Crop Prediction using Machine Learning

TUSHAR GUPTA[1], SUNIL MAGGU[2], BHASKAR KAPOOR[3]
[1] *Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India.*
[2, 3] *Faculty, Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, 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 much about the state of the soil, farmers simply keep growing the same crops time and time again by applying arbitrary amounts of fertiliser. 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 scenarios. By recommending the ideal crop to plant based on the weather and soil conditions, our algorithm aids farmers. Therefore, using our model, farmers can learn about the various crops they need to grow in order to increase production, which then increases profit. Therefore, a crop prediction model using machine learning is used to solve the problem. For this, it requires input from various factors such as soil quality, weather variables, and historical crop data. Using this data, a model can then predict the future crop that should be grown, assisting users and farmers in deciding which crop is best to grow given the current situation. The model is trained using historical crop data as well as the pertinent variables, including water and soil characteristics. Our dataset is completely dependent upon the training and testing of the data, and it is this dataset that allows us to determine the model's accuracy. 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.***

Indexed Terms- ***Crop prediction, Machine Learning Models, Soil Checking, Crop Recommendation, Rainfall prediction.***

I. INTRODUCTION

India's most crucial industry is agriculture because it has been instrumental in the country's overall growth. India has a population of roughly 1.429 billion, therefore to meet their needs, 60.22% of the country's total land is used for agriculture. But as the soil deteriorates over time as a result of farmers' ignorance, it's critical to adopt new agricultural technologies. Farmers will benefit because they can improve production and profit, and people will benefit because they can purchase commodities for less money. Before the advent of the new methods, farmers would simply assess the needs of society and grow crops in response, without having a clear understanding of the state of their soil. They would also apply fertilisers haphazardly, which would cause the soil to become more acidic, which would reduce crop yields since the soil's top layer would be harmed. In light of these issues, we make the decision to create an ML model for the benefit of farmers.

Research and development into crop prediction using machine learning has been extensive. Numerous machine learning techniques, such as decision trees, random forests, and KNN, have been used to solve this problem. Using these models, predictions have been made about crop prediction. Researchers have also developed methods for merging data from many sources, such as sensors, Government Data, Private Data, and other technologies, in order to improve the accuracy of crop prediction models. Overall, crop prediction using machine learning has the potential to greatly increase the productivity and profitability of farming operations.

The created model will suggest the crop that should be cultivated based on the weather and soil parameters, such as temperature, rainfall which is predicted by another dataset by entering the state, soil PH and humidity . The desired inputs from the farmers, such as rainfall, nitrogen content, phosphorus content, potassium content, etc., will be used by the model. All

of the inputs were then applied to the model, including KNN, Decision Tree, and Random Forest, in order to find patterns within the data and predict the crop in accordance with those patterns.

1.1 Organisation of Paper : The paper is organised as follows: Section 1 provides a brief overview of the need for an ML model due to farmers' limited knowledge, Section 2 summarises what has been done thus far to ensure a good yield, and Section 3 provides an overview of the model itself and how it will function to address the problem.

II. LITERATURE SURVEY

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 utilise 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. By anticipating the finest compatible crops, Ashwani Kumar Kushwaha [3] explains how to maximise the profit of farmers. To forecast the farmers' ideal harvest, they employ the Hadoop platform and big data. By considering factors such as weather conditions, soil parameters, and rainfall, Nischitha K et al.[4] describe which crop should be grown along with the amount of nutrition it requires and the market price for the same crop. He uses decision trees for crop prediction and SVM for rainfall prediction. In order to boost the amount and quality of the crops, Archana Gupta et al.[5] discuss crop prediction combining IOT and ML by using both live and historical data. In order to reduce the suicide rate, Rushika Ghadge et al.[6] describe crop prediction using two algorithms and how they choose the optimal system by comparing how accurate it was. After comparing 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. With a 97% accuracy rate, S. Agarwal et al.[9] predict crops using both machine learning (SVM) and deep learning (LSTM,RNN). They also predict the best potential cheaper crop for good yield and profit. Using data mining techniques, S. Veenadhari et al.[10] predict crops with an accuracy of above 75% based on the input parameter of climate. In order to predict crop prices, R. Dhanapal et al.[11] used machine learning to train a Decision Tree Regressor model on a variety of Kharif and Rabi crops. The model was then used to estimate crop prices for the following 12 months.

III. PROPOSED SYSTEM

The proposed system will forecast the best suited crop based on soil characteristics and weather variables including PH, rainfall, humidity, and temperature.

