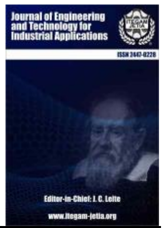
Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.10 n.48, p. 63-68. July/August., 2024.
DOI: https://doi.org/10.5935/jetia.v10i48.1186

ISSN ONLINE: 2447-0228



RESEARCH ARTICLE OPEN ACCESS

CROP RECOMMENDATION SYSTEM USING MACHINE LEARNING

Vandana Kale[1] Badri Narayan Mohapatra[2]

[1,2] Assitant Professor, Department of Instrumentation AISSMS IOIT, Pune, INDIA.

[1]http://orcid.org/0000-0002-7724-9977 ,[2]http://orcid.org/0000-0003-1906-993 .

Email:[1] badri1.mohapatra@gmail.com,[2] vandana.kale@aissmsioit.org

ARTICLE INFO

Article History
Received: July 07th, 2024
Received: July 08th, 2024
Accepted: July 08th, 2024
Published: July 18th, 2024

Keywords:
Crop recommendation system,
correlation,
random forest,
support vector machine,
k-nearest neighbor.

ABSTRACT

Agriculture is the main field of employment in India. Farmers are faced with many problems when evaluating the yield of their crops. The production of crops plays an important role in our Indian economy. This proposed system helps the farmers choose suitable crops based on rainfall, humidity, type of soil, pH of soil, and temperature Accurate crop prediction results in increased crop cultivation. It will help farmers by reducing the losses they face and improving yield. Machine learning plays an important role in the area of crop cultivation. This work proposes a crop recommendation system using machine learning techniques such as k-nearest neighbor (KNN), artificial neural network (ANN), random forest (RF), and support vector machine (SVM). The models are simulated comprehensively on an Indian data set. The SVM predictive model had an accuracy of 97.85% and a training time of 218.691ms. The K-NN predictive model gave an accuracy of 97.95% a training time of 218.691ms, and the RF gave an accuracy of 99.22% a training time of 138.021ms. This model is beneficial to farmers because it allows them to know the type of crop before cultivating the agricultural field and thus encourages them to make suitable decisions.

Copyright ©2024 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Modern techniques are now being used and put into practice by other countries for financial gain. One can only imagine how much of an advantage they already have when it comes to using scientific and technological ways to improve agriculture and farming. India, on the other hand, continues to use farming's conventional methods and technologies. As we all know, a significant portion of our nation's income comes from agriculture alone. When discussing the gross domestic product value, the income is quite useful. As we get closer to globalization, the need for food has multiplied. To fulfill food demand, crop yields increase if farmers can accurately predict crop growth.

In this work, we have compiled a dataset that includes data on rainfall, meteorological conditions, and various soil nutrients. Input parameters such as nitrogen (N), phosphorus (P), potassium (K), soil PH, humidity, temperature, and rainfall are used to suggest the best crop. The most appropriate crop is suggested based on the parameters. This will help farmers better understand agricultural production patterns while taking regional and environmental aspects into account. Additionally, our approach foresees the dearth of any specific inputs needed to develop a certain crop. The agricultural industry might greatly benefit from our predictive method. With the aid of our predictive technology, the issue of nutrient deficiency in certain areas—caused by the reason of planting the improper crop at the wrong time is eliminated. Farmers' production efficiency decreases as a result. Agriculture will undoubtedly advance to new heights as more and more science is applied to it is used a random forest algorithm [1] to predict the best crop yield. They achieved the most accurate value. The KNN algorithm was used by [2]. For the implementation of the predictive model, real-time data from Tamil Nadu was used, and the architecture was checked using samples. For [3] says that in this paper, SVM, naive Bayes, RF, and DT are used. With an accuracy of 96%, RF is more accurate than DT and has the highest accuracy value when compared to SVM. used supervised machine learning methods in this paper. The main algorithms employed were decision tree learning, K-Nearest Neighbour Regression, and RF algorithms. The suggested system's accuracy for DT, KNN, and the random forest was

