2022 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) | 978-1-6654-6490-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICWITE57052.2022.10176256

A Creative Use of Machine Learning for Crop Prediction and Analysis

Ruchira C. Mahore
M.Tech. Student, Dept. of CSE,
Government College of Engineering Amravati,
Maharashtra, India
Email: mahore170899@gmail.com
(Corresponding Author)

Naresh G. Gadge
Assistant Professor, Dept. of CSE,
Government College of Engineering Amravati,
Maharashtra, India
Email: naresh.gadge@gmail.com

***Abstract*— In an agricultural nation like India, farming is significant. Various aspects, including production rate, environmental circumstances, soil type, market value, and unique government rules, must be taken into consideration when deciding which crop to cultivate and farm. Although we have been producing for generations, many advancements in the field of agriculture are still required to enhance improvements in our Indian economy. Artificial intelligence techniques can be used on agricultural areas to make these gains. This study focuses on an overview of current methods for crop prediction and advice, followed by a description of our own prototype for a crop production analysis and advice system that covers a variety of agricultural problems. We've created a system employing machine learning to focus on two primary modules in our suggested strategy: crop production analysis and crop recommendation system. We initially gathered the data for the crop production analysis module from official websites. The general and exploratory analyses of the crop data are then completed. The information on agricultural production is methodically grouped by year, season, and state. Next, we performed the same analysis on a different dataset. The correlation matrix that is produced reveals a lot of commonalities. Then, using SVM and more than 25000 records grouped in 8 columns, we trained and evaluated our system. The system's accuracy and standard deviation are determined, and NB, SVM, Random Forest, and Decision Tree Machine Learning Classifiers are compared. In light of this, the crop is advised while taking crucial factors like temperature, humidity, ph, and rainfall into account. Python is used for programming, and the libraries Pandas, NumPy, Scikit-Learn, Matplotlib, and Seaborn are employed.**

***Keywords*— Artificial Intelligence, Crop Production, Crop Recommendation, Machine Learning, Python**

I. INTRODUCTION

Agriculture is the foundation of all economies. Agriculture must improve to keep up with demand in a country like India where the population is expanding and there is a continual need for food. Since it was so common, agriculture has been seen as India's fundamental cultural practise since the dawn of time. Because ancient people raised crops on their own land, they needed accommodations to suit their needs. Because of this, natural crops are grown and used by people, animals, birds, and other living creatures [1]. The main goal of agricultural planning is to maximise crop output rates while making use of scarce natural resources. Several machine learning techniques may be used to increase agricultural productivity rates.

When losses happen under unfavourable circumstances, crop selection procedures should be employed to reduce the losses. Furthermore, crop production rates are typically boosted in perfect conditions. Increasing the rate benefits the country's economy. the factors influencing agriculture output rates. They are seed quality and crop selection [2]. The quality of the seeds should be examined before to sowing. Using high-quality seeds boosts agricultural productivity rates. In addition, two parameters are used to choose crops. The main factors include both favourable and unfavourable conditions. Techniques for hybridization could be useful in this situation. A lot of research is done to enhance agricultural planning. The goal is to achieve the maximum crop output possible. different methods of categorisation. Many classification strategies are applied to crops to help them generate their best. The model's fundamental goal is to give farmers with an optimum advice for producing crops while taking into account aspects such as soil composition, environmental factors such as temperature, humidity, rainfall, and geographical effect.

Agriculture is the foundation of the Indian economy. The majority of crops in India are weather dependent. However, the condition of the soil also has a significant impact on crop output. Rice agriculture, for example, is heavily reliant on rainfall. All seasonal moments are no longer the same as they once were. We couldn't even forecast if there would be future floods or water shortages. Furthermore, farmers are not technologically advanced enough to anticipate crop for each given crop if it is selected to be cultivated. However, it is unavoidable that soil health status will be utilised to propose a crop variety to be farmed the next season. So, in order to optimise crop , crop forecast of numerous elements is necessary based on local meteorological conditions. prediction is a significant agricultural issue.

