International Journal of Advanced Science and Technology
Vol. 29, No. 03, (2020), pp. 12738 - 12747

Crop Recommendation System through Soil Analysis Using Classification in Machine Learning

[1]Dr.A.K.Mariappan, [2]Ms C. Madhumitha, [3]Ms P. Nishitha, [4]Ms S. Nivedhitha
[1]Professor, Information Technology, Easwari Engineering College, Chennai, Tamilnadu.
[2]Student, Information Technology, Easwari Engineering College, Chennai, Tamilnadu.
[3]Student, Information Technology, Easwari Engineering College, Chennai, Tamilnadu.
[4]Student, Information Technology, Easwari Engineering College, Chennai, Tamilnadu.
[1]maris2612@yahoo.com
[2]98cmadhumitha@gmail.com
[3]nishitha.prakash@gmail.com
[4]nivedhitha2aim@gmail.com

Abstract

India is the land of agriculture and it is the major source of economy.70% of Indian population directly relies on agriculture. The common problem existing among the young Indian farmers is to choose the right crop based on the soil requirements. Due to this, they face a serious setback in productivity. Our work proposes to help farmers determine the soil quality by doing analysis on its various parameters and to suggest crops based on the results obtained using machine learning approach. The system uses the Classification algorithm of K nearest neighbour to improve the efficiency of Crop Recommendation System. The system maps the soil and crop data to predict the list of suitable crops for the soil and it also provides the information about nutrients which are deficient in soil for the particular crop . Hence it leaves upon the user to decide on the crop to be sown. Thus, the system helps to provide knowledge to the dilettante farmers.

Keywords:*KNN algorithm, croprecommendation, soilrecommendation, nutrients, classification, machine learning.*

1. Introduction

Agriculture is the backbone of India. As we known, food stands first in the basic need of survival, agriculture sector needs to be given the highest preference in development. Indian agriculture sector accounts for 18 per cent of Indian agriculture gross domestic product (GDP) and provides employment to 50% of the country's workforce. India is a global agricultural powerhouse. It is the world's largest producer of milk, pulses, and spices, and has the world's largest cattle herd (buffaloes), as well as the largest area under wheat, rice and cotton. It is the second largest producer of rice, wheat, cotton, sugarcane, farmed fish, sheep & goat meat, fruit, vegetables and tea. The main reason for considered Agriculture sector is because it plays a crucial role in developing the country's economy. The proposed System uses the Crop Selection as the area of research since it the first and most important step in the process of agricultural development and success of this step guarantees increase in result of production. Agriculture development provides

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

assistance to the crop producers with the help of various agricultural resources. As a result, it provides high productivity with low consumption of resources. Despite the overwhelming size of the agricultural sector, yields per hectare of crops in India are generally low compared to international standards. There are several factors that affect agriculture. Some of them are soil, climate and land relief. Soil is a critical part of successful agriculture and is the original source of the nutrients that we use to grow crops. The nutrients move from the soil into plants that we eat like tomatoes. Nutrients are also a part of the food animals (like cows) eat. In the end, we benefit from healthy soil. The healthiest soils produce the healthiest and most abundant food supplies .Farmers must use certain practices to make sure they are taking good care of the soil. Following best practices for crop production is essential, but prior to that, the selection of right crop suitable for the fertile land must be ensured to make the best use of the practices followed. Hence the System focuses on the crop selection as it directly contributes to the crop production rate.

2. Literature Survey

The related works underwent in recent time includes Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques [1] which uses the Maharashtra soil data. It uses Random forest Algorithm for crop Prediction. The system predicts suitable crop for the field under consideration based on region in Maharashtra state of India and type of soil. It also provides proper recommendation of fertilizers to the farmers.

Another Related study includes Comparative Analysis of Supervised Machine Learning Algorithms for GIS-Based Crop Selection Prediction Model [7] which predicts the crop along with its yield by using the historical data. By performing the computations on the historical data and using AI, the system predicts the crop for cultivation. In 2018, Sadia Afrin, Abu Talha Khan, MahrinMahia, Rahbar Ahsan and Mahbubur Rahman Mishal presented a paper on the title Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Different Agricultural Regions of Bangladesh [9] which mainly focuses on the analysis to predict Bangladesh's four most yielding crops; wheat, jute, T-Aman and mustard. To carry out the whole experiment, they have analyzed soil properties of medium high land and high land from different sub districts of Bangladesh and also their respective climatic data and crop production using data mining techniques such as K-means, PAM, CLARA and DBSCAN for clustering and four linear regression methods to predict crop yields.

