INTERNATIONAL JOURNAL OF RESEARCH CULTURE SOCIETY ISSN(O): 2456-6683
Monthly Peer-Reviewed, Refereed, Indexed Journal [Impact Factor: 6.834]
Volume - 7, Issue - 3, March - 2023 Publication Date: 31/03/2023

DOIs:10.2017/IJRCS/202303014 --:-- Research Paper / Article

Crop Recommendation System Using Random Forest Algorithm

[1]G. Buvaanyaa, [2]Dr. S. Radhimeenakshi
[1]PhD., Research Scholar, Department of Computer Science, Tiruppur Kumaran College for Women, Tiruppur, TamilNadu, India.
[2]Research Supervisor, Associate Professor, Department of Computer Science, Tiruppur Kumaran College for Women, Tiruppur, TamilNadu, India.
Email - [1] buvaanyaa220497@yahoo.com

Abstract: *India is the place where there is agribusiness and it is the significant wellspring of economy.70% of Indian populace straightforwardly depends on farming. The regular issue existing among the youthful Indian ranchers is to pick the correct yield dependent on the dirt prerequisites. Because of this, they face a genuine difficulty in efficiency. Arising advancements can be utilized to further develop efficiency of the harvests by changing conventional cultivating over completely to accuracy cultivating. The significant issue yet to be settled is developing exact harvest at exact time. This should be possible with the assist with machining learning calculations which is viewed as a powerful strategy for anticipating the reasonable harvest. The crop recommendation parameters such as NPK, temperature, humidity, rainfall and pH are collected from the benchmark repository. The framework created utilizing Machine Learning greatly assists the farmers to take a valuable decision.*

Key Words: *Agriculture, Crop Recommendation, Crop Yield, Precision Farming, Prediction, Machine Learning.*

1. INTRODUCTION:

Farming is the foundation of India. As we known, food stands first needing perseverance cultivating region ought to be given the most raised tendency being created. Indian agribusiness region addresses 18% of Indian cultivating absolute public result (Gross Domestic Product) and gives work to half of the country's workforce. The standard clarification behind considered Agribusiness region is in light of the fact that it expects a huge part in developing the country's economy. The proposed Framework uses the Harvest Determination as the zone of investigation since it is the first and most huge development in the communication of rustic new development and the achievement of this movement guarantees the outcome of creation.

Yield forecast is an essential issue in agribusiness. Any farmer is enthusiastic about realizing how much yield he will expect. Research the different related credits like region, pH regard from which alkalinity of the soil is settled. Close by it, level of enhancements like Nitrogen (N), Phosphorous (P), and Potassium (K), Area sort of soil, supplement assessment of the soil in that region can be settled. All of these credits of data will be inspected, train the data with various proper Machine Learning estimations for making a model. The structure goes with a model to be careful and exact in predicting crop yield and convey the end client with genuine recommendations about required compost extent subject to soil limits of the land which move up to grow the reap yield and augmentation farmer pay. Sort of soil expects a huge part in the collect yield. Suggesting the usage of composts might help the farmers with making the best decision for their managing situation.

By totally separate the previous data we can propose the farmer for a prevalent reap for the improved yield. Collect yield estimate is the critical investigation which helps with getting food. For the better perception of the gather yield, we really want to inspect massive data with the help of Machine Learning computation and suggest the farmer for a predominant reap.

2. OBJECTIVES:

- Design a recommendation system for accurate crop selection based on the various NPK, temperature, humidity, rainfall,pH.
- To improve crop productivity by providing predictions of high accuracy and efficiency through the machine learning algorithm.
- To reduce the wrong choice on a crop by application of principles of precision agriculture.

3. PROBLEM CONTEXT :

Productivity of a harvest significantly relies upon weather patterns, yield of the harvest and expenses of development and creation. Financial government assistance of rancher relies upon yield of the harvest as well as interest for the harvest. Horticulture being the essential vocation of the work force in India, many factors should be tended to while pursuing a yield choice as it influences the rancher's financial welfare [3]. In the event that a rancher can be suggested regardless of whether a decision of harvest will be productive in light of key elements like yield, weather conditions figure, market interest and barely any others costs then it would advance financial government assistance of ranchers.

