International Interdisciplinary Virtual Conference on
'Recent Advancements in Computer Science, Management and Information Technology'
International Journal of Scientific Research in Computer Science,
Engineering and Information Technology | ISSN : 2456-3307 (www.ijsrcseit.com)

Crop Recommendation Using Machine Learning Classification Techniques

Kranti G. Sapkal[1], Avinash B. Kadam[2]
[1]Research Scholar, [2]Assistant Professor
Sant Gadge Baba Amravati University, Amravati, Maharashtra, India

ABSTRACT
Agriculture plays an important role in the Indian economy. It is challenging to deal with various climactic changes affecting the yield production in this field. Use of technologies can help to improve the agricultural yield production. Ma-chine learning and IoT are emerging technologies which are proving best in the almost all areas. By using machine learning it becomes easier for farmers to pre-dict the crop for cultivation based on various factors such as contents of soil, cli-mate, and water resources. In this paper various Machine Learning classification algorithms such as decision tree, support vector machine, KNN, random forest are implemented on soil dataset to recommend the suitable crop. The accuracy of these algorithms is also tested and compared
Keywords: Agriculture, machine Learning, IoT, cultivation.

I. INTRODUCTION

To feed the growing population is becoming challenging as the amount of yield produced in the farm is less compared to the requirement of food. Farmers are required to cultivate more and more crops because the environmental conditions and requirement of particular food items are changing drastically. But most of the farmers don't have enough knowledge about the new crops and the suitable environmental conditions required for these kind of crops which leads to take wrong decisions about the crop selection. With the help of machine learning techniques, farming can be done more smart and efficient way. Machine learning refers to the set of techniques meant to deal with huge data in the most intelligent way in order to derive actionable insights [1]. Farming practices would change into with the so-called knowledge-based agriculture that would be able to increase production levels and products quality. The ultimate view point of ML is to automate the data analysis process with the help of algorithms that are enabled with continuous learning skill [2]. Hence ML refers to the set of techniques meant to deal with huge data, collected from IoT sensors in the most intelligent way in order to derive actionable insights [3]. There are three major types of ML algorithms (i) Supervised (Task driven) (ii) Unsupervised (Data Driven) (iii) Reinforcement learning (learns to react to an environment) [4]. The dataset is used as an experimental basis. After the data processing it is divided into training dataset and testing dataset.

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

In the proposed system the environmental parameters such as temperature, humidity, rainfall with soil characteristics like ph, N, P,K contents of soil are being considered to suggest a suitable crop to the farmer. The proposed system is implemented in python using machine learning classification algorithms. The aim of this is model is to provide more accurate information about the crop cultivation for the particular soil type and climatic conditions.

II. LITERATURE SURVEY

Champaneri, Mayank implemented a model for predicting the crop yield in advance of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks [5]. Jeevan Nagendra Kumar, Y. et al. implemented a system to predict crop production from the collection of past data. Using data mining techniques crop yield is predicted. Random Forest algorithm is used for predicting the best crop yield as output. In agriculture field, the crop yield prediction is mostly appropriate[6]. B S, Anisha, and Ramakanth P. Kumar implemented smart agriculture for maximizing agricultural farm water supplies, crop prediction, and wild animal prevention. Depending on the level of soil moisture, the system can be used to turn the water sprinkler on / off, thereby making the process easier to use[7]. Kalimuthu, M., P. Vaishnavi, and M. Kishore used Naive Bayes, a supervised learning algorithm to predict the crop at high accuracy. Using seed data of crop is used with the appropriate parameters like temperature, humidity and moisture content, which helps the crops to achieve a successful growth. In addition the authors developed the software, a mobile application for Android is being developed[8]. Yamaç et al. evaluated the performance of deep learning (DL), artificial neural network (ANN) and k-nearest neighbour (kNN) models to estimate field capacity (FC) and permanent wilting point (PWP) using four combinations of soil data. The DL, ANN and KNN models are compared with the previous published pedotransfer functions (PTF)[9]. Kavita, and Pratistha Mathur presented research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction. In the experimental study they showed the combination of ML with the agricultural domain field for improving the advancement in crop prediction. Most of the existing models utilized neural networks, random forests, KNN regression techniques for CYP and a variety of ML techniques were also used for best prediction[10]. The research work done by Pant J et.al. shows the different machine algorithms are used to predict crop yield in India. Researchers have used the data set for making prediction for four primary crops such as potatoes, rice, wheat and maize. The decision tree Regressor achieves highest accuracy to predict crop yield[11]. M. Keerthana et al. have implemented a system for crop yield prediction from formerly collected data. This has been settled with usage of some of the machine learning techniques. In this study Ensemble of Decision Tree Regressor with AdaBoost Regressor is used to predict the outcome with increased accuracy rate[12]. D. J. Reddy and M. R. Kumar explored various ML techniques utilized in the field of crop yield estimation and provided a detailed analysis in terms of accuracy using the techniques. The research shows several existing models that consider elements such as temperature, weather condition, performing models for the effective crop yield prediction[13]. Paper presented by S. Vaishnavi et al. depicts many Machine Learning techniques have been used to analyse the agriculture parameters. Proper prediction of crops can be informed to agriculturists in time basis[14].

