Machine learning approach: Recommendation of suitable crop for land using meteorological factors

*Srilakshmi A[1]
[1]Sastra Deemed to be university

[1*]srilakshmi@cse.sastra.edu

*Madhumitha K
[1]Sastra Deemed to be university

madhumithakathiravan@gmail.com

* Geetha K
Sastra Deemed to be university
geetha@cse.sastra.edu

Abstract. Increasing population increases the need for food. As most population migrates towards cities for employment, the cultivable lands are turning into factories and apartments. The landlords are selling the plots due to the loss they face after cultivation. This loss occurs due to improper selection of crop. The loss could be rectified if they are suggested with a suitable crop based on the meteorological factors over the land area like testing soil quality, humidity, pH, etc. The farmers in interior places face difficult in consulting with the experts. To overcome this problem, ANN came to play role and also gave an effective solution. After knowing the suitable crop for the field, it is getting easier to decide the fertilizers and intercrop alongside. The profit rate is considerably high using this method. It is also cost efficient. This paper discusses the model for crop prediction using Machine learning algorithms. The model is compared with different approaches such as random forest, decision tree and SVM aiming to get a complete solution for the crop prediction and recommendation problem.

Keywords: Artificial neural networks(ANN),Decision Tree(DT),Support Vector Machine (SVM)

1 Introduction

The model arrives at the solution for increasing crop production and profit for the farmers using ANN models like Random forest, SVM and decision tree. However, the Random forest arrives at the required prediction results with 95% accuracy than the rest of them. This model approaches the problem with all possible ANN models and each having its own way to classify images. Instead of giving images, model is trained with dataset of factors influencing crop yield. the model is easier to train and arrive at the prediction results thereby making it cost efficient and smarter method than consulting scientists and getting solutions. More the data given, more accurate the result is. The training dataset consists of around 3100 data. Thus, a lot of data can be trained in the same time. There is a lot of procedures involved in recommending crop, if this model is used efficiently, farmers can get the required profit along with market demanding crops. The model could be utilized for all crop varietieshence there is no restrictions in this model for crop prediction. The dataset is trained by the machine using SVM and Random forest algorithms. The SVM uses Gridsearch which is used to identify suitable hyperparameter for the training model. The kernels are actual replacement of the neural networks in this algorithm. The results are analyzed based on the accuracy values of each algorithms and whether they are providing the required crop prediction answers. Based on that values the crop prediction and plotted graph, the result proves to be effective in the field of agriculture.

2 Related Work

Images can be cropped from its vivid background to increase system efficiency. CNN after training a preprocessed image sends it to SVM classifier along with latitude and longitude values, which can be used for predicting the suitable crop [1]. Data digging systems can be used to collect surrounding conditions instead of conventional methods. This dataset when used with SVM, NAÏVE Bayes and Random Forest can give a better recommendation [2].DNN can be used for better prediction as it can cause only low prediction errors.[4] Crop sequencing algorithm helps in sequencing of plants based on the yield and market price at that time of the year. The crop sequencing algorithm used along with crop selection using ANN can provide farmers a huge profit [6].Crop selection algorithm consists of three parts. It includes choosing a crop, verifying its market rate and crop variety selection. By using CSVM and ANN ,the prediction could be made efficient[7]. By using c-means clustering and random forest architecture an efficient model for crop recommendation based on soil condition, weather, due factor, etc. taken into account [8].An application made of hardware and web applications which is used to test the climatic conditions in a plot of land and storing it in cloud and comparing with the trained data thereby producing the user, crop to be planted and fertilizers, pesticides to be used with them for high yield [9].SVM,Random Forest,NAÏVE Bayes,ANN are the best methods to be applied for crop prediction. Ensemble which is a datamining technique can be used in order to increase the efficiency [10].