4. LITERATURE REVIEW :

There are a few horticultural yield suggestion frameworks accessible thinking about different boundaries by utilizing ML calculations. Different ML methods are applied in agribusiness area to concentrate on the verifiable information which can be useful to ranchers as well as country's economy. This part surveys not many examinations done in horticulture for crop proposal and productivity of harvest to rancher.

Taj at el [1] applied both order and relapse procedures to construct a harvest proposal framework. Information utilized comprised of boundaries related soil condition and climate conditions. Beginning characterization is finished utilizing K-Nearest Neighbors (KNN) to find the best boundaries that altogether affect the harvest yield. Consequently, a relapse model utilizing Artificial Neural Networks (ANN) is utilized to anticipate a harvest for proposal. This study was done principally to address food security issue in Egypt. Banavlikar et al. [2], thought of a thorough, precise and strong yield suggestion framework worked by utilizing brain networks with an intention of assisting ranchers with picking a right harvest for a specific area of land. Soil and temperature are thought of and Components like soil dampness sensor, mugginess sensor and a temperature sensor are sent to quantify the water content in soil, measure of fume present in encompassing air.

Devadhe et al [4] proposed a specialist framework for crop determination considering boundaries like area wise month to month precipitation and efficiency of harvests during 2000 to 2014 in Maharashtra. Linear regression, Decision tree and Random forest calculations are applied to anticipate the yield pace of harvests which might assist with working on the choice of grouping of occasional harvests to be planted. Be that as it may, because of thought of just two or three boundaries, the review is restricted exclusively for occasional choice of harvests. Random Forest calculation gave better forecast results when contrasted with straight relapse promotion choice tree.

Jain et al [5] proposed a harvest determination strategy in light of different elements like natural, financial and yield rate to boost the yield creation which can help ranchers and economy of country to defeat the food supply interest. To accomplish the expansion in yield, determination of harvest in view of specific elements assumes a significant part. The element of cost is added to different boundaries like soil type, precipitation, temperature for crop choice. Weka classifiers and relapse techniques are utilized to foresee the proper determination of harvest and afterward a yield sequencing strategy is proposed by utilizing crop sequencing calculation in view of yield rate and market cost. Additionally proposed AI models can be utilized in various ways in agribusiness area like water system, sickness identification, design discoveries which can additionally propel horticulture area. Rajak et al. [6] concocted a harvest proposal framework to boost crop yield by considering soil explicit properties gathered from soil testing research facilities from pune, Maharashtra. Likewise consolidated general yield information. Soil assumes a significant part in efficiency of harvest. The review is more unambiguous towards soil ascribes like PH, shade of soil, surface and so forth. An Ensembling method called greater part casting a ballot procedure is utilized to foresee the yield.

5. METHODOLOGY:

A brief step by step procedure of designing the crop recommendation system is explained as follows:

Step 1: Input

The info dataset is a yield suggestion document containing the dataset, which must be exposed to preprocessing.

Step 2: Preprocessing of information

Input dataset is dependent upon different preprocessing strategies like filling of missing qualities, encoding of straight out information and scaling of values in the suitable reach

Step 3: Splitting into preparing and testing dataset

The preprocessed dataset is then parted into preparing and testing dataset in view of the predefined split proportion. The split proportion considered in the proposed work is 75:25, and that implies 75% of the dataset is utilized for the preparation the troupe model and the rest 25% is utilized as test dataset.

Step 4: Building individual classifiers on the training dataset

The training dataset is taken care of to every one of the independent base learners and the singular classifiers are assembled utilizing the training dataset.

Step 5: Testing the information on every one of the classifiers
The testing dataset is applied on every one of the classifiers, and the singular class names are gotten.
Step 6: Ensembling the singular classifier yield
The class marks got from the singular classifiers is exposed to get an ensembled class name as the last forecast.

A. Ensemble Framework

The ensemble framework is of utmost importance. The ensemble framework is explained as follows.

Before diving into the details of the ensemble framework, the actual meaning of ensembling and the reason for its usage[11]. Ensembling is a technique of building a prescient model by incorporating multiple models. The main reason for using an ensemble framework is that it provides a classifier that outperforms each of the individual classifiers.

Ensembling uses two frameworks, dependent framework and independent framework: In the dependent framework, the yield of one classifier is utilized in the development of the following classifier. The second method involves independent method, that is each classifier produces a class label in an independent fashion[10]. All the classifiers work in a parallelized manner. The output of one classifier is independent of the other. The independent method has been used in the proposed work since it reduces the execution time.

