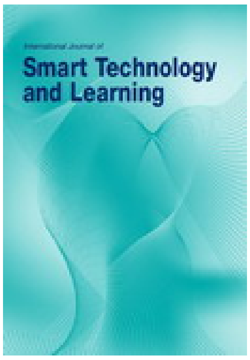


International Journal of Smart Technology and Learning

ISSN online: 2056-4058 - ISSN print: 2056-404X
https://www.inderscience.com/ijsmarttl

A machine learning based crop recommendation system and user-friendly android application for cultivation

Kaniz Fatima Tonni, Mahfuzulhoq Chowdhury

DOI: 10.1504/IJSMARTTL.2023.10054698

Article History:

Received:	03 October 2022
Last revised:	21 January 2023
Accepted:	25 January 2023
Published online:	17 March 2023

A machine learning based crop recommendation system and user-friendly android application for cultivation

Kaniz Fatima Tonni and
Mahfuzulhoq Chowdhury*

Computer Science and Engineering Department,
Chittagong University of Engineering and Technology,
Chittagong 4349, Bangladesh
Email: kaniztonni8@gmail.com
Email: mahfuz_cse@cuet.ac.bd
*Corresponding author

Abstract: Bangladesh is essentially an agricultural nation, and its economy is heavily dependent on it. A farmer could plant a crop if he knew which one would yield more. The existing literature works fail to provide a user-friendly mobile application for cultivation as well as machine learning-based crop recommendation by taking different factors into account. This paper creates a mobile application that enables farmers to forecast viable crops based on climate factors like humidity, rainfall, and temperature as well as soil characteristics. The suggested model is used to forecast agricultural production using crop records of diverse crops with various properties of soil and climate parameters. The suggested model offers farmers a comprehensive list of recommendations to help them choose crops that are best for them based on particular considerations like production costs and fertiliser recommendations. The user's feedback shows satisfactory remarks in terms of its usefulness.

Keywords: crop recommendation; machine learning; prediction; evaluation; cultivation; Bangladesh; Android application.

Reference to this paper should be made as follows: Tonni, K.F. and Chowdhury, M. (2023) 'A machine learning based crop recommendation system and user-friendly android application for cultivation', *Int. J. Smart Technology and Learning*, Vol. 3, No. 2, pp.168–186.

Biographical notes: Kaniz Fatima Tonni received her BSc in Computer Science and Engineering at the Department of CSE, Chittagong University of Engineering and Technology in 2022. Her research interests include machine learning and mobile application development.

Mahfuzulhoq Chowdhury received his PhD in Telecommunications at the University of Quebec, INRS, Montreal, Canada. He is a faculty member at the Department of Computer Science and Engineering, Chittagong University of Engineering and Technology since 2010. His research interests are related to machine learning, optimisation, mobile application development, game theory, resource allocation in future generation networks, cloud computing, IoT, age of information, and blockchain. He has published several research papers at highly cited IEEE journals, Transactions, magazines, and conference proceedings as well as chapters of books.

1 Introduction

Bangladesh has long been renowned as an agricultural country due to its huge amount of cash crop production, such as jute, rice, etc. The major factor behind this prodigious agricultural development is rich delta soil, plenty of water, and several agricultural seasons. Bangladesh's position as a market leader has changed over time due to several challenging factors, such as a dense population and environmental calamities. The nation is always looking for ways to boost overall agricultural productivity through a variety of means. Technology is one such instrument that can help anybody to achieve their goals. It's even more effective when utilised in a local setting. Also, it has reached every corner of the world.

Nowadays, rural people also use digital devices and smartphones. One of the most significant benefits of smart technology and the digital era is the ability to check factors and market prices/demand from distant locations while also receiving useful technical agricultural advice. Using low-tech techniques, Bangladeshi farmers have succeeded for decades. Therefore, any new technology offered to aid in improving their standard of living must be simple to use. Apart from accessibility, many of these farmers still adhere to their 'old methods'. It's important to investigate why some of them could still be viable choices in the future. Farmers in remote locations who may benefit greatly from affordable, high-quality farming technology usually lack the resources, know-how, and training necessary to put it to use. The expense must also be taken into account. Since the expense of growing both short-term and long-term plants relies on the region's size.

Currently, various study recommends a crop selection strategy depending on the features of the soil, the climate, and places throughout the world (Reddy et al., 2019). Several machine learning-based frameworks have been developed to help farmers with best crop selection for cultivation based on the properties of the soil. Basic machine learning frameworks were developed, and the effort also included the Android operating system (Salpekar et al., 2019). Bangladesh has also worked on agriculture, with proposals for projects based on the environment and crops specific to different regions (Sadia et al., 2021). The ability to choose which crop to plant depending on factors such as soil type, characteristics, season, and yielding cost would be of great use to those in the agricultural industry. This sort of system may offer a comprehensive notion for cultivating any plants, such as approximate profit for a particular budget, cultivation method, and acceptable plants for a specific type of soil.

Kulkarni et al. (2018) proposed a crop recommendation system that makes use of machine learning's ensembling technique. The ensembling technique is used to create a model that accurately recommends the best crop depending on the unique characteristics of the soil by combining the predictions of machine learning models. The authors used Naive Bayes, Random Forests, and Linear Support Vector Machine as base learners in an ensemble framework. When determining the crop, the authors emphasised the importance of soil type, pH value, porosity of the soil, average rainfall, temperature, and planting season. Bepery et al. (2020) dealt with several remote soil monitoring systems based on IoT protocols to improve crop yields. They also provided an overview of sensors, technologies, advantages, disadvantages, and the system's future aspects.

However, the existing works could not provide an intelligent crop recommendation system that utilises real-time data, soil parameters, production time, user preference, and cost at the same time. Moreover, information regarding fertiliser usage and usage of both supervised and unsupervised learning were out of their investigations. To empower the agriculture system, this paper presents an Android application that uses a recommendation system using machine learning based on soil factors, and production time and gives approximate knowledge about the cost of the current time specified by the authority.

The primary contribution of this paper is listed as follows:

1 This paper develops a system using machine learning that can predict plants based on multiple factors. This paper investigates both supervised and unsupervised learning to predict crops.

2 This paper develops an android application that offers some suggestions on how to apply fertilisers to the recommended plant.

3 This paper assists users in locating the cost field and an estimated cost for those fields in advance.

4 This paper develops a data set for crop recommendation and collects cost of production data from the agricultural department of an *Upazila Parishad*.

5 This paper provides a performance comparison of different machine learning algorithms and chooses the best prediction model for crop recommendation.

This work is detailed as follows. Section 2 provides some discussion regarding literature works. Section 3 gives a brief discussion, methodology, and results regarding the proposed scheme. Section 4 offers the results regarding the android-based crop recommendation application features. Section 5 provides a summary of this work and future works.

