https://doi.org/10.46335/IJIES.2022.7.8.1 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 1-6

International Journal of Innovations in Engineering and Science, www.ijies.net

Crop Info Web Application Using Machine Learning

Piya Kaware[1], Saurabh Rithe[2], Nishad Kathane [3], Prof. K. R Gavhale [4]

Department of Information Technology
Yeshwantrao Chavan College of Engineering, Nagpur, India

piyakawarepk@gmail.com

***Received on**: 11 June, 2022* ***Revised on**: 26 July, 2022,* ***Published on**: 01 August, 2022*

Abstract – *With the rapid advancement of technology and data, the agricultural domain is one of the most significant study fields in the present times. Farmers keep to farm the same crops without trying with new varieties of crops. Based on the soil composition, different crops can be cultivated in different types of soil. An analytic insight on crops may help to reduce the workload and give a proper recommendation of crops. Machine learning can be very useful to increase the efficiency and accuracy of crop prediction. In the current times, there is a need for efficient machine learning models that can be used in the agricultural system to recommend crops and cultivate a healthy crop. Here we proposed a system that allows users to predict the suitable crop based on soil compositions like NPK, ph value, rainfall in mm, temperature, and humidity along with the weather analysis for different districts in Maharashtra. We used four different classification algorithms and techniques that increase crop prediction efficiency and accuracy.*

Keywords- *Crops, Soil, Agriculture, crop prediction, Machine learning, Classification, Supervised learning.*

I – INTRODUCTION

Agriculture is crucial toa country's growth and prosperity. In many countries, agriculture is considered one of the major sources of employment as well as it is also related to the progress and prosperity of any nation. The agricultural sector is one of the most important research subjects in today's times, due to tremendous advancements in technology and data. Indian Economy is known for its agricultural sector, more than 45 percent of the population is involved in agricultural activities. A lack of information causes new problems or worsens existing ones in every aspect of agriculture, raising the cost of farming. The studies on agriculture have proven that there is more to crop production than it seems. Soil quality and composition, such as nitrogen, potassium, phosphorus, and pH value, as well as environmental factors such as temperature, humidity, and land fertility levels, can all be used to determine the best crop for a given soil. However, predicting such ideal crops necessitates the use of effective machine learning model techniques that produce positive outcomes. The main purpose of creating this project is to help farmers to know the nature and quality of their soil and to recommend the most suitable crop based on soil proposition and environmental conditions which will help the farmerto plan adequately their course of action. The datasetused in this project included the various chemical compositions present in the soil and some environmental characteristics such as precipitation, temperature, and humidity which has very important role in crop growth. We used a machine learning model to identify the optimum crops for growth based on soil composition and environmental factors in this study. As a result, farmers may determine the most beneficial crops to produce in a specific soil based on certain qualities that boost agricultural output, and farmers can reap greater benefits. Various machine learning algorithms and other methodologies have been

https://doi.org/10.46335/IJIES.2022.7.8.1 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 1-6

implemented and evaluated to improve the accuracy and efficiency of crop recommendations

II- RELATED WORK

Dr. A. K. Mariappan, Ms. C. Madhumitha, Ms. P. Nishitha, and Ms. S. Nivedhitha [1] employ the K nearest-neighbor classification algorithm to increase the Crop Recommendation System's efficiency. The system maps soil and crop data to anticipate a list of crops that are acceptable for the soil, as well as information on nutrients that are insufficient in the soil for such specific crops. As a result, it is up to the user to choose which crop to plant. Nischitha K, Dhanush Vishwakarma, Mahendra N, Ashwini, and Manjuraju M.R presented a[2] system that displays the required seed for cultivation in Kg per acre for the recommended crop and offers details about required fertilizers like Nitrogen(N), Phosphorus(P), and potassium(K) in Kg per hectare. The Decision Tree algorithm will predict the crop based on list data. M.V.R. Vivek, D.V.V.S.S. Sri Harsha and P. Sardar Maran's paper [3] provides an overview of how using Data Mining Systems for climate prediction produces excellent results and should be viewed as an alternative to traditional metrological approaches. This system displays the capabilities of various calculations in predicting a few climate wonders, such as temperature, rainstorms, precipitation, and deduced that real systems. Urvashi, Dr. Kanwal Garg's [4] review describes the data mining techniques like DBSCAN, and MLR in concern of farming. The review sketches several parameters such as Policy decisions, Recommendation Systems, and Accuracy of Soil Fertility. From all the above, consolidate that much work is to be needed to be done to know how these techniques can be implemented to improve the Decision Making and Productivity. In 2022 Madhuri Shripathi Rao, N.V. Subba Reddy, Arushi Singh andDinesh U Acharya [5]presented three different model [6]estimatese crop prices and anforecastsst the pricefor the following 12 months. For increased accuracy, algorithms such as the Decision Tree Regressor were used.[7]This research by M. Kalimuthu, P. Vaishnavi, and M. Kishore uses machine learning, one of the most advanced crops prediction technologies, to assist beginner farmers in sowing appropriate crops. Naive Bayes, a supervised learning algorithm, demonstrates how to do so. Users are encouraged to enter data such as temperature and location to begin the prediction process, which will be taken automatically by this application. This research [8] by Jain, Sonal, and Ramesh, Dharavath discusses using seasonal weather predictions to determine the best sowing period for compatible crops. Weather prediction is done using machine learning methods like the recurrent neural network, and crop selection is done using the Rand Forestest classification algorithm. In this paper [9], S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha find a common problem among Indian farmers: they don't choose the correct crop for their soil. This issue is tackled by introducing an ensemble model with a majority voting technique that employs Random tree, CHAID, Naive Bayes, and KNN was used to accurately and efficiently select a crop for site-specific factors

