	78	42	42	20.13017482	81.60487	7.628473	262.7173	rice
7	69	37	42	23.05804872	83.37012	7.073454	251.055	rice
8	69	55	38	22.70883798	82.63941	5.700806	271.3249	rice
9	94	53	40	20.27774362	82.89409	5.718627	241.9742	rice

Fig. 2. Dataset for Crop recommendation.

analyzing historical data on crop performance, weather patterns, and soil conditions, the model predicts the most suitable crops for cultivation in specific regions. This approach assists farmers in making informed decisions regarding crop selection, ultimately leading to higher yields and improved agricultural productivity.

III. Proposed Approach

For training the crop recommendation model and the fertilizer recommendation model, we follow this architecture, which has four major parts:

1) Dataset data collection
2) Preprocessing and Feature extraction
3) Recommendation System
4) Recommendation of crops or fertilizer

A. Dataset

Crop Recommendation Model Dataset:

- Parameters: Nitrogen(N), pH value of soil, Phosphorous(P), Stickiness, Potassium(K), Temperature, Rainfall

- Crops: Rice, Lentil, Maize, Kidney beans, Moth peas, Pigeon beans, Mung bean, Black gram, Watermelon, Pomegranate, Papaya, Orange, Banana, Apple, Mango, Muskmelon, Grapes, Coconut, Cotton, Jute, Coffee

Fertilizer Recommendation Model Dataset:

- Parameters: Temperature, Humidity, Moisture, Soil Type, Crop Type, Nitrogen, Potassium, Phosphorous

- Fertilizers: 10-26-26, 14-35-14, 17-17-17, 20-20, 28-28, DAP, Urea

Here is the snapshot of the both datasets:

	A	B	C	D	E	F	G	H	I
1	Temparatu	Humidity	Moisture	Soil Type	Crop Type	Nitrogen	Potassium	Phosphorous	Fertilizer Name
2	26	52	38	Sandy	Maize	37	0	0	Urea
3	29	52	45	Loamy	Sugarcane	12	0	36	DAP
4	34	65	62	Black	Cotton	7	9	30	14-35-14
5	32	62	34	Red	Tobacco	22	0	20	28-28
6	28	54	46	Clayey	Paddy	35	0	0	Urea
7	26	52	35	Sandy	Barley	12	10	13	17-17-17
8	25	50	64	Red	Cotton	9	0	10	20-20
9	33	64	50	Loamy	Wheat	41	0	0	Urea

Fig. 3. Dataset for Fertilizer recommendation.

B. Preprocessing and Feature extraction

Preprocessing lays the foundation for accurate crop and fertilizer recommendation systems by filtering incomplete or inconsistent data and standardizing information. Through techniques like outlier detection and data cleaning, noise is minimized, ensuring the integrity of the dataset. Features crucial for crop and fertilizer recommendation, such as NPK levels, temperature, humidity, and soil type, are then extracted. These features enable the identification of suitable crops based on nutrient requirements and environmental preferences. Additionally, they aid in recommending the right type and dosage of fertilizers tailored to specific soil and crop conditions. Overall, robust preprocessing and feature extraction are indispensable for optimizing agricultural practices and maximizing crop yields while minimizing resource wastage.

C. Recommendation System

In this part, with the help of the training dataset, we will train our machine learning models like KNN, GaussianNB, Decision Tree, SVM, Random Forest (RF), Logistic Regression (LR), etc, to recommend the crop and the fertilizer.

D. Architecture of the recommendation system

System architecture serves as our project's blueprint, encapsulating its structure and behavior. It delineates the system's logic and inter-process dependencies, comprehensively representing its functionality. Visualizing the relationship between processes offers invaluable insights into the system's operation. Our proposed model's system architecture, depicted in Figure 4, embodies this conceptual framework, guiding the development and understanding of our project. A robust system architecture not only defines the system's essence but also facilitates its examination and comprehension graphically, paving the way for efficient design and implementation.