III. METHODOLOGY

The farmers get suggestions about the crop selection based on contents of the soil and the environmental conditions. In this paper four different classification algorithms such as decision tree, support vector machine, KNN, random forest are used to develop the crop recommendation model. Dataset of soil with the features is used to train the model. Further model is evaluated and performance analysis of model against these algorithms is done. The overall working of the proposed system is depicted in Fig. 1

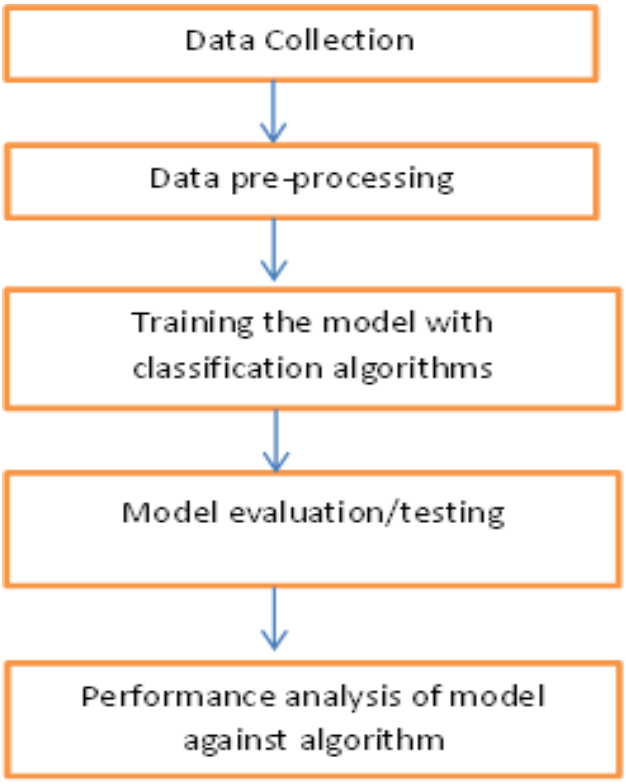

Fig. 1. Workflow of the proposed system

1.1 Data Collection

The experimental data is collected from Kaggle repository. The dataset contains the N, P,K contents of soil, temperature, humidity, ph and rainfall as independent features. Label is a dependent variable showing 22 types of crops . Each label has 100 values . Dataset contains total 2200 instances.

1.2 Data Pre-processing and Exploratory data Analysis

Preprocessing of data is very important in machine learning. Preprocessing involves dealing with the missing values, incomplete, inconsistent and duplicate data values. Exploratory Data Analysis is an approach in analyzing datasets to summarize their main characteristics, often using statistical graphics and other data visualization methods[15].In this phase dataset is pre-processed by checking null values and applying scaling and normalization on it. After pre-processing the data, it is divided into training and testing parts. Training data is used to train the model and testing data is used to evaluate the model. Table 1 shows the few samples from dataset.