Farmers used to forecast their production based on the previous year's crop. As previously stated, we were unable to anticipate production based on last year's results owing to a variety of variables such as crop stress, soil impurity, floods, pesticides, pests, and illnesses. I'm going to employ some pre-existing mathematical models in this case. Farmers are cultivating hybrid items on soil that is typically unsuitable, but they are utilising pesticides and growing those. As a result, the soil's quality deteriorates. The structure of this essay is as follows. The motivation behind identified problem is presented in Section 3. The reason for the formulation of the problem and the implemented methodology is provided in Section 5. The operation of the suggested technique is

thoroughly explained in Section 6. The results are reported in Section 7, and the conclusion is provided in Section 8.

II. LITERATURE SURVEY

Farmers in the agriculture sector, particularly in India, are concerned about making decisions throughout the cropping season. Farmers' goal is to achieve high crop production productivity throughout cropping season. To do this, it should need professional advice, on which farmers may base crop production decisions. Amount of research conducted in this sector by agricultural institutions and agriculture research centres. However, owing to a lack of contact between academics and farmers, most recommendations do not reach farmers. To address this issue, a suggestion or advice system is available to farmers.

The recommendation system has a collection of input datasets drawn from prior years' real data, following which they begin processing. They process the data step by step as guided by the algorithm to get the outcome and draw a conclusion. This algorithm is crucial. As a consequence, the outcomes of an algorithm constitute knowledge. Previous crop recommendation system research used a variety of methodologies or procedures over a long period of time.

The team of Thomas van Klompenburg and others [12], in order to anticipate agricultural production, machine learning is a crucial decision-support technology that helps farmers decide which crops to cultivate and what to do throughout the crop's growing season. Studies on agricultural production prediction have benefited from the application of many machine learning methods. In this study, we used a Systematic Literature Review (SLR) to extract and synthesise the techniques and traits used in research that predicted crop yield. Using inclusion and exclusion criteria, we selected 50 of the 567 relevant publications we found using our search parameters from six internet databases. We carefully examined the methodology and other factors used in these selected publications, evaluated the results, and provided suggestions for further research.

SA Bhat and others [13] Due to the use of technology and communication (ICT) in precision agriculture, which created new methods to make farming more productive, competent, and well-regulated while preserving the environment, sustainable agricultural development is an essential response to fast population expansion. Due to its massive data analytical capabilities to abstract important information and support agricultural practitioners in understanding well farming practises and making precise decisions, big data (machine learning, etc.) is one of the critical ICT technologies used in precision agriculture. This essay's main goal is to increase knowledge of the current Big Data uses in smart agriculture. K.L.M. Ang and associates [14] Processing of hyperspectral and multispectral data using machine learning has been described. Additionally, the potential of scalable parallel discriminant analysis and ensemble machine learning in agricultural information processing has not been explored. On hyperspectral data from agriculture, validation studies and data analytics have been conducted. Our method worked successfully, as seen by the results..

III. MOTIVATION

Typically, every farmer makes an attempt to predict how closely the output will match his goals. In the past, forecasts were produced by looking at farmers' prior experience with a certain crop. Weather, insect activity as the seasons change, and harvest operation planning all have a significant impact on agricultural . Making decisions pertaining to agricultural risk management requires accurate knowledge about crop production history. The proposed method therefore proposes a paradigm to anticipate crop . The farmer will gauge the crop's per acre before cultivating the field to boost output.

IV. AIM AND OBJECTIVE

1) Establish a clear model capable of forecasting agricultural sustainability in India for particular soil types and weather scenarios.
2) Make suggestions for the best appropriate crops in the area so that the farmer does not lose money.
3) Seasonal Analysis For agricultural evaluations based on seasonal production characteristics
4) Time Series Analysis To assess agricultural production over time

V. METHODOLOGY

Machine learning is also helpful for calculating agricultural output generation. In general, machine learning is the process of analysing data from a variety of perspectives and distilling it into knowledge that will produce superior outcomes. Users can categorise and summarise the relationships discovered by analysing data from a range of dimensions or views using machine learning tools. Machine learning is frequently used to discover correlations or patterns among hundreds of variables in enormous relational databases. Additionally, the patterns, relationships, and connections in all of this data may have information. Information is commonly turned into understanding of both past and potential future patterns. An overview of agricultural output statistics, for instance, can help farmers spot crop losses and stop them from happening again.