2 Literature review

At present, a handsome amount of literary work has been done on crop recommendation. To accomplish these tasks, machine learning was a key study. In Rajak et al. (2017), support vector machines, Naive Bayes, Multilayer Perceptrons (Artificial Neural Networks), and Random Forests were used as part of an ensemble model. The characteristics assessed in the parameters include depth, texture, PH, soil colour, permeability, drainage, water holding, and erosion.

Rajeswari et al. (2020) suggested an approach that makes use of the soil's micronutrients to forecast which crops will grow best in a given area. The fuzzy rules are used to anticipate if the crop is acceptable using a rudimentary set-based rule induction approach. Suresh et al. (2021) suggest helping ranchers determine the quality of the soil by examining its various boundaries. The authors also recommended crops based on the findings using an information mining approach. The system maps the soil and yields data to predict a list of suitable harvests for the soil. It also provides information about supplements that are insufficient in the soil for the particular harvest.

Talukder et al. (2020) have developed a machine learning-based prediction and recommendation model that assesses productivity based on the variables temperature, humidity, and precipitation. They have used techniques for collaborative filtering, multi-condition filtering, k-nearest neighbour (KNN), support vector machines (SVM), random forest, naive Bayes' classifier, and logistic regression for the prediction. They used Multi-Condition Filtering and Collaborative Filtering algorithms for the recommendation.

Bepery et al. (2020) suggest a general structure for a current technology-based (e.g., Internet of Things) soil monitoring system. They displayed an overview of twenty-nine IoT-based soil monitoring publications (i.e., published between 2016 and 2020). The most prevalent issue farmers have is that they do not choose their crops based on the requirements of the soil, which has a significant negative impact on their output. Precision agriculture can be used to solve this issue.

In Reddy et al. (2019), to select a good crop based on soil data with high specific accuracy and efficiency, the authors have developed an ensemble model with majority voting approaches. The authors employed random tree, K Nearest Neighbour, and Naive Bayes as the learner. To estimate the yield of a certain crop under a given set of weather conditions, the categorised image is created by these algorithms. They used statistical data as its parameters, including weather, crop yield, state-by-state crop data, and district-by-district crop data. Three factors are taken into account by this method such as soil types, soil characteristics, and crop yield data gathering. Based on these factors, an appropriate crop is recommended to cultivate.

Salpekar et al. (2019) developed an app that enables the forecasting of the best crop in a given region based on environmental factors like rainfall and temperature. The suggested model is used to estimate crop yield using a dataset of different crops from different regions of India, as well as rainfall and temperature data for the same locations. The suggested model offers a comprehensive list of recommendations regarding the best crop selection based on details like location, farm size, weather, rainfall, and other crop datasets. Sadia et al. (2021) attempted to address the issue by suggesting a system that advises the user to assist farmers in identifying the most suitable crops for a given soil and geological information.

Mokarrama et al. (2017) provide a recommendation system for farmers that might suggest the best crops to grow in various regions. The system uses the Pearson co-relation similarity algorithm to calculate the similarity between *upazilas* after first determining the user's location and utilising various agroecological and agro-climatic data at the *upazila* level.

Agriculture automation is a mechanical process that can be done with or without human involvement. Due to the limited amount of domestic land, it has become crucial to select the crops that are most suited to the local conditions. In Bandara et al. (2020), a theoretical and conceptual framework for the recommendation system was provided using integrated models for gathering environmental information. It is constructed with Arduino microcontrollers, machine learning methods, and Artificial Intelligence methods to recommend a crop for the chosen land with site-specific parameters, high accuracy, and effectiveness. By gathering environmental data that affect plant growth and combining them with the trained sub-models of the system's main model, this crop suggestion solution forecasts which crop variety would be best suited for the chosen area. Doshi et al. (2018) have described an intelligent system called AgroConsultant that aims to help farmers decide which crop to cultivate based on the sowing season, the location of their farm, the properties of the soil, as well as environmental elements like temperature and rainfall.

Thilakarathne et al. (2022) investigated a cloud-based machine learning-driven crop recommendation system using the nitrogen, phosphorus, and potassium requirements of different crops. However, the authors did not develop any mobile application for cultivation and did not take the total cost of production for the crop recommendation. Gill et al. (2021) investigated a crop factors prediction system (e.g., raw seeds, time, meteorological data) related to a crop using a Long Short-Term Memory or LSTM-based framework. Pande et al. (2021) presented an area and soil type parameter-based suitable crop recommendation system using machine learning.

Villanueva et al. (2022) developed a soil health-based crop recommendation system using a cloud computing-based deep learning approach. Vaishnavi et al. (2021) utilised the season and previous year productivity value-based crop suggestion system for the Indian farmers. However, they did not use any other parameters like cost and soil parameters for crop prediction. Gupta et al. (2021) presented a weather-based crop recommendation system using the map-reduce and k means clustering scheme. However, they did not propose any mobile application for farmers or cultivation assistance.

Most of the research work only utilises soil parameters to suggest a crop. There is no detailed cost estimate in different sectors for crop cultivation. Also, most of the papers don't recommend the best fertilisers for the output crops. Our paper is presented to overcome these limitations.

3 Proposed model

This section predicts the best crop for the users' farm using a machine-learning approach. We have collected datasets for the machine learning model and data on the cost of production for each crop. Figure 1 represents the overview of the machine learning models of our work. We have used two approaches for prediction: the supervised and unsupervised approaches. The prediction from both approaches is displayed in the android application.

Supervised learning is characterised by the use of labelled datasets to educate computers that can precisely identify data or forecast events. To choose the best-fit model for prediction, we employed the ensemble model. For our dataset, this model's output provides a single crop. The supervised learning system works on an ensemble model. The model will be trained to utilise the dataset and tested with the input. This model consists of three base learners: Random forest, K nearest neighbours (KNN), and Decision tree. Next, the system determines the plants that are suitable for yielding based on the soil parameters, climate, cost factor, and other user inputs.

Unsupervised Learning makes inferences from unlabeled datasets. By using the crop dataset the system first determines the optimum value of clusters. Then, we selected the best clustering model and predicted the cluster for user input. Next, we obtained the crop list and visualised crop details. Figure 2 shows the workflow diagram of supervised and unsupervised learning.