According to their clustering analysis, PAM gives better result compared to K-means, DBSCAN and CLARA. Among the four regression techniques we have used, generalized linear model has shown the lowest RMSE. This proposed work can appear to be useful for agricultural research or knowledge to learn to avoid adversity and to get the best yield from variant soil nutrients and climatic factors. Although this proposed work focuses on medium highland and highland for soil nutrients and crop production, the same work can be extended by considering other land types and additional soil contents like water level, soil moisture, soil type and fertilizers too

In 2018, Sae-Han Suh, Ji-EunJhang, KwangheeWon,Sung-Y. Shin and Chang Oan Sung presented a paper on Development of Vegetation Mapping with Deep Convolutional Neural Network [10] which uses precision agriculture (PA) for improving the decision-making process. It aims to optimally allocate the resources

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

to maintain the sustainable productivity of farmland and reduce the use of chemical compounds.

P. S. Vijayabaskar, Sreemathi.R, and Keertanaa.E presented a paper on Crop Prediction Using Predictive Analytics [13]. They constructed a model for testing the soil fertility. It also suggests the crop which has to be planted depending upon the value obtained from the sensor. It also provides the regional wise information about the crop in the form of graph.

PankajBhambri, Inderjit Singh Dhanoa, Vijay Kumar Sinha, and Jasmine Kaur presented a paper on Paddy Crop Production Analysis Based on SVM and KNN classifier [6] which uses the KNN algorithm to predict paddy production rate. It was also found that KNN provides better accuracy than SVM for crop datasets.

From the results and findings, it was found that KNN machine Learning algorithm provides better accuracy and efficiency than the other classification algorithms.

3. Proposed System

The existing system predicts the crop yield by using the soil parameters and also recommends fertiliser. It uses the crop yield in-formation to make the end users decide on the crop to be sown. Hence the system is not simple enough for dilettante farmers to understand.

The proposed system obtains the soil and crop parameters and maps those to list the suitable crops. It passes the various inputs to the controller which uses the KNN algorithm for classification. The proposed system provides easy accessibility to the users. They are also easy to use and under- stand by the dilettante farmers. It improves the visualisation and understandability.

The Crop recommendation system incorporates three modules, the crop recommender through city-wise analysis, crop recommender through detailed analysis and soil recommender. Crop recommender through city wise analysis uses the crop and city dataset, the second modules recommends crop for a particular soil nutrient values and the third module provides information about the soil nutrients which are deficient for a particular crop.

1. Data Set Description

The following are the dataset used to test and train the pre- diction model of classification built by using KNN algorithm. The system uses two datasets: crop dataset and soil dataset. The crop dataset contains the corresponding pH, N, P and K values of crops and the soil dataset contains the corresponding pH, N, P and K for soils for various cities in Chennai and other districts of Tamil Nadu. The data are obtained from the Agriculture department of Tamil Nadu in Kancheepuram district.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

Table 1. Sample Dataset of Crop Data

Crop	N	P	K	pH
Rice	80	40	40	5.5
Wheat	100	40	0	5.5
Maize	80	40	20	5.5
Ragi	50	40	20	5.5

Table 2. Sample Dataset of Soil Data

Location	Y	X	N	P	K	pH
Yercaud	14.4212044	76.5545525	104	23	106	6.98
Vridhachalam	14.4251505	76.5226052	125	28	125	6.77
Vellore	14.3233453	75.7592645	95	9	198	5.26

2. System Architecture

The system architecture as shown in Figure 1 below consists of two datasets which includes crop dataset and soil dataset. The crop dataset contains the data samples of crops along with its parameters. The crop parameters which are used to describe the crop are pH, Nitrogen (N), Phosphorus (P) and Potassium (K).A vast range of crops grown in fertile lands of India are taken and characterised by using their crop parameters. The soil dataset contains the data samples taken from various places in Chennai and other district of TamilNadu. The soil parameters which are used to describe the soil are pH, Nitrogen (N), Phosphorus (P) and Potassium (K). The data samples are obtained from the authorised horticulture department of Tamilnadu-Kancheepuram district. The system uses an interface to acquire input from the users. The input type may differ according to the module selected by them.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