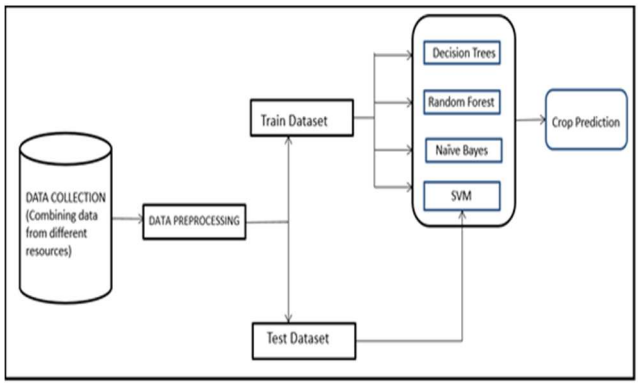


Fig. 1: Architecture Diagram of Our Machine Learning Model for Crop Prediction and Recommendation

In the last several years, Artificial Intelligence (AI) has advanced significantly, and it now has branches in every industry. As a crucial component of AI, machine learning is beneficial for taking into account incoming data, looking for relevant patterns, improving upon itself, and displaying the results. Crop production prediction is frequently overlooked in the new computerised world as controlling the economy takes precedence.

Programming with Python and the libraries pandas, NumPy, Scikit-Learn, Matplotlib, and Seaborn are utilised in

our suggested strategy. Because of its general-purpose character, which makes it applicable in practically every area of software advancement, Python is highly recognised and well appreciated. Python today has a significant presence in every sector that is emerging. Python has emerged as the programming language with the quickest recent growth and is capable of creating any application. We created a system focused on agricultural data analysis and forecasting using Python. The architecture diagram of our machine learning model for crop classification and recommendation are shown in Fig. 1 & Fig. 2.

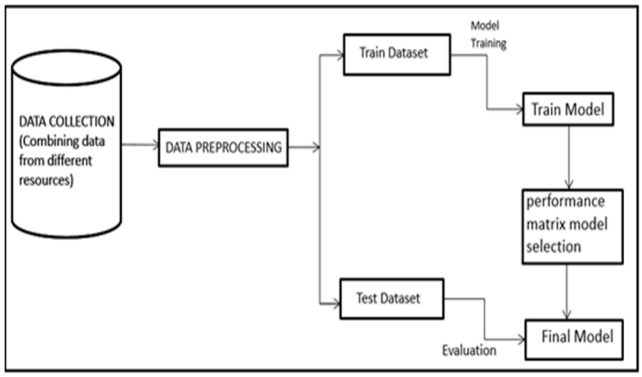
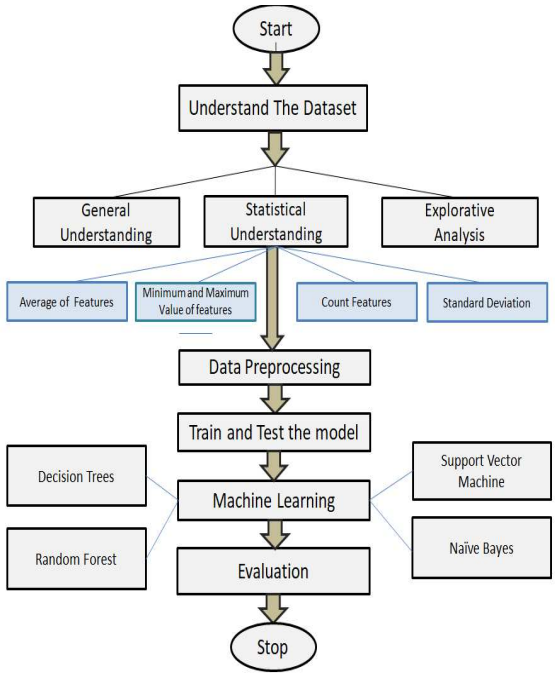


Fig. 2: Architecture Diagram of Our Machine Learning Model for Crop Classification

The creation of the machine learning model comes next. We must divide our dataset into training and test sets before beginning to develop the machine learning model. The data has been divided in a 70-30 ratio. We use the training data as a starting point and then apply our machine learning algorithms to the dataset's characteristics. On our training dataset, we employed 4 machine learning methods; on the test dataset, we will choose the algorithm that provides the best accuracy. The system flowchart is provided in Fig. 3



.Fig. 3: System Flowchart

Here are the steps for SVM training:

Step 1: Import the dataset

Step 2: Explore the data to figure out what they look like Step 3: Pre-process the data

Step 4: Split the data into attributes and labels

Step 5: Divide the data into training and testing sets.

