Krashignyan: A Farmer Support System

Pragati Kanchan,
Dept. of CSE, MIT School of Engineering,
MIT ADT University
Pune, India
pragatikanchan04@gmail.com

NikhilKumar Shardoor,
Professor, Dept. of CSE, MIT School of Engineering,
MIT ADT University
Pune, India

***Abstract*—Agriculture is the primary component of the Indian economy. It is the primary source of food supply and is essential to our livelihoods. The majority of Indians rely on agriculture for their employment. Agriculture production declines as a result of unpredictable weather, wrong selection of crops, unbalanced fertilizer use, and a lack of market awareness. Farmers face numerous challenges in traditional farming, and many times, farmers fail to select the appropriate crop for cultivation. Crop growth is affected by a variety of factors such as weather, soil parameters, and fertilizers. A crop recommendation system is proposed in this paper to assist farmers in selecting the appropriate crop based on the location, weather data, crop sowing season, and soil parameter. Various Machine Learning techniques, such as Decision Tree (DT), Random Forest (RF), Gaussian Naive Bayes, and XGBoost Classifier methods, were used for recommendation. The XGBoost classifier gives the best results with a 97% accuracy, hence the final model was developed using the XGBoost classifier. This system will help farmers in selecting the best crop for their fields while increasing agricultural yield.**

Keywords— Crop Recommendation, Decision Tree, Random Forest, Naive Bayes, XGBoost Classifier.

I. INTRODUCTION

India is an agricultural country. Agriculture is considered the primary occupation of the majority of Indians. It contributes about 17% to the total GDP and employs more than 50% of the workforce[16]. Agriculture is the prime source of food production and also supplies raw materials to various industries. Population growth drives up food demand[2]. Farming needs to be done in an effective way to increase food production. Farmers face numerous challenges in traditional farming, and many times, farmers fail to select the appropriate crop for cultivation. A farmer selects a crop based on the recommendations of other farmers without knowing whether or not that crop is appropriate for his farm, which sometimes resulting in crop loss. There are different factors like weather conditions, soil parameters, use of fertilizers that affect crop growth. With uncertain nature, it gets difficult for farmers to grow a healthy crop. If the farmer knows in advance which crop is best for his farm based on these factors, he will definitely benefit in farming. With the help of technology, it is possible to recommend the best crop for a farmer's field based on environmental, soil, and other parameters. Nowadays, technological advancements in the agriculture field help farmers in increasing crop yields. In this paper, a crop recommendation system is proposed using machine learning technology to help farmers in choosing the best crop for their farm. Machine Learning (ML) is one of the most widely used technologies in the current technological era. It is the subfield of AI, and its widely used for analysis, classification, regression, and forecasting purposes. Nowadays, in various disciplines such as healthcare, finance, agriculture, automotive, machine learning is used. Machine learning techniques assist in the development of an intelligent system that learns from experience and makes data-driven decisions. In the agriculture field, different machine learning techniques, such as supervised, unsupervised techniques, are used to analyze different data patterns for better decision-making. Krashignyan: A Farmer Support System is proposed in this paper to help farmers in selecting the best crop based on their location, weather data, crop sowing season, and soil parameter. Various Machine Learning techniques, such as Decision Tree (DT), Random Forest (RF), Gaussian Naive Bayes, and XGBoost Classifier methods, were used for recommendation. XGBoost classifier gives the best results, so the final model was built using an XGBoost classifier. This system will help farmers in selecting the best crop for their fields while increasing agricultural yield.

II. LITERATURE SURVEY

Archana, K. et al.[1] in this paper developed a system using a voting-based ensemble classifier to recommend the most appropriate crop, fertilizer, and to predict crop yield. The proposed system recommends a suitable crop for cultivation based on Nitrogen(N), Potassium(K), Phosphorus(P), soil type, soil texture, land type, pH, electric conductivity, and temperature parameters. The proposed system also suggest fertilizers to farmers for recommended crop and predict crop yield. The author also developed a crop rotation module, in which alternative crops except predicted crops were suggested to farmers based on crop cultivation seasons.

Chaudhari, A. et al.[2] in this paper proposed a platform for crop recommendation and its optimal seed price using shopbot. A dataset with the following parameters - crop, rainfall, temperature, season, year, production, area, and the location used for crop recommendation. Dataset was normalized and split into a 7:3 ratio. Various data mining techniques were applied to the dataset, and there accuracy score was calculated. A suitable crop is recommended using the classification algorithm. In their work, for optimal pricing of seeds, web scraping was performed using the selenium framework. The authors found that the Decision Tree Classifier gave better accuracy for crop recommendation.