Figure 1 Workflow diagram of machine learning model

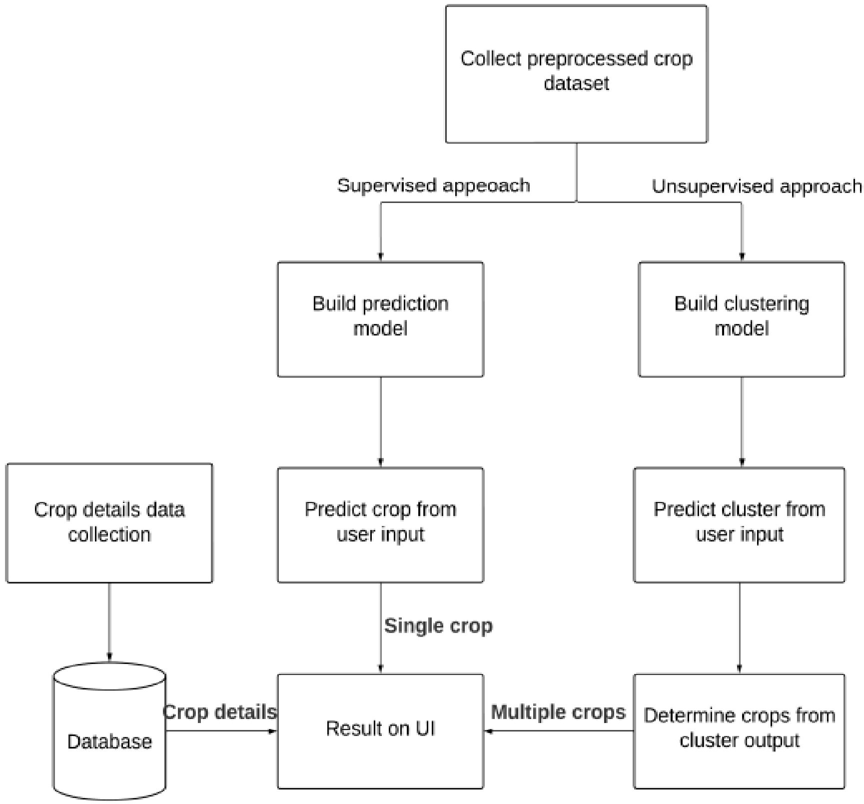
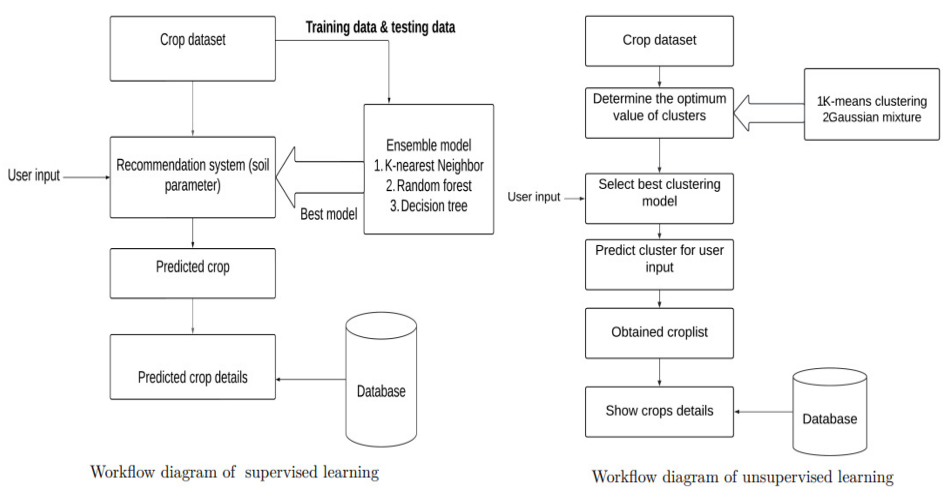


Figure 2 Workflow diagram of supervised and unsupervised learning



3.1 Crop dataset

This paper collects a dataset of crop-specific soil factors, such as temperature, humidity, rainfall, etc, for the crop recommendation system. The dataset is divided into training datasets and testing datasets. The dataset needs to be big enough to train the system, which will be trained using the training dataset and evaluated using the testing dataset (Pudumalar et al., 2017). Our dataset has been collected from Kaggle (2009).

Our dataset contains a total of 28 crops, vegetables, and fruits, and consists of eight columns where seven columns denote the parameters, which are *N, P, K, temperature, humidity, pH, rainfall,* and another column is for crop names. We have taken the features *'N', 'P', 'K', 'temperature', 'humidity', 'ph', and 'rainfall'* and the target variable *'label'* to train the model. A small fraction of our dataset is given in Figure 3.

Figure 3 A small fraction of dataset

	A	B	C	D	E	F	G	H
1	N	P	K	temperature	humidity	ph	rainfall	label
2	90	42	43	20.87974371	82.00274423	6.502985292	202.9355362	rice
3	85	58	41	21.77046169	80.31964408	7.038096361	226.6555374	rice
4	60	55	44	23.00445915	82.3207629	7.840207144	263.9642476	rice
5	74	35	40	26.49109635	80.15836264	6.980400905	242.8640342	rice
6	78	42	42	20.13017482	81.60487287	7.628472891	262.7173405	rice
7	69	37	42	23.05804872	83.37011772	7.073453503	251.0549998	rice
8	69	55	38	22.70883798	82.63941394	5.70080568	271.3248604	rice
9	94	53	40	20.27774362	82.89408619	5.718627178	241.9741949	rice
10	89	54	38	24.51588066	83.5352163	6.685346424	230.4462359	rice
11	68	58	38	23.22397386	83.03322691	6.336253525	221.2091958	rice
12	91	53	40	26.52723513	81.41753846	5.386167788	264.6148697	rice
13	90	46	42	23.97898217	81.45061596	7.50283396	250.0832336	rice
14	78	58	44	26.80079604	80.88684822	5.108681786	284.4364567	rice
15	93	56	36	24.01497622	82.05687182	6.98435366	185.2773389	rice
16	94	50	37	25.66585205	80.66385045	6.94801983	209.5869708	rice
17	60	48	39	24.28209415	80.30025587	7.042299069	231.0863347	rice
18	85	38	41	21.58711777	82.7883708	6.249050656	276.6552459	rice
19	91	35	39	23.79391957	80.41817957	6.970859754	206.2611855	rice
20	77	38	36	21.8652524	80.1923008	5.953933276	224.5550169	rice
21	88	35	40	23.57943626	83.58760316	5.85393208	291.2986618	rice
22	89	45	36	21.32504158	80.47476396	6.442475375	185.4974732	rice
23	76	40	43	25.15745531	83.11713476	5.070175667	231.3843163	rice
24	67	59	41	21.94766735	80.97384195	6.012632591	213.3560921	rice
25	83	41	43	21.0525355	82.67839517	6.254028451	233.1075816	rice

3.2 Cost of production data

It is required to get production cost data from an *Upazila Parishad* agricultural division for the suggested crops. By having comprehensive data, users should be able to determine which fields they should invest in for the crop. Depending on the industry, cost factors can vary, including labour expenses per person and transportation costs.