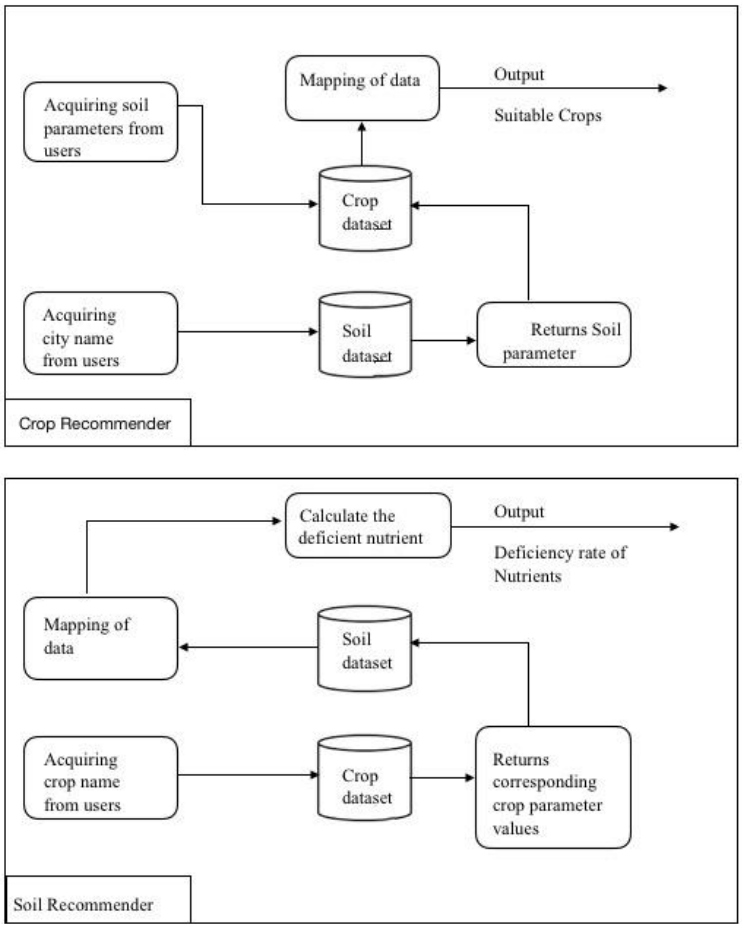

Figure 1. System Architecture diagram

3. Tools and Necessary Packages

The Crop recommendation system is build supportive to the windows platform on version 10 and above. The system must also be installed with python version 3.3. Any IDE like PyCharm can be used to deploy the system successfully. It can also be deployed on other operating system provided the version of tools and packages may vary. The necessary packages installed in python library includes: Scikit-learn, Numpy, pandas, flask, matplotlib and Scipy. For web development, some of the basic tools like HTML, CSS and javascript are used.

4. Crop Recommender Through City-wise Analysis

In the first module, the system requires the input of the region or city. The input is mapped with the city and crop data stored. The user inputs the city name through an interface. The acquired input is stored in a variable. The variable reaches out to the soil dataset to gather the soil parameters of the corresponding city. The parameters are then stored into an array. The list of parameters reaches out the crop dataset where the mapping takes place. The array list trains itself and with the help of KNN classification algorithm in machine learning, the array locates itself into the suitable class of crop. As a result of mapping, the dataset returns the suitable crop

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

for the particular soil parameter. The obtained result is displayed to the user through an interface.

Thus the farmer obtains the list of crops that are best suited for the city or region the farmer sows

Pseudocode

1. Assign city_name as x
2. Assign N, P, K, pH taken from the dataset as X
3. Map the X with the parameter values X1 on crop dataset
4. Create an empty list lt
5. Apply classification to the inputs X, X1
6. Apply clf.fit()
7. Apply clf.shuffle()
8. Apply the predict function to the variables X and X1
9. The function returns the suitable crop class Y for the city
10. For i in range(10):
11. append Y in lt
12. printlt[]

5. Crop Recommender Through Detailed Analysis

The second module helps the farmer to know what kind of crop grows best according to the soil nutrients that his farmland has.