Kelvin Tom Thomas et al.[3] in his work proposed a recommendation system, which recommends the most suitable crop based on soil parameters to the farmers. The different machine learning techniques like Decision Tree(DT), K-Nearest Neighbor(KNN), KNN with cross-validation, Naive Bayes were used to train a model. A dataset with the following soil parameters - soil NPK values, soil pH values, and most suitable crop was used for prediction. The author found that KNN with cross-validation

gives better results. The final system has built using KNN. Based on the record present in their dataset value for 'k' has taken as 10. The proposed model predicts the most suitable crop by taking soil NPK and soil pH values.

Lavanya B et al.[4] in this paper presented a crop prediction system using data analytics techniques in the form of an android application. The dataset used in this study includes the following parameters: weather, soil type, soil pH, previous three harvests, fertilizers, season, and market demand. Initially proposed system take input from farmers like details of previous three harvest, availability of water and the data of weather conditions, temperature and soil condition are taken based on users location In their system, based on collected data from users and the dataset available, a suitable crop for cultivation was predicted by applying machine learning techniques.

Nischitha K. et al.[5] in this paper designed a crop recommendation system for particular land using machine learning techniques. In their work, developed system that predict the crop based on weather data and soil parameters, like temperature, humidity, soil pH, and rainfall. The dataset includes parameters such as temperature, humidity, rainfall, crop data, soil pH, and NPK values. For rainfall prediction, separate rainfall prediction model was built using Support Vector Machine Classifier with Radial Basis Function (RBF) kernel to predict rainfall. For crop recommendation, input parameters were entered manually from the user or taken by sensors, and the suitable crop predicted using the Decision Tree algorithm.

Kulkarni, N. H.[6] et al. in this paper designed a crop recommendation system using the Ensembling Technique. The proposed model predicts suitable crop based on soil dataset. The dataset was preprocessed and split into a 75:25 ratio. The ensemble model was build using Random Forest, Naive Bayes, and linear Support Vector Machine. The result of individual learners was combine based on the majority voting technique.

Mythresh A et al.[7] in this paper developed a crop prediction application using machine learning technology. The proposed system predicts the crop based on land suitability. Various parameters such as soil parameters, weather parameters, temperature, rainfall, and so on were considered. The naive Bayes technique was used for crop prediction.

Pudumalar, S. et al.[8] in this paper proposed a recommendation system using ensemble learning technique for crop recommendation based on soil characteristics. The ensemble model was build using Random Forest, Naive Bayes, K- Nearest Neighbors, and Chi-squared Automation Interaction Detection (CHAID). The result of individual learners was combine based on the majority voting technique. In their work for crop recommendation, various soil parameters like depth, texture, pH, soil color, permeability, water holding, drainage, and erosion were considered.

Shinde, M. et al.[9] designed a system to recommend a crop and fertilizers to farmers to increased crop yield. Farmers can use this system through an android based mobile application. This system suggests fertilizers to farmers based on past purchased history using the Apriori algorithm. For crop recommendation, the Random Forest algorithm was used. The designed application allows the farmer to purchase fertilizer from the shopping portal.

Kumar, A. et al.[10] in this paper proposed a recommendation system to predict the suitable crop and to suggest pest control techniques to farmers. In his work Decision Tree Algorithm, Support Vector Machine Classification Algorithm, and Logistic Regression Algorithm applied to crop prediction model. The author found that SVM Classification gave better accuracy as compared to other algorithms used. The proposed model predicts the most suitable crop based on soil parameters.

Banavlikar, T. et al.[11] proposed a neural network-based crop recommendation system. The following parameters like soil moisture, temperature, and humidity are used for crop recommendation. Three sensors, DHT11, humidity sensor, and soil moisture sensor, were used in the proposed system to measure temperature, humidity, and soil moisture.

Zeel Doshi et al.[12] presented an intelligent crop recommendation system. The proposed system recommend crop based on environmental parameters, farm geolocation, and soil characteristics. Also implemented another model for rainfall prediction using Linear Regression, which predicts rainfall for the next 12 months. Decision Tree, K Nearest Neighbors, Random Forest, and Neural Network techniques were used for crop recommendation. Their proposed model provides a map visualization feature that helps the farmer decide which crop to grow.

Kalimuthu, M. et al.[13] in this paper proposed a crop prediction system using machine learning technique in the form of an android application. The dataset used in this study includes the following parameters: soil moisture, soil pH, rainfall, temperature, humidity. Dataset was split into a 7:3 ratio. A Naive Bayes Gaussian classifier with boosting algorithm was used for prediction.

Patil, A. et al.[14] introduced crop recommendation system using a machine learning classification algorithm. The dataset used in this study includes the climatic parameters, soil parameters, and production parameters. Three classification algorithms like Decision Tree, KNN, and Naive Bayes algorithm, were used and their performance was compared.