III – WORKDONE

1] Data: Project's data set is publicly available on the Kaggle platform, it contains soil contents such as nitrogen and phosphorus, potassium, and ph value and certain environmental features such as temperature in degrees Celsius, humidity in percentage, and rainfall in mm and 22 crops associated with mentioned features. The dataset used for the past year's weather analysis was collected from Weather and Climate [10]. The rainfall analysis data was used from the Open Government Data Platform India [11] which contains the average monthly rainfall from the year 1951 to 2000.

2] Pre-Processing: Structuring is the process of organizing raw data to make it easier to obtain data and execute various operations on it. Data Frame, a function in the Python pandas library, is used to structure the data in this project. Pre-processing is a method of cleaning, organizing, and preparing data to be used in a machine learning model. Pre-processing done on the dataset we used includes:

a. Identifying a relationship between features and targets:

The rectangular data was plotted as a color-encoded matrix using the heat map function from the sea born python library. It accepts a 2D dataset as a parameter. A nd array can be created from that dataset. It can display the relationship between features, and the target value, this is a wonderful way to depict data. We plotted and analyzed correlation matrix for better understanding of features. This helped our proposed model work more efficiently.

b. Transforming categorical classes to labels in integer format:

To encode or to transform the target label with values between 0 and n-1, we used Label Encoder transformer

https://doi.org/10.46335/IJIES.2022.7.8.1 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 1-6

from Scikit Learn. Label encoding is the process of transforming labels into a numeric format so that they can be read by machines. In supervised learning, it is a crucial pre-processing step for the structured dataset.

b. Normalizing the data in the dataset:

We have used StandardScaler from sklearn for feature scaling through standardization which involves rescaling the features to make them fit into a conventional normal distribution witha standard deviation of one and mean of zero.

3] Methodology:

a. Support Vector Machine:

SVC classifier supports multi-class classification. It's a C-support vector classification system based on libsvm. This class is accountable for multi-class support using a one-vs-one mechanism. This classification model splits multiple classes into pairs of two to perform binary or one vs one classification.

b. Logistic regression

Multinomial logistic regression algorithm that extends the logistic regression model by changing the loss function to cross-entropy loss and the predicted probability distribution to a multinomial probability distribution. To modify Logistic Regression to address multiple classes we did some hyper tuning on classifier hyper parameters such as multi_class = 'multinomial', solver = 'lbfgs' so that it will address multi-class classification problems natively.

c. Random Forest Classifier:

The main purpose to use this classifier is to increase predicted accuracy and control over-fitting by fitting several decision tree classifiers on various sub-samples of the original dataset. Random forest takes an entropy like measure and it tries to provide best split by reducing it. Gini Index is commonly used metric value close to 0 denotes perfect equality and value close to 1 denotes inequality. In fig (1) sample tree represent the branch splitting based on Gini value. The vote for a certain class from all samples is stored in the value array.