3 Materials and Methods

The data has been collected from online github repository. It is stored in the form of csv file with 3100 rows x 5 columns. The meteorological factors including temperature, humidity, ph and rainfall is considered as features and result is taken as label which denotes the crop suitable for that particular atmospheric condition. These features are most dominating paramaters which is used to make prediction accurately. There are two categories in soil, such as Macronutrients and mironutrients. Among this macronutrients which is available in the soil is used to predict crop. The other part is prediction based on meterlogical factors.The Label consists of all types of crops. Training a huge amount of data gives the proper solution for the problem. The dataset hasn't required any preprocessing as it has already been preprocessed. Now the preprocessed dataset is ready to be trained by using machine learning algorithms.

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...
3095	25.287846	89.636679	6.765095	58.286977	watermelon
3096	26.638386	84.695469	6.189214	48.324286	watermelon
3097	25.331045	84.305338	6.904242	41.532187	watermelon
3098	26.897502	83.892415	6.463271	43.971937	watermelon
3099	26.986037	89.413849	6.260839	58.548767	watermelon

3100 rows × 5 columns

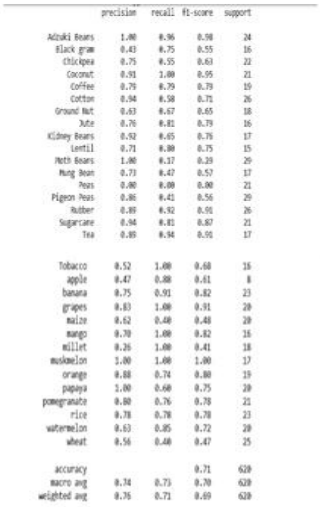
Fig.3.1 Parameters for learning

The above figure Fig.3.1 shows the arrangement of data in the dataset, the factors affecting crop growth are mentioned and suitable crop name is mentioned in label column. Dataset can contain some unfilled places or NaN values in them. some machine learning algorithms could not arrive at result if there are blank spaces in the provided dataset so the spaces are found out using code and the left-out values are defined as 0.The dataset would be complete now and it is ready for training

using deep learning algorithms giving an increased accuracy.The model is trained using Support Vector Machine (SVM), decision tree and Random forest algorithms. This model is used for multiclass classification. Out of the algorithms mentioned, Random forest gave the best result of 95% accuracy.

3.1 SVM hyper parameter tuning using Grid search

Support-vector machine performs the supervised learning with algorithms to analyze and classify data. It can be used for both linear and non-linear classification. Here the data is labelled so supervised learning is possible. The examples are plotted to a space and gap created separates the samples based on the features into categories.SVM has hyper-parameters namely C or gamma values. The suitable hyperparameter can be found by training using all possible combinations. There comes the Gridsearch to play its role in finding the best parameter. Here the svm classifier trained using three types of kernels namely polynomial, gaussian, sigmoid and respective confusion matrix using GridSearchCV.This model is trained for all possible kernel combination to find the best parameter .