Dhruv Piyush Parikh et al.[15] in this paper, a crop recommendation model has been built using machine learning technology. The proposed system recommends a suitable crop based on Nitrogen(N), Potassium(K), Phosphorus(P), temperature, humidity, pH, and rainfall parameters. The different machine learning techniques like Support Vector Machine(SVM), Logistic Regression, and Random Forest Classifier were used to train a model. The author found that Random Forest Classifier gives better results. The final system has been built using Random Forest Classifier.

We learned from the survey that both soil and weather parameters are important for crop prediction, and the survey also helped us in selecting the best machine learning techniques for prediction.

III. PROPOSED MODEL

Figure 1 depicts the system architecture of the crop recommendation model. Whereas in the first stage, the dataset is collected and preprocessed before being used to

train the machine learning model. The crop dataset with the following parameters - Temperature, Humidity, Soil Moisture, Soil Type, Crop Season, and Crop label is considered for a recommendation. The machine learning model was trained using various machine learning algorithms, and the model with the best prediction results was chosen. Once training is completed and the model built, the built model is deployed using the web application framework Flask. These all process is done at the backend. The user enters data into a web application that serves as the user interface. The web app makes an API request, and the loaded model gives a response.

TABLE I. DIFFERENT MACHINE LEARNING TECHNIQUES USED FOR CROP RECOMMENDATIONS AND THE RESULTS OF VARIOUS PAPERS.

AUTHOR	DATASET	TECHNIQUES	PARAMETERS	ACCURACY
Archana, K. et al. [1]	Agriculture Crop Soil Dataset.	Voting-based Ensemble Classifier: Chi-square Automatic Interaction Detection, Random Forest, Naive Bayes.	Soil Parameters - N, P, K, soil type, soil texture, land type, temperature, pH, electric conductivity.	92%
Chaudhari, A. et al. [2]	Climatic conditions dataset. data.gov.in	Data Mining Techniques : Classifiers - Decision Tree, Extra Tree, ExtraTreesClassifier (with ensemble), Random Forest, K-Neighbors, Web Scraping.	Year, area, location, season, crop, rainfall, temperature, production.	DecisionTreeClassifier: 87%, RandomForestClassifier: 86%
Kelvin Tom Thomas et al.[3]	–	Machine Learning Techniques - SVM, KNN, KNN with cross validation, DT, Naive Bayes.	Soil value - N, P, K, and pH , Corresponding crops.	KNN: 85%, DT: 81% , Naive Bayes: 82% , SVM: 78%
Lavanya, B. et al. [4]	–	Data Analytics Technique.	Previous three harvests, Soil type, soil pH, weather, season , fertilizers, market demand.	–
Nischitha K. et al. [5]	Government website, V C Farm Mandya, Weather department.	SVM Classifier, Decision Tree.	Soil pH, and NPK values, temperature, humidity, rainfall, crop data.	–
Kulkarni, N. H. et al. [6]	Dataset consists of climatic details, soil chemical and physical properties. data.gov.in	Ensemble model with majority voting technique. Random Forest, Naive Bayes, and Linear SVM.	Soil type, soil pH values, NPK content of soil, Surface temperature, Avg. rainfall, Surface temperature, and sowing season.	Average accuracy 99.91%
Mythresh A, et al.[7]	Soil site suitability dataset, Crop requirement dataset.	Naive Bayes.	Slope, soil texture, depth, gravel, rocky, erosion, drainage and pH mapped to survey IDs, conditions required for crop growth.	60%
Pudumalar, S. et al. [8]	Madurai district.	Ensemble model with majority voting technique: KNN, Random tree, Naive Bayes.	Depth, texture, pH , soil color, permeability, drainage, water holding and erosion, crop.	–
Kumar, A. et al. [10]	Soil, Crop Dataset.	Logistic regression, SVM classification, Decision Tree.	Avg. rainfall, temperature, soil color, pH .	SVM: 89.66%
Banavlikar, T. et al.[11]	Dataset consists of Soil Moisture, Humidity, Temperature.	Neural Network.	Soil Moisture, Humidity, Temperature.	-
Doshi, Z. et al. [12]	India Agriculture and Climate Data Set.	Machine learning techniques: Multi-label classification KNN, Decision Tree,	Precipitation, temperature, location parameters, soil and meteorological parameters,	Rainfall Predictor linear regression: 71% Decision Tree: 90.20% K-NN: 89.78%

		Random Forest, Neural Network, Linear Regression.	soil type, aquifer thickness, soil pH, thickness of topsoil.	Random Forest: 90.43% Neural Network: 91.00%
Kalimuthu, M. et al.[13]	Seed Data, Soil Nature, and Climatic conditions Data.	Naive Bayes Gaussian classifier.	Soil moisture, soil pH, rainfall, temperature, humidity.	97%
Dhruv Piyush Parikh et al.[15]	Crop Recommendation Dataset. Kaggle	SVM , Logistic regression, Random Forest classifier.	Temperature, humidity, N, P, K, pH, rainfall.	-