90.20%, 89%, and 90.43%, respectively. In this paper, [4] describes the important work done by data mining techniques in the agricultural sector. They have incorporated KNN, SVM, RF, and others. The harvests were mostly predicted based on climate features, giving an accuracy score of 95%. According to [5] in this paper, KNN, SVM, RF, neural networks, and decision trees have all been employed. Compared to neural networks, random forests perform significantly better. P.K. Ramesh in this work [2],[6] propose a yield prediction system for farmers in which algorithms such as SVM, MLR, and RF are used. The random forest among them demonstrated the best outcome with 95% accuracy. In this, [7] worked on the SVM algorithm. It is taken into consideration for classification to evaluate the correctness of the suggested technique, and a confusion matrix is created [8]. To determine the best crop to be farmed, this paper uses the CNN and Random Forest models. The accuracy provided by the Random Forest algorithm was 75%, whereas the accuracy provided by the CNN architecture was 95.21%. (Chaudhary)illustrate the application of the Random Forest and Decision Tree models to forecast which crop will grow best in each type of soil depending on the available data. According to [9] use decision trees, Nave Bayes, RF, and SVM. (Namgiri Suresh)in this article [2], crop yield prediction is done by using the random forest algorithm. Here, data collected from an online source and RF gave an accuracy of 75%. According to [10], using a neuro-fuzzy-based machine learning algorithm, detail this to assist farmers in assessing soil quality by measuring the levels of urea, potassium, magnesium, pH, and nitrogen. The result yields an effective crop rotation sequence with an accuracy of 80%. Robots are also used in agriculture [11]. Crop recommendation is very useful [12-15].

The current systems, which employ various machine learning methods, are useful and accurate when employed in their respective regions. This [2] model is useful for the Tamil Nadu region. It is noted that in some existing recommendation models, environmental conditions are used as primary input (data, n.d.)and soil properties are predominantly used as input for predicting crops [10]. Existing systems used different predictive models like Random Forest [3],[16], [5] K-Nearest Neighbour [16],[9],[6],[2], Support Vector Machine [3], Decision Tree [2], Neuro-Fuzzy [2], etc. RF has achieved the highest accuracy of 99.22%. Climate and soil characteristics are the input parameters to suggest the best crop to increase agricultural output for local farmers. The existing system's accuracy is poor. The proposed approach offers greater accuracy in KNN, RF, and SVM.

II. METHODOLOGY

We have proposed a methodology that is separated into various stages as shown in Figure 1.

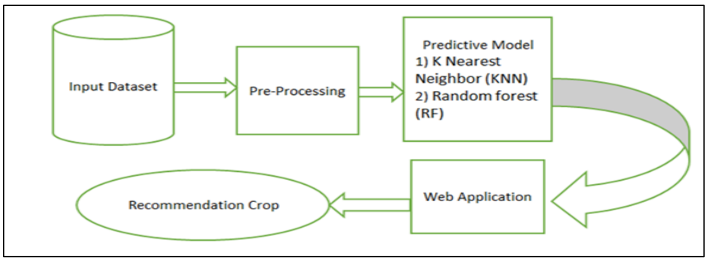


Figure1: Block Diagram of Methodology Of The Proposed System.

Source: Authors, (2024).

II.1 DATA COLLECTION

In this work, the dataset is collected from an open-source platform called Kaggle [8]. This dataset is a combination of two more complex datasets.

a) Soil content dataset:

It contains information about the ratios of nitrogen (N), phosphorus (P), and potassium (K) and the pH of the soil.

b) Climatic condition dataset:

It contains information about rainfall, humidity, and temperature.

c) Data Pre-processing

After collecting data for our machine learning model, we performed pre-processing on our dataset. In this stage, label encoding is done on the dataset. We have the 22 crop types in string format, which needs to be converted into numeric form to be understood by the computer. For this purpose,

d) Label Encoding

It converts the string format into the numeric format. As compared with the collected dataset, there are 22 different crop types, and it gives a number to every crop type in alphabetical order of crop names.

Table 1: Label Encoding.