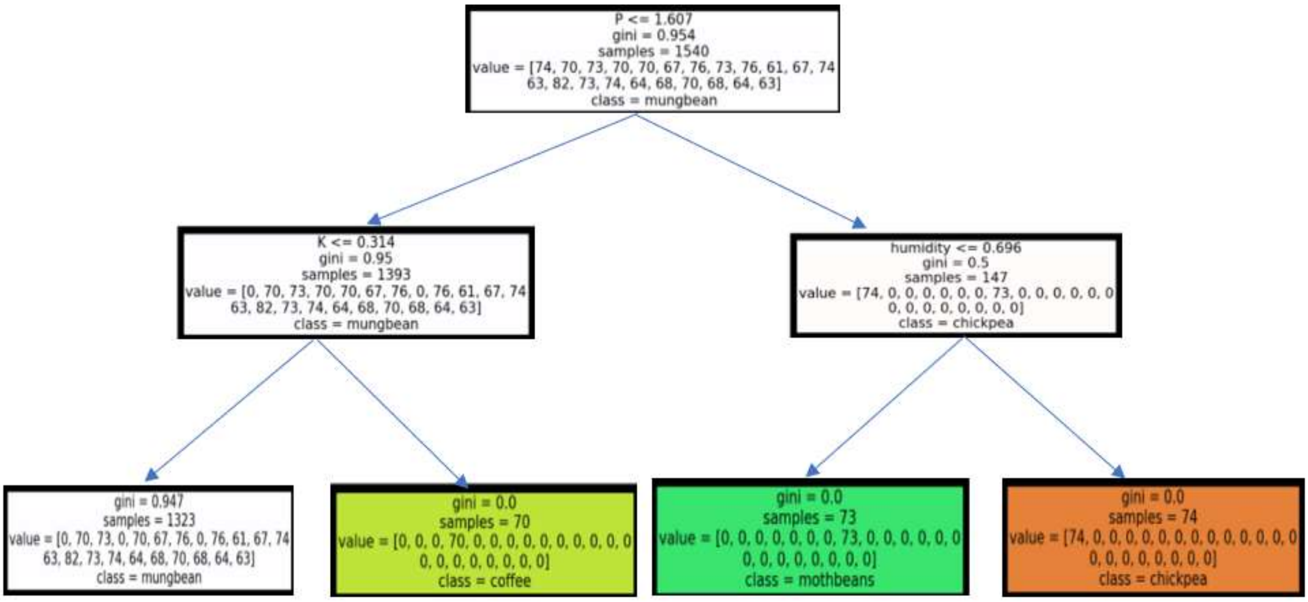

Fig (1) Random Forest Tree

c. KNN Classifier:

It works on the assumption principle that every data point that is close to one-another belongs to the same class. In other words, it classifies a new data point based on similarity and this similarity is decided as per the distance (like Euclidean, Manhattan, etc.) between two points.

$$Euclidean\ distance = d(b,a) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2}$$

The class with the highest votes will be target label for that particular input.

We plotted the graph of Error rate vs value K, and we got the elbow point at K=3, after that the error rate was continuously increasing as we can see in the below Fig. (2), so we trained our KNN model with K=3

https://doi.org/10.46335/IJIES.2022.7.8.1 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 1-6

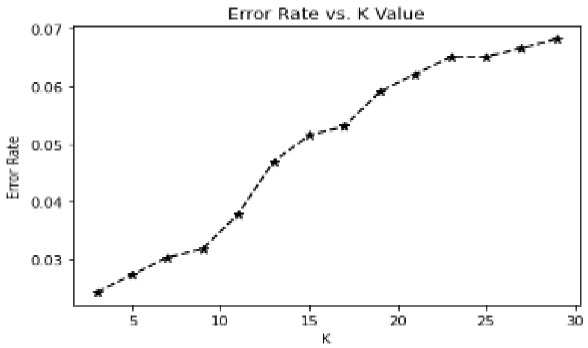

Fig 2- Error Rate vs Value of K

4] Testing and evaluating the model:

We have tested and evaluated the performance of our model for testing data by using various evaluation metrics in machine learning such as accuracy score, F1-score, Confusion matrix, and Classification report. We used the mentioned evaluation metrics to identify which model is best suited for prediction.

5] Saving the model:

The trained and tested model with the highest accuracy rate and F1 score is saved in the pickle file format for later usage, allowing the model to be loaded immediately as needed.

6] Visualizing the past data:

We provided live weather conditions as well as the visualization of climate conditions in different regions by using past year weather data.

7] Deployment:

For the deployment, we used Flask, a Python-based micro web framework. We have created an interface using HTML and CSS that enables the user to provide the feature values and predict the best suitable crop.