E. Different Machine Learning Models

K-Nearest Neighbors (KNN): KNN is a simple and effective supervised learning algorithm used for classification and regression tasks. It predicts the class of a sample by a majority vote of its k nearest neighbors in the feature space.
Equation (for classification):

$$y = \max(y_{i1}, y_{i2}, ..., y_{ik})$$

Gaussian Naive Bayes (GNB): GNB is a probabilistic classifier based on Bayes' theorem, assuming independence between features. It's particularly useful for text classification

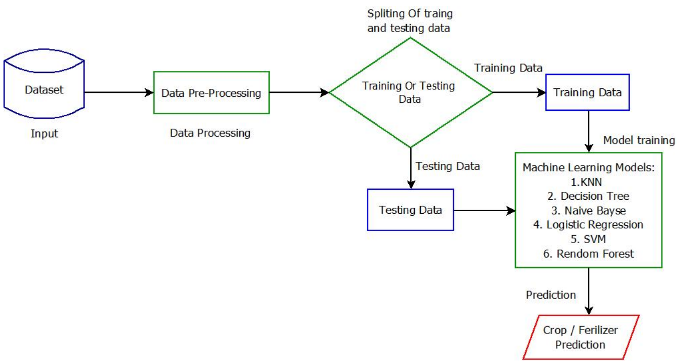

Fig. 4. System Architecture .

tasks.

Equation (for classification):

$$P(y|x_1, x_2, ..., x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1, x_2, ..., x_n)}$$

Decision Tree with Entropy: Decision trees are a popular supervised learning algorithm for classification and regression tasks. Decision trees with entropy use information gained to split nodes to minimize disorder within each node.

Calculate Entropy: Calculate the entropy of the dataset before any split. Entropy measures the randomness or impurity of the dataset.

$$H(S) = -\sum_{c \in C} p_c \log_2 p_c$$

where:
- $H(S)$ is the entropy of the dataset S.
- p_c is the probability that an element in S belongs to class c.

Calculate Information Gain: For each feature F, calculate the Information Gain $IG(S, F)$.

$$IG(S, F) = H(S) - \sum_{f \in F} \frac{|S_f|}{|S|} H(S_f)$$

where: - $IG(S, F)$ is the Information Gain of splitting dataset S based on feature F. - $|S_f|$ is the number of elements in S where feature F has value f. - $|S|$ is the total number of elements in S. - $H(S_f)$ is the entropy of the subset of S where feature F has value f.

Select Feature with Maximum Information Gain: Choose the feature F that yields the highest Information Gain as the root node of the decision tree.

The feature selected as the root node is the one that best separates the dataset into subsets with the least amount of entropy (i.e., the most homogeneous subsets). This process is repeated recursively for each subset until the tree is fully grown or a stopping criterion is met.

Support Vector Machine (SVM): SVM is a powerful supervised learning algorithm used for classification and regression tasks. It finds the hyperplane that best separates classes in the feature space by maximizing the margin between them.

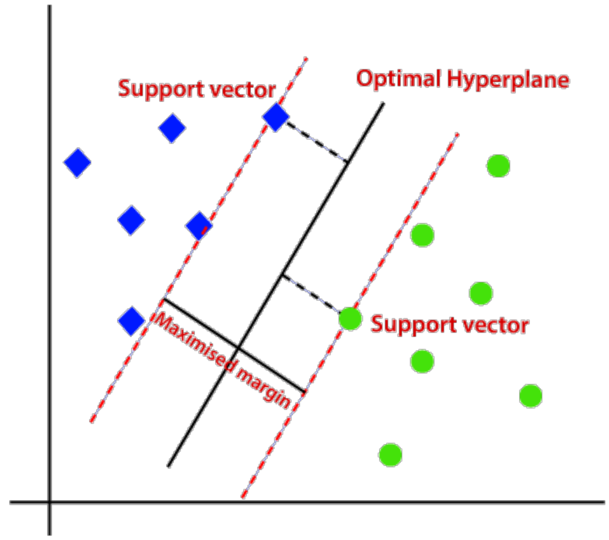

Fig. 5. SVM Classification.