CROP TYPE	LABEL
Apple	0
Banana	1
Black gram	2
Chickpea	3
Coconut	4
Coffee	5
Cotton	6
Grapes	7
Jute	8
Kidney beans	9
Lentil	10
Maize	11
Mango	12
Moth beans	13
Mungbean	14
Muskmelon	15
Orange	16
Papaya	17
Pigeon peas	18
Pomegranate	19
Rice	20
Watermelon	21

Source: Authors, (2024).

III. VISUAL CORRELATION PLOT

It is possible to determine how strongly or weakly the traits are related to one another by looking at the correlation between them. The number 1 indicates that there is a positive association between the two variables. In this work (Figure 2), It is observed that phosphate (P) and potassium (K) values are highly correlated with one another. Temperature and humidity are moderately correlated with each other. Rainfall and pH show less correlation between them. Individually, P and K will contribute similar levels of variance to the model. pH and humidity show less correlation with each other. Nitrogen and phosphorus show a low correlation between them. The crop's 'label' target vector is most correlated to humidity and least correlated to ph. Humidity will have a large effect on our model, while Ph may have a negligible effect.

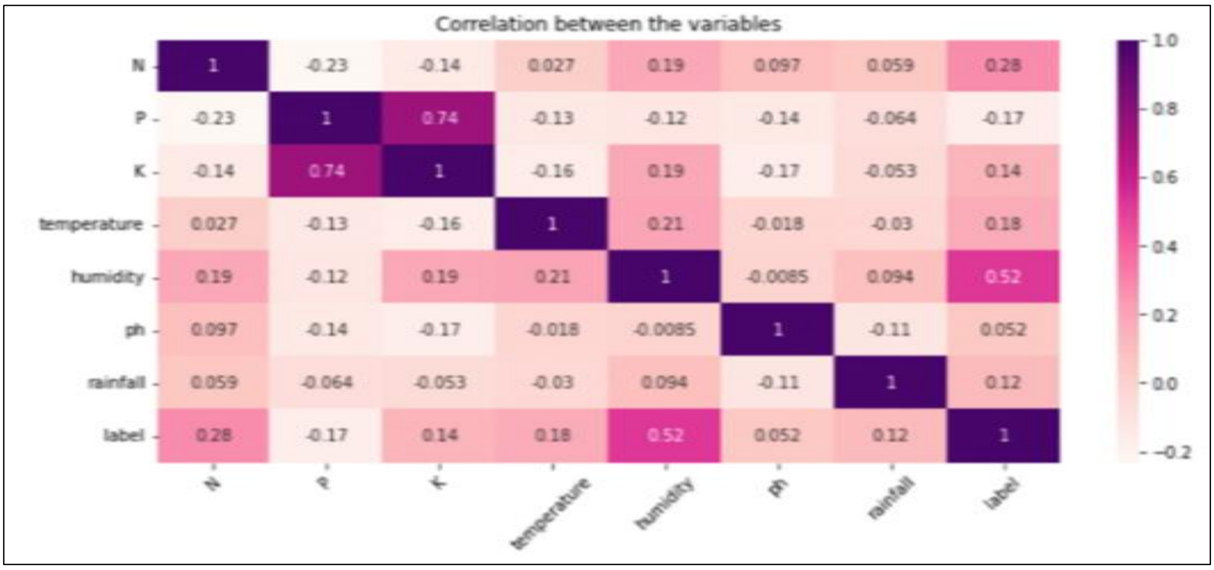


Figure 2: Correlation Plot.
Source: Authors, (2024).

Choosing the perfect algorithm for a crop dataset is a difficult task. To solve this problem, a pair plot (Figure 3) with the Seaborn library is used. The data appears to be too overlapping, according to the data visualizations. In this scenario, we can use algorithms like random forests, SVM, K-NN, and ANN.