We have collected our cost of production data from the Savar Upazila Agriculture Office. The collected data consists of several fields where users might have to invest in yielding crops. *Land preparation, plantation, irrigation, weeding, pesticides, fertilisers, seeds, transport, harvesting, threshing, overhead costs, and other fields* are among them. The cost value in these fields may vary from place to place, such as labour cost per head may vary during *transport, harvesting, weeding,* etc. All these costs are measured in terms of per hectare area. A small fraction of our cost dataset is given in Figure 4.

Figure 4 A small fraction of cost dataset

Cost Of Production Mustard

Sl No	Production ingredient	Quantity	Price Per unit	Total Cost(Tk) per ha
01	Land Preparation(Tractor)			2500/-
02	Seed	7.5 kg	150/-	1,125/-
03	Plantation	5 Labour	700/-	3500/-
04	Weeding	5 Labour	700/-	3500/-
05	Irrigation	1 times		15000/-
06	Pesticides			1500/-
07	Fertilizer			14200/-
08	Harvesting& Treshing	10 labour	700/-	7000/-
09	Transport	10 labour	700/-	10,000/-
10	Over head Cost			40,800/-
11	Others			18000/-
	Total			117125/-

Cost Of Production Maize

Sl No	Production ingredient	Quantity	Price Per unit	Total Cost(Tk) per ha
01	Land Preparation(Tractor)			2500/-
02	Seed	20 kg	500/-	10,000/-
03	Plantation	25 labour	700/-	17,500/-
04	Weeding	5 labour	700/-	3500/-
05	Irrigation	4 times	12000/-	48,000/-
06	Pesticides			2,000/-
07	Fertilizer			17400/-
08	Harvesting& Treshing	10 labour	700	7000/-
09	Transport	15 labour	700/-	13,000/-
10	Over head Cost			40800/-
11	Others			10000/-
	Total			1,71,700/-

3.3 Building machine learning model

Utilising learning algorithms, machine learning tries to produce statistical models for data analysis and prediction. The ML algorithms should be capable of learning on their own (depending on the input data) and producing precise predictions without having been specially designed for a given task. In our machine learning model for prediction, we have used both supervised and unsupervised learning.

The supervised approach recommends one crop for our dataset. We have taken the features *'N', 'P', 'K', 'temperature', 'humidity', 'ph', and 'rainfall'* and the target variable *'label'* to train the model. In our dataset, we recommended various crops using an unsupervised methodology. Each row in this dataset represents a crop and is part of a cluster.

3.4 Supervised learning

To determine the model that best fits our dataset, we trained an ensemble model on our training data. Our dataset includes both training and testing datasets. The learner for the system model is decided to be K-Nearest Neighbour, Random Forest, and Decision Tree. For each algorithm or learner in the ensemble model, *the precision, recall, f1-score, and* *support* are calculated using metrics from the Scikit-learn library, the classification report function. Precision measures the accuracy of a positive prediction provided by the model. The *recall* computes ratio of Positive samples that were correctly categorised as Positive to the total number of Positive samples.

The *F1 score* summarises a model's prediction performance by combining two contradictory criteria – *accuracy and recall*. The number of samples of the genuine response that fall into each class of goal values can be described as *support*. Using metrics from the Scikit-learn library, the classification report function calculated the *precision, recall, f1-score, and support* for KNN, random forest, and decision tree model as shown in Figure 5.

Figure 5 Output of KNN, random forest, and decision tree

knn's Accuracy is: 82.85714285714286

	precision	recall	f1-score	support
Broccoli	0.30	0.54	0.39	26
Cabbage	0.25	0.32	0.28	22
Cauliflower	0.07	0.06	0.06	18
Potato	0.25	0.08	0.12	24
apple	1.00	1.00	1.00	17
banana	1.00	1.00	1.00	14
blackgram	0.94	1.00	0.97	17
chickpea	1.00	1.00	1.00	23
coconut	1.00	1.00	1.00	19
coffee	1.00	1.00	1.00	13
cotton	0.96	0.96	0.96	23
grapes	1.00	1.00	1.00	23
jute	0.89	0.89	0.89	18
kidneybeans	1.00	1.00	1.00	17
lentil	1.00	1.00	1.00	14
maize	0.95	0.95	0.95	21
mango	1.00	1.00	1.00	23
mothbeans	0.92	0.92	0.92	26
mungbean	0.81	1.00	0.90	13
muskmelon	1.00	1.00	1.00	21
mustard	0.15	0.08	0.11	24
orange	1.00	1.00	1.00	20
papaya	0.97	1.00	0.98	28
pigeonpeas	1.00	1.00	1.00	14
pomegranate	1.00	1.00	1.00	20
rice	0.95	0.90	0.93	21
watermelon	1.00	1.00	1.00	17
wheat	1.00	1.00	1.00	24
accuracy			0.83	560

(a)

RF's Accuracy is: 0.8696428571428572

	precision	recall	f1-score	support
Broccoli	0.26	0.23	0.24	26
Cabbage	0.46	0.50	0.48	22
Cauliflower	0.29	0.39	0.33	18
Potato	0.36	0.33	0.35	24
apple	1.00	1.00	1.00	17
banana	1.00	1.00	1.00	14
blackgram	1.00	1.00	1.00	17
chickpea	1.00	1.00	1.00	23
coconut	1.00	1.00	1.00	19
coffee	1.00	1.00	1.00	13
cotton	1.00	1.00	1.00	23
grapes	1.00	1.00	1.00	23
jute	0.95	1.00	0.97	18
kidneybeans	1.00	1.00	1.00	17
lentil	1.00	1.00	1.00	14
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	23
mothbeans	1.00	1.00	1.00	26
mungbean	1.00	1.00	1.00	13
muskmelon	1.00	1.00	1.00	21
mustard	0.48	0.42	0.44	24
orange	1.00	1.00	1.00	20
papaya	1.00	1.00	1.00	28
pigeonpeas	1.00	1.00	1.00	14
pomegranate	1.00	1.00	1.00	20
rice	1.00	0.95	0.98	21
watermelon	1.00	1.00	1.00	17
wheat	1.00	1.00	1.00	24
accuracy			0.87	560

(b)

DecisionTrees's Accuracy is: 85.35714285714285

	precision	recall	f1-score	support
Broccoli	0.00	0.00	0.00	26
Cabbage	0.48	0.50	0.49	22
Cauliflower	0.26	0.33	0.29	18
Potato	0.23	0.50	0.32	24
apple	1.00	1.00	1.00	17
banana	1.00	1.00	1.00	14
blackgram	0.94	0.94	0.94	17
chickpea	1.00	1.00	1.00	23
coconut	1.00	1.00	1.00	19
coffee	1.00	1.00	1.00	13
cotton	1.00	1.00	1.00	23
grapes	1.00	1.00	1.00	23
jute	0.94	0.94	0.94	18
kidneybeans	1.00	1.00	1.00	17
lentil	0.93	1.00	0.97	14
maize	1.00	1.00	1.00	21
mango	1.00	1.00	1.00	23
mothbeans	0.92	0.88	0.90	26
mungbean	1.00	1.00	1.00	13
muskmelon	1.00	1.00	1.00	21
mustard	0.69	0.38	0.49	24
orange	1.00	1.00	1.00	20
papaya	1.00	1.00	1.00	28
pigeonpeas	1.00	1.00	1.00	14
pomegranate	1.00	1.00	1.00	20
rice	0.95	0.95	0.95	21
watermelon	1.00	1.00	1.00	17
wheat	1.00	1.00	1.00	24
accuracy			0.85	560

(c)

Next, the *cross value score* of the KNN, random forest and decision tree model was measured using the built-in function Scikit-learn. Using the ensemble technique, we have found the *accuracy* of each algorithm and compared them. As a result, we have found the best-fit model for our dataset. The data for plotting them in the graph, where the x-axis denotes the model and the y-axis denotes the *accuracy score* for each model, has been plotted using pyplot of the matplotlib library.

