A Survey on Machine Learning Algorithms In Crop Yield Prediction

Dr.Thara.L[1]. Rahul Prasad.M[2]
[1]Associate Professor, Dept of MCA
[2]Dept of MCA
[1, 2] PSG College of Arts & Science, Coimbatore,India

Abstract- *The majority of Indians choose agriculture as their vocation, and the country ranks second in terms of agricultural production. Because they don't know anything about the state of the soil, farmers simply repeat producing the same crops over and time again by applying arbitrary amounts of fertilizer. This gradually destroys the top soil layer and causes the soil to become more acidic. Therefore, we created a machine learning model for farmers to address these problems. By recommending the ideal crop to plant based on the weather and soil conditions, our algorithm aids farmers. Therefore, utilizing our model, farmers may learn about the various crops they need to produce in order to enhance productivity, which then increases profit.*

In order to resolve the problem, a crop prediction model employing machine learning is used, and for that purpose, it accepts input from numerous parameters such as weather and soil conditions. parameters and previous crop data, as well as by Using this data, a model is able to forecast a crop that should be planted in the future that benefits the users/farmers to choose the optimum crop to grow in the present situation. The model is educated using historical information on crops and pertinent various factors, including water characteristics and soil properties. The data's testing and training depends entirely on our dataset and the Using that, we may learn about the accuracy of our design.

Overall, farmers lack literacy and knowledge of weather patterns and soil conditions, which causes the soil to deteriorate due to excessive pesticide and insecticide use or causes crops to not yield as well as they should due to nutrient deficiencies in the soil. To solve these problems, farmers can use this crop prediction model to increase profits and reduce soil acidification.

I. INTRODUCTION

Agriculture is the essential to India's overall development, agriculture is the most significant industry in India. India has a population of roughly 1.39 billion, therefore to meet their needs, 60.3% of the country's total land is used for agriculture. But since the soil deteriorates over time as a result of farmers' ignorance, it's critical to adopt new agricultural methods. Farmers will benefit because they can improve production and profit, and people will benefit because they can purchase commodities for less money. Prior to the development of new farming methods, farmers would simply assess the needs of their customers and grow crops accordingly without considering the state of their soil or the amount of fertilizer they would need to use.The use of machine learning to predict crops has been the subject of extensive study. Numerous machine learning algorithms, such as decision trees, random forests, and neural networks, have been used to solve this problem.

Using these models, predictions have been made about crop production, quality, and disease resistance. Researchers have also developed methods for merging data from many sources, such as sensors, satellites, and other remote sensing technologies, in order to improve the accuracy of crop forecast models. Overall, crop prediction using machine learning has the potential to greatly increase farming enterprises' productivity and profitability.

By examining factors such as temperature, soil PH, soil moisture, rainfall, and other weather variables, the created model will suggest the crop that should be produced most successfully.

The model will use the farmers' desired inputs, including rainfall, nitrogen content, phosphorus content, etc. All of the model's inputs, including KNN, Decision Tree, and Random Forest, were then applied in order to find patterns in the data and predict the crop in accordance with those patterns.

II. LITERATURE SURVEY

A.Balasubramanian[1] provides a description of the many soil types that are present. In this essay, he travels to many locales to inform farmers about various soil types and which soils are best for specific types of cultivation. Machine learning is used by Girish L et al.[2] to estimate rainfall and crop productivity. They utilize a variety of machine learning (ML) methods to estimate crop yields and rainfall, including

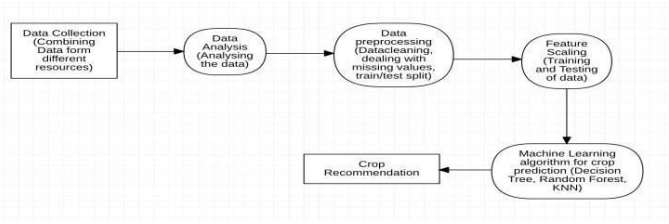
KNN, SVM, linear regression, and decision trees. In the end, they choose SVM since it makes predictions more accurately.