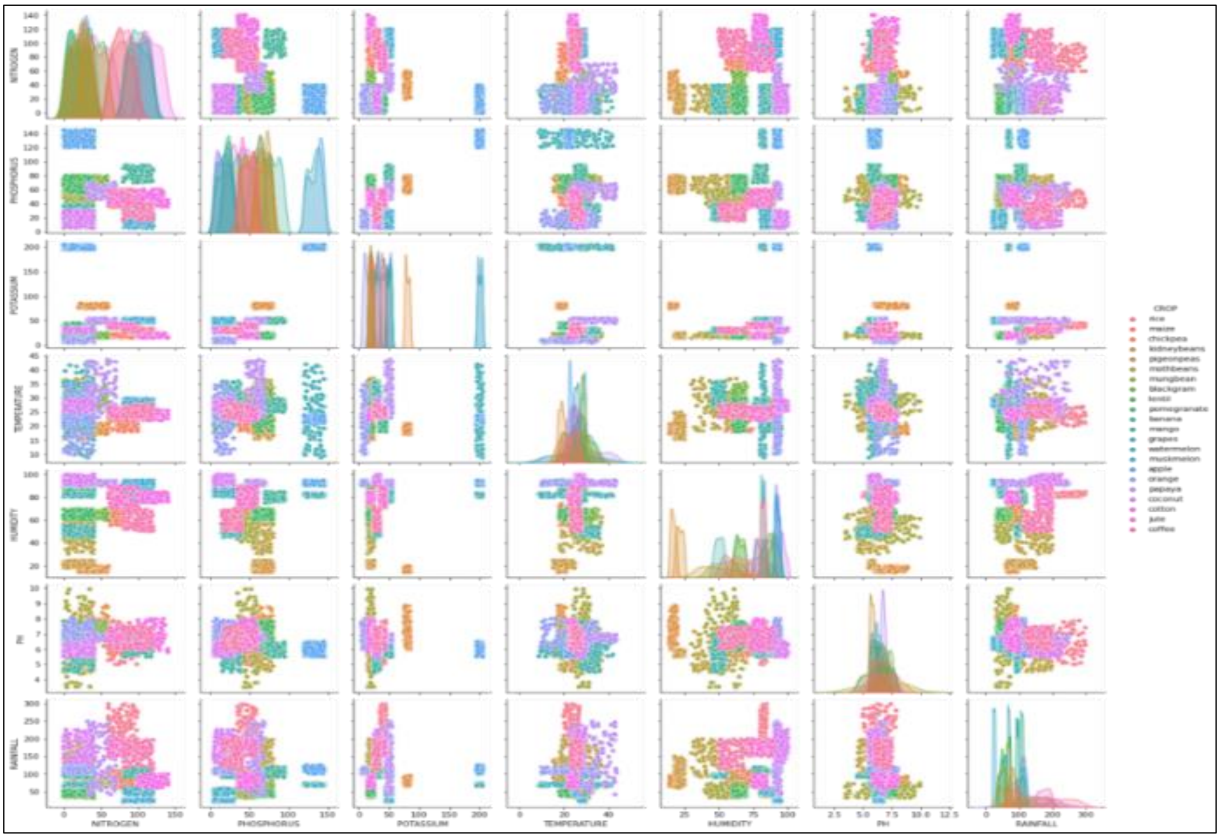


Figure 3: Pair Plot.
Source: Authors, (2024).

IV. VISUAL CORRELATION PLOT

IV.1 PREDICTIVE MODEL

In this stage, the dataset is split into training and testing sets to understand model performance by using the function train test split (). In this work, three parameters must be passed: features, goal, and test set size. You may also use a random state to choose records at random. The dataset was divided into 80% training data and 20% testing data from an existing dataset. The following predictive model is used in this work:

IV.II RANDOM FOREST

The random forest algorithm is a popular supervised machine learning algorithm. It is based on the random forest algorithm principle. The presence of excessively random trees in a random forest network leads to the high accuracy of the model. It finalizes the decision from the tree with the most votes. As a result, the greater the number of decision trees in the forest, the greater the accuracy of the model.

IV.III SUPPORT VECTOR MACHINE

SVM employs optimum lines or decision boundaries to partition the n-dimensional space into classes, allowing new data points to be conveniently placed in the appropriate category. A hyperplane represents the ideal choice boundary. The hyperplane equation is as follows:

$$w.x + b = 0 \quad (1)$$

If the value of $wx + b > 0$ then it is a positive point otherwise negative.