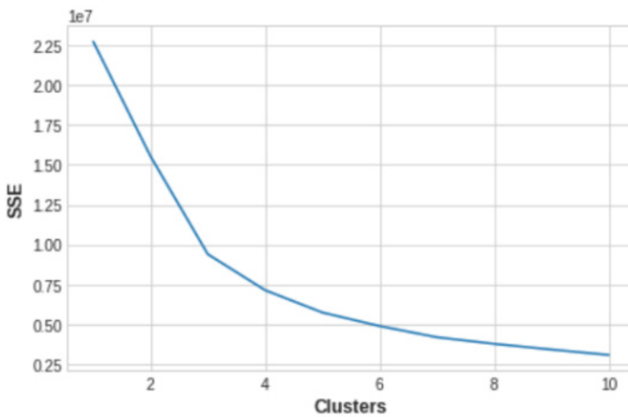
3.5 Unsupervised learning

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to evaluate and group unlabeled datasets. Without the assistance of a human, these algorithms detect hidden patterns or data clusters. A crop is represented by each row in our dataset. They, too, are part of a cluster. If we can anticipate the cluster based on the user input, we can locate the crops that are associated with that cluster. We employed k-means clustering and a Gaussian mixture for these approaches.

To lower the objective function, the k-means approach is similar to a gradient descent procedure that iteratively updates the starting cluster centroids. K-means are always able to find a local minimum. To determine the optimum value of k we have used the elbow method. We have taken the features *'N', 'P', 'K', 'temperature', 'humidity', 'ph', and 'rainfall'*.

The elbow approach is a heuristic used in cluster analysis to determine the number of clusters in a data set. To find the elbow point, we have taken cluster numbers ranging from 1 to 11. We can visualise the elbow point in a graph plotting SSE against cluster number. The squared Euclidean distances between each point and its nearest centroid are added up to form the SSE. Here, the features are *'N', 'P', 'K', 'temperature', 'humidity', 'ph', and 'rainfall'*. According to the graph in Figure 6, the optimal cluster number is 3. So we decide to fit the dataset using this cluster number and determine which cluster each point belongs to.

Figure 6 Elbow method graph and silhouette score for k means clustering algorithm



Elbow method graph

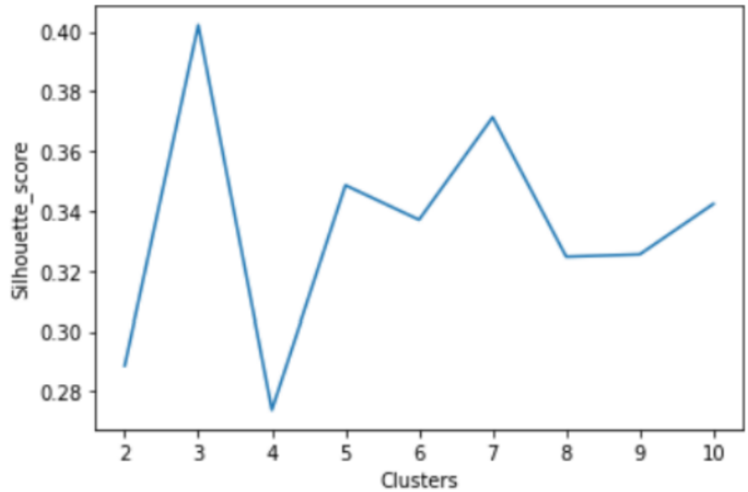
```
from sklearn.metrics import silhouette_score
score = silhouette_score(X,y)
print(score)
```

```
0.43696535384079954
```

silhouette_score for k-means clustering algorithm

A way of interpreting and validating consistency within data clusters is referred to as silhouette. The method offers a brief graphic representation of each object's classification accuracy. Here, we have calculated the silhouette score by Scikit-learn metrics (see Figure 6). GMM clustering can be quite effective in some circumstances, even though it might not be the fastest solution. Agglomerative GMM clustering, the form of GMM clustering we have examined is a bottom-up approach to clustering. To find the optimum value of n_component, we have taken the values ranging from 2 to 11 as prime numbers. The ideal n_component number, as shown by the graph in Figure 7, is 3, where the silhouette score is highest. Therefore, we choose to fit the dataset with this component value to identify the cluster to which each point belongs.

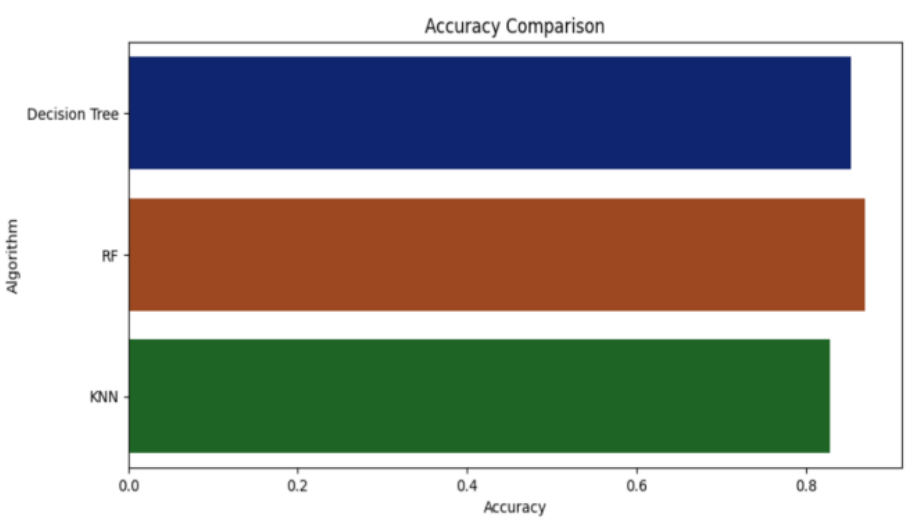
Figure 7 GMM cluster determination graph using silhouette score



3.6 Result and analysis of machine learning model

To build the recommendation in our supervised technique, we used an ensemble model. Random Forest was the best model we identified utilising ensembles in our work, with an accuracy score of 86.96%. As a result, we chose this prediction model. Figure 8 shows the accuracy comparison figure. For unsupervised learning, we used silhouette score to determine which model is best for clustering. We can observe from the silhouette score in the figure silhouette score that k-means clustering outperforms Gaussian mixing.