The user inputs the soil parameters through an interface. The acquired input is stored in a variable. The variable reaches out to the crop dataset. The input and the data samples are mapped and the input locates itself into a crop class by using the KNN algorithm of classification in machine learning. As a result of mapping, the dataset returns the suitable crop for the particular soil parameter. The obtained result is displayed to the user through an interface. The data sample includes the N, P, K and pH for wide range of crops and fruits grown on the fertile soil of India. The KNN generates train and test variable X and Y which are also included into the dataset for increasing the accuracy

Thus, this feature requires the farmer to have the soil tested for obtaining the nutrient values of his land which is to be inputted into the system. The input is mapped with the soil and crop database and the output of list of crops that matches the particular soil is obtained.

Pseudocode

1. Assign soil parameters as x
2. Assign N, P, K, pH taken from the dataset as X
3. Create empty lists lt[]
4. Apply classification to the inputs x and X
5. Apply clf.fit ()

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

6. Apply clf.shuffle ()
7. Apply the predict function to the input variables x and X
8. The function returns the suitable crop class Y for the given parameters
9. Fori in range (10):
10. Append Y in lt[]
11. printlt[]

6. Soil recommender

The third module is for the farmers who have already decided upon what crop to be sown in his land but want to make sure that his farmland is nourished enough for it.

The user inputs the soil parameters and the name of particular crop to be sown through an interface. The acquired inputs are stored in two different variables. The variable reaches out to the crop dataset. The input and the data samples are mapped and the crop parameters for the particular crop are obtained. The obtained value is stored in another variable. Now, the difference between the soil and crop parameters is calculated which provides us the deficient nutrition rate. As a result, the dataset returns the difference in nutrition to the user through an interface.

Thus this feature enables the farmer understand the deficiency of soil in his farmland for the required crop.

Pseudocode

1. Assign soil parameters as x and particular crop name as x1.
2. Create empty lists lt[] lst[] and final[]
3. Apply classification to the inputs x and x1
4. Apply clf.fit()
5. Apply clf.shuffle()
6. Apply the predict function to the input variables x and x1
7. The list val[] contains the N,P,K and Ph for the particular crop
8. Convert the parameters into float and load onto val1[]
9. The val2[] contains the tuples of val1[].
10. For i n range(3):
11. final[]=val1[i]-val2[i]
12. print final

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

4. Experimental Result

The Recommendation System requires Classification and Clustering Algorithms to perform mapping of Datasets. The proposed system uses KNN Algorithm to perform Machine Learning. On an analysis conducted within various algorithms, the KNN was found to provide highest efficiency and precision compared to Decision tree, Random Forest etc. for crop datasets. Hence the KNN algorithm is used in the pro- posed system to find the suitable crop list. The existing Crop Recommendation System using KNN algorithm uses few parameters like temperature, rainfall, moisture etc., But, the Proposed System uses the KNN algorithm with the addition- al parameters like phosphorous, nitrogen, potassium, pH which are the major nutrients of the crop which helps in listing the more suitable crops according to the soil with the more accuracy than the existing system.

KNN algorithm: Pseudocode

1. X-train data (crop or soil dataset)
2. Y-class label of X
3. x- test data (input data)
4. knn-classify(X, Y, x)
5. for any input x do
6. calculate distance(X, x)-//distance function
7. end for
8. compute smallest distance for x using distance(X,x)
9. return Y //class label with minimum distance

From analysis done on the classification report of crop prediction system, KNN was observed to provide better accuracy for production rate than the SVM algorithm. The accuracy obtained as a result of using KNN algorithm was 89% and by using SVM algorithm was 80%.The accuracy is calculated based on the results obtained from test and train data. The dataset is split into 80% for training data and 20% for testing data. Then the dataset is fit into the model to perform prediction. The precision, recall and f1 factors obtained from the classification report were used to determine the accuracy of the system. Since, the KNN algorithm results in higher accuracy, the KNN algorithm for classification in Machine Learning was used to implement Crop recommendation system.