IV.IV K-NEAREST NEIGHBOR

All of the available data is stored by the K-NN algorithm, which compares each new data point to the existing data set. Rainfall and soil type have been designated as input parameters, while other parameters may also be taken into account. The values of the closest known neighbors can be used to predict the crop yield, which is an unknown value. By using Euclidian calculations, this is attainable.

$$D(a, b) = \sqrt{(b1 - a1)^2 + (b2 - a2)^2} \quad (2)$$

V. WEB APPLICATION

In this stage, a web application for a crop recommendation system was developed. The front end of this application was developed using HTML and CSS. For the backend the python module flask framework is used. Web application include a login page (Figure 4), a registration page (Figure 5), and a crop predictor page (Figure 6).

The farmer has to enter all the necessary details to predict the crop, like the ratio of nitrogen in the soil, the ratio of phosphorus in the soil, the ratio of ph, the temperature, the humidity, the ratio of potassium, the rainfall rate, etc. All this information is very important for our model to predict which crop will be the best for the farmer to grow. At the backend of the web application, we have connected the machine learning model and database.

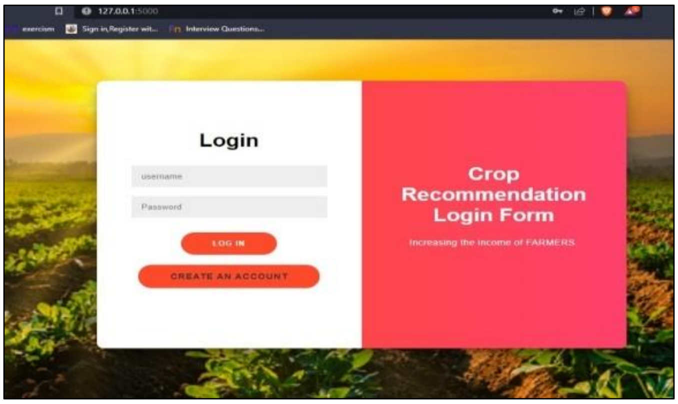


Figure 4: Login Page.
Source: Authors, (2024).

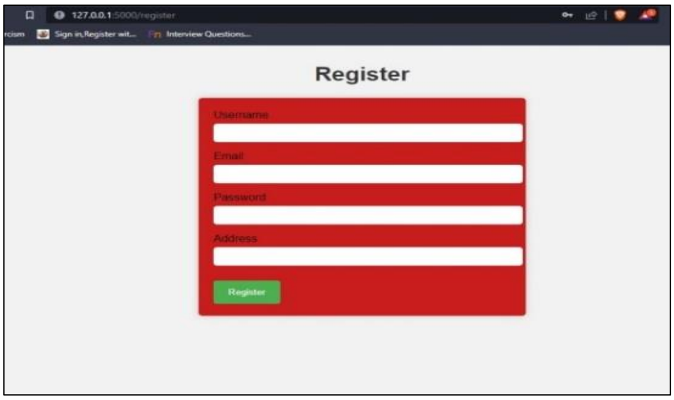


Figure 5: Registration Page.
Source: Authors, (2024).

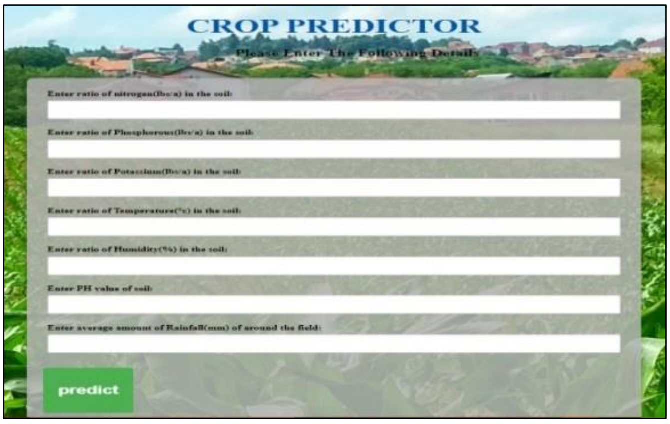


Figure 6: Crop predictor page.
Source: Authors, (2024).