The ensemble framework comprises of the basic components that are explained as follows:

a) **Training set** – A labelled set of instances that is utilized in training the ensemble model. Each example in the training set is potrayed as attribute-value vectors.

b) **Base Inducers** - Inducer is an inducing algorithm that produces a classifier on feeding a labelled set of instances to the inducer as input. The resulting classifier shown in figure 1 gives a distinctive potrayal of the generalized relationship between the input attribute and the target attribute.

$$M=I(S)$$

M => Classifier , I => Inducer, S => Training set

c) **Diversity generator** – Generation of diverse classifiers[7].

d) **Combiner** – Responsible for combining the class labels obtained from the individual classifiers shown in table 1.

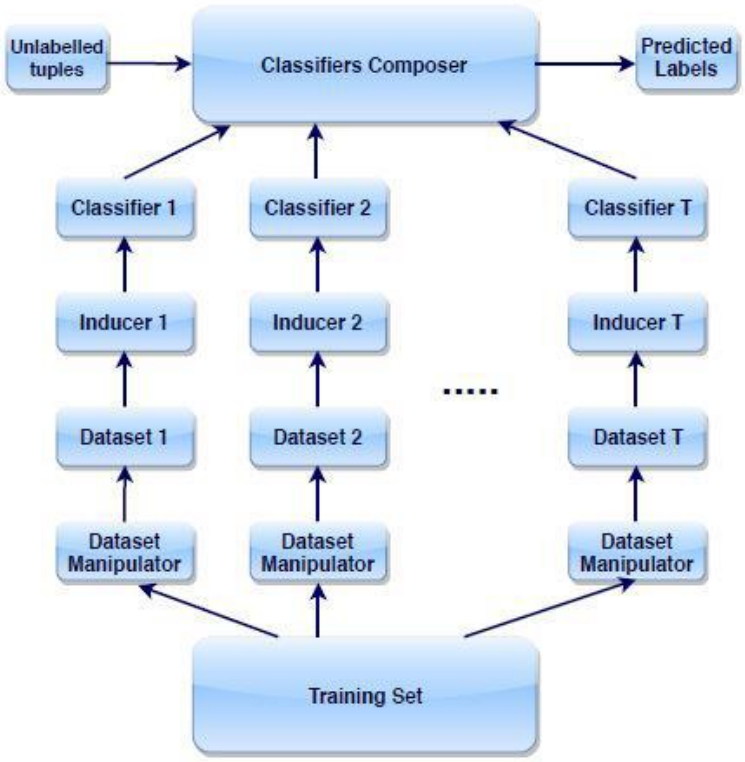

Figure 1: System Architecture

Advantage of using the independent framework of ensembling:

a) Enhances the prescient intensity of the classifiers.

b) Decreases the aggregate execution time.

Table 1: Experimental work for Crop Recommendation System

S.No	N	P	K	Temperature	Humidity	pH	Rainfall	Label
1	90	42	43	20.87974	82.00274	6.502985	202.9355	Rice
2	85	58	41	21.77046	80.31964	7.038096	226.6555	Rice
3	60	55	44	23.00446	82.32076	7.840207	263.9642	Rice
4	71	54	16	22.6136	63.69071	5.749914	87.75954	Maize
5	61	44	17	26.10018	71.57477	6.931757	102.2662	Maize
6	80	43	16	23.55882	71.59351	6.657965	66.71995	Maize
7	40	72	77	17.02498	16.98861	7.485996	88.55123	Chickpea
8	23	72	84	19.02061	17.13159	6.920251	79.92698	Chickpea
9	39	58	85	17.88776	15.4059	5.996932	68.54933	Chickpea
10	13	60	25	17.13693	20.59542	5.685972	128.2569	Kidneybeans
11	25	70	16	19.63474	18.90706	5.759237	106.3598	Kidneybeans
12	31	55	22	22.9135	21.33953	5.873172	109.2256	Kidneybeans
13	3	72	24	36.51268	57.92887	6.031608	122.654	Pigeonpeas
14	40	59	23	36.89164	62.73178	5.269085	163.7267	Pigeonpeas
15	33	73	23	29.23541	59.38968	5.985793	103.3302	Pigeonpeas
16	3	49	18	27.91095	64.70931	3.692864	32.67892	Mothbeans
17	22	59	23	27.32221	51.27869	4.371746	36.50379	Mothbeans
18	36	58	25	28.66024	59.31891	8.399136	36.9263	Mothbeans
19	19	55	20	27.43329	87.80508	7.185301	54.73368	Mungbean
20	8	54	20	28.33404	80.77276	7.034214	38.79764	Mungbean
21	36	55	20	27.0147	84.34263	6.635969	55.29635	Mungbean

