Application of Machine Learning Models for Crop Recommendation

S. Manikanda Prabu[1]

[1]Associate Professor, Department of Computer Science and Engineering, Adithya Institute of Technology, Coimbatore, Tamil Nadu, India.
E-mail: smaniprabume@gmail.com

Abstract--- Optimizing agricultural yields is essential for nations whose agricultural industries are strong economic drivers. Farmers could increase their yields if they knew what to plant based on their location. This paper builds a crop recommendation system using machine learning algorithms that can aid farmers' in increasing crop yield. The proposed system classify soil using a variety of machine learning algorithms, which enables the system to suggest crops to grow based on the characteristics of the specific soil. Only four soil characteristics are considered, potassium, nitrogen, phosphorus, and pH value. Several experiments are conducted in order to select the best machine learning algorithm for crop recommendation. Experimental results show that the proposed system can recommend crop to increase yearly yield efficiently.
Keywords--- Agriculture, crop recommendation, machine learning, soil characteristics and yield.

I. INTRODUCTION

AGRICULTURAL is important and significant in the economies of most countries, particularly in less developed countries where agricultural activity contributes significantly to the economy. As a result, high levels of agricultural production are critical for most countries' economic well-being and the welfare of their people. The most common crops used to anticipate crop yields include wheat, rice, peas, pulses, sugar cane, tea, cotton, greenhouse crops, corn, and soybeans. Crop yield prediction is crucial to the world's food production. Seed firms anticipate that new hybrids will be employed in a range of soil conditions in order to breed superior types for various locales. Farmers gain from crop forecasting by preventing financial losses. Farmers need to be aware of which crops are suitable for a given season and what safety measures should be implemented based on the soil's and environment's conditions. Additionally, farmers need to understand the kinds of crops that need to be planted by taking into account past experiences.

The relevance of a timely crop yield forecast model on a regional scale is highlighted by the difficulties exacerbated by climate change and the increased food demand caused by the continuously expanding population. Better crop management, food security, and better policy and agricultural decision-making must all be goals of the model. For this reason, numerous strategies have been created and used to forecast agricultural yield, ranging from process-based models to data-driven statistical algorithms. Statistical approaches, however, offer intriguing alternatives and complementary tools because process-based models for simulating the physiological systems are bound by the availability of data for parameterization, model calibration, and validation.

The primary contributions of our research are as follows: 1) we build a crop recommendation system that is more user-friendly for farmers utilising simply soil properties, such as nutrients, nitrogen, potassium, phosphorus, and pH level.2) We compare different machine learning techniques to determine which one is most appropriate for creating a crop recommendation system utilising only soil attribute data. c) To determine which features will produce greater accuracy, we evaluate the strength of the features we employ in this study to that of other features that have been previously used.

II. LITERATURE SURVEY

Savla et al. [1] investigated the agricultural accuracy of the dataset collected for soybean crop production prediction throughout numerous years in order to examine the effectiveness of classification algorithms. The authors of this study used Bayes classification, vector machines, random forest, and neural networks. Ahmed et al. [2] predicted cereal prediction in Bangladesh's key districts using data mining techniques. To determine agricultural production in various districts, the data set included two area-related variables, three biotic variables, and five environmental parameters. The research suggested future geospatial analytical initiatives to boost precision. Paul et al. [3] used K Nearest Neighbor (KNN) and Naïve Bayes (NB) algorithms to predict crop yield. Regression approaches were successfully used to estimate yields in a specific site by Shastry et al. [4]. This study demonstrated that the regression model may be used to accurately predict crop yields within a certain geographic area. Kumar et al. [5] employed descriptive analytics concepts when it comes to agricultural data. This study includes details on the data analytics techniques used on sugarcane harvest datasets, depending on the findings of the investigation. Renuka et al. [6] discussed the implications of using machine learning techniques to analyse large datasets related to sugar cane agriculture in Karnataka, India, as well as the implications of

ISSN: 2321-2381

their findings. In an effort to determine the most successful methodology, many machine learning techniques have been applied to agricultural data. Three different supervised learning techniques were used to examine the sugar cane dataset: Support Vector Machine (SVM), KNN, and Decision Tree (DT).