N	P	K	temperature	humidity	Ph	rainfall	label
51	56	18	28.127878	64.209777	6.706506	70.863408	blackgram
63	50	52	28.645556	93.226426	6.751748	115.816394	papaya
94	26	27	26.366299	52.257385	7.456460	177.317616	coffee
86	37	16	20.517168	59.212355	5.561511	67.610137	maize
29	78	25	19.959917	59.331578	5.982855	195.787103	pigeonpeas
57	64	55	26.683865	92.958541	6.583760	62.506897	papaya
69	51	23	22.217382	72.854628	6.801639	106.621316	maize
26	80	18	19.325096	23.333479	5.581022	104.778395	kidneybeans
1	35	34	30.793757	46.695368	6.273398	92.213186	mango
90	86	52	25.850370	81.955805	5.793260	119.085617	banana

Table 1. Sample data from dataset

1.3 Implementation of Machine Learning Algorithm on Dataset

Before deciding the best model for crop recommendation it is necessary to evaluate each model by its performance. In this phase supervised machine learning-classification algorithms such as Decision Tree, K-nearest neighbor, Random Forest and support vector machine are used to train the model.

1) k-Nearest Neighbors.

The k-NN algorithm is the simplest machine learning algorithm. Here 'k' is the number of neighbors. Building the model consists only of storing the training dataset [16]. To make a prediction for a new data point, the algorithm finds the closest data points in the training dataset—its "nearest neighbors" using Euclidian distance measures Formula as eq.1

$$\text{Dis(A,B)}=\sqrt{(x2-x1)^2+(y2-y1)^2}$$

Fig. 1 depicts the training and testing accuracy of model for 1 to 10 value of k for the given dataset. From Fig. it is clear that if the value of k is 5 then better accuracy can achieved with minimum training and testing accuracy difference.

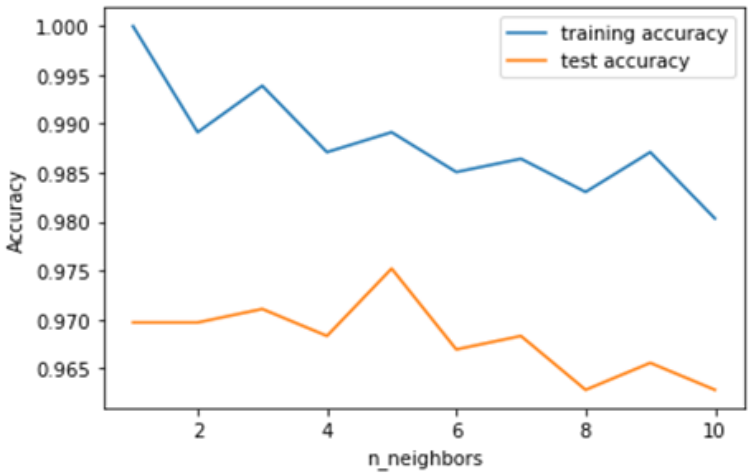

Fig. 2. Training and testing accuracy of k-Nearest Neighbours

2) Decision tree.

Decision trees are widely used models for classification and regression tasks. Essentially, they learn a hierarchy of if/else questions, leading to a decision [17]. Using decision tree algorithm on the experimental dataset ,the accuracy on the training set is 100%, while the test set accuracy is much worse. This is an indicative that the tree is overfitting and not generalizing well to new data. Therefore, we need to apply pre-pruning to the tree. If max_depth=3 is set then it decreases overfitting. This leads to a lower accuracy on the training set, but an improvement on the test set.

Feature importance rates how important each feature is for the decision a tree makes. It is a number between 0 and 1 for each feature, where 0 means "not used at all" and 1 means "perfectly predicts the target." The feature importance always sum to 1 fig.3 shows the feature importance of the seven features from the dataset.