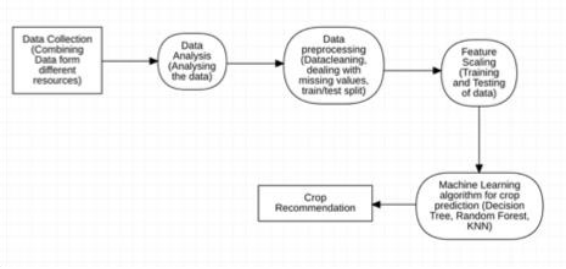

Figure 1. Proposed System Architecture

As mentioned in the flow chart above, we first gathered the data from a variety of sources, including public data, kaggle, and some private data. We then analysed the data to determine how each parameter affected the crop prediction, and in the third step, we preprocessed our data by cleaning, identifying data gaps, and splitting our main dataset into two separate tables, one of which contained the resultant crops with their unique id and the other contained every other parameter. After that, we move on to the machine learning method for crop prediction, where we employ numerous different algorithms for doing so. Once each step is complete, we move on to the next stage, which is forecasting the crop. Here is the proposed system:

3.1. 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 might not be able to generate precise predictions or choose the right course of action if the data is of poor quality or is not indicative of the issue the model is attempting to solve. Therefore, we gather information about soil nutrients like potassium, phosphorus, and nitrogen as well as the soil's pH value and weather variables like temperature, humidity, and rainfall from a variety of sources, including government websites, Kaggle, and some private data, and combine it into one dataset by understanding each dataset and combining it when each dataset is ready so that each has the same exact column name and contains no duplicate data.

3.2. 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 thoroughly analysing 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 assess the correlation between them and learn more about the significance of each parameter, we study the data using a variety of Python tools, including Matplotlib, Seaborn, Numpy, and pandas.

3.3. 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 thoroughly analysing 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 it is able to predict crop growth accurately by carefully preparing the data. 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.

3.4. 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 utilised, such as normalisation, standardisation, 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.

3.5. Model Selection :By carefully analysing the data and the problem we are aiming to solve, we can decide which algorithms to use for the crop prediction project. Depending on how well they perform on the data, how smart they are, and how well they generalise 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.

3.6 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.

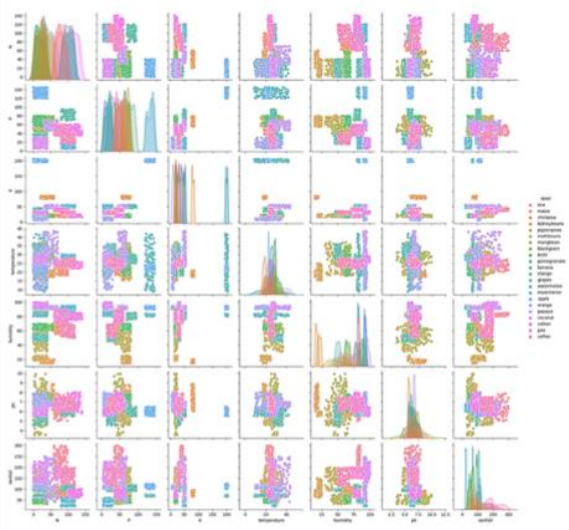

Figure 2.Pairplot for each of the inputs we need to make a forecast

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 features. 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! Visualising how distinct classes are from one another in a certain environment is quite beneficial.

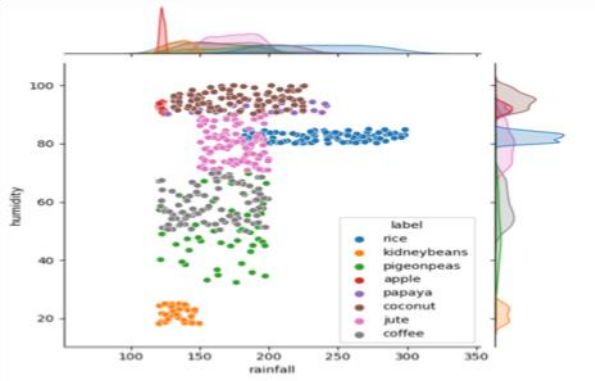

Figure 3. Joinplot of humidity vs rainfall

By looking at this plot, we can observe which crops need temperatures below 30°C and rainfall totals of over 120 mm, indicating that they require a lot of water and cold weather. Out of the 21 crops, only 8 require cold temperatures and significant rainfall.

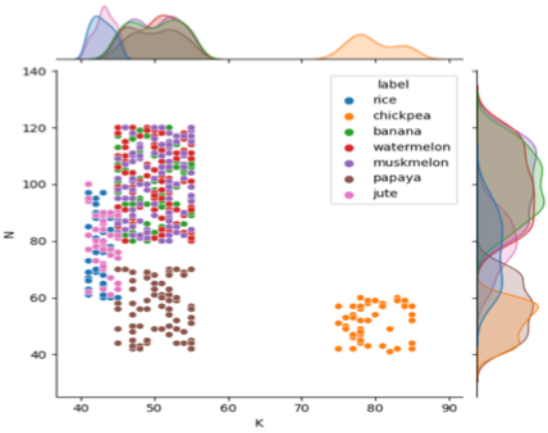

Figure 4. Joinplot of Nitrogen vs Potassium

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.

