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- ***India is the second in agriculture production and majority of people in India opt agriculture as their profession. Farmers just keep growing the same crops over ans over as they don't know much about soil condition so they use random amount of fertilisers to grow crops. This will eventually increase the acidification of soil and damages the upper layer of soil. So, to overcome these scenarios we, design a Machine Learning model for farmers. Our model helps the farmer by suggesting the best crop to be grown according to the weather ans soil conditions. Hence by using our model farmers can know about different type of crops they have to grow so that they can increase their production which leads to increase their profit and minimise the soil pollution.***

So a crop prediction model using machine learning is used to solve the issue and for that it takes input of various factors such as Soil condition, weather parameters and historical data of the crops and by using this information a model can predict about future crop that should be grown, which helps the users/farmers to decide that which crop is best to grow in current scenario. The model is trained with the historical data of the crops and the relevant different parameters such as water parameters and soil conditions. The training and testing of the data is totally dependent upon our dataset and by the help of that we are able to know about the accuracy of our model.

Overall, all farmers are not literate and have less knowledge of the soil condition and weather condition by which the soil degrades due to the excess use of pesticides and insecticides or the crop is not yield as good as it has to be due to lack to nutrients in the soil so to remove all the issues farmers can use this crop prediction model to gain more profit and minimising the acidification of the soil.

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

I. INTRODUCTION

Agriculture is the most important sector in India as it is responsible for the overall development of India. India's population is about 1.39 billion so to satisfy the needs of that many people about 60.3% off the total land is used for agriculture purposes. But as the soil is degraded over time due to farmers lack of knowledge adopting new agriculture technologies is important. This will help both the farmers and the people as the farmers can increase the production and profit whereas the people are able to get the crops at less prices. Before the new techniques farmers just sees the need of the people and grow crops accordingly without the proper knowledge of their soil condition and uses random amount of fertilisers which leads to soil acidification which leads to less production as the upper layer of the soil get damaged. So, by seeing these problems we decide to design a ML model for the betterment of the farmers.

Crop prediction using machine learning has undergone a lot of research and development. This issue has been addressed using a wide range of machine learning techniques, including decision trees, random forests, and neural networks. Predictions concerning the production, quality, and disease resistance of crops have all been made using these models. In order to increase the precision of crop prediction models, researchers have also created techniques for combining data from a number of sources, including sensors, satellites, and other

remote sensing technologies. Overall, using machine learning to crop prediction has the potential to significantly raise farming operations' productivity and profitability.

The designed model will recommend the most suitable crop to be grown by checking the soil condition and weather conditions like Temperature, PH value of soil, Humidity of the soil, Rainfall, etc. The model will take the desired inputs from the farmers such as Rainfall, Nitrogen Content, Phosphorus Content, etc. Then all the inputs applied to the model like (KNN, Decision Tree and Random Forest) to identify the pattern among the data and then predict the crop according to those inputs.

1.1 Organisation of Paper : The paper is organised as: Section 1 is the brief introduction of the need for the ML model because of the farmers less knowledge, ans section 2 is the brief of what is done till now for the sake of the good yield and section 3 is the brief of the model here that how it is going to work to solve the issue.

II. LITERATURE SURVEY

A. Balasubramanian[1] describes the different kinds of soils that are present. In this paper he moves across different regions to educate farmers about different soils and which soil is good for which production. Girish L et al.[2] works upon the rainfall prediction and crop yielding using machine learning. They uses many algorithms of ML for predicting the crop yields and rainfall and the algorithms they used are KNN, SVM, linear regression and decision tree and in the end they took SVM as it predicts more accurately. Ashwani Kumar Kushwaha [3] describes how to increase the profit of farmers by predicting the best suitable crops. They uses Hadoop platform with big data to predict the best suitable crop for the farmers. Nischitha K et al.[4] describes which crop should be grown with the amount of nutrition it requires and also the market price for the same crop by taking some parameters like weather condition, soil parameters and the rainfall in which he uses decision tree for crop prediction and SVM for rainfall prediction. Archana Gupta et al.[5] describes crop prediction using IOT and ML by using both live and historical data to increase the quantity and quality of the crops. Rushika Ghadge et al.[6] describes crop prediction by using two algorithms and selected the best algorithm by comparing their accuracy so that the suicide rate goes down. P. S. Nishant et al.[7] uses stacked regression after comparing with different models they build a web application and in the future they wanted to make an app for the sake of the farmers in different regional languages. D. A. Bondre [8] predicts the crop yield on the basis of location, he uses two algorithms random forest with an accuracy of 86.35% and support vector machine with an accuracy of 99.47% and crop disease detection using image processing by which a user can get the certain pesticides from the market according to the the disease. S. Agarwal et al.[9] uses both machine learning (SVM) and deep learning (LSTM,RNN) for prediction of crop with an 97% accuracy and they predict the best possible cheaper crop for good yield and profit. S. Veenadhari et al.[10] uses data mining techniques to predict the crops on the basis of the climate input parameter with an accuracy of over 75%. R. Dhanapal et al.[11] is predicting the price of the crops and the prices for the next 12 months are also proposed by using machine learning and model which is implemented is Decision Tree Regressor they trained the model on several Kharif and Rabi crops for price predicting.