III. Methodology

The introduced crop recommendation system is illustrated in Figure.1. It consists of five major processes are as follows:

- Data collection
- Data normalization
- Data division
- Model training and
- validation

Data Collection

We used a cross-sectional data set from Kaggle that included soil properties as well as other environmental parameters that affect crop growth, such as temperature, humidity, and precipitation, to construct our crop recommendation system. The dataset contains information on the soil's pH, phosphorus, potassium, nitrogen, and magnesium levels. Since soil pH impacts the amount of nutrients available in the soil, it is also included in the data set. Sample of collected data is given in Table.1.

Data Normalization

Data preprocessing is a data mining technique that comprises putting raw data into an understandable format. All of the missing values should be removed at first because the original dataset might contain a significant number of them. Missing values are indicated by a dot in the dataset, and their presence might reduce the data's total value and hinder performance. In order to solve this issue, we swap out these integers for their mean values. The second step is to make the class labels. Since we are using supervised learning, each entry in the dataset should have a class label, which is created during the preparation stage. Data is scaled into a common range between 0 and 1.

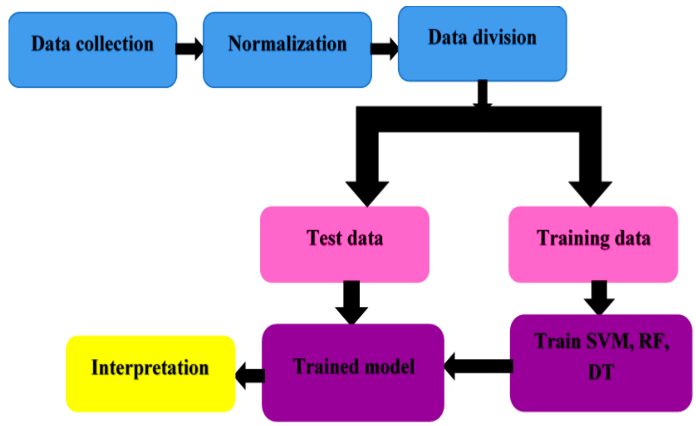

Figure 1: Proposed Crop Recommendation System

Table 1: Sample Data Set

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.87974	82.00274	6.502985	202.9355	rice
85	58	41	21.77046	80.31964	7.038096	226.6555	rice
60	55	44	23.00446	82.32076	7.840207	263.9642	rice
74	35	40	26.4911	80.15836	6.980401	242.864	rice
98	35	18	23.79746	74.82914	6.252798	91.76337	maize
76	57	18	18.98027	74.52601	6.092726	94.26249	maize
99	56	17	24.10859	73.13112	6.23433	71.07562	maize
60	44	23	24.79471	70.04557	5.72258	76.7286	maize
74	48	17	21.63163	60.27766	6.430616	69.21803	maize

Data Division

Data is partitioned into two groups: training data and test data. 80% of the samples are used for training the system and remaining 20% of the samples are used for validation.

Model Training and Validation

Three machine learning algorithms such as SVM, DT, and RF are used to develop a crop recommendation system. In this study we also compare the performances of the algorithms to establish which ones are best suited for crop recommendation. A single hyper-plane or a collection of hyper-planes are used by SVM for various tasks, relapses, or characterization. These hyper-planes were created in high-dimensional or limitless space. The RF algorithm, which may be applied to classification and regression issues, falls within the category of supervised machine learning algorithms. This technique is based on ensemble learning, and one of its benefits is that it uses many trees and only uses a portion of the input for each tree, making it objective.