Figure 8 The graph of accuracy measure of machine learning algorithm



We can observe that between these two models, k-means clustering produces the superior clusters using both Gaussian mixture and k-means clustering. As a result, we use k-means clustering to determine which data in our dataset belongs to which cluster (see Figure 9). Figure 9 also shows a sample output using k-means clustering for specific input. We get one crop as an output for the input provided to the supervised learning algorithm in our dataset (see Figure 10). However, there may be other crops that can be produced in the soil with the conditions we supplied. In our dataset, each row represents a crop. They, too, are members of a cluster. If we can anticipate the cluster based on the user input, we can locate the crops that are associated with that cluster. We can get more crops that can be recommended by using this concept. From Figure 10, we can see that the clustering algorithm also gives the same crop in its lists.

Figure 9 Sample output of unsupervised model's selected algorithm

```
from sklearn.metrics import silhouette_score
score = silhouette_score(X,y)
print(score)

0.43696535384079954
```

Silhoute score of k-means clustering

```
from sklearn.metrics import silhouette_score
score = silhouette_score(X,cluster)
print(score)

0.40214098129083126
```

silhoute score of gaussian mixing

```
data['cluster'] = y
data.to_csv(r'Cluster_added.csv', index=False)
output_example = kmeans.predict([[104,18, 30, 23.603016, 60.3, 6.7, 140.91]])
print(output_example)

[1]
```

Figure 10 Sample output of supervised and unsupervised model's selected algorithm

```
output_example = kmeans.predict([[104,18, 30, 23.603016, 60.3, 6.7, 140.91]])
output_data = output_data.loc[:, ["label", "cluster"]]
df2=output_data.loc[output_data['cluster'] == output_example[0]]
print(df2['label'].unique())

['rice' 'kidneybeans' 'pigeonpeas' 'banana' 'papaya' 'coconut' 'jute'
 'coffee']

data = np.array([[104, 18, 30, 23.603016, 60.3, 6.7, 140.91]])
prediction = RF.predict(data)
print(prediction)

['coffee']
```

4 Android application features and user evaluation

We have created an android application based on the recommended system that we decided to establish using machine learning. Our application will be user-friendly in the sense that it will be simple, convenient to use, and reliable. We have used react-native as our android application development platform. The application includes both the Basic and Advanced approaches. To predict crops, an advanced method employs a machine learning model. The user must submit *'N', 'P', 'K', 'temperature', 'humidity', 'ph', and 'rainfall'* data for the soil. The model receives these parameters via the API and returns suitable crops. We take the suggested crop from our supervised model and use a clustering model to determine which cluster the user's inputs belong to. We may get the crops that belong to that cluster by using the cluster number. We filter out the crops in the cluster based on the soil type of the recommended crop before showing it. When we click for more detail on a certain crop, the app takes us to the detail page. We have added two tabs to the detail page namely production cost and fertiliser.

4.1 Home page and crop recommender system

Now, we will discuss the different mobile application features of our crop recommendation application. First, the user can see the homepage with two options: basic and advanced methods. The basic method provides a crop list based on *soil type, season, and period*. The advanced method provides crop lists based on *'N', 'P', 'K', 'temperature', 'humidity', 'ph', and 'rainfall'* factors using a machine learning model. Figure 11 shows the homepage of our application. After clicking the type, the user needs to complete the input field. Next, when you press the 'Continue' button, the program will take you to the input parameter page.

Figure 12 shows the application screen regarding input field fulfilment for the basic method. To obtain the output, all fields in the advanced method must be filled (see Figure 13). When you press the predict button app, you will be taken to the crop list page. In the crop list, we can see the multiple crops listed based on the user input (see Figure 14). Pressing on the details or the name of the crop, the app moves to the details page. On this page, there are two tabs where the first tab shows the cost of production in a tabular format on the detail page. The other tab shows the fertiliser required for the crop and when to use it. The fertiliser information is collected from Bangladesh Agricultural Research Council (2018). Figure 15(a) shows the cost of the production screen and Figure 15(b) shows the fertiliser usage time information.