According to Ashwani Kumar Kushwaha [3], farmers can boost their profits by foreseeing the ideal crops to grow. To forecast the farmers' ideal harvest, they employ the Hadoop platform and big data. According to Nischitha K et al.[4], He utilizes a decision tree for crop prediction and a support vector machine (SVM) for rainfall prediction to determine which crop should be planted based on the quantity of nutrition it needs and the market price for the same crop. In order to improve the quantity and quality of data, Archana Gupta et al.[5] explain crop prediction using IOT and ML by using both live and historical data the plants. In order to reduce the suicide rate, Rushika Ghadgeet al.[6] describe crop prediction using two algorithms and how they choose the optimal systemby comparing how accurate it was. After thecomparing their results with those of other models, P. S. Nishant et al.[7] used layered regression to create a web application. In the future, they hoped to create an app in several regional languages for the benefit of farmers. In order to predict crop yield based on location, D.

A. Bondre [8] employs two algorithms: random forest, which has an accuracy of 86.35%, and support vector machine, which has an accuracy of 99.47%. He also uses image processing to detect crop diseases so that users can purchase specific pesticides based on the disease.

III. PROPOSED SYSTEM

The most suited crop would be predicted by the proposed system based on soil characteristics and climatic variables as PH, rainfall, humidity, and temperature.We initially gathered the data from a variety of sources, including public data, kaggle, and some private data.



We next analyzed the data to determine the importance of each parameter in the crop forecast and other tasks.

The third step involves cleaning, identifying data gaps, and splitting our main dataset into two separate tables, one of which contains the resultant crops with their unique ID and the other of which contains every other parameter. The fourth step, feature scaling, involves splitting our main dataset into two datasets, one of which is used to train our dataset and the other for testing so that we can estimate our accuracy. suggested system follows :

a. **Collection of Data:** The gathering of data is crucial for machine learning. A machine learning model must be trained on a sizable and varied dataset in order to be successful. The model may not be able to produce precise predictions or take the necessary steps if the data is of poor quality or is not indicative of the issue the model is attempting to solve. Therefore, we gather data from various sources, including government websites, Kaggle, and some private data, that includes soil nutrients such as (Potassium, Phosphorus, and Nitrogen) along with the Ph value of the soil and weather parameters such as Temperature, Humidity, and Rainfall. We then combine these data into one dataset by understanding each dataset and doing so when each dataset is complete.

b. **Analysing of Data :** Data analysis is a critical stage in the crop prediction project because it allows researchers to discover patterns and trends that can be used to train the model. By carefully analyzing the data, we can identify the most relevant and useful components to include in the model as well as any potential problems or biases that can compromise the model's performance. This enables us to develop a model that is more accurate, effective, and capable of reliably forecasting crop growth. In order to analyze the correlation between them and learn more about the significance of each parameter, we analyze the data using a variety of Python tools, including Matplotlib, Seaborn, Numpy, and pandas.

c. **Data Preprocessing :** Data analysis is a critical stage in the crop prediction project because it allows researchers to discover patterns and trends that can be used to train the model. By carefully analyzing the data, we can identify the most relevant and useful components to include in the model as well as any potential problems or biases that can compromise the model's performance. This enables us to develop a model that is more accurate, effective, and capable of reliably forecasting crop growth. It is essential to creating a model that is precise, trustworthy, and effective. We may enhance the model's performance and make sure that it is able to generate precise predictions about by carefully preparing the data about crop growth. In order to preprocess our data, we split it into two separate tables, one of which contains all the parameters besides the crop name and the other of which just contains the crop name and its unique ID.

d. **Feature Scaling :** Using feature scaling, we can alter the data so that each feature has a comparable scale. To do this, a variety of methods can be utilized, such as normalization, standardization, and min-max scaling. By scaling the features, we may improve the data's compatibility for machine learning methods and the model's accuracy and reliability. After data preprocessing, the dataset must be scaled into training and testing data, which determine the model's accuracy.