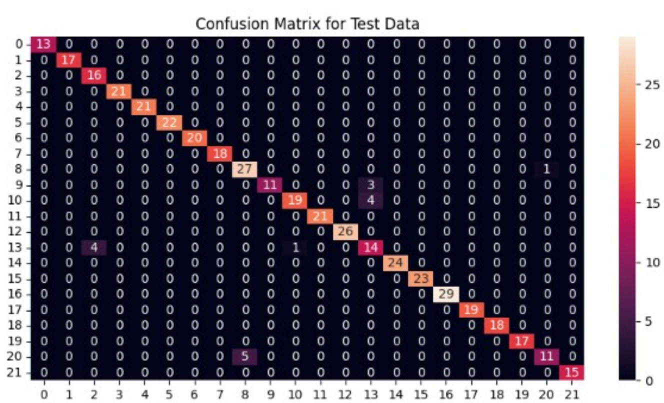

Fig. 6. Confusion Matrix for crop recommendation System.

Equation (for linear SVM):

$$\min_{w,b} \frac{1}{2}||w||^2 \text{ subject to } y_i(w \cdot x_i + b) \geq 1$$

Random Forest: Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees.

No single equation captures Random Forest; it's an ensemble of decision trees.

IV. Result

After training both models, we obtain results based on the training data and analysis. These results include statistical metrics such as accuracy, F1 score, precision, recall, and the confusion matrix.

A. Crop Recommendation Results

Confusion Matrix for crop recommendation system:

For different machine learning models, we get these statistical results:

TABLE I
PERFORMANCE METRICS OF DIFFERENT MODELS

Model	Accuracy	Precision	Recall	F1 Score
KNN	0.9891	0.9892	0.9891	0.9891
Decision Tree	0.9591	0.9626	0.9546	0.9563
Naive Bayes	0.9909	0.9943	0.9886	0.9905
Support Vector Machine	0.9773	0.9781	0.9776	0.9777
Logistic Regression	0.9523	0.9540	0.9487	0.9504
Random Forest	0.9909	0.9929	0.9891	0.9904

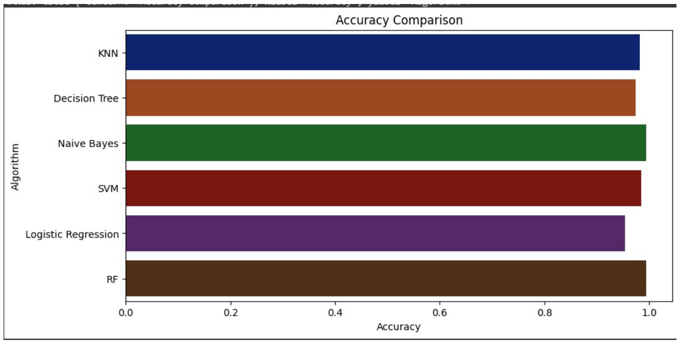

Fig. 7. Accuracy for crop recommendation System.

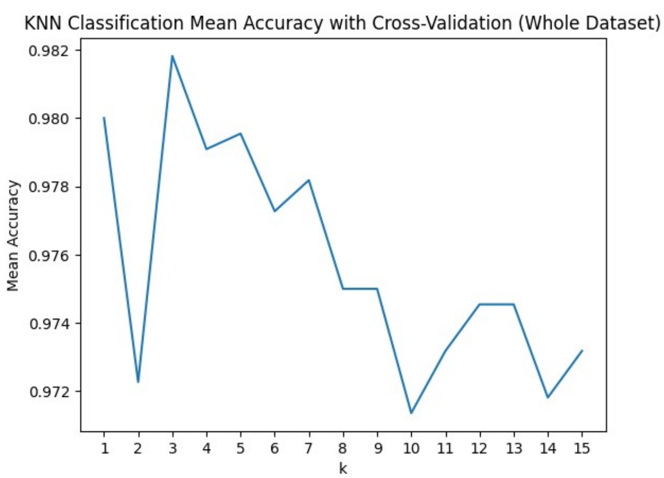

Fig. 8. KNN algorithm accuracy vs k.

For the KNN classification model, we consider various closest neighbors, and then the results show that best k=3, where we get the best accuracy.

For the Decision Tree classification model, we consider various tree depths, and then the results show that the best d=6, where we get the best accuracy for d=7, we get overfit.

B. Fertilizer Recommendation System

Confusion Matrix for fertilizer recommendation system:

For different machine learning models, we get these statistical results:

For the KNN classification model, we consider various closest neighbors, and then the results show that best k=1, where we get the best accuracy.

For the Decision Tree classification model, we consider various tree depths, and then the results show that the best d=4, where we get the best accuracy for d=5, we get overfit.