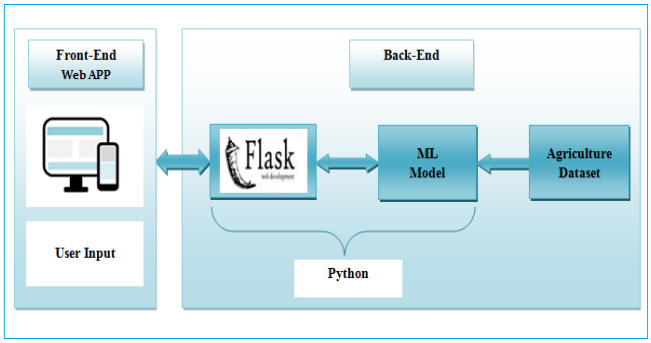

Fig. 1. System Architecture of Crop Recommendation System

Machine Learning Model :

As illustrated in Figure 2, In the first step, the crop dataset was collected. Two distinct datasets were collected and merged based on similar crop labels. The prepared dataset contains the following parameters - temperature, humidity, soil moisture, soil type, crop sowing season, and crop label for the Pune district. After finalizing data in the next step data is preprocessed. Data Pre-processing is the most important task because the collected dataset mostly contains missing values, which affect the accuracy of the model. Data were preprocessed and then divided into two datasets: training and testing. 75% of data was used for training, while 25% was kept for testing model performance. The four machine learning classification algorithms like the Decision Tree, Random Forest, Gaussian Naive Bayes, and XGBoost Classifier were used for training the prediction model. After training performance of these four techniques was compared. The final model has been built using XGBoost Classifier. Following the model training, a predictive model was developed, which will predict a suitable crop for cultivation based on user input.

Techniques:

Decision Tree (DT):

The most often used supervised machine learning technique for solving both regression and classification problems is a decision tree. It created a tree-like structure and divided the entire data set into smaller sub-parts based on certain conditions. The root node in the decision tree represents the whole sample, while the leaf node represents the outcome based on the decision made. Initially, all the independent variables are the candidate for the root node in the decision tree. Different attribute selection criteria, such as Information Gain and the Gini Index, are used to determine which attribute to chose as the root node and subsequent node. The root node is compared with record attributes to predict label or class. The below equation shows the formula for calculating the Gini Index.

$$\text{Gini}(D) = 1 - \sum_{i=0}^{n} P_i^2$$

Where P_i is a relative frequency of class i in dataset D. Attribute with lowest Gini index should be preferred.

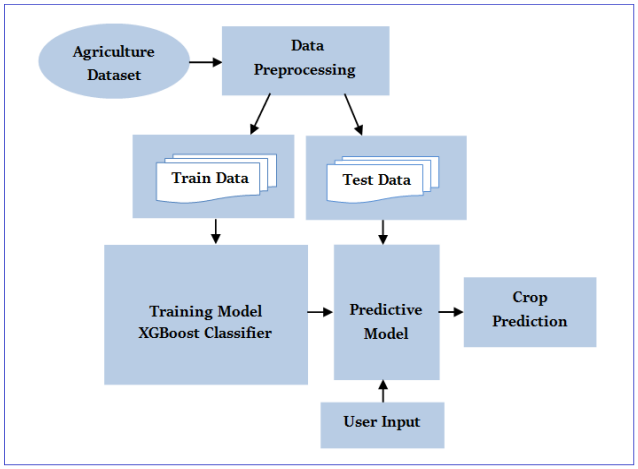

Fig. 2. Machine Learning (ML) Model

Random Forest (RF):

Random Forest is also one of the most popular supervised machine learning techniques used for solving both regression and classification problems. It's consists of multiple decision trees and uses decision trees in a randomized way. It uses decision trees of different shapes and sizes. Random Forest is an ensemble classifier as it comprises multiple decision trees, and the final decision is taken based on the majority of votes.

Naive Bayes (NB):

Naive Bayes is a supervised machine learning technique used for solving classification problems. It is a probabilistic classifier that can solve binary as well as multiclass classification problems. It is based on the Bayes theorem and assumes that the features in the dataset are independent of one another. In this study, the Gaussian Naive Bayes algorithm has used. Gaussian Naive Bayes algorithm assumes that features follow a normal distribution. The below equation shows the formula for calculating the Gaussian Probability Density Function.

$$P(x \mid y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Where σ represents standard deviation and μ is the mean.

XGBoost Classifier:

XGBoost stands for extreme gradient boosting. It's a decision tree-based boosting algorithm. The XGBoost technique is used to solve problems involving classification,

regression, and prediction. It is implemented on top of the gradient boosting framework. This technique runs at a fast speed and with high accuracy. It performs better on small to medium datasets. It uses the parallelization concept. It prevents the model from overfitting and also prevents trees from growing beyond a certain size.

