

Crop Prediction using Machine Learning

¹Mrs V. Nikitha(Assistant Professor),
²Ankitha, ³Bharagavi, ⁴Ambika, ⁵Kalyani,

Department of CSE,
MALLA REDDY INSTITUTE OF TECHNOLOGY AND SCIENCE, Telangana,
hyderabad

ABSTRACT

When it comes to the growth of national economies, agribusiness is crucial. Agricultural workers are especially at risk from natural disasters and the effects of global warming. An essential tactic for overcoming this restriction is machine learning (ML). The Harvest production prediction method uses a variety of easily accessible past and future data, including weather and soil factors, yield predictions, and other similar tools, to anticipate crop yields. The major emphasis of this study is on the misuse of the Random Forest Algorithm for projecting harvest outcomes using historical data. We constructed this model using real-world state data and evaluated its efficacy using metrics derived from data. A farmer may use this assumption to forecast harvest yield before investing in a certain agricultural location. Random Forest is the most well-known and innovative machine learning technology, and researchers rely on it to make accurate predictions about crop yield.

Terms of importance: Support Vector Machines, RFA, and Hadoop

INTRODUCTION

For all living things, the natural environment is fundamental to their survival. Farming is the best way to compare and contrast nature. One indicator of a country's prosperity in the past was its level of agricultural growth. Agriculture is vital to the economics of every country. Helping countries like India, whose population is growing rapidly, need agricultural innovations. One of the things that has long been considered to distinguish India from other countries is its horticultural heritage. Younger farmers have no option but to cater to the needs of their more senior counterparts, who tend to their own agricultural areas. As a result, a wide variety of creatures, including birds, depend on foraging for food in the wild. Although the animal has appropriated land, the lifestyle it advocates is backed by the government and is built on common goods and environmentally conscious products that are imported to that area. An essential part of human flourishing is agriculture, the study of growing food for human consumption. Modern farming practices are suffering as a result of technological advancements. These factors have contributed to a rise in the production of fake goods, which may have devastating effects on the wealthy individuals involved. More over two million people work as ranchers in the Indian state of Maharashtra, according to 2016 figures. Due to a growing lack of support, the suicide rate among farmers is rising. Therefore, in order for our

technological setup to generate a high harvest yield, it is crucial that farmers understand the significance of having a reliable yield forecast, that they can prosper with basic soil quality data, and that they can comprehend the insightful climatic needs specific to their region. The significance of farming has diminished throughout the years. The weather has a role in this. The weather was used to estimate the crop. As a result of the country's increasing population, famine is becoming more prevalent in India. Therefore, it is critical to boost agricultural output. The main goal of this crop yield forecast is to provide farmers the most recent data on weather patterns and the best crops to grow so that the nation's growing population can be fed. Most existing frameworks rely on hardware, which may be expensive and a hassle to maintain. Furthermore, precise results are required. Organization of crops according to production rate and market cost is suggested by certain models. The suggested approach uses structured data analysis to try to get around constraints and predict agricultural production. The agricultural landscape is the focal point of the "Forecast of soil quality utilizing data mining approach" project. Because it is entirely code-based, it doesn't give maintenance issues any consideration. As an added bonus, the accuracy would be head and shoulders above that of machines.

robust strategies, since a variety of variables are tracked to provide an accurate prognosis, including soil area, soil type, pH value, weather projections, and more. Data mining is the practice of extracting useful information from large datasets by use of specialized algorithms and software programs. Statistical analysis and visualisation of data are both made possible by this tool. To put it more precisely, it's the process of searching through massive social datasets for cross-domain modeling and correlation patterns. The relationships between people, for example, can throw some insight on the numbers. Insights gained from this data are derived from past events and anticipated trends. For instance, farmers might potentially prevent production losses by identifying their causes with the use of outlined data on the cultivation process. Predicting crop yields is a major challenge for farmers. For ranchers, ROI (return on investment) is king. Ranchers used to

review past data in a particular field to predict how well a harvest would turn out. Weather, pests, and who gets to do what during harvest are the top three most important variables influencing agricultural output. Without thorough and precise