VI. EXPERIMENTAL RESULTS

The proposed system was successfully implemented on software having specifications as follows:
operating system: 64-bit Windows 10
Python library: Tensorflow, Keras, Matplotlib.
Python Framework: Flask
Frontend: CSS, HTML

In random forest, we fitted the random forest algorithm to the training set and checked the accuracy of the model by varying the value of the n_estimator (required number of trees) parameter.

The results of taking ten different values for the number of trees are shown in the table. It is observed that if the value of estimator is 10 or 20, then it will give high accuracy. It is important to take care of the overfitting issue.

Table 2: RF Accuracy by Varying N_Estimator Value.

No.of trees	Accuracy
3	0.9836
5	0.9886
7	0.9927
9	0.9909
10	0.9940
11	0.9922
12	0.9931
13	0.9909
15	0.9931
16	0.9927

Source: Authors, (2024).

In this work, we plotted the confusion matrix of each predictive model to observe the performance of the classification algorithm. The confusion matrix of RF shows all diagonal elements, which represent true prediction, and others that show false prediction. RF shows a lower number of

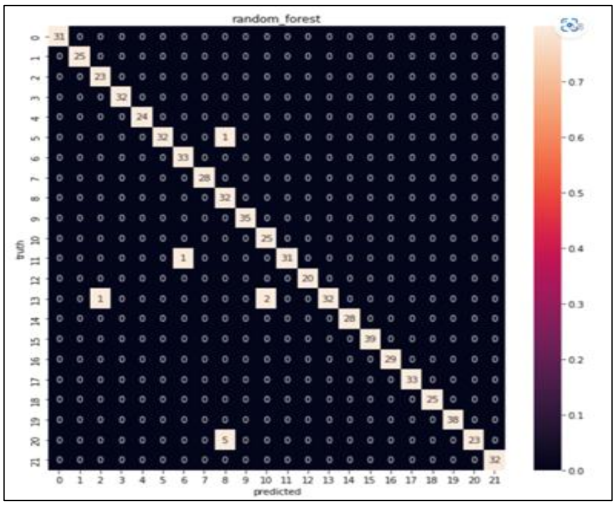


Figure 7: Confusion Matrix of RF.
Source: Authors, (2024).

In support vector machine work, we fitted the SVM algorithm to the training set and checked the accuracy of the model. The plotted confusion matrix of SVM, Figure 8 shows more off-diagonal elements, which indicates more incorrect predictions.

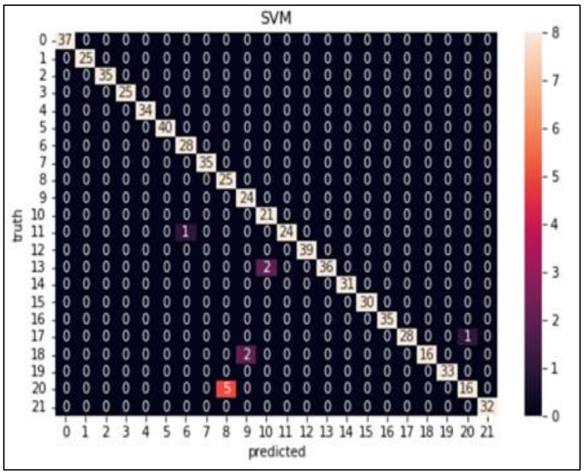


Figure 8: Confusion matrix of SVM
Source: Authors, (2024).

In K-Nearest Neighbor work, we fitted the KNN algorithm to the training set and checked the accuracy of the model by varying the value of the k_estimator (required number of trees) parameter. It is observed that if the value of n_estimator is 10 or 20, then it will give high accuracy.

Table 3: KNN Accuracy by Varying K_Estimator Value.

K_estimator	Accuracy
1	96.8
2	97.5
3	97.9
4	97.7
5	98.4
10	97.73
15	96.8
6	98
7	97.9
20	97.7

Source: Authors, (2024).