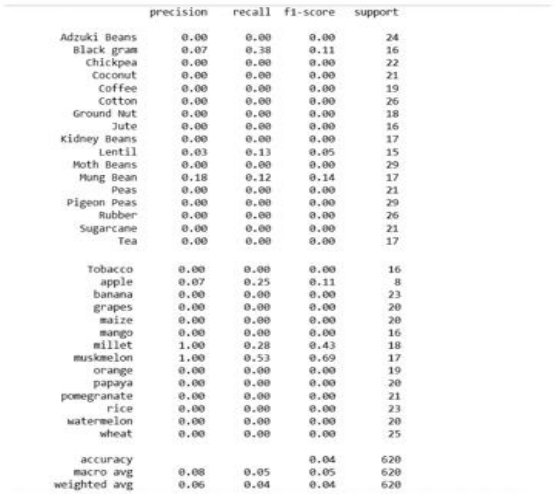


	precision	recall	f1-score	support
Adzuki Beans	1.00	0.96	0.98	24
Black gram	0.43	0.75	0.55	16
Chickpea	0.75	0.55	0.63	22
Coconut	0.91	1.00	0.95	21
Coffee	0.79	0.79	0.79	19
Cotton	0.94	0.58	0.71	26
Ground Nut	0.63	0.67	0.65	18
Jute	0.76	0.81	0.79	16
Kidney Beans	0.92	0.65	0.76	17
Lentil	0.71	0.80	0.75	15
Moth Beans	1.00	0.17	0.29	29
Mung Bean	0.73	0.47	0.57	17
Peas	0.00	0.00	0.00	21
Pigeon Peas	0.86	0.41	0.56	29
Rubber	0.89	0.92	0.91	26
Sugarcane	0.94	0.81	0.87	21
Tea	0.89	0.94	0.91	17
Tobacco	0.52	1.00	0.68	16
apple	0.47	0.88	0.61	8
banana	0.75	0.91	0.82	23
grapes	0.83	1.00	0.91	20
maize	0.62	0.40	0.48	20
mango	0.70	1.00	0.82	16
millet	0.26	1.00	0.41	18
muskmelon	1.00	1.00	1.00	17
orange	0.88	0.74	0.80	19
papaya	1.00	0.60	0.75	20
pomegranate	0.80	0.76	0.78	21
rice	0.78	0.78	0.78	23
watermelon	0.63	0.85	0.72	20
wheat	0.56	0.40	0.47	25
accuracy			0.71	620
macro avg	0.74	0.73	0.70	620
weighted avg	0.76	0.71	0.69	620

Fig.3.2.a.Polynomial kernel

	precision	recall	f1-score	support
Adzuki Beans	1.00	1.00	1.00	24
Black gram	0.52	0.81	0.63	16
Chickpea	0.81	1.00	0.90	22
Coconut	0.70	1.00	0.82	21
Coffee	0.23	0.37	0.29	19
Cotton	0.63	0.46	0.53	26
Ground Nut	0.50	0.39	0.44	18
Jute	0.54	0.94	0.68	16
Kidney Beans	0.92	0.65	0.76	17
Lentil	0.68	1.00	0.81	15
Moth Beans	1.00	0.55	0.71	29
Mung Bean	0.46	1.00	0.63	17
Peas	0.95	1.00	0.98	21
Pigeon Peas	1.00	0.17	0.29	29
Rubber	1.00	0.54	0.70	26
Sugarcane	1.00	0.62	0.76	21
Tea	0.39	0.53	0.45	17
Tobacco	0.46	1.00	0.63	16
apple	0.15	1.00	0.26	8
banana	0.61	0.74	0.67	23
grapes	0.59	1.00	0.74	20
maize	0.67	0.20	0.31	20
mango	0.59	1.00	0.74	16
millet	1.00	1.00	1.00	18
muskmelon	1.00	1.00	1.00	17
orange	0.00	0.00	0.00	19
papaya	1.00	0.10	0.18	20
pomegranate	0.00	0.00	0.00	21
rice	0.58	0.65	0.61	23
watermelon	0.00	0.00	0.00	20
wheat	0.38	0.20	0.26	25
accuracy			0.61	620
macro avg	0.62	0.64	0.57	620
weighted avg	0.65	0.61	0.57	620

Fig.3.2.b.Gaussian kernel



	precision	recall	f1-score	support
Adzuki Beans	0.00	0.00	0.00	24
Black gram	0.07	0.38	0.11	16
Chickpea	0.00	0.00	0.00	22
Coconut	0.00	0.00	0.00	21
Coffee	0.00	0.00	0.00	19
Cotton	0.00	0.00	0.00	26
Ground Nut	0.00	0.00	0.00	18
Jute	0.00	0.00	0.00	16
Kidney Beans	0.00	0.00	0.00	17
Lentil	0.03	0.13	0.05	15
Moth Beans	0.00	0.00	0.00	29
Mung Bean	0.18	0.12	0.14	17
Peas	0.00	0.00	0.00	21
Pigeon Peas	0.00	0.00	0.00	29
Rubber	0.00	0.00	0.00	26
Sugarcane	0.00	0.00	0.00	21
Tea	0.00	0.00	0.00	17
Tobacco	0.00	0.00	0.00	16
apple	0.07	0.25	0.11	8
banana	0.00	0.00	0.00	23
grapes	0.00	0.00	0.00	20
maize	0.00	0.00	0.00	20
mango	0.00	0.00	0.00	16
millet	1.00	0.28	0.43	18
muskmelon	1.00	0.53	0.69	17
orange	0.00	0.00	0.00	19
papaya	0.00	0.00	0.00	20
pomegranate	0.00	0.00	0.00	21
rice	0.00	0.00	0.00	23
watermelon	0.00	0.00	0.00	20
wheat	0.00	0.00	0.00	25
accuracy			0.04	620
macro avg	0.08	0.05	0.05	620
weighted avg	0.06	0.04	0.04	620