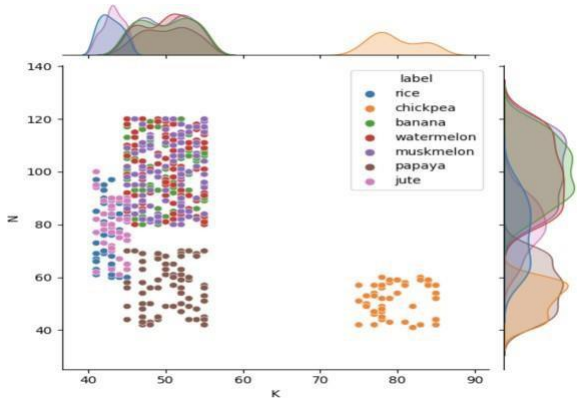
e. **Model Selection:** By carefully examining the data and the problem we are trying to solve, we can decide which algorithms to use for the crop prediction project. Depending on how well they perform on the data, how sophisticated they are, and how effectively they generalize to new data, different algorithms may need to be compared. By selecting the most appropriate method, we can make sure that the model is capable of providing an accurate estimate of crop growth. We choose KNN, Decision Tree, and Random Forest as our models. After deciding on three models, we implement each one, choosing the Decision Tree since it predicts outcomes more precisely than the other two algorithms.

f. **Crop Prediction :** Based on the projected rainfall, the soil's composition, and meteorological conditions, the system will recommend the ideal crop for cultivation. This method also shows, in parts per million, how much seed is required to grow a suggested crop.

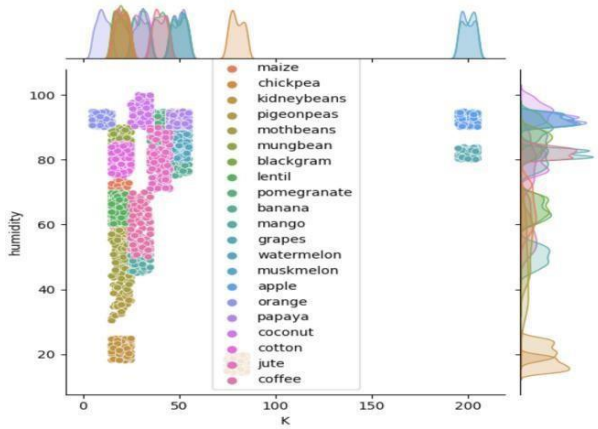
IV. RESULTS & DISCUSSIONS

By considering a variety of elements, including nitrogen, phosphorus, and potassium, as well as environmental elements, such rainfall and temperature, our method recommends the best crop for farmers. And in order to get the desired outcome, the user needs enter the parameters. We split the main dataset into two pieces for training and testing data, and we evaluated our model on those in order to get the desired outcome.

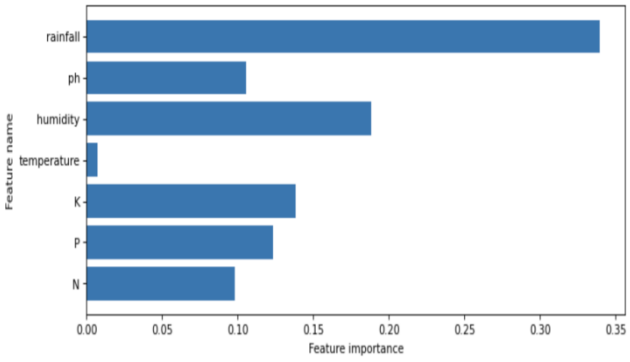
When attempting to describe a relationship between two variables, pair plots are used to identify the most distinct clusters or the most effective arrangement of characteristics. It also aids in the development of certain simplistic classification models by establishing some simple linear separations or fundamental lines in our data set. The diagonal distribution of two qualities across all potential combinations is shown in this key graphic! Visualizing how distinct classes are from one another in a certain environment is quite beneficial.



Only 7 out of the 21 crops that require a high concentration of both nitrogen and potassium in the soil are present in this plot, which shows crops with more than 40 ppm of both nitrogen and potassium.



By looking at this plot, we can see that just two crops, apples and grapes, required more potassium than the majority of other crops, which ranged from 0 to 50 parts per million.