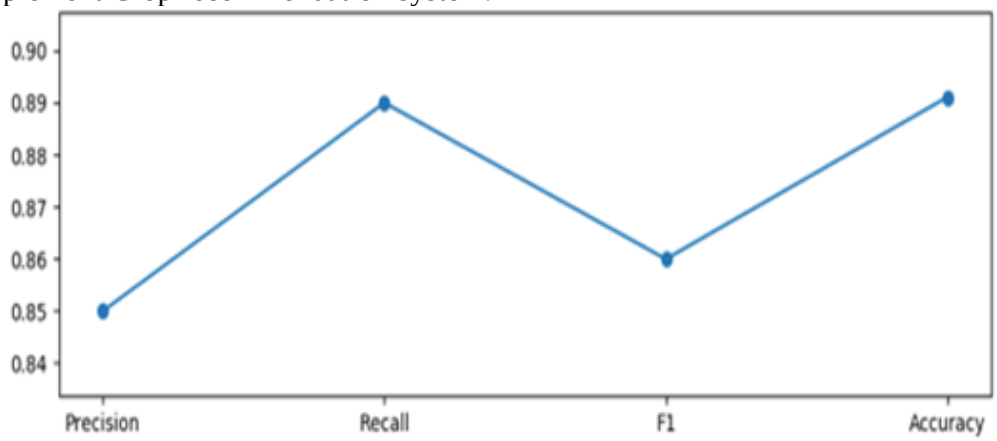

Figure 2. Prediction Accuracy using KNN algorithm.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

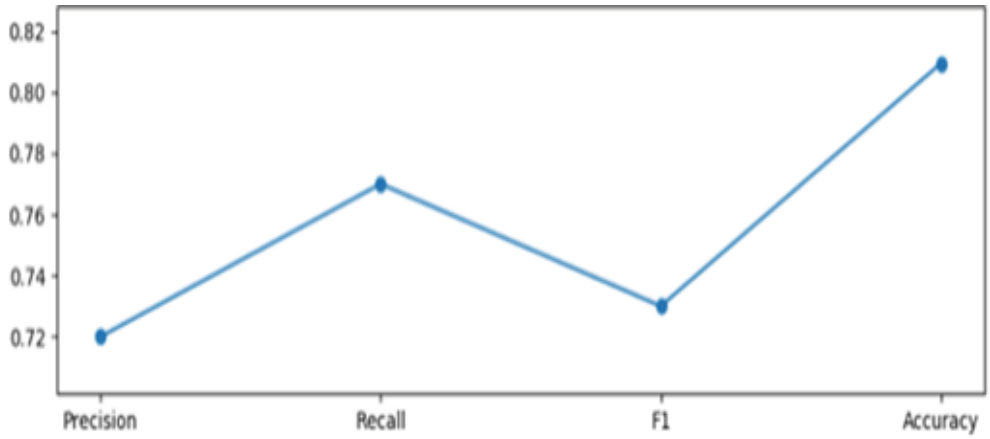

Figure 3. Prediction Accuracy using SVM algorithm

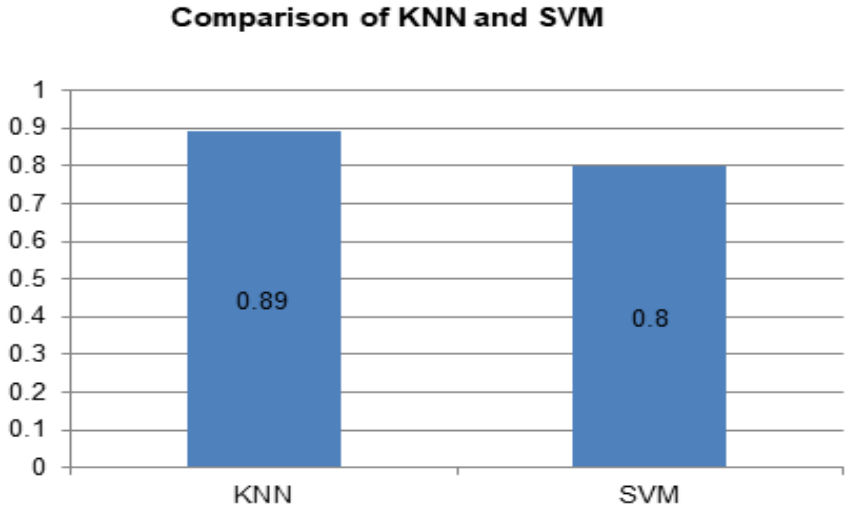

Figure 4. Comparison between two models with respect to their accuracy score.

5. Conclusion

The Crop Recommendation system uses KNN classification in Supervised Machine Learning Algorithm to recommend suitable crops with higher accuracy and efficiency. The system lists the suitable crops based on the soil and leaves it upon the farmers to decide on the crop to be sown.