IV. Results and Discussion

Table 2: Evaluation of Proposed Crop Recommendation System

Model	Accuracy (%)	Recall	RMSE	MAE
SVM	89.34	0.875	0.121	0.137
RF	93.45	0.932	0.11	0.01
DT	85.55	0.856	0.135	0.164

Table .2 presents the results of the developed crop recommendation system using three classifiers SVM, RF, and DT. From the Table.2, it is observed that the best performing machine learning model is RF whose accuracy is 93.45% which is significantly higher than the 89.34% accuracy of SVM and 85.55% accuracy of DT. Recall value of SVM, RF, and DT are 0.875,0.932, and 0.856 respectively. The designed system with SVM, RF, and DT achieved Root Mean Square Error (RMSE) values are 0.121, 0.11, and 0.135. Similarly, the developed system with RF attained Mean Absolute Value of 0.01 which lower than that of 0.137 of SVM and 0.164 of DT. Pictorial representation of Table .2 is given in Figure.2 and Figure.3. It is inferred from the results that the RF is suited for crop recommendation.

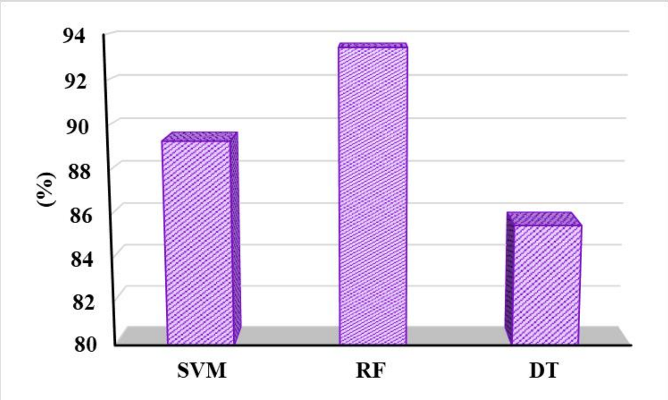

Figure 2: Performance of the Proposed System in Terms of Accuracy

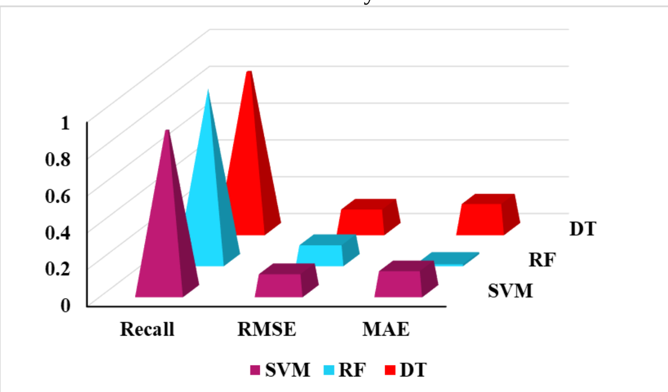

Figure 3: Performance of the Proposed System in Terms of Recall, RMSE, and MAE

V. Conclusion

This study developed a crop recommendation system that can determine which crop to grow based on the attributes of the soil. In the process of developing a crop recommendation system, we used SVM, RF, and DT, and we discovered that the RF algorithm performs the best for crop recommendation. The four attributes such as nitrogen, phosphorus, and potassium and the soil's pH level, were considered. Our approach can be more easily used by the majority of farmers by limiting its consideration to these four soil properties, whose data is more easily accessible and affordable to get. In the discussion of our findings, it is revealed that the proposed system uses the RF algorithm and produces an accuracy of 93.45%.

References

[1] A. Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia, Sanya Bhardwaj (2015), Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture, Innovations in Information, Embedded and Communication Systems (ICIIECS).

[2] A.T. M. S. Ahamed et al., "Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh," 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD),2015, pp. 1-6,

[3] M.Paul, Santosh K. Vishwakarma, Ashok Verma (2015), Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach, International Conference on Computational Intelligence and Communication Networks.

[4] Shastry, A., Sanjay, H. A., & Bhanushree, E. (2017). Prediction of crop yield using Regression Technique. International Journal of Computing, 12(2), 96–102.

[5] Kumar, A., Kumar, N., & Vats, V. (2018). Efficient crop yield prediction using machine learning algorithms. International Journal of Research in Engineering and Technology, 05(6), 3151–3159.

[6] Renuka & Terdal, S. (2019). Evaluation of Machine learning algorithms for Crop Yield Prediction. International Journal of Engineering and Advanced Technology, 8(6).