Fig.3.2.c.Sigmoid kernel

From the above represented images, i.e Fig.3.2.a the confusion matrix obtained for each labels are depicted using support vector machine Grid search with polynomial kernel Fig.3.2.b indicates the output of SVM Gridsearch using gaussian kernel. This kernel function is used when there is no idea of what the dataset consists of as shown in figure. Fig.3.2.c indicates the output of SVM Gridsearch using sigmoid kernel which act as a proxy for neural networks. The best suitable kernel is chosen

based on the Gridsearch values and it is used for the model. the linear and radial basis function type of kernels are also used and accuracy is greater when using them both for this model than the rest of the types and it also works faster.

3.2 SVM decision tree (Hybrid approach)

SVM in order to train large datasets it takes a lot of time. When SVM combined with decision tree it gives the better result. Decision tree can be used for both regression and prediction of the given dataset.If the dataset needs to be classified into different classes, decision tree for classification can be used. Here the problem is to predict the crop variety given the features so that decision tree for regression is used to find the accuracy. And it proved to be time efficient and gave better accuracy of 91.8%.

3.3 Random forest (Grid versus Randomized Search)

Random forest is the most helpful machine learning algorithm almost used in building new models for classification. Random forest algorithm is best in classifying large datasets which is a huge advantage. Disadvantage is that it is not so good at regression problems mainly incase of linear regression. The dataset consists of heterogenous datatypes features are in float values and label is in string format, thus Random forest comes to picture. Random forest uses a method called ensemble. That is the main tree is divided into smaller subtrees each has an estimator. All estimators produce their own predictions. finally,random forest combines all the predictions into single and accurate prediction. The proposed model is working better with this algorithm giving 95% of accuracy.Random forest using randomized search also worked out. But the amount of time taken to find the suitable parameters is very long. but the model proves to have accuracy of 95%. randomized search and Gridsearch both aim on finding suitable hyperparameters. random search uses grid to find the parameters. To use a correct model hyperparamter selection is very important.The grid dearch works by computing many trials based on cross validation , whereas the randomized search , from the grid of total number of hyper parameters ,it selects randomly the combinations for prediction .

4 Results and discussions

The above-mentioned algorithms which depicts the classification model that suits well for crop prediction. SVM classifier's result when plotted using seaborn pairplot shows the prediction as follows.