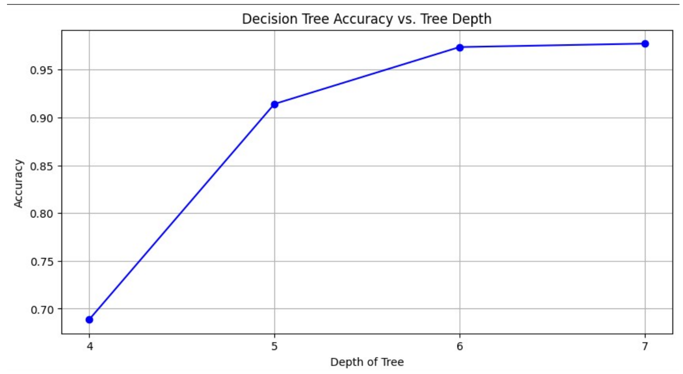

Fig. 9. Decision Tree algorithm accuracy vs d.

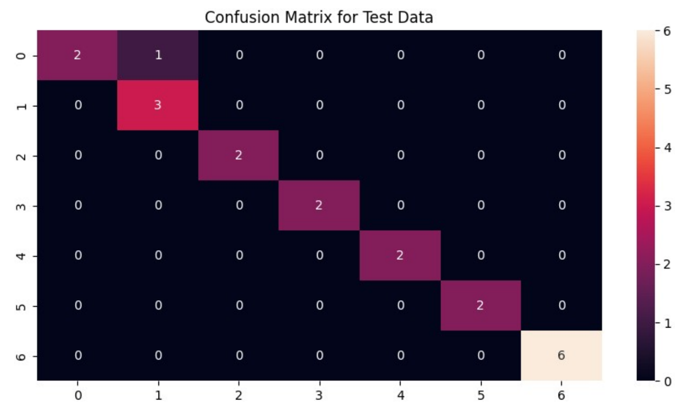

Fig. 10. Confusion Matrix for fertilizer recommendation System.

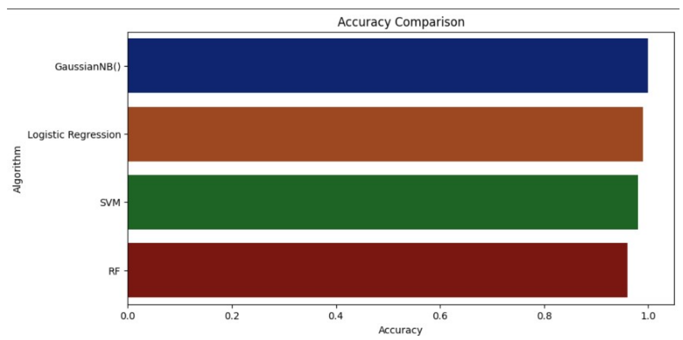

Fig. 11. Accuracy for Fertilizer recommendation System.

V. FUTURE WORK

Our project aims to integrate a comprehensive crop recommendation system with our existing fertilizer recommendation system to provide farmers with holistic guidance for optimizing crop yield. This integration will involve analyzing factors such as soil characteristics, climate conditions, and market demand to offer personalized crop suggestions tailored to each farmer's unique circumstances. Additionally, we plan to expand the project scope to include an irrigation management component, leveraging sensor data and weather forecasts to optimize irrigation scheduling and minimize water wastage.

TABLE II
PERFORMANCE METRICS OF DIFFERENT MODELS

Model	Accuracy	Precision	Recall	F1 Score
KNN	0.9184	0.9905	0.9796	0.9841
Decision Tree	0.9700	0.6429	0.7143	0.6667
Naive Bayes	1.0000	1.0000	1.0000	1.0000
SVM	0.9500	0.9643	0.9524	0.9510
Logistic Regression	1.0000	1.0000	1.0000	1.0000
Random Forest	0.8500	0.8714	0.8333	0.8214