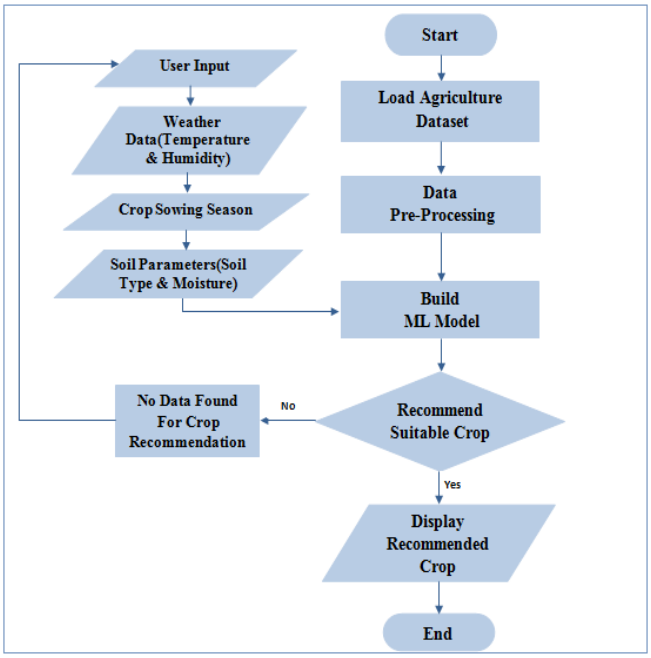

Fig. 3. Flowchart of Crop Recommendation System

Crop Recommendation:

Based on the soil parameters, crop season, and weather data, the proposed model will help to predict the best crop for cultivation for the Pune district. In this study, the user's input for crop recommendation includes location, soil moisture, soil type, and crop season. Based on data entered by the user suitable crop is predicted with the help of machine learning techniques. The XGBoost classifier was used for crop recommendation.

IV. RESULTS AND DISCUSSION

Experimental Setup:

For crop recommendations, the datasets were downloaded from Kaggle and data.gov.in websites. The exact dataset was not found, so the dataset was prepared using collected datasets. Based on available datasets, the following parameters are considered in the proposed model; Climatic Parameters: Temperature, Humidity, Soil Parameters: Soil Moisture, Soil Type, and Crop Season. Two distinct datasets were collected and merged based on similar crop labels.

The first dataset collected (District-wise-season-wise-crop-production dataset) includes the following parameters: State Name, District Name, Crop Year, Season, Crop, Area, and Production, as shown in figure 4. This dataset has been downloaded from the Kaggle and data.gov.in website. The data of the Pune district was taken from the given dataset. There are a total of 495 records for the Pune district in the given dataset.

	A	B	C	D	E	F	G
1	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
2	Maharashtra	PUNE	1997	Autumn	Maize	4	4570
3	Maharashtra	PUNE	1997	Kharif	Arhar/Tur	3900	2900
4	Maharashtra	PUNE	1997	Kharif	Bajra	135700	112000
5	Maharashtra	PUNE	1997	Kharif	Gram	43400	26100
6	Maharashtra	PUNE	1997	Kharif	Jowar	4200	5200
7	Maharashtra	PUNE	1997	Kharif	Maize	4400	4700
8	Maharashtra	PUNE	1997	Kharif	Moong(Gr	2800	1200
9	Maharashtra	PUNE	1997	Kharif	Pulses tot	203	91
10	Maharashtra	PUNE	1997	Kharif	Ragi	12200	9800
11	Maharashtra	PUNE	1997	Kharif	Rice	59400	66000
12	Maharashtra	PUNE	1997	Kharif	Sugarcane	40900	39227
13	Maharashtra	PUNE	1997	Kharif	Total fooc	2456	2113
14	Maharashtra	PUNE	1997	Kharif	Urad	2300	1200
15	Maharashtra	PUNE	1997	Rabi	Jowar	510800	235000
16	Maharashtra	PUNE	1997	Rabi	Maize	6900	10300
17	Maharashtra	PUNE	1997	Rabi	Other Rat	790	33
18	Maharashtra	PUNE	1997	Rabi	Wheat	57300	71400
19	Maharashtra	PUNE	1997	Summer	Maize	1200	2200
20	Maharashtra	PUNE	1997	Whole Yea	Cotton(lir	2	600
21	Maharashtra	PUNE	1998	Kharif	Arhar/Tur	3900	1100
22	Maharashtra	PUNE	1998	Kharif	Bajra	150000	110900
23	Maharashtra	PUNE	1998	Kharif	Castor see	200	100
24	Maharashtra	PUNE	1998	Kharif	Cotton(lir	300	500
25	Maharashtra	PUNE	1998	Kharif	Groundnu	45000	66100

Fig. 4. District-wise-season-wise-crop-production dataset

The second dataset (Fertilizer prediction dataset) downloaded from the Kaggle, has the following parameters - Temperature, Humidity, Moisture, Soil Type, Crop, Nitrogen, Potassium, Phosphorous, Fertilizer Name, as shown in figure 5. The given dataset contains a total of 99 records.