B. Random Forest

A Random Forest is a classifier comprising of accumulation of tree-organized classifiers where independent random vectors are disseminated indistinguishably and each tree make a unit choice for the most mainstream class at input x is shown in figure 2. A random vector is produced which is autonomous of the past arbitrary vectors with same dissemination and a tree is created by utilizing the training set[8,9]. The main advantages of considering the Random Forest algorithm is that it provides better accuracy, vigorous to the outliers, quicker than bagging and boosting, basic and easy to parallelized.

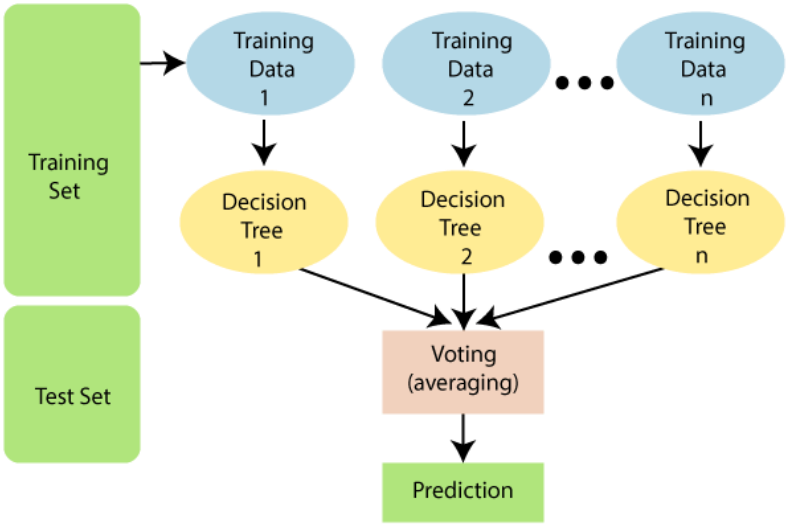

Figure 2: Working of Random Forest Algorithm

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

Algorithm for Random Forest in Machine Learning:
Step 1: Select random K data points from the training set.
Step 2: Build the decision trees associated with the selected data points (Subsets).
Step 3: Choose the number N for decision trees that you want to build.
Step 4: Repeat Step 1 &2.
Step 5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

DATASET DETAILS

The dataset considered for usage in the given proposed work is a crop recommendation dataset primarily comprising of soil properties, along with the temperature, humidity, rainfall details. An open source dataset is obtained from the benchmark repository.

The dataset attributes that are of prime importance are

- NPK of the soil
- pH value
- Humidity
- Rainfall
- Temperature

6. RESULTS :

The collected data is initially subjected to pre-processing and then architectural flow diagram has been shown in figure 3. Post dataset pre-processing, the dataset is divided into training set and test set samples. Out of the 100% samples, 75% samples are used as training samples, and the rest 25% samples are used as test samples. Each of the samples is trained and tested on the Random Forest algorithm. The average accuracy of crop recommendation is 91.99%.