Fig (3) User Input

Fig (4) -Predicted Crop

https://doi.org/10.46335/IJIES.2022.7.8.1 e-ISSN: 2456-3463
Vol. 7, No. 8, 2022, PP. 1-6

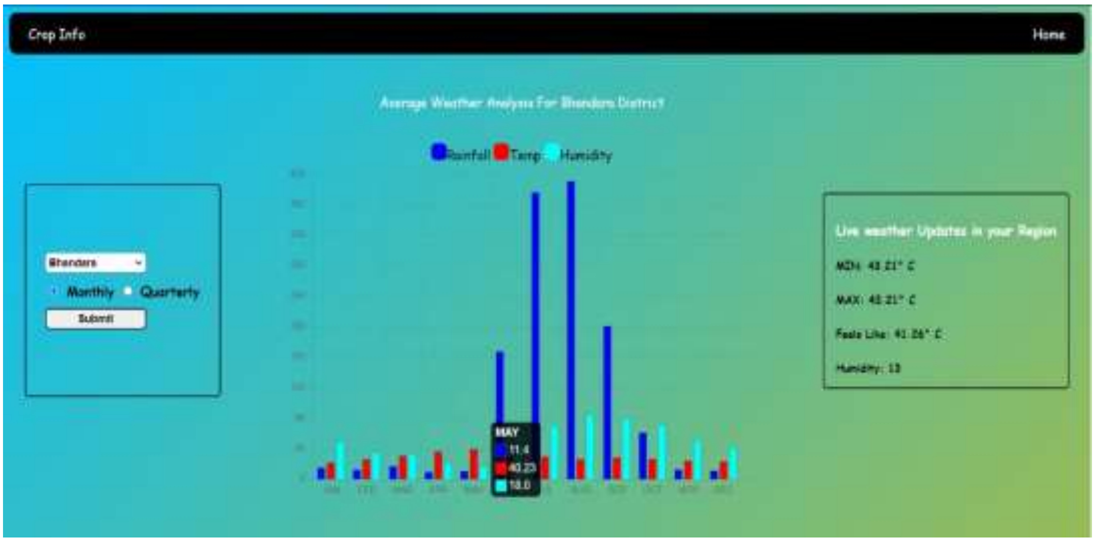

Fig 5 -Historical and Liveclimatecondition

IV - RESULT AND DISSCUSSION

1] Confusion Matrix:

Table (1) displays the confusion matrix for the trained model and how each class is classified.

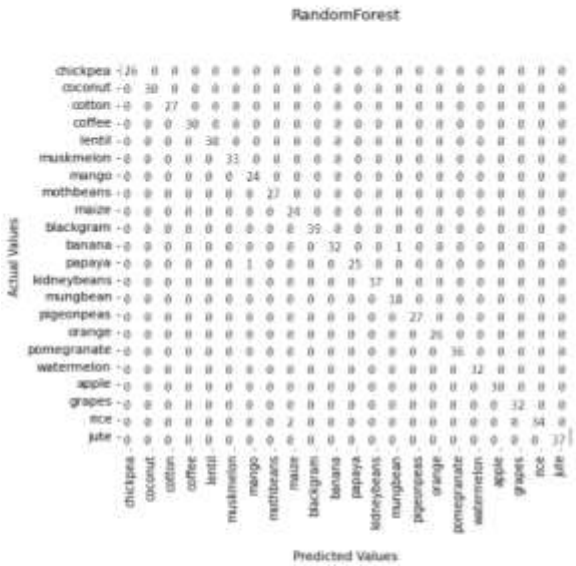

Table (1) -Confusion Matrix

2] Classification report

Table (2) displays the classification report for the trained model, where precision, recall, and f1-score values are 1 for most of the classes which shows that the trained model is good for predicting the suitable crops against the unknown values.

	precision	recall	f1-score	support
chickpea	1.00	1.00	1.00	26
coconut	1.00	1.00	1.00	30
cotton	1.00	1.00	1.00	27
coffee	1.00	1.00	1.00	30
lentil	1.00	1.00	1.00	30
muskmelon	1.00	1.00	1.00	33
mango	0.96	1.00	0.98	24
mothbeans	1.00	1.00	1.00	27
maize	0.92	1.00	0.96	24
blackgram	1.00	1.00	1.00	39
banana	1.00	0.97	0.98	33
papaya	1.00	0.96	0.98	26
kidneybeans	1.00	1.00	1.00	37
mungbean	0.95	1.00	0.97	18
pigeonpeas	1.00	1.00	1.00	27
orange	1.00	1.00	1.00	26
pomegranate	1.00	1.00	1.00	36
watermelon	1.00	1.00	1.00	32
apple	1.00	1.00	1.00	30
grapes	1.00	1.00	1.00	32
rice	1.00	0.94	0.97	36
jute	1.00	1.00	1.00	37
accuracy			0.99	660
macro avg	0.98	0.98	0.98	660
weighted avg	0.99	0.99	0.99	660