Step 6: Train the SVM algorithm

Step 7: Test the model for Crop recommendation

Step 8: Evaluate the results of the algorithm

I compiled a dataset for India by supplementing existing databases of rainfall, temperature, and fertiliser data. This will provide us with a better understanding of crop patterns while taking into account various environmental and geographical aspects. I use this dataset to train a Machine Learning model to estimate the best crop to plant in a given location. Machine learning has the potential to be a game changer in the agricultural business. By forecasting the best crop to cultivate, the suggested approach would assist farmers in determining the raw materials and other resources needed far sooner than they would have otherwise. This would eliminate the issue of nutrient deficit in fields caused by planting the incorrect crops, which may reduce output efficiency significantly. India is still lagging behind in terms of developing technical solutions for agriculture, which is the major source of income for almost half of the country's population. More scientific solutions must be promoted in order to propel India's agricultural business to higher heights.

VI. PROPOSED APPPROACH

The crop production data is systematically arranged into year wise, season wise and state wise. Fig. 4 shows the year wise production of crop, state wise production of crop, season-wise production of crop, season-wise production of rice.

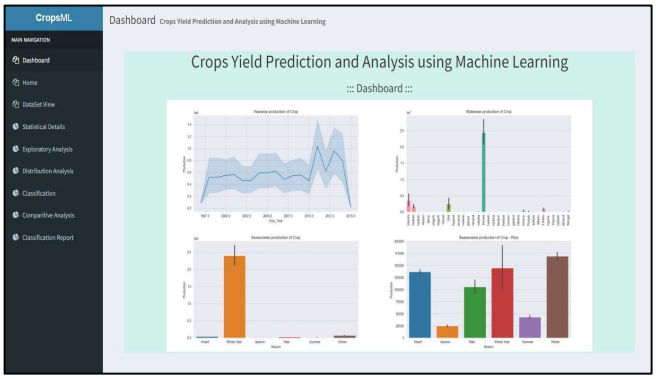


Fig. 4: Dashboard of Crop Prediction and Analysis by using Machine Learning

Now, the dataset general information can enable. The data set files are crop production and crop recommendation having large amount of data. The attributes with datatypes of crop production and crop recommendation can also be viewed. Now, the sample data set view of our proposed approach can be seen in Fig. 5. Here, we have given the gist of crop production data and crop recommendation data. Out of the large data in our data set, we have shown the first 7 rows of data for each.

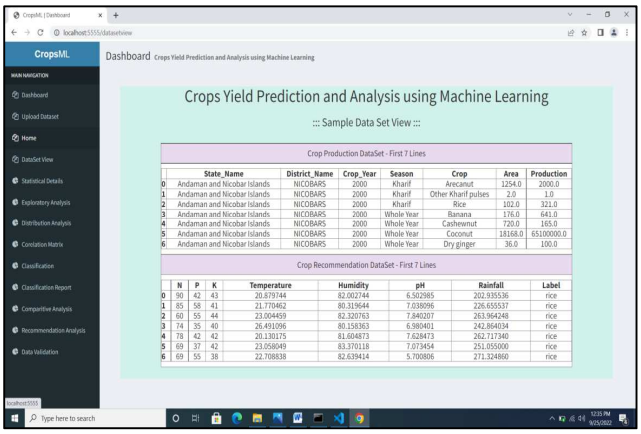


Fig. 5: Sample Data Set of Proposed Approach

The statistical details of the data set of our proposed approach can also be allow. Here, we have given the gist of crop production dataset and crop recommendation dataset. The crop production dataset and the parameters of area and production, whereas the crop recommendation dataset utilizes parameters of temperature, humidity, ph and rainfall. Finally, the exploratory data analysis of our proposed approach is used. We can see the summary of year wise production of crop and season-wise production of crop.