III. PROPOSED SYSTEM

The Proposed System will predict the most suitable crop on the basis of soil condition and weather parameters like PH, Rainfall, Humidity, and Temperature.

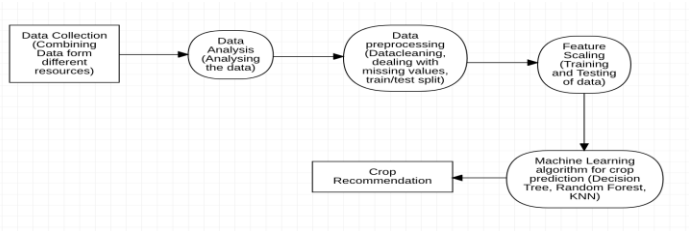

Figure 1. Proposed System Architecture

In the above flow chart as mentioned we first collected the data from various places such as government data, kaggle and some private data, after collecting the data we analyse it that how every parameter matters in the prediction of the crop and on

the third step we preprocess our data i.e cleaning, data missing problem and diving our main dataset into two separate tables in which one contains the resultant crops with its unique id and the other contains every other parameter after that we move on to the forth section i.e the feature scaling in which we separate our main dataset into two datasets in which now of them is to train our dataset and other is for testing so that we can predict our accuracy after that we move on to the machine learning algorithm for crop prediction where we use many different algorithms for predicting the crops and when every step is finished we move on out final phase i.e predicting the crop. Proposed System follows :

3.1. Collection of Data :The collection of data is extremely important in machine learning. In order for a machine learning model to be effective, it needs to be trained on a large and diverse dataset. If the data is of poor quality or is not representative of the problem the model is trying to solve, the model may not be able to make accurate predictions or take appropriate actions. So We collect the data which contains soil nutrients such as (Potassium, Phosphorus & Nitrogen) along with the Ph value of the soil and weather parameters such as Temperature, Humidity and Rainfall from various places such as government websites, Kaggle, and some private data and place them into one dataset by understanding every dataset and put them together when every data set is ready to have the same exact column name with no duplicated data.

3.2. Analysing of Data :In order to find patterns and trends that may be utilised to train the model, data analysis is a crucial phase in the crop prediction project. We can choose the most pertinent and practical elements to incorporate into the model by carefully examining the data, and we can also see any potential issues or biases that can affect the model's performance. This makes it possible for us to create a model that is more precise and efficient and that can anticipate crop growth with greater reliability. We analyse the data by using many python libraries like Matplotlib, Seaborn, Numpy and pandas by plotting different graphs with many parameters to analyse the correlation between them and to know more about the importance of every parameter.

3.3. Data Preprocessing : In order to find patterns and trends that may be utilised to train the model, data analysis is a crucial phase in the crop prediction project. We can choose the most pertinent and practical elements to incorporate into the model by carefully examining the data, and we can also see any potential issues or biases that can affect the model's performance. This makes it possible for us to create a model that is more precise and efficient and that can anticipate crop growth with greater reliability. It plays a key role in developing a model that is accurate, reliable, and efficient. By carefully preprocessing the data, we can improve the performance of the model and ensure that it is able to make accurate predictions about crop growth. We preprocess our data by dividing them into two different tables in which one contains all the parameters except the crop name and the other table consist of only crop name with its unique id.

3.4. Feature Scaling : We can change the data so that each feature is on a comparable scale by using feature scaling. Several techniques, including normalisation, standardisation, and min-max scaling and, can be used to accomplish this. We can increase the data's suitability for machine learning techniques and the model's accuracy and dependability by scaling the features. After Data Preprocessing we have to scale the dataset into two parts : training and testing data which help to find out our accuracy of the model.

3.5. Model Selection : We can choose the best algorithms to utilise for the crop prediction project by conducting a thorough examination of the data and the issue we are attempting to solve. This may entail contrasting various algorithms according to how well they work on the data, how sophisticated they are, and how well they generalise to new data. We can make sure that the model is able to estimate crop growth accurately by choosing the most suitable algorithm. In Model selection we select KNN, Decision Tree and Random forest. After selecting 3 different models we implement every single one and at the end we take Decision Tree as it predicts more accurately than the other two algorithm.