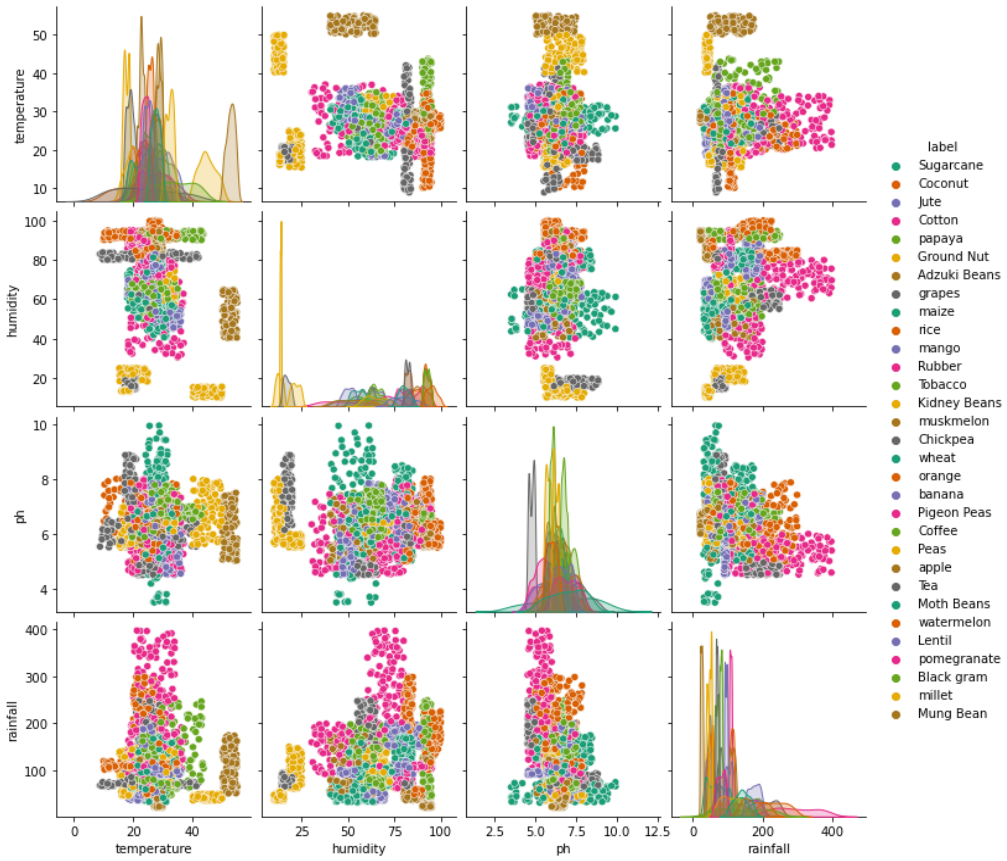


Fig.4.1.SVM using sea born pair plot

From Fig.4.1 it is clear that how SVM plots the features in the space and the gap separating them to classify and predict the results. The colors plotted in the graph indicates crop varieties.The graph compares the relationship between two variables.as the transparency changes the graph would become more readable and user can interpret the required results.Crowding of points in the graph indicates that on that particular condition both types of crops can be cultivated as both can adopts to that particular atmospheric condition.

	precision	recall	f1-score	support
Adzuki Beans	1.00	1.00	1.00	17
Black gram	0.85	1.00	0.92	23
Chickpea	1.00	0.94	0.97	16
Coconut	1.00	0.94	0.97	18
Coffee	0.71	0.63	0.67	19
Cotton	0.80	0.84	0.82	19
Ground Nut	0.94	0.89	0.92	19
Jute	0.63	0.86	0.73	14
Kidney Beans	0.91	1.00	0.95	20
Lentil	0.79	0.95	0.86	20
Moth Beans	0.86	0.71	0.77	17
Mung Bean	0.96	0.96	0.96	23
Peas	1.00	1.00	1.00	21
Pigeon Peas	0.67	0.44	0.53	18
Rubber	1.00	0.82	0.90	28
Sugarcane	0.78	0.74	0.76	19
Tea	0.77	1.00	0.87	23
Tobacco	0.94	0.67	0.78	24
apple	0.88	0.79	0.83	19
banana	0.76	0.76	0.76	17
grapes	0.83	1.00	0.90	19
maize	0.57	0.57	0.57	21
mango	0.77	0.94	0.85	18
millet	1.00	1.00	1.00	19
muskmelon	1.00	1.00	1.00	17
orange	0.67	0.80	0.73	20
papaya	0.94	0.76	0.84	21
pomegranate	0.58	0.61	0.59	18
rice	0.95	0.80	0.87	25
watermelon	0.96	1.00	0.98	26
wheat	0.59	0.59	0.59	22
accuracy			0.84	620
macro avg	0.84	0.84	0.84	620
weighted avg	0.85	0.84	0.84	620