A horizontal pargraph called Barh is displayed here because it displays the most accurate forecast among the other models. This model demonstrates the significance of the feature, and it is clear that most of our predictions are based on rainfall, which has the highest value, followed by humidity in

second place, temperature in third place, and so on until the final and least dependent variable.

V. CONCLUSION & FUTURE SCOPE

By using machine learning to predict farming operations Agriculture May become substantially more efficient and profitable. By developing a model that takes into account a number of factors that can influence crop growth, such as soil nitrogen, phosphorus, and potassium levels, temperature, humidity, pH, and rainfall, we can make more accurate predictions about how a particular crop will perform in a particular location. The use of machine learning techniques allows the model to learn from historical data and improve over time at making predictions. This will allow farmers to choose when and how to plant, water, and harvest their crops, which will ultimately lead to better yields and higher profits. We can improve our model so that farmers no longer need to enter rainfall data because we can obtain that data from another source. To make farmers' work easier, we can create an app that is simple to use and available in a variety of languages so that farmers can use it easily. Wecan also suggest fertilisers to farmers so that soil acidification is almost completely eliminated and yield will grow properly.

REFERENCES

[1] A. Balasubramanian "Earth"s , Atmospheric and Ocean Sciences in August" 2017.

[2] Girish L "Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning" 2019.

[3] Ashwani Kumar Kushwaha, Sweta bhattachrya "crop yield prediction using agro algorithm in Hadoop" 2015.

[4] N. K, Manju raju M.R, Ashwini, M. N, and D. Vishwakarma, "Crop prediction using machine learning approaches," International Journal of Engineering Research and, vol. V9, no. 08, 2020.

[5] Archana Gupta, Dharmil Nagda, Pratiksha Nikhare, Atharva Sandbhor "Smart Crop Prediction using IoT and Machine Learning" 2020.

[6] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L "Prediction of Crop Yield using Machine Learning" 2018.

[7] P. S. Nishant, P. Sai Venkat, B. L. Avinash, and B. Jabber, "Crop yield prediction based on Indian agriculture using machine learning," 2020 International Conference for Emerging Technology (INCET), 2020.

[8] D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning an algorithms," International Journal of Engineering Applied Sciences and Technology, vol. 04, no. 05, pp. 371–376, 2019.

[9] S. Agarwal and S. Tarar, "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms," Journal of Physics: Conference Series, vol. 1714, no. 1, p. 012012, 2021.

[10] S. Veenadhari, B. Misra, and C. D. Singh, "Machine Learning Approach for forecasting crop yield based on climatic parameters," 2014 International Conference on Computer Communication and Informatics, 2014.

[11] R. Dhanapal, A. Ajan Raj, S. Balavina yagapragathish, and J. Balaji, "Crop price prediction using supervised machine learning algorithms," Journal of Physics: Conference Series, vol. 1916, no. 1, p. 012042, 2021.

[12] N. Bali and A. Singla, "Emerging trends in machine learning to predict crop yield and study its influential factors: A survey," Archives of Computational Methods in Engineering, vol. 29, no. 1, pp. 95–112, 2021.

[13] K. Peng, R. Jiao, J. Dong, and Y. Pi, „„A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter,"" Neuro computing, vol. 361, pp. 19–28, Oct. 2019.

[14] J. Zhang, H. Wang, J. Chu, S. Huang, T. Li, and Q. Zhao, „„Improved Gaussian– Bernoulli restricted Boltzmann machine for learning discriminative representations,"" Knowl.-Based Syst., vol. 185, Dec. 2019, Art. no. 104911.

[15] H. Jahangir, M. A. Golkar, F. Alhameli, A. Mazouz, A. Ahmadian, and A. Elkamel, „„Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ANN,"" Sustain. Energy Technol. Assessments, vol. 38, Apr. 2020, Art. no. 100601.

[16] T. U. Rehman, S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, „„Current and future applications of statistical machine learning algorithms for agricultural machine vision systems,"" Comput. Electron. Agricult., vol. 156, pp. 585–605, Jan. 2019