A web framework is a structure that includes tools, libraries, and functions for quickly and efficiently building and maintaining large online applications. They are intended to simplify programming and encourage code reuse. A server-side language is required to develop the server side of the web application. Python is home to a plethora of such frameworks, the most well-known of which being Django and Flask. Python Flask Framework is a small micro-framework built on Werkzeug and Jinja2. It is referred to as a micro framework because it strives to keep its core functionality minimal while being flexible to accommodate a wide range of small and big applications. Flask Framework is dependent on two third-party libraries:

VII. RESULTS AND ANALYSIS

In our proposed approach, we've developed a system using machine learning to focus on two main modules viz., crop production analysis and crop recommendation system. In the crop production analysis module, we've first collected the data from the government websites. Then the general analysis and exploratory analysis of crop data is done.

The crops are analysed while considering the important parameters of humidity, ph, temperature and rainfall as shown in Fig. 6, Fig. 7, Fig. 8 & Fig. 9 respectively. Then we did the same analysis with another dataset. The correlation matrix is generated, as shown in Fig. 10 & Fig. 11, which shows significant similarities.

We have then trained and tested our system with more than 25000 records arranged in 8 columns using SVM. The accuracy and standard deviation of the system is calculated and comparison of Random Forest, Decision Tree Machine Learning Classifiers and NB, SVM.