Fig.4.2.SVM hyper parameter tuning by Grid search

This Fig.4.2 shows the output of SVM classifier using grid search with the accuracy of 84%. Then the model is trained using SVM by Decision tree algorithms.The accuracy of this model is: 91.8 when combined with decision tree it gave a better accuracy value than SVM classifier. The below result shows that random forest shows 95.09% accuracy.

```
accuracy                0.95 1023
macro avg      0.95 0.95 0.95 1023
weighted avg   0.95 0.95 0.95 1023
=== crop prediction ===
[0.9516129 0.9483871 0.97096774 0.93548387 0.94193548 0.96774194
0.94516129 0.9483871 0.94516129 0.95483871]
=== predictions === Random Forest: 0.9509677419354838
```

Fig.4.3. Random forest (Grid Search)

Fig.4.3 clearly displays the efficiency of the random forest classifier for predicting the crops to be planted as per the atmospheric condition over there with a greater efficiency of 95% which is comparably better than SVM.

```
Accuracy 0.94 1023
macro avg 0.95 0.94 0.94 1023
weighted avg 0.95 0.94 0.94 1023
=== prediction ===
[0.93870968 0.94516129 0.97096774 0.93225806 0.93225806 0.95806452
0.93870968 0.94516129 0.94516129 0.96451613] === prediction===
Random Forest: 0.9470967741935483
```

Fig.4.4. Random forest (Randomized Search)

The above results shown in Fig.4.4 shows that Randomized search applied to random forest classifier achieved the result with the accuracy of 94.7%.When compared with other models such as hybrid SVM and decision tree which gives reasonable accuracy still the random search is able to predict more accurate than other models.

5 Conclusion and Future Work

Various machine learning techniques are applied in the agriculture fields in the area of crop prediction, monitoring, weed classification, fertilizer recommendation etc. Random forest gives better accuracy when compared with other models. Still in future this could be extended to predict more accurately the crop for any type of field and appropriate fertilizer recommendation also to be added with this .Also by recommending more crops for different fields neural network would be suggested to implement more accurately. In case if the dataset size is increased then a better preprocessing technique could be added with this to implement the same. Still more parameters will be added for more precise prediction of the crop, i.e. by considering soil parameters we try to extend more attributes and finally performing the crop prediction.

6 References

1. SK Prasad, BS Sreedharan, S Jaıshanth. (2017)Crop monitoring and recommendation system using machine learning techniques.
2. M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran. (2019). A Survey on Crop Recommendation Using Machine Learning,7, pp-120-125.
3. S. Pudumalar, E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, J. Nisha(2017) Crop recommendation system for precision agriculture.
4. Nathaniel K. Newlands,Azar Ghahari,Yulia R. Gel,Vyacheslav Lyubchich,Tahir Mahdi (2019)Deep Learning for Improved Agricultural Risk Management.
5. Rakesh Kumar , M.P. Singh , Prabhat Kumar , J.P. Singh (2015)Crop Selection Method to

maximize crop yield rate using machine learning technique.
6. Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni(2017) Crop Selection Method Based on Various Environmental Factors Using Machine Learning,4, pp-1530-1533
7. G. Vishwa, J. Venkatesh, Dr. C. Geetha (2019) Crop Variety Selection Method using Machine Learning,12, pp-35-38.
8. R. Karthikeyan, M. Gowthami, A. Abhishhek, P. Karthikeyan (2018) Implementation of Effective Crop Selection by Using the Random Forest Algorithm,7, pp-287-290.
9. Prof.K.D. Yesugade, Aditi Kharde, Ketki Mirashi, Kajal Muley, Hetanshi Chudasama (2018) Machine Learning Approach for Crop Selection based on Agro-Climatic Conditions,7, pp-103-106.
10. Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke.Pooja Shinde, Suresh Rathod, Avinash Devare (2017) Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique,4, pp-952-953.