3.6 Crop Prediction : The system will suggest the best crop for cultivation based on the amount of anticipated rainfall, the composition of the soil, and

weather parameters. This approach also displays the amount of seed needed to cultivate a recommended crop in parts per million.

IV. RESULTS & DISCUSSIONS

Our system suggests the optimum crop for farmers by taking into account many factors, such as nitrogen, phosphorus, and potassium, as well as environmental factors, such as rainfall and temperature. And the user must enter the parameters in order to obtain the desired result. In order to acquire the desired result, we divided the main dataset into two sections as training and testing data and tested our model on those.

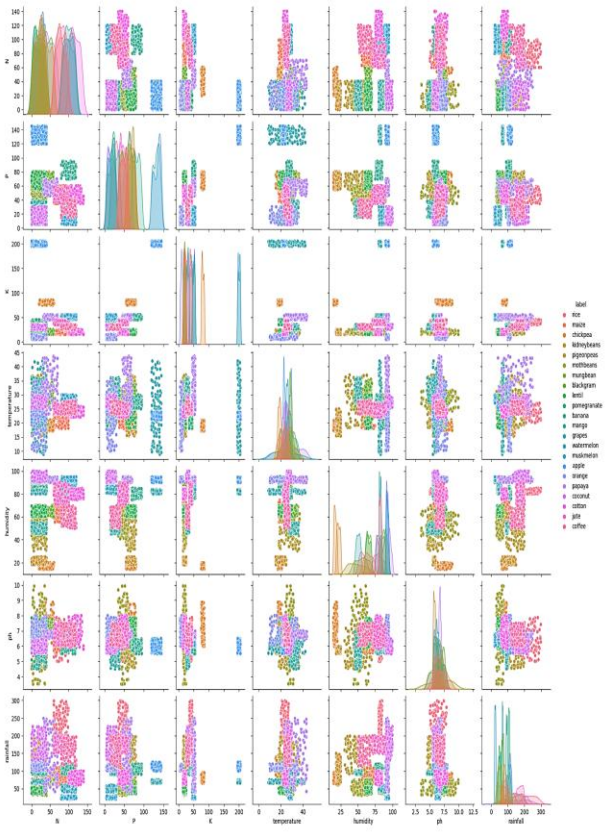

Figure 2. Pairplot for all the parameters that we have to give for the prediction

Pair plots are used to determine the most distinct clusters or the best combination of features to describe a connection between two variables. By creating some straightforward linear separations or basic lines in our data set, it also helps to create some straightforward classification models. It is a crucial figure that shows the diagonal distribution of two attributes across all possible combinations! It is quite helpful to picture how different classes are from one another in a specific setting.

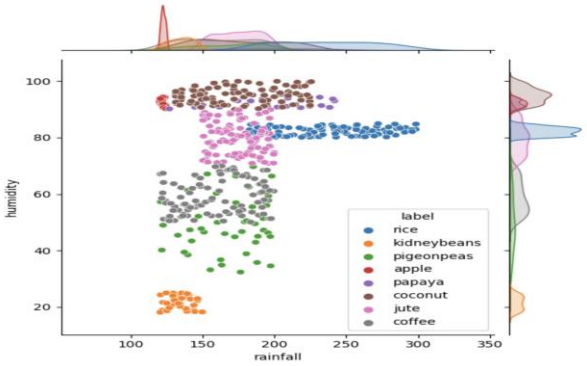

Figure 3. Joinplot of humidity vs rainfall

By this plot we see the crops whose requirements of rainfall is over 120 mm and temperature below 30°C which implies that large amount of water and cold weather so by this we can see that there are only 8 crops out of 21 crops need cold weather and high rainfall.

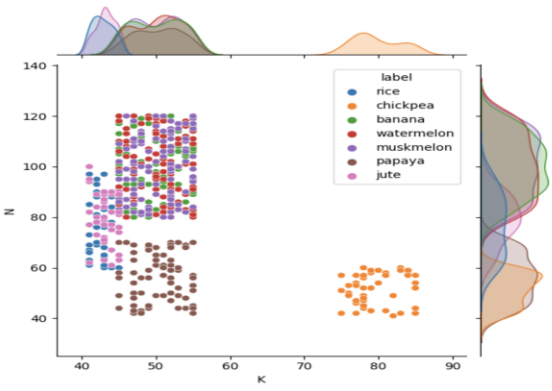

Figure 4. Joinplot of Nitrogen vs Potassium

In this plot we see the crops with more than 40 ppm of both Nitrogen and Potassium and as we can see that there are only 7 crops out of 21 crops are the present which requires high concentration on both nitrogen and potassium in the soil.