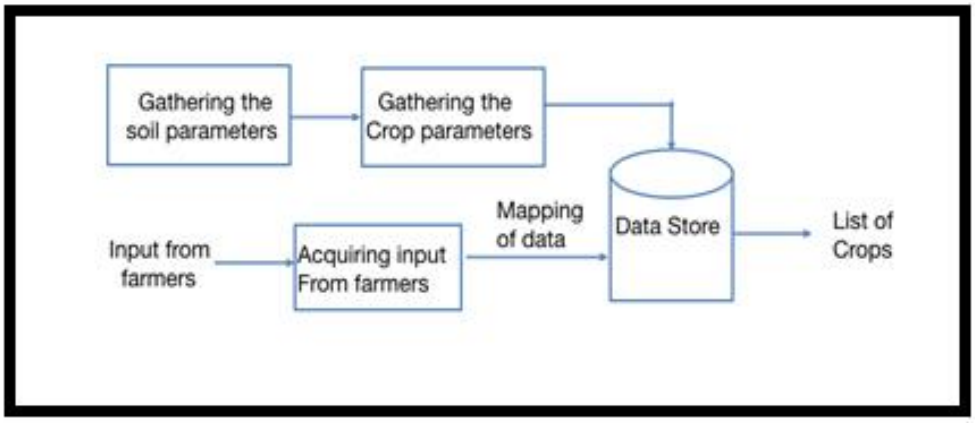

Figure 3: Architectural diagram for crop recommendation system

A. Data Pre-Processing

Information pre-handling is a procedure that is utilized to change over the crude information into clean dataset. Here the harvest and area datasets are assembled in crude arrangement which isn't possible for the investigation. So the crude information in the datasets is cleaned is shown in figure 4.

- Load data in Pandas
- Drop columns that aren't useful
- Drop rows with missing values
- Create dummy variables
- Take care of missing data
- Convert the data frame to NumPy
- Divide the data set into training data and test data

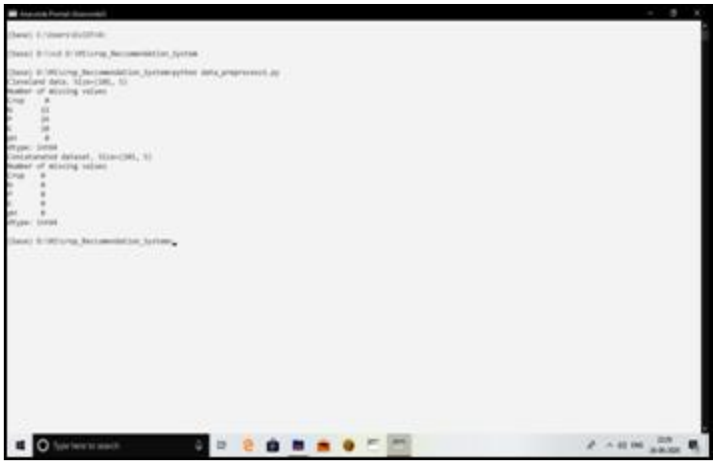

Figure 4: Data Pre-processing of crop dataset

B. Applying Machine Learning Algorithm

AI is perhaps the most compelling and amazing innovations in this day and age. AI is an apparatus for transforming data into information. AI encourages PCs in building models from test information robotize dynamic cycles dependent on information inputs.

The suggestion framework requires grouping and bunching calculations to perform planning of datasets. The proposed framework utilizes Random Forest machine calculation to perform AI shown in figure 5. On an examination led inside different calculations, the Random forest was found to give most noteworthy effectiveness and exactness contrasted with choice tree and so forth subsequent the Random forest calculation is utilized in the proposed framework to locate the appropriate yield list.