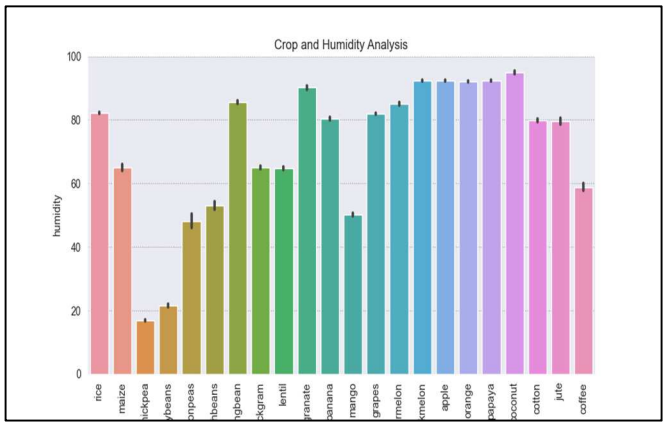


Fig. 6: Graph showing the Crop and Humidity Analysis

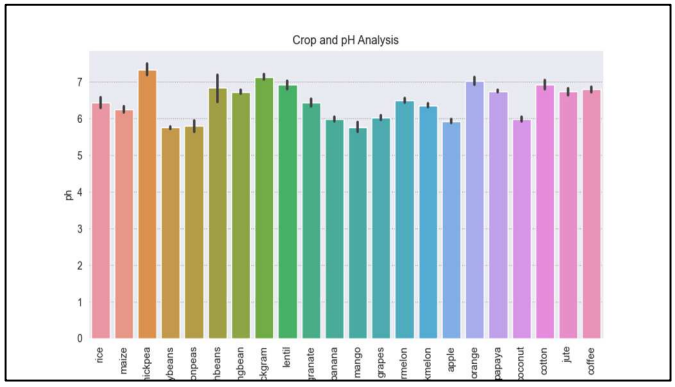


Fig. 7: Graph showing the Crop and pH Analysis

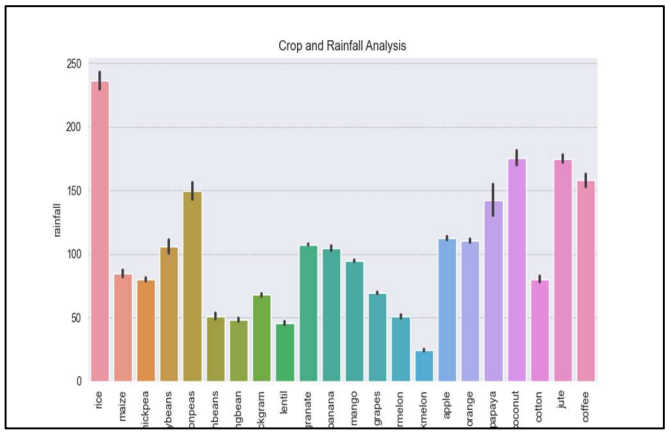


Fig. 8: Graph showing the Crop and Rainfall Analysis

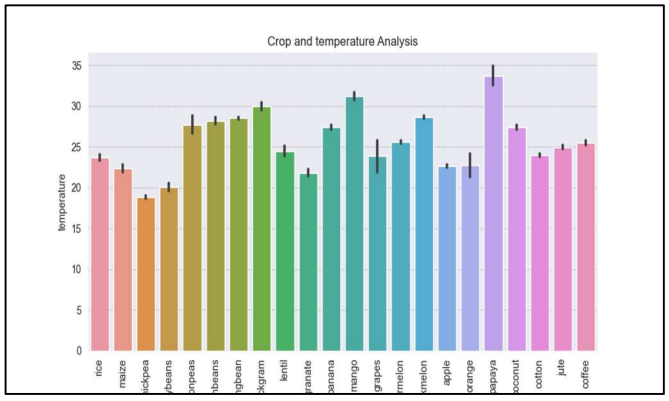


Fig. 9: Graph showing the Crop & Temp. Analysis

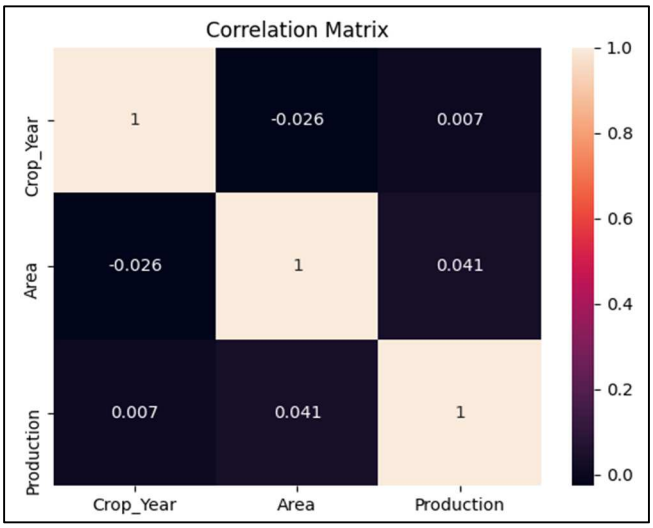


Fig. 10: Correlation Matrix between Crop production year, Are and Production Quantity