6. Acknowledgement

We express our sincere thanks and gratitude to our Founder Chairman Dr.T.RParivendhar, and our beloved Chairman Dr.R.Shivakumar, for their support and inspiration. We would like to convey our due respect and regards to our Principal Dr.R.S.Kumar and Head Dr.N.Ananthi , for their constant encouragement and guidance. We owe our profound gratitude to our project supervisor Dr.A.K.Mariappan, Professor, who took keen interest in our project work and guided us all along in bringing out this project in complete shape by providing all the necessary information for developing a

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC

good system. We express our sincere thanks to all those who guided us throughout the process.

7. References

1. Archana Chougule, Vijay Kumar Jha and Debajyoti Mukhopadhyay, "Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques", Journal of Springer Nature Singapore Pte Ltd. (**2019),** pp. 1-2.
2. Aswathy S, S.Manimegalai, M.MariaRashmi Fernando, and Dr.J.Frank Vijay, "Smart Soil Testing",International Journal of Innovative Research in Science, Engineering and Technology Vol. 7, Special Issue 4, April **(2018)**, pp. 41-42
3. Bhanumathi, M.Vineeth and N.Rohit , "Crop Yield Prediction and Efficient use of Fertilizers" , IEEE International Conference on Communication and Signal Processing, April 4-6, **(2019)**, pp. 0769-0771.
4. ChandlaEllis,Vishal J, Saravanan L S and Vignesh B, "Realtime Data Analysis For Crop Prediction Using IOT",International Journal of Engineering and Techniques - Volume 4 Issue 2, Mar – Apr **(2018)**, pp. 296-297.
5. Excel Philip B. Guidang, "Inception-v3-Based Recommender System for Crops", ICMLSC 2019, January 25–28, Da Lat, Viet Nam , Association for Computing Machinery, **(2019)**, pp. 107-108.
6. PankajBhambri, Inderjit Singh Dhanoa, Vijay Kumar Sinha, and Jasmine Kaur, "Paddy Crop Production Analysis Based on SVM and KNN classifier", International Journal of Recent Technology and Engineering, Volume-8 Issue-5, January **(2020)**, pp.2791-2793.
7. Preetam Tamsekar, Nilesh Deshmukh, Parag Bhalchandra, Govind Kulkarni, Kailas Hambarde and Shaikh Husen ," Comparative Analysis of Supervised Machine Learning Algorithms for GIS-Based Crop Selection Prediction Model ", Journal of Springer Nature Singapore Pte Ltd. **(2019)**, pp. 310-312.
8. Pudumalar S, E.Ramanujam, R.HarineRajashree, C.Kavya, T.Kiruthika and J.Nisha.," Crop Recommendation System for Precision Agriculture", IEEE Eighth International Conference on Advanced Computing **(2016)**, pp. 32-34.
9. Sadia Afrin, Abu Talha Khan, MahrinMahia, Rahbar Ahsan, Mahbubur Rahman Mishal, Wasit Ahmed and Rashedur M. Rahman, "Analysis of Soil Properties and Climatic Data to Predict Crop Yields and Cluster Different Agricultural Regions of Bangladesh", Journal of IEEE **(2018)**, pp. 80-82.
10. Sae-Han Suh, Ji-EunJhang, KwangheeWon,Sung-Y. Shin and Chang Oan Sung, "Development of Vegetation Mapping with Deep Convolutional Neural Network", Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems ,October **(2018)**, pp. 1-1.
11. Suchithra M.S and Maya L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters", China Agricultural University. Production and hosting by Elsevier **(2019)**, pp. 1-6.
12. Surya A. Venkaiah and Kondajji Swati Sunitha, "A new approach for predicting crop yield prediction using data mining techniques ", International Scientific Journals JEISR **(2019)** Vol. 3–no.1, pp. 42-44
13. Vijayabaskar P.S, Sreemathi.R and Keertanaa.E ," Crop Prediction Using Predictive Analysis ", International Conference on Computation of Power, Energy, Information and Communication, **(2017)**, pp. 370-371.
14. YogeshGandge,Sandhya, "A Study on Various Data Mining Techniques for Crop Yield Prediction ", International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques, **(2017)**, pp. 420-422.
15. Zingade,Omkar Buchade,Nilesh Mehta,Shubham Ghodekar and Chandan Mehta, " Crop Prediction System using Machine Learning" , International Journal of Advance Engineering and Research Development Special Issue on Recent Trends in Data Engineering , Volume 4, Special Issue 5, Dec, **(2017)**, pp. 34-35.

ISSN: 2005-4238 IJAST
Copyright ⓒ 2020 SERSC