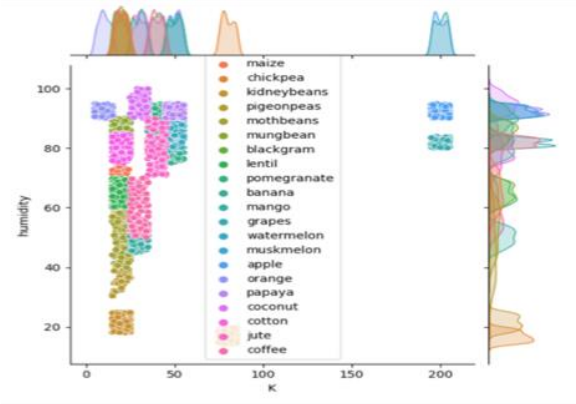

Figure 5. Joinplot of Humidity vs Potassium

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.

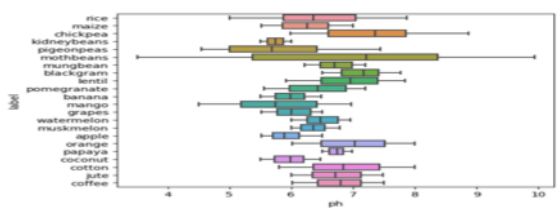

Figure 6. Boxplot shows all the crops against the soil's pH value

By using this box plot, we can see that the soil has a pH range of 6-7, meaning that for good crop yields, we require somewhat acidic soil.

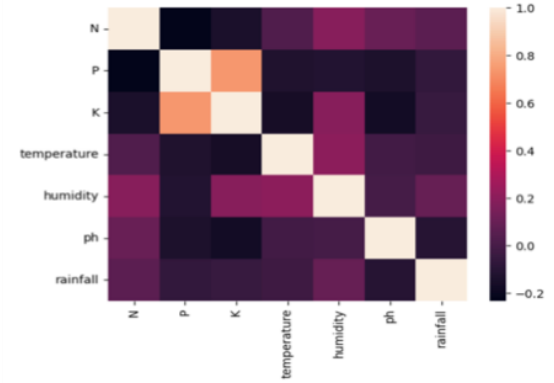

Figure 7. Heat maps for all parameters except from label parameter

This heat map reveals that potassium and phosphorus have a stronger correlation with one another than the other factors.

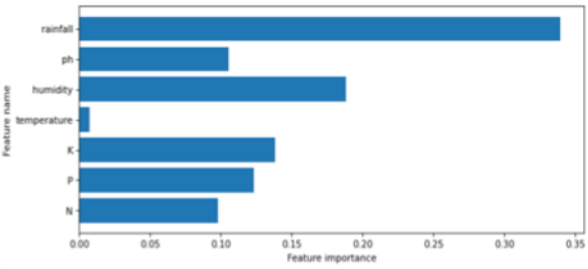

Figure 8. Barh for Decision Tree Prediction

A horizontal bargraph 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 dependant variable.

```
from sklearn.tree import DecisionTreeClassifier

clf = DecisionTreeClassifier(random_state=42).fit(X_train, y_train)
clf.score(X_test,y_test)

place = input("Enter the place : ")
place_rainfall = dict[place]
check = clf.predict([[90,42,43,20.8,82,6.5,place_rainfall]])
print(targets[check[0]])
```

```
Enter the place : ARWAL
banana
```

Figure 9 : District Wise Prediction

In this, we anticipate the crop based on the inputs we are supplying to our model, which are: N-90ppm, P-42ppm, K-43ppm, Temperature-20.8°C, Humidity-6.5 g.m-3, Ph-6.5, and District as 'ARWAL' by which the rainfall is projected from different datasets. After processing, our model indicates that the best crop to be cultivated there is "banana," which is our outcome.

V. CONCLUSION AND FUTURE SCOPE

By using machine learning to predict crops, farming operations can become substantially more efficient and profitable. By developing a model that takes into account a variety of factors that can affect crop growth, such as nitrogen, phosphorus, and potassium levels in the soil, temperature, humidity, pH, and rainfall, we can make more accurate predictions about how a specific crop will perform in a specific location. By applying machine learning methods, the model may learn from historical data and improve its ability to predict outcomes over time. Thanks to this, farmers will have more control over 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 we can create an app that is simple to use and available in a variety of languages for the farmers, enabling them to utilise it with ease. We can also propose fertilisers to the farmers so that soil acidity is nearly completely eliminated and yield will grow as intended.

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 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.