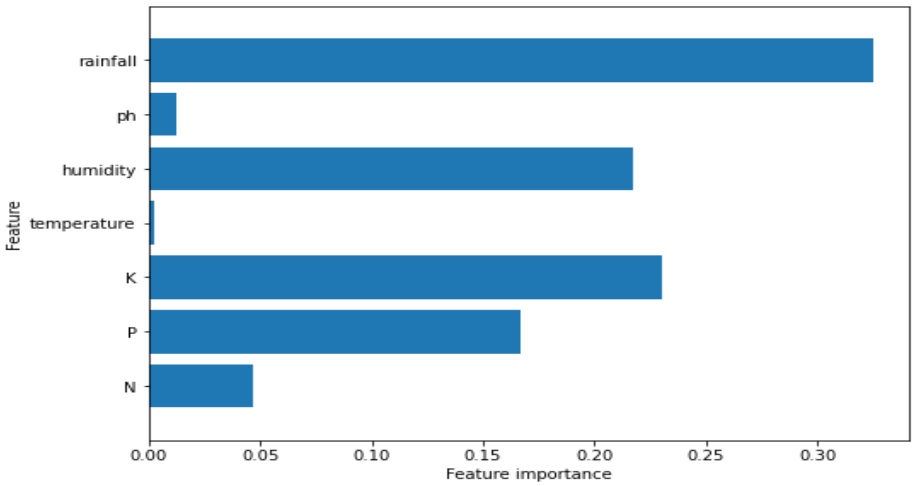

Fig. 3. Feature importance for decision tree

3) Support Vector Machine.

SVM model is basically a representation of different classes in a hyperplane in multidimensional space[18]. Support vector machines classify data by finding the hyperplane that maximizes the margin between the classes in the training data. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH)[19]. For the given dataset the Linear Kernel Accuracy is 0.99, Rbf Kernel Accuracy is 0.31 and Poly Kernel Accuracy is 0.96. To increase SVC Linear model accuracy parameter tuning is used which gives 0.97 accuracy. From this it is clear that Linear kernel also gives satisfactory results but fine tuning increases the computation and might be inefficient in some cases. The accuracy can be increased in poly kernel by tweaking parameters but might lead to intensive overfitting.

4) Random Forest.

Accuracy of the model for given dataset is improved by use of Random Forest classifier. Random forest contains multiple decision trees and compute the average to improve the accuracy[20].The final output depends on the maximum votes of predictions from each tree, instead of relying on one decision tree. Fig. 4 depicts the feature importance of the seven features. Less important feature can be omitted to improve the accuracy of the model. For the given dataset random forest gives maximum accuracy that is 99 percent. The overfitting problem is prevented. The greater number of trees in the forest leads to higher accuracy.

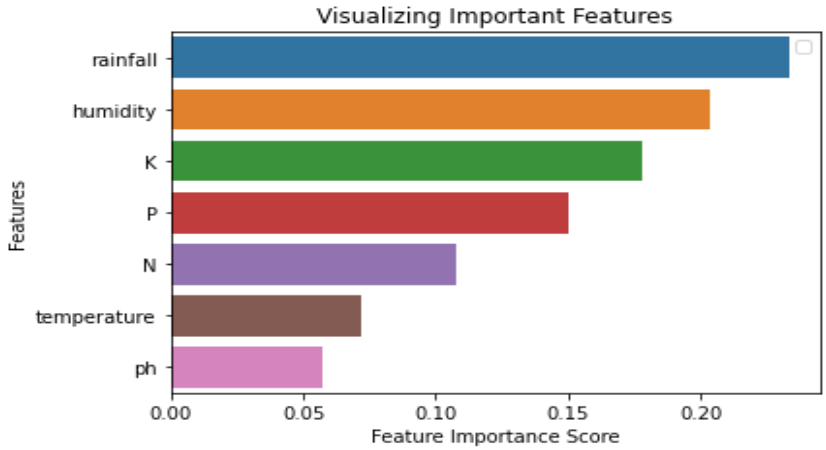

Fig. 4. Feature importance for Random Forest classifier

Fig.5 depicts the accuracies achieved on the soil dataset using four types of classification algorithm such as SVM, Decision tree, Random Forest and KNN classifier.

	Model	Accuracy
0	SVM	0.975207
1	Decision Tree	0.972452
2	Random Forest	0.993113
3	KNN	0.975207

Fig. 5. Accuracies of four classifiers

IV. CONCLUSIONS

The study carried out for prediction of suitable crop by using machine learning models shows the accuracy of models and how to control model complexity. For many of the algorithms, setting the right parameters is important for good performance. From the dataset it is seen that KNN gives better accuracy for optimal value of K . For Decesion tree, model performs well if max depth and n_estimator are important to fine tune otherwise trees will be densely graphed which will be a classic case of overfitting. Max_depth=4 and n_estimators=10 gives pretty much satisfying results by making sure model is able to generalize well. For SVM, linear kernel has better results than rbf kernel and poly kernel SVM . The random forest gives us an accuracy of 99%, better than the linear models or a single decision tree, without tuning any parameters.