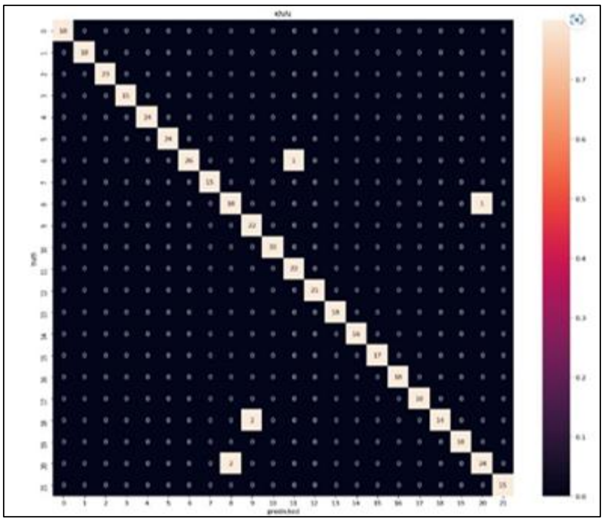


Figure 9: Confusion matrix of KNN
Source: Authors, (2024).

The confusion matrix of KNN Figure 9 shows all diagonal element that represents true prediction and other off-diagonal shows false prediction. KNN shows less number of false predictions than SVM. In the proposed system, the prediction of a crop by putting different parameter values on the precisely predicted crop displayed on the crop predictor page (Figure 10 and Figure 11).

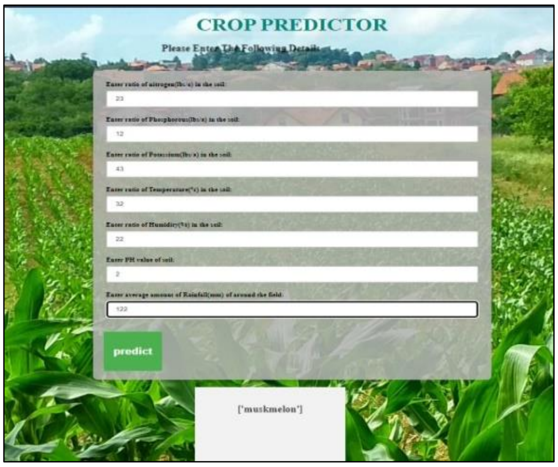


Figure10: Prediction Snapshot 1
Source: Authors, (2024).

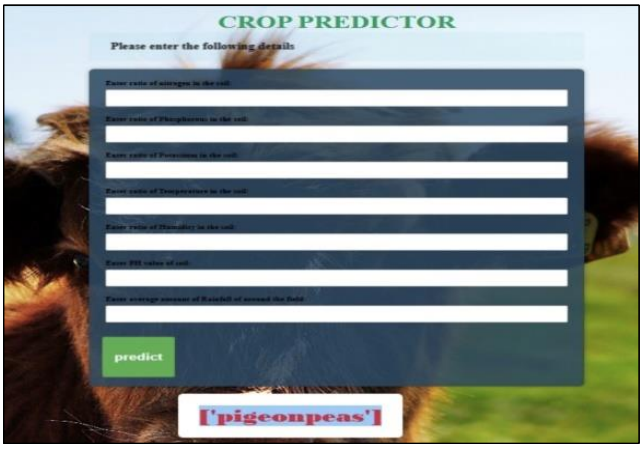


Figure 11: Prediction Snapshot 2
Source: Authors, (2024).

VII. RESULT ANALYSIS

The result obtained from RF, KNN, and SVM models, it can be concluded that RF performs better compare to KNN and SVM. The average accuracy of used predictive models is shown in Table 4.

Table 4: Average Accuracy Of Different Algorithms In Percentage.

Predictive Model	Average Accuracy	Training Time
Random Forest	99.22%	138.021 ms
K-Nearest Neighbor	97.95%	218.691 ms
Support Vector Machine	97.85%	218.691 ms

Source: Authors, (2024).