	A	B	C	D	E	F	G	H	I
1	Temparature	Humidity	Moisture	Soil Type	Crop	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
10	30	60	42	Sandy	Millets	21	0	18	28-28
11	29	58	33	Black	Oil seeds	9	7	30	14-35-14
12	27	54	28	Clayey	Pulses	13	0	40	DAP
13	31	62	48	Sandy	Maize	14	15	12	17-17-17
14	25	50	65	Loamy	Cotton	36	0	0	Urea
15	32	62	41	Clayey	Paddy	24	0	22	28-28
16	26	52	31	Red	Ground Nu	14	0	41	DAP
17	31	62	49	Black	Sugarcane	10	13	14	17-17-17
18	33	64	34	Clayey	Pulses	38	0	0	Urea
19	25	50	39	Sandy	Barley	21	0	19	28-28
20	28	54	65	Black	Cotton	39	0	0	Urea
21	29	58	52	Loamy	Wheat	13	0	36	DAP
22	30	60	44	Sandy	Millets	10	0	9	20-20
23	34	65	53	Loamy	Sugarcane	12	14	12	17-17-17
24	35	68	33	Red	Tobacco	11	0	37	DAP
25	28	54	37	Black	Millets	36	0	0	Urea

Fig. 5. Fertilizer prediction dataset

These two datasets were merged based on similar crop labels. The Unique() command was used to search for unique crop labels in both datasets. After examining both datasets, it was found that there are a total of eight crop labels that are common among both datasets. Maize, Sugarcane, Cotton, Wheat, Millets, Oilseeds, Pulses, and Groundnuts are the crop labels that are similar in both datasets. So datasets have been merged based on similar crop labels.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	State_Name	District_Name	Crop_Year	Area	Production	Season	Temparature	Humidity	Moisture	Soil Type	Crop	Nitrogen	Potassium	Phosphorous	Fertilizer Name
2	Maharashtra	PUNE	1997	4	4570	Autumn	26	52	38	Sandy	Maize	37	0	0	Urea
3	Maharashtra	PUNE	1997	4	4570	Autumn	31	62	48	Sandy	Maize	14	15	12	17-17-17
4	Maharashtra	PUNE	1997	4	4570	Autumn	26	52	44	Sandy	Maize	23	0	20	28-28
5	Maharashtra	PUNE	1997	4	4570	Autumn	33	64	51	Sandy	Maize	5	9	29	14-35-14
6	Maharashtra	PUNE	1997	4	4570	Autumn	30	60	47	Sandy	Maize	22	0	21	28-28
7	Maharashtra	PUNE	1997	4	4570	Autumn	28	54	25	Sandy	Maize	9	10	30	14-35-14
8	Maharashtra	PUNE	1997	4400	4700	Kharif	26	52	38	Sandy	Maize	37	0	0	Urea
9	Maharashtra	PUNE	1997	4400	4700	Kharif	31	62	48	Sandy	Maize	14	15	12	17-17-17
10	Maharashtra	PUNE	1997	4400	4700	Kharif	26	52	44	Sandy	Maize	23	0	20	28-28
11	Maharashtra	PUNE	1997	4400	4700	Kharif	33	64	51	Sandy	Maize	5	9	29	14-35-14
12	Maharashtra	PUNE	1997	4400	4700	Kharif	30	60	47	Sandy	Maize	22	0	21	28-28
13	Maharashtra	PUNE	1997	4400	4700	Kharif	28	54	25	Sandy	Maize	9	10	30	14-35-14
14	Maharashtra	PUNE	1997	6900	10300	Rabi	26	52	38	Sandy	Maize	37	0	0	Urea
15	Maharashtra	PUNE	1997	6900	10300	Rabi	31	62	48	Sandy	Maize	14	15	12	17-17-17
16	Maharashtra	PUNE	1997	6900	10300	Rabi	26	52	44	Sandy	Maize	23	0	20	28-28
17	Maharashtra	PUNE	1997	6900	10300	Rabi	33	64	51	Sandy	Maize	5	9	29	14-35-14
18	Maharashtra	PUNE	1997	6900	10300	Rabi	30	60	47	Sandy	Maize	22	0	21	28-28
19	Maharashtra	PUNE	1997	6900	10300	Rabi	28	54	25	Sandy	Maize	9	10	30	14-35-14
20	Maharashtra	PUNE	1997	1200	2200	Summer	26	52	38	Sandy	Maize	37	0	0	Urea
21	Maharashtra	PUNE	1997	1200	2200	Summer	31	62	48	Sandy	Maize	14	15	12	17-17-17
22	Maharashtra	PUNE	1997	1200	2200	Summer	26	52	44	Sandy	Maize	23	0	20	28-28
23	Maharashtra	PUNE	1997	1200	2200	Summer	33	64	51	Sandy	Maize	5	9	29	14-35-14
24	Maharashtra	PUNE	1997	1200	2200	Summer	30	60	47	Sandy	Maize	22	0	21	28-28
25	Maharashtra	PUNE	1997	1200	2200	Summer	28	54	25	Sandy	Maize	9	10	30	14-35-14