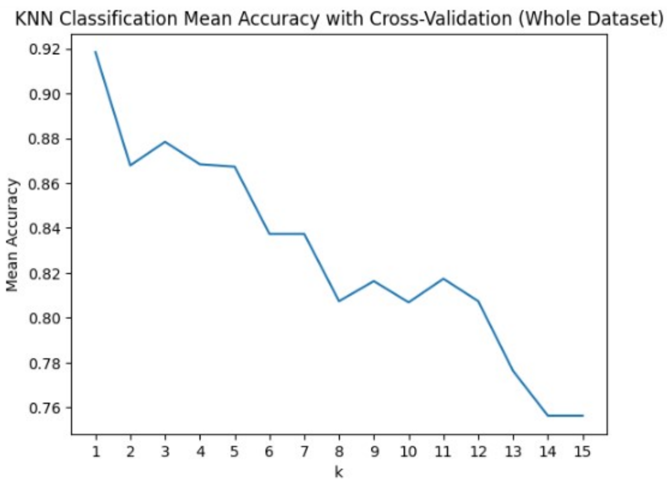

Fig. 12. KNN algorithm accuracy vs k.

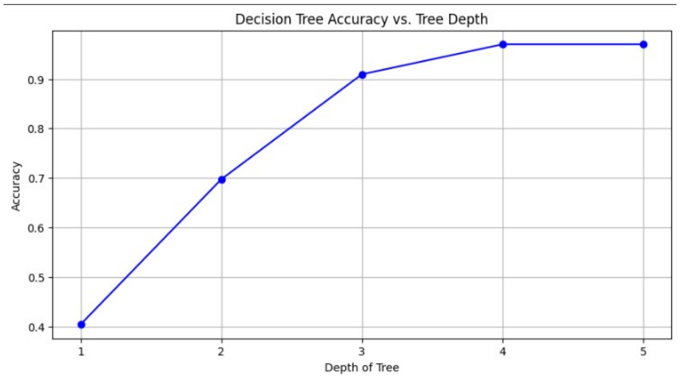

Fig. 13. Decision Tree algorithm accuracy vs d.

VI. CONCLUSION

In this study, we employed a machine learning approach to optimize crop yield through tailored fertilizer management across different seasons. Our results demonstrate the efficacy of machine learning models in predicting optimal fertilizer usage, leading to enhanced crop yield.

Key findings reveal that our models, particularly Naive Bayes and Random Forest, exhibit remarkable accuracy and precision in predicting the most suitable fertilizer application for varying seasonal conditions. This suggests that machine learning algorithms can effectively adapt to the dynamic nature of agricultural environments, thereby assisting farmers in making informed decisions to maximize yield.

Furthermore, our study underscores the importance of season-specific fertilizer management strategies. By accounting for seasonal variations in climate, soil conditions, and crop requirements, our models offer personalized recommendations that align with the unique needs of each growing period.

Combining machine learning techniques in crop and fertilizer management holds significant promise for sustainable agriculture. By leveraging data-driven insights, farmers can optimize resource allocation, reduce environmental impact, and ultimately increase productivity to meet the growing demands of global food security.

VII. REFERENCE

1) Aditya Motwani, Param Patil, Vatsa Nagaria, Shobhit Verma, Sunil Ghane. ”Soil Analysis and Crop Recommendation using Machine Learning”. In *IEEE Conference*, 2022.
2) Kamatchi, S. Bangaru, and R. Parvathi. ”Improvement of Crop Production Using Recommender System by Weather Forecasts.” *Procedia Computer Science* 165 (2019): 724-732.
3) Shilpa Mangesh Pande, Dr. Prem Kumar Ramesh, Anmol, B.R Aishwarya, Karuna Rohilla, Kumar Shaurya. ”Crop Recommender System Using Machine Learning Approach”. In *IEEE Conference*, 2021.
4) Savla, Anshal, Nivedita Israni, Parul Dhawan, Alisha Mandholia, Himtanaya Bhadada, and Sanya Bhardwaj. ”Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture.” In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pp. 1-7. IEEE, 2015.
5) Doshi, Jash, Tirthkumar Patel, and Santosh kumar Bharti. ”Smart Farming using IoT, a solution for optimally monitoring farming conditions.” *Procedia Computer Science* 160 (2019): 746-751.
6) Kumar, Rakesh, M. P. Singh, Prabhat Kumar, and J. P. Singh. ”Crop Selection Method to maximize crop yield rate using machine learning technique.” In *2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)*, pp. 138-145. IEEE, 2015.