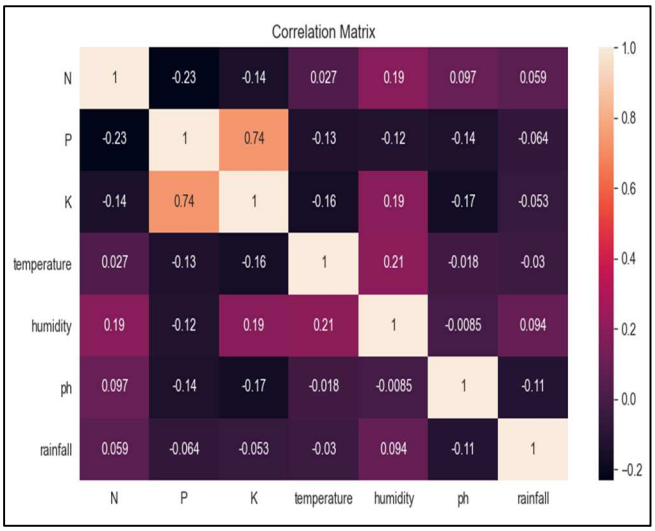


Fig. 11: Correlation Matrix between humidity, ph, temperature and rainfall

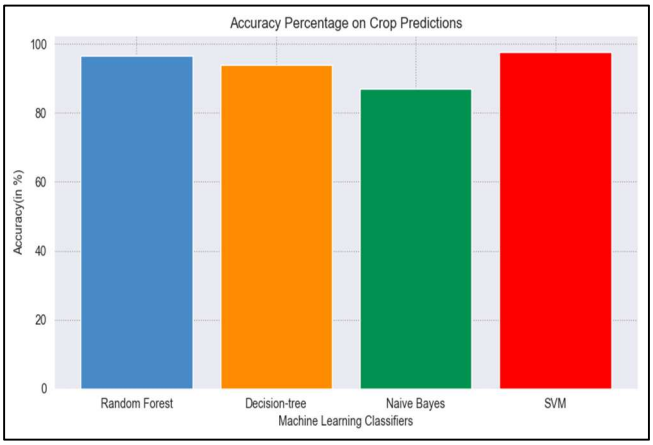


Fig. 12: Accuracy Comparison

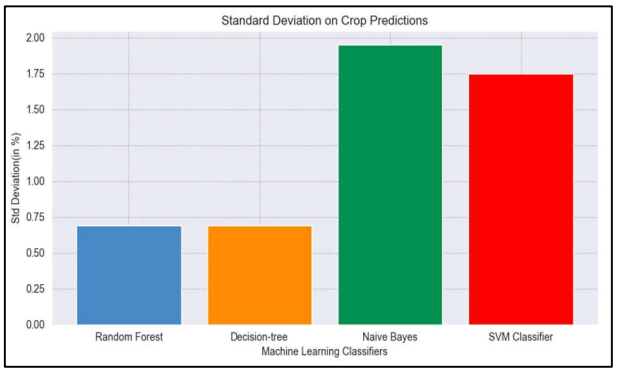


Fig. 13: Standard Deviation of the system using different ML Algorithms and compared

VIII. CONCLUSION

The primary goal of this study is to provide an overview of existing approaches for crop prediction and advise, which is followed by a description of our own prototype for a crop production analysis and advising system that addresses a range of agricultural issues. In order to concentrate on the two key components of our proposed strategy—crop production analysis and crop recommendation system—we have developed a system using machine learning. The information for the agricultural production analysis module was first acquired from official websites.

The final step is to finish the general and exploratory evaluations of the crop data. By year, season, and state, the data on agricultural productivity are meticulously arranged. The same analysis was then run on a new dataset. Many similarities may be seen in the correlation matrix that is generated. Our system was then trained and tested using SVM, NB, SVM, Random Forest, and Decision Tree Machine Learning Classifiers are contrasted as well as the system's accuracy and standard deviation. As a result, the crop is suggested while important parameters like temperature, humidity, ph, and rainfall are taken into consideration. Programming is done with Python, and the libraries NumPy, Scikit-Learn, Matplotlib, and Seaborn are utilised.

References

[1] F. H. Tseng, H. H.Cho, and H. T. Wu, "Applying big data for intelligent agriculture-based crop selection analysis," IEEE Access, vol. 7, pp. 116965-116974, 2019.

[2] P. Tiwari, and P. Shukla, "Crop prediction by modified convolutional neural network and geographical indexes," International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[3] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deeplearning-based detector for real-time tomato plant diseases and pests' recognition," Sensors, vol. 17, no. 9, pp. 2022, 2017.

[4] J. Sun, L, Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean prediction using deep CNN-LSTM model," Sensors, vol. 19, no. 20, pp. 4363, 2019.

[5] D. A. Bondre, and S. Mahagaonkar, "Prediction of Crop and Fertilizer Recommendation Using Machine Learning Algorithms," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 5, pp. 371-376, 2019.

[6] B. Devika, and B. Ananthi, "Analysis of crop prediction using data mining technique to predict annual of major crops," International Research Journal of Engineering and Technology, vol. 5, no.12, pp. 1460-1465, 2018.

[7] V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, "Performance Evaluation of Machine Learning Techniques for

Mustard Crop Prediction from Soil Analysis," Journal of Scientific Research, vol. 64, no. 2, 2020.
[8] P. M. Gopal, and R. Bhargavi, "A novel approach for efficient crop prediction," Computers and Electronics in Agriculture, vol. 165, pp. 104968, 2019.
[9] S. Khaki, and L. Wang, "Crop prediction using deep neural networks." Frontiers in plant science, vol.10, pp. 621, 2019.
[10] T. Vijayakumar, "Posed Inverse Problem Rectification Using Novel Deep Convolutional Neural Network," Journal of Innovative Image Processing (JIIP), vol. 2, no. 03, pp. 121-127, 2020.
[11] T. Senthil Kumar, "Data Mining Based Marketing Decision Support System Using Hybrid Machine Learning Algorithm," Journal of Artificial Intelligence, vol. 2, no. 03, pp. 185-193, 2020.
[12] Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture. 1;177:105709, 2020.
[13] Bhat, S.A. and Huang, N.F.. Big data and ai revolution in precision agriculture: Survey and challenges. IEEE Access, 9, pp.110209-110222. 2021.
[14] Ang, K.L.M. and Seng, J.K.P.,. Big data and machine learning with hyperspectral information in agriculture. IEEE Access, 9, pp.36699-36718. 2021.