Fig. 6. Final Dataset

As shown in figure 6, the final prepared dataset contains the following parameters- State_Name, District_Name, Crop_Year, Season, Crop, Area, Temperature, Humidity, Moisture, Soil Type, Nitrogen, Potassium, Phosphorous, Fertilizer Name for the Pune district. Out of all these parameters for crop recommendation, State_Name, District_Name, Crop_Year, Season, Temperature, Humidity, Moisture, Soil Type, Crop parameters were used.

Algorithm Performance:

Four classification algorithms such as Decision Tree, Random Forest, Naive Bayes, and XGBoost classifiers were used for crop prediction.

TABLE II. ALGORITHM PERFORMANCE

Techniques	Training Accuracy	Testing Accuracy
Decision Tree Classifier	100%	91.89%
Decision Tree Classifier With Hyperparameter Tuning	100%	91.89%
Random Forest Classifier	100%	91.89%
Random Forest Classifier With Hyperparameter Tuning	100%	94.59%
Gaussian Naive Bayes Classifier	88.07%	78.38%
XGBoost Classifier	100%	94.59%
XGBoost Classifier With Hyperparameter Tuning	100%	97.3%

As shown in Table II, except Gaussian Naïve Bayes Classifier other techniques such as Decision Tree, Random Forest, and XGBoost classifiers gave 100\% accuracy on training data with Hyperparameter Tuning and without Hyperparameter Tuning. Where Gaussian Naïve Bayes Classifier gave 88.07\% accuracy on the training dataset. The accuracy of the Decision Tree classifier with and without hyperparameter tuning and Random Forest algorithm without hyper parameter tuning was 91.9\% and Random Forest algorithm with hyper parameter tuning was 94.59\% on test data. The accuracy of the Naive Bayes algorithm was 78.38\% on test data. The accuracy of XGBoost Classifier without hyper parameter tuning was 94.59\% and with hyper parameter tuning was 97.3\% on test data. As the XGBoost classifier gave the highest accuracy as compared to other algorithms, hence the final model was developed using the XGBoost classifier. Same is depicted in figure 7.

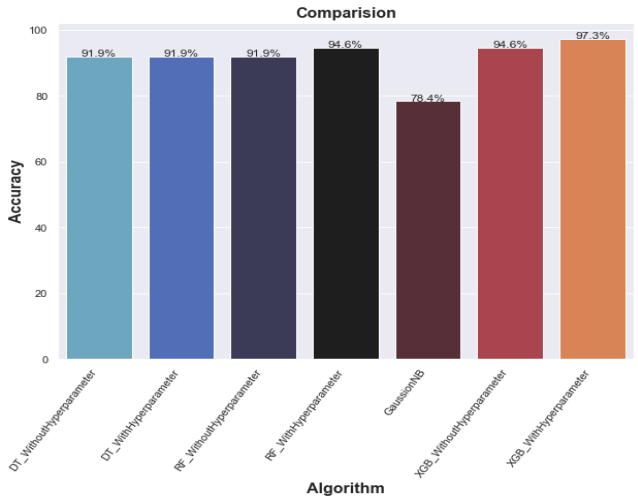

Fig. 7. Performance Analysis

For the recommendation system, the web application is designed using Flask Framework, Python, HTML, and CSS technology.

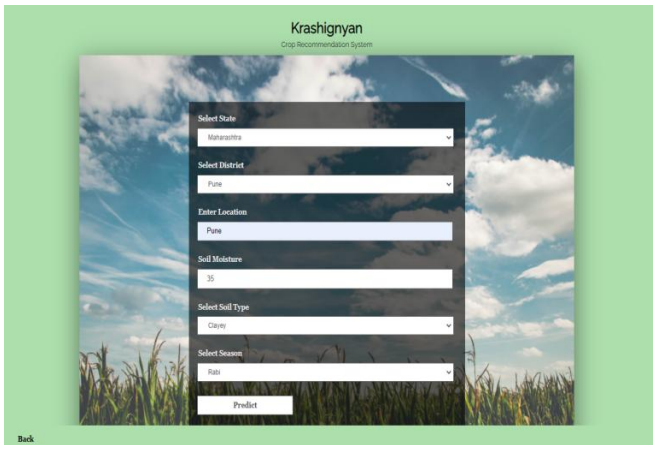

Fig. 8. User Interface

At the front end, the user interface is designed to take user inputs. As shown in Figure 8, the user enters values for location, type of soil, percentage of soil wetness, and crop sowing season. With the help of weather web API, the temperature and humidity of that specific location are fetch. The model then predicts most suitable crops for cultivation based on the user's input. Figure 9 shows the output of the recommended crops.