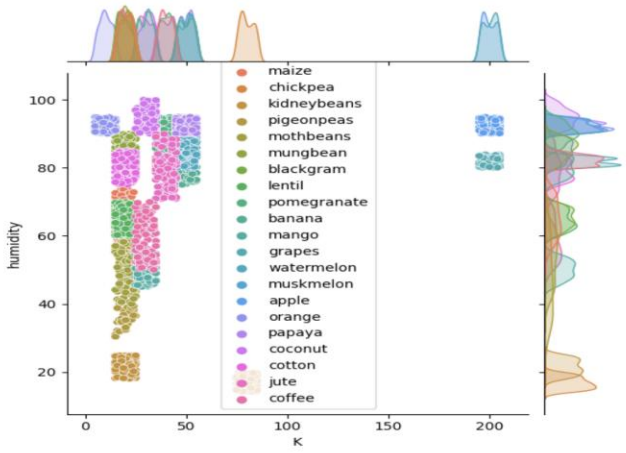

Figure 5. Joinplot of Humidity vs Potassium

By this plot we can see that there are only two crops apple and grapes that required around 200 ppm of potassium which is a lot as compare to other crops as most the crops required 0-50 ppm of potassium.

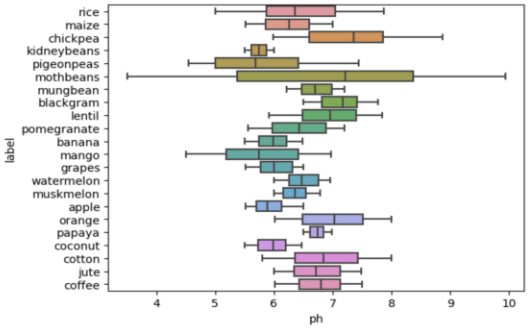

Figure 6. Boxplot of all the crops to the PH value of the soil

By this box Plot we can see that the soil generally of ph 6-7 that implies we need slightly acidic soil for the good crop yielding.

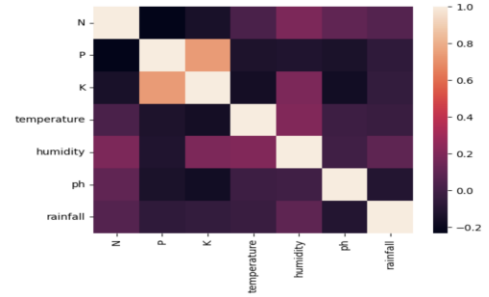

Figure 7. Heat map for every parameter except the label parameter

By this heat map we can see that phosphorus and potassium are highly correlated to each other than the other parameters.

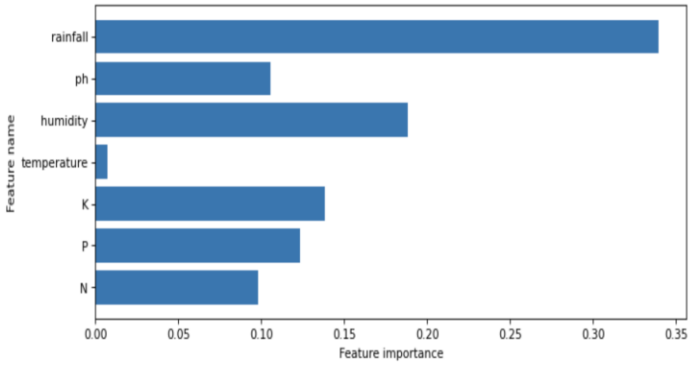

Figure 8. Barh for Decision Tree Prediction

Barh is a horizontal bargraph and we show it here because it shows the best prediction among the other models. This model shows the importance of the feature and as we can see that most of our predictions are based on rainfall as rainfall has the highest value and as for the humidity to the second and so on until the last and least dependent variable that is the temperature.

CONCLUSION AND FUTURE SCOPE

The effectiveness and profitability of farming operations can be significantly increased by applying machine learning to crop prediction. We can create more precise predictions about how a specific crop will perform in a specific place by creating a model that takes into consideration a variety of parameters that can affect crop growth, such as nitrogen, phosphorous, potassium levels in the soil, temperature, humidity, pH, and rainfall. The model can learn from past data and get better at making predictions over time by utilising machine learning algorithms. Farmers will be able to decide when and how best to plant, water, and harvest their crops thanks to this, which will ultimately result in better yields and greater earnings. We can upgrade our model in such a way that farmers do not have to put the rainfall information as we can gather that information from some other source and to make the work of the farmers more easy we can make a easy to use app and in many different languages for the

farmers so that the farmers are able to use it with ease or we can even recommend fertilisers to them so that soil acidification is almost reduced to zero 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 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.