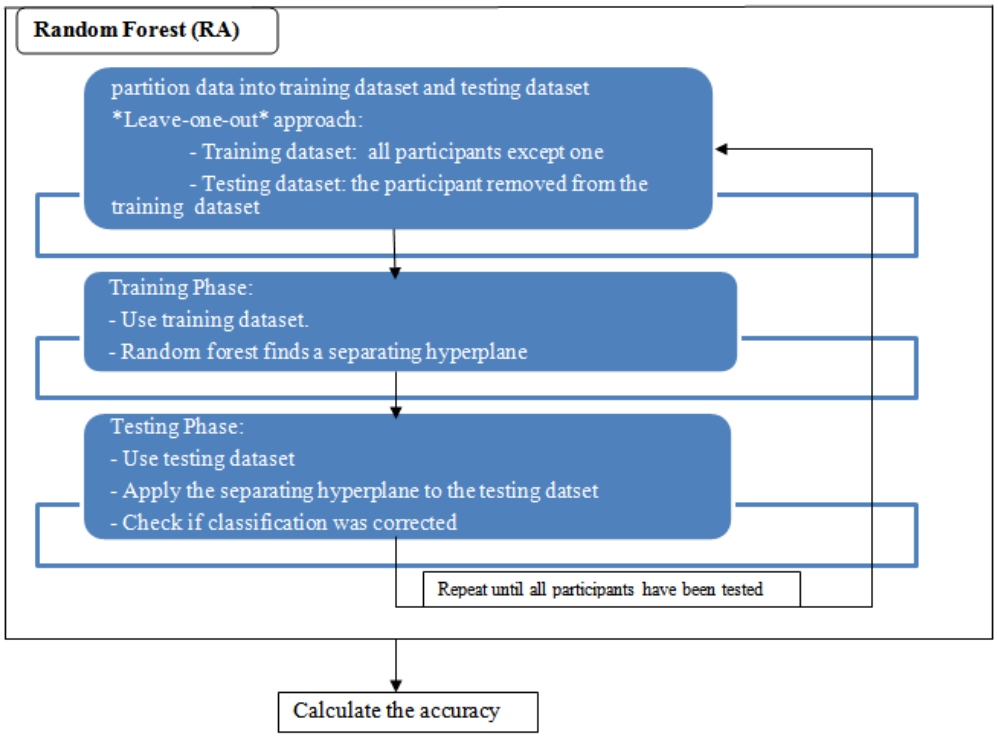

Figure 5: Flowchart for Random Forest Algorithm

C. Listing the Suitable Crops

The framework utilizes supervised machine learning algorithm to suggest reasonable harvests with higher precision and productivity. The framework records have been shown in table 2 that the, appropriate harvests dependent on the dirt and leaves it upon the ranchers to settle on the yield to be planted.

Table 2: Output for Crop Recommendation System

Nutrients				Recommended Crop
N	P	K	pH	
85	42	44	6.502985	Rice
71	44	16	6.931757	Maize
40	72	85	5.996932	Chickpea
13	70	22	5.685972	Kidneybeans
19	54	20	7.034214	Mungbean

7. CONCLUSION:

A crop recommendation system has been designed that takes into consideration the crop recommendation dataset with respect to the few crops. The crop recommendation dataset is first preprocessed and then the ensembling technique performs a critical function in the classification of the few crops. The individual base learners used in the ensemble model is Random Forest provide the best accuracy. The accuracy obtained using the ensembling technique is 91.99%. Hence, the proposed work provides a helping hand to the farmer in the accurate selection of the crop for cultivation. This creates an exponential gain in the crop productivity which in turn boosts the economy of the country. In future work, various machine learning algorithm can take place for further work.

REFERENCES :

1 Taj, M. B.N., Kavya, H.C., Nayana, R. R., Bindu, H.S., Meghana, D. P. , " A Crop Recommendation System for Precision Agriculture," International Journal of Engineering Research & Technology, 6(15), 2018.
2 T. Banavlikar, A. Mahir, M. Budukh, S. Dhodapkar, "Crop recommendation system using Neural Networks," International Research Journal of Engineering and Technology (IRJET), vol. 5(5), pp.1475-1480, 2018.
3 G.Buvaanyaa, Dr.S.Radhimeenakshi "The Impact Of Data Mining In Machine Learning Algorithm For Crop Recommendation To Yield – A Review" 2022 JETIR July 2022, Volume 9, Issue 7.
4 S. Devadhe, A. Kausadikar, P. Daphal, A. Joshi, A. 2017, "Expert System for Crop selection. International Journal of Scientific Research in Science and Technology, Volume 3(3), pp. 436-438, 2017.
5 N. Jain, A. Kumar, S. Garud, V. Pradhan, P. Kulkarni, "Crop Selection Method Based on Various Environmental Factors Using Machine Learning", International Research Journal of Engineering and Technology (IRJET),Volume (04), 2017.
6 R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique", Int Res J Eng Technol, vol. 4(12), pp. 950–953, 2017.
7 LiorRokach, "Ensemble-based classifiers", ArtifIntell Rev (2010) 33:1–39DOI 10.1
8 D. Bhattacharjee et al. (eds.),"A Fuzzy Logic-Based Crop Recommendation System" Proceedings of International Conference on Frontiers in Computing and Systems, Advances in Intelligent Systems and Computing, Springer Nature Singapore Pte Ltd. 2021
9 Dr.G.Suresh (2021), "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming". International Journal of Modern Agriculture.
10 Anguraj.K (2021), "Crop Recommendation on Analyzing Soil Using Machine Learning". Turkish Journal of Computer and Mathematics Education.
11 Soumya Sri Attaluri (2020), "Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques". Journal of Applied Technology and Innovation.