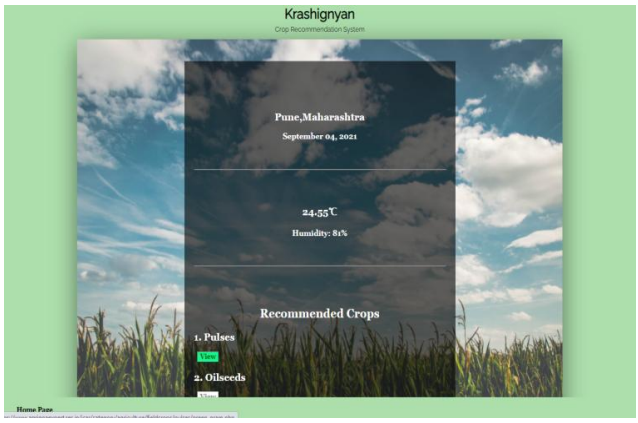

Fig. 9. Crop Recommendation Output

V. CONCLUSION

Agriculture has always been an important activity in India because it's not only a source of food but also a source of income for some people. As a result, in order to boost agricultural production, it must be done efficiently. There are various factors that affect crop growth, so choosing the right crop will always increase crop productivity. A crop recommendation system is proposed to help farmers in selecting the best crop for cultivation based on their location, soil factors, and weather parameters. Various machine learning approaches were used, such as decision tree classification, random forest classification, Naive Bayes classifier, and XGBoost classifier. Out of them, the XGBoost classifier gives the best results with a 97% accuracy, hence the final model was developed using the XGBoost classifier. In the future, for crop recommendation, more soil parameters are going to be considered and the system is further extended with IoT technology to take real-time values of soil parameters for more precise results.

REFERENCES

[1] Archana, K., & Saranya, K. G. (2020). Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier. SSRG Int. J. Comput. Sci. Eng, 7.

[2] Chaudhari, A., Beldar, M., Dichwalkar, R., & Dholay, S. (2020, September). Crop Recommendation and its Optimal Pricing using ShopBot. In 2020 International Conference on Smart Electronics and Communication (ICOSEC) (pp. 36-41). IEEE.

[3] Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese, Er. Jinu Thomas. Crop Prediction Using Machine Learning. International Journal of Future Generation Communication and Networking. Vol. 13, No. 3, (2020), pp. 1896–1901.

[4] Lavanya, B., Nisarga, B., Meghana, B. S., & Mythresh, A. CROP PREDITION USING MACHINE LEARNING.

[5] Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R. Crop Prediction using Machine Learning Approaches. International Journal of Engineering Research & Technology (IJERT). Vol. 9 Issue 08 (August-2020).

[6] [6] Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique. In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) (pp. 114-119). IEEE.

[7] Mythresh A, Lavanya B, Meghana BS, Nisarga B (2020). Crop Prediction using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 7(6).

[8] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.

[9] Shinde, M., Ekbote, K., Ghorpade, S., Pawar, S., & Mone, S. (2016). Crop recommendation and fertilizer purchase system. International Journal of Computer Science and Information Technologies, 7(2), 665-667.

[10] Kumar, A., Sarkar, S., & Pradhan, C. (2019, April). Recommendation system for crop identification and pest control technique in agriculture. In 2019 International Conference on Communication and Signal Processing (ICCSP) (pp. 0185-0189). IEEE.

[11] [11] Banavlikar, T., Mahir, A., Budukh, M., & Dhodapkar, S. (2018). Crop recommendation system using Neural Networks. International Research Journal of Engineering and Technology (IRJET), 5(5), 1475-1480.

[12] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE.

[13] Kalimuthu, M., Vaishnavi, P., & Kishore, M. (2020, August). Crop prediction using machine learning. In 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 926-932). IEEE.

[14] Patil, A., Kokate, S., Patil, P., Panpatil, V., & Sapkal, R. (2020). Crop Prediction using Machine Learning Algorithms. Kapila Journals, 1(1), 1-8.

[15] Dhruv Piyush Parikh, Jugal Jain, Tanishq Gupta and Rishit Hemant Dabhade. Machine Learning Based Crop Recommendation System. International Journal of Advanced Research in Science, Communicationand Technology (IJARSCT), Volume 6, Issue 1, June 2021.

[16] https://en.wikipedia.org/wiki /Agriculture_in_India

[17] https://www.kaggle.com/gdabhishek/fertilizer-prediction

[18] https://www.kaggle.com/anjali21/agricultural-production-india

[19] https://data.gov.in/catalog/district-wise-season-wise-crop-production