V. REFERENCES

[1]. Mekonnen, Yemeserach et al. 2020. “Review—Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture.” Journal of The Electrochemical Society.

[2]. Van Klompenburg, Thomas, Ayalew Kassahun, and Cagatay Catal. 2020. “Crop Yield Pre-diction Using Machine Learning: A Systematic Literature Review.” Computers and Elec-tronics in Agriculture 177: 105709.

[3]. Elijah, Olakunle et al. 2018. “An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges.” IEEE Internet of Things Journal 5(5).

[4]. Nishant, Potnuru Sai, Pinapa Sai Venkat, Bollu Lakshmi Avinash, and B. Jabber. 2020. “Crop Yield Prediction Based on Indian Agriculture Using Machine Learning.” 2020 Inter-national Conference for Emerging Technology, INCET 2020: 5–8.

[5]. Champaneri, Mayank, and Information Technology. 2020. “Crop Yield Prediction Using Regression Model.” International Journal of Innovative Technology and Exploring Engi-neering 9(10): 269–73.

[6]. Jeevan Nagendra Kumar, Y. et al. 2020. “Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector.” Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020 (Icces): 736–41.

[7]. B S, Anisha, and Ramakanth P. Kumar. 2020. “Real-Time Monitoring of Agricultural Land with Crop Prediction and Animal Intrusion Prevention Using Internet of Things and Ma-chine Learning at Edge.” Proceedings of CONECCT 2020 - 6th IEEE International Confer-ence on Electronics, Computing and Communication Technologies.

[8]. Kalimuthu, M., P. Vaishnavi, and M. Kishore. 2020. “Crop Prediction Using Machine Learning.” Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020 (Icssit): 926–32.

[9]. Yamaç, Sevim Seda, Cevdet Şeker, and Hamza Negiş. 2020. “Evaluation of Machine Learning Methods to Predict Soil Moisture Constants with Different Combinations of Soil Input Data for Calcareous Soils in a Semi Arid Area.” Agricultural Water Management 234(December 2019).

[10]. Kavita, and Pratistha Mathur. 2021. “Satellite-Based Crop Yield Prediction Using Machine Learning Algorithm.” 2021 Asian Conference on Innovation in Technology, ASIANCON 2021 (Iciccs): 1466–70.

[11]. Pant, J., Pant, R. P., Kumar Singh, M., Pratap Singh, D., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings, 46, 10922–10926. https://doi.org/10.1016/J.MATPR.2021.01.948

[12]. M. Keerthana, K. J. M. Meghana, S. Pravallika and M. Kavitha, "An Ensemble Algorithm for Crop Yield Prediction," 2021 Third International Conference on Intelligent Communica-tion Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 963-970, doi: 10.1109/ICICV50876.2021.9388479.

[13]. D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using Machine Learning Algo-rithm," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1466-1470, doi: 10.1109/ICICCS51141.2021.9432236.

[14]. S. Vaishnavi., M. Shobana., R. Sabitha. and S. Karthik., "Agricultural Crop Recommenda-tions based on Productivity and Season," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 883-886, doi: 10.1109/ICACCS51430.2021.9441736.

[15]. Jeevan Nagendra Kumar, Y. et al. 2020. “Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector.” Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020 (Icces): 736–41.

[16]. Kumar, V., Kumar, R., Kumar, S., & Jorvekar, P. P. (2021). Agriculture Soil Analysis for Suitable Crop Prediction. PP, Agriculture Soil Analysis for Suitable Crop Prediction (May 24, 2021).

[17]. Elavarasan, D., & Durairaj Vincent, P. M. (2020). Crop Yield Prediction Using Deep Rein-forcement Learning Model for Sustainable Agrarian Applications. IEEE Access, 8, 86886–86901. https://doi.org/10.1109/ACCESS.2020.2992480

[18]. PANDE, SHILPA MANGESH, et al. "Crop recommender system using machine learning approach." 2021 5th International Conference on Computing Methodologies and Commu-nication (ICCMC). IEEE, 2021.

[19]. Bhat, Showkat Ahmad, and Nen-Fu Huang. "Big data and ai revolution in precision agricul-ture: Survey and challenges." IEEE Access 9 (2021): 110209-110222.

[20]. Sharma, Robin. "Artificial Intelligence in Agriculture: A Review." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.