V. CONCLUSIONS

Farmers in our nation frequently struggle to choose the best crops to plant since there is a lack of scientific understanding of the various aspects affecting crops. As a result of their lower output, they suffer a loss in earnings. In this paper, a proposed crop recommendation system is proposed to suggest the best crop to the farmer by using soil parameters like N, P, K, pH, and environmental conditions such as rainfall, humidity, and temperature. This proposed system predicts one of the 22 crops by using different algorithms like RF, K-NN, SVM, etc. Out of these algorithms, RF gives the highest accuracy (99.22%) and less training time(138.021ms); hence, the RF algorithm was used and additionally, we designed a user-friendly web page using HTML, CSS, and the Flask Python framework. The webpage incorporates essential features such as registration, login, and a prediction form, enabling users to easily access and utilize the crop recommendation system, these recommended crops will bring in the maximum profit.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Dr. V. V. Kale
Methodology: Dr. V. V Kale.
Investigation: Dr. V. V. Kale and Dr. B. N. Mohapatra.
Discussion of results: Dr. V. V. Kale and Dr. B. N. Mohapatra
Writing – Original Draft: Dr. V. V. Kale and Dr. B. N. Mohapatra
Writing – Review and Editing: Dr. V. V. Kale and Dr. B. N. Mohapatra.
Resources: Dr. V. V. Kale and Dr. B. N. Mohapatra **Supervision:** Dr. V. V. Kale
Approval of the final text: Dr. B. N. Mohapatra.

VII. REFERENCES

[1] Y. J. N. Kumar, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi, 'Supervised machine learning approach for crop yield prediction in agriculture sector', in 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020.

[2] A. Suresh, P. Ganesh Kumar, and M. Ramalatha, 'Prediction of major crop yields of Tamilnadu using K-means and Modified KNN', in 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2018.

[3] A. Lokhande and M. Dixit, 'Crop Recommendation System Using Machine Learning', International Research Journal of Engineering and Technology (IRJET), vol. 9, pp. 2395–0056, 2022.

[4] Geetha, V., Punitha, A., Abarna, M., Akshaya, M., Illakiya, S., & Janani, A. P. (2020, July). An effective crop prediction using random forest algorithm. In 2020 international conference on system, computation, automation and networking (ICSCAN) (pp. 1-5). IEEE.

[5] M. Ramu and J. T. Sri, 'Wheat yield prediction using Artificial Intelligence models and its comparative analysis for better prediction', in 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021.

[6] S. M. Pande, P. K. Ramesh, A. Anmol, B. R. Aishwarya, K. Rohilla, and K. Shaurya, 'Crop recommender system using machine learning approach', in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021.

[7] D. Modi, A. V. Sutagundar, V. Yalavigi, and A. Aravatagimath, 'Crop recommendation using machine learning algorithm', in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021.

[8] Motwani, P. Patil, V. Nagaria, S. Verma, and S. Ghane, 'Soil analysis and crop recommendation using machine learning', in 2022 International Conference for Advancement in Technology (ICONAT), Goa, India, 2022.

[9] R. K. Ray, S. K. Das, and S. Chakravarty, 'Smart crop recommender system-A machine learning approach', in 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2022.

[10] E. E. Vigneswaran and M. Selvaganesh, 'Decision support system for crop rotation using machine learning', in 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2020.

[11] B. N. Mohapatra and R. K. Mohapatra, 'Design of an automated agricultural robot and its prime issues', 2020.

[12] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, 'Improving crop productivity through A crop recommendation system using ensembling technique', in 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018.

[13] B. J. Kushal, N. J. Sp, N. S. Raaju, K. G. Gv, A. R. Kp, and S. Gowrishankar, 'Real Time Crop Prediction based on Soil Analysis using Internet of Things and Machine Learning', in 2022 International Conference on Edge Computing and Applications (ICECAA), IEEE, 2022, pp. 1249–1254.

[14] N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, 'Improving crop productivity through A crop recommendation system using ensembling technique', in 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018.

[15] B. J. Kushal, N. J. Sp, N. S. Raaju, K. G. Gv, A. R. Kp, and S. Gowrishankar, 'Real Time Crop Prediction based on Soil Analysis using Internet of Things and Machine Learning', in 2022 International Conference on Edge Computing and Applications (ICECAA), IEEE, 2022, pp. 1249–1254.

[16] M. Ramu and J. T. Sri, 'Wheat yield prediction using Artificial Intelligence models and its comparative analysis for better prediction', in 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2021.