Figure 11 Mobile application homepage

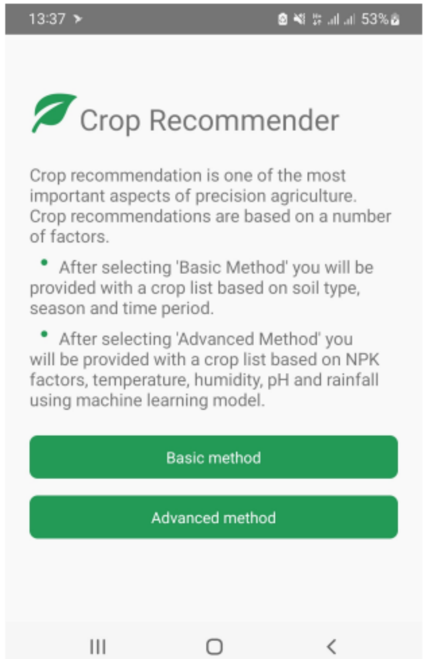


Figure 12 Basic method screen

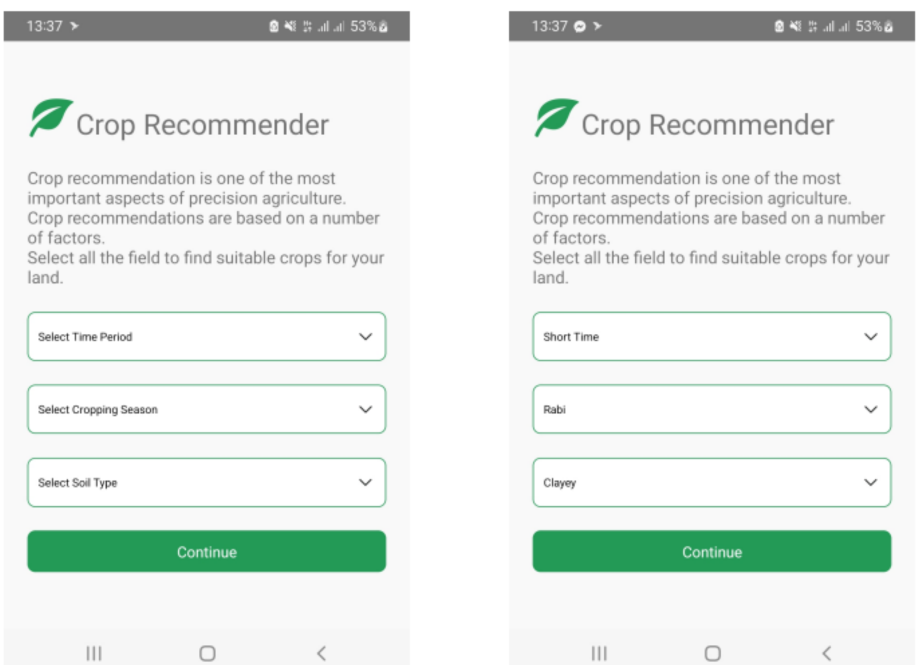


Figure 13 Advanced method screen

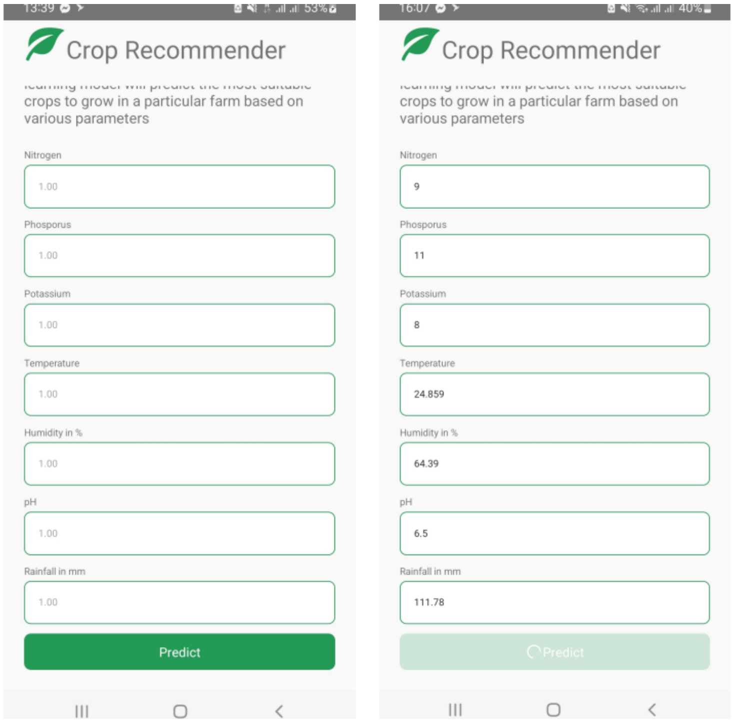
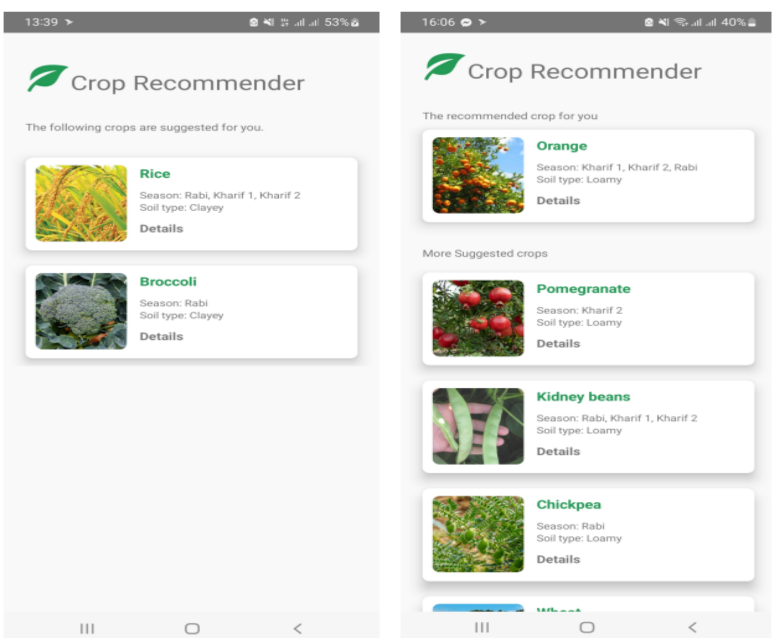
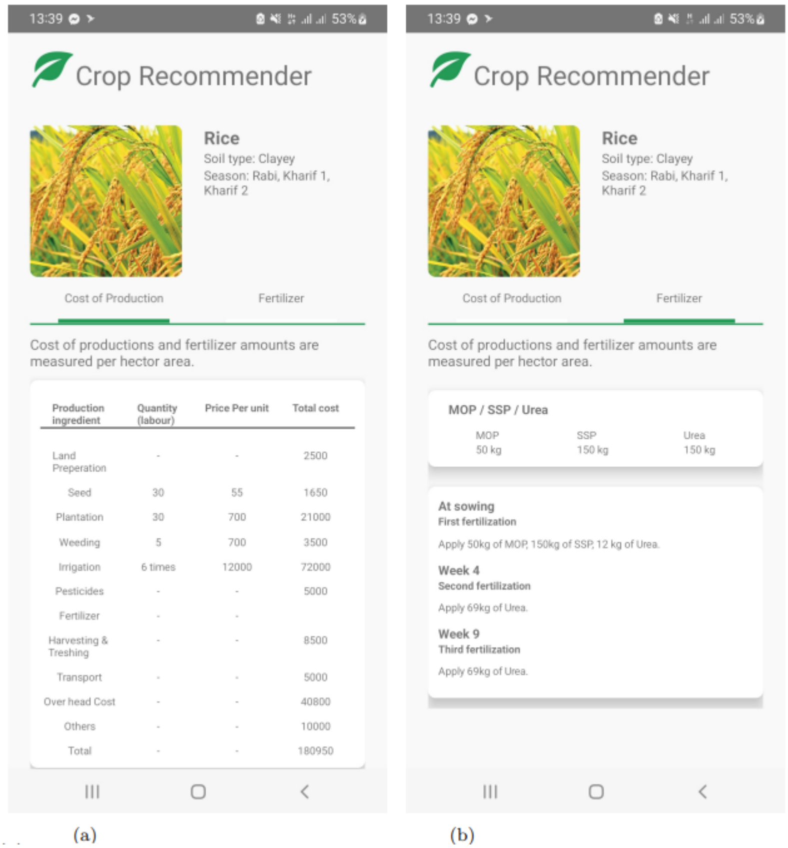