Table (2) -Classification Report

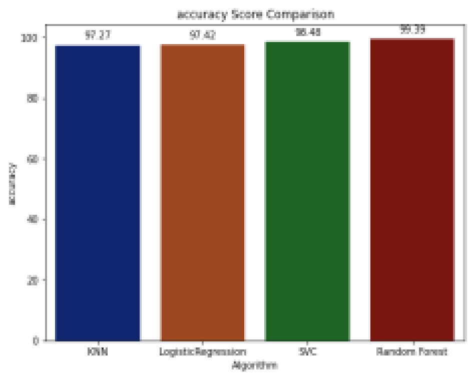

Fig (6) Accuracy Comparision

https://doi.org/10.46335/IJIES.2022.7.8.1 **e-ISSN: 2456-3463**
Vol. 7, No. 8, 2022, PP. 1-6

International Journal of Innovations in Engineering and Science, www.ijies.net

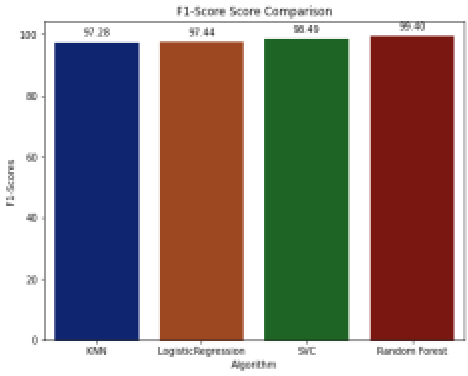

Fig (7) - F1-Score Comparison

K	Accuracy	F1-Score
1	100	100
2	99.77	99.69
3	99.31	99.31
4	98.86	98.81
5	99.77	99.74

Table (3) Cross Validation

We used the Random Forest algorithm for the final trained model, in Fig (6), and Fig (7) we can see the accuracy and f1 scores plotted for different algorithms. The random forest algorithm was more efficient than the other three algorithms in terms of both accuracy and f1-score, so we saved the model trained with Random Forest algorithm.

V- CONCLUSION

This project can be used to predict the most suitable crop from particular soil composition and can assist the user in making better decisions for farming. We used four different machine learning classifiers, with the Random Forest classifier being the most efficient one. We also performed the hyper parameter tuning and the result obtained an accuracy of 99% and an F1 score of 99.4% shows that the model is good and performs well. We saved the trained model and also deployed it to use for crop prediction. Normal users can also use this project without any prior knowledge and helps them to increase the crop yield, through the crop info web application. In future work, a large-sized dataset can be used to train the model with some more crops which will make it more efficient for increasing the profit of different crops

REFERENCES

[1] *Dr. A. K. Mariappan, Ms. C. Madhumitha, Ms. P. Nishitha, and Ms. S. Nivedhitha. (2020). Crop Recommendation System through Soil Analysis Using Classification in Machine Learning. International Journal of Advanced Science and Technology, 29(3), 12738 - 12747.*

[2] *N, Mahendra. (2020). Crop Prediction using Machine Learning Approaches. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS080029.*

[3] *M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran. A Survey on Crop Recommendation Using Machine Learning. International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue5C February 2019, ISSN: 2277-3878.*

[4] *Urvashi1, Dr. Kanwal Garg.A Review Techniques of Data Mining in the Agronomy field. International Journal of Scientific Research in Computer Science Applications and Management Studies, Volume 7, Issue 3 May 2018, ISSN 2319 – 1953.*

[5] *Madhuri Shripathi Rao et al 2022 J. Phys.: Conf. Ser. 2161 012033 Crop Prediction using Machine Learning Approaches.*

[6] *RanjaniDhanapal et al 2021 J. Phys.: Conf. Ser. 1916 012042, Crop price prediction using supervised machine learning algorithms..*

[7] *M., Kalimuthu& Vaishnavi, P. & Kishore, M.. (2020). Crop Prediction using Machine Learning. 926-932. 10.1109/ICSSIT48917.2020.9214190..*

[8] *Jain, Sonal& Ramesh, Dharavath. (2020). Machine Learning convergence for weather-based crop selection. 1-6. 10.1109/SCEECS48394.2020.75.*

[9] *S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), 2017, pp. 32-36, doi: 10.1109/ICoAC.2017.7951740.*

[10] *https://data.gov.in/catalog/district-rainfall-normalmm-monthly-seasonal-and-annual-data-period-1951-2000.*