Figure 14 Crops list screen



(a) Crop list for basic method (b) Crop list for advanced method

Figure 15 Fertiliser recommendation and cost of production screen

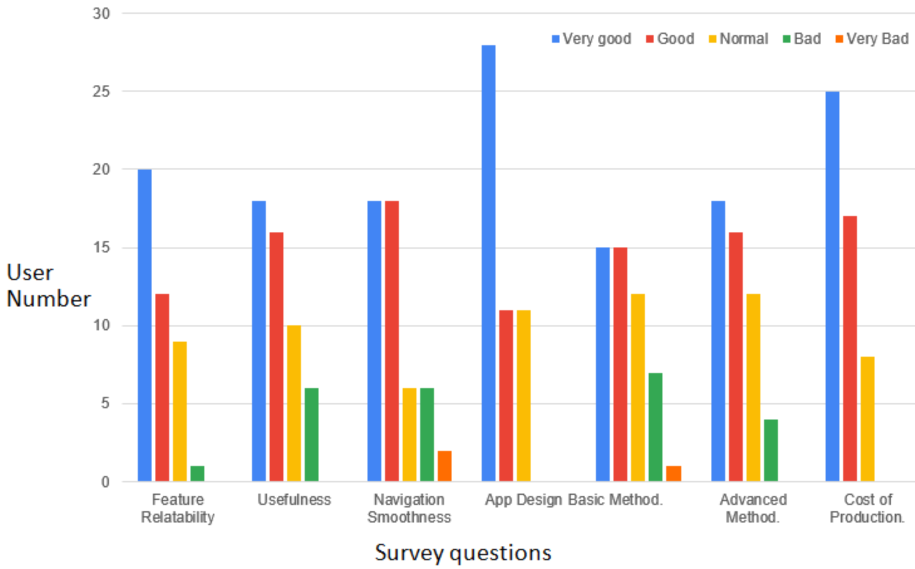


4.2 *User review or evaluation*

With the help of 50 reviewers via an interview process, we analysed the user evaluation performance of our Android application. Figure 16 shows the user assessment results. The survey questions are given on the x-axis. The y axis shows the user number regarding the evaluation remarks.

To assess the app's performance we have asked several survey questions to users such as *did you find the features of the application relatable with cultivation assistance? (feature relatability), how much do you think the application is useful? (usefulness), was the windows of the app navigating smoothly? (navigation smoothness), how attractive do you find the application? (app design), rate the performance of the basic method (basic method performance), rate the performance of the advanced method (advanced method performance), and rate the representation of the cost of production (cost of production feature*). It can be observed from Figure 16 that most of the reviewers rated our application as 'very good' in terms of usefulness, attractive design, and user-friendliness.

Figure 16 Users application evaluation



5 Conclusion

This paper presents a system using machine learning that predicts suitable crops based on multiple important factors such as soil type, features, season, and yielding cost. This paper develops an android application that forecasts crops based on elements like soil type, traits, season, and yielding cost. The application incorporates the cost of production data from Upazilla Parishad's agriculture department and fertiliser recommendations for the crops. The cost of production data reveals the fields where the farmer should make investments as well as the total cost per hectare. Those in the agricultural sector would benefit greatly from having the option to select which crop to grow under parameters such as soil type, features, season, and yielding cost. This type of system may provide a thorough understanding of how to cultivate any plants, including an estimate of profit for a certain budget, the cultivation process, and the appropriate plants for a given type of soil. In the future, the mobile application can use more parameters for crop recommendation and include more features based on user recommendations. Further, we will improve the security of our mobile application and expand the size of our dataset to give the model more data to train on.

References

Bandara, P. et al. (2020) 'Crop recommendation system', *International Journal of Computer Applications*, Vol. 975, p.8887.

Bangladesh Agricultural Research Council (2018) *Fertilizer Recommendation Guide 2018*. Available online at: http://www.bfa-fertilizer.org /wp-content/uploads/2019/09/Fertilizer-Recommendation-Guide2018-English.pdf (accessed in June 2022).

Bepery, C. et al. (2020) 'Framework for internet of things in remote soil monitoring', *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, IEEE Explore, pp.1–6.

Doshi, Z. et al. (2018) 'Agroconsultant: intelligent crop recommendation system using machine learning algorithms', *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE Explore, pp.1–6.

Gill, H.K. et al. (2021) 'A context aware recommender system for predicting crop factors using LSTM', *2021 Asian Conference on Innovation in Technology (ASIANCON)*, India, pp.1–4.

Gupta, R. et al. (2021) 'WB-CPI: weather based crop prediction in India using big data analytics', *IEEE Access*, Vol. 9, pp.137869–137885.

Kaggle (2009) *Crop recommendation dataset*. Available online at: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset (accessed on 16 June 2022).

Kulkarni, N.H. et al. (2018) 'Improving crop productivity through a crop recommendation system using ensembling technique', *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pp.114–119.

Mokarrama, M.J. et al. (2017) 'Rsf: a recommendation system for farmers', *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pp.843–850.

Pande, S.M. et al. (2021) 'Crop recommender system using machine learning approach', *5th International Conference on Computing Methodologies and Communication (ICCMC)*, India, pp.1066–1071.

Pudumalar, S. et al. (2017) 'Crop recommendation system for precision agriculture', *2016 Eighth International Conference on Advanced Computing (ICoAC)*, pp.32–36.

Rajak, R.K. et al. (2017) 'Crop recommendation system to maximize crop yield using machine learning technique', *International Research Journal of Engineering and Technology*, Vol. 4, No. 12, pp.950–953.

Rajeswari, A. et al. (2020) 'Fuzzy decision support system for recommendation of crop cultivation based on soil type', *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp.768–773.

Reddy, D.A. et al. (2019) 'Crop recommendation system to maximize crop yield in Ramtek region using machine learning', *International Journal of Scientific Research in Science and Technology*, Vol. 6, No. 1, pp.485–489.

Sadia, S. et al. (2021) 'A fruit cultivation recommendation system based on Pearson's correlation co-efficient', *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, pp.361–365.

Salpekar, H. et al. (2019) 'Design and implementation of mobile application for crop yield prediction using machine learning', *Global Conference for Advancement in Technology (GCAT)*, IEEE Explore, pp.1–6.

Suresh, G. et al. (2021) 'Efficient crop yield recommendation system using machine learning for digital farming', *International Journal of Modern Agriculture*, Vol. 10, No. 1, pp.906–914.

Talukder, S. et al. (2020) 'Enhancing crops production based on environmental status using machine learning techniques', *2020 International Conference on Computer Science and Its Application in Agriculture (ICOSICA)*, pp.1–5.

Thilakarathne, N.N. et al. (2022) 'A cloud enabled crop recommendation platform for machine learning-driven precision farming', *Sensors, MDPI*, Vol. 22, No. 16, pp.1–21.

Vaishnavi, S. et al. (2021) 'Agricultural crop recommendations based on productivity and season', *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, pp.883–886.

Villanueva, A. et al. (2022) 'Real-time best-fitted crops recommendation system based on agricultural soil health', *3rd International Informatics and Software Engineering Conference (IISEC)*, Ankara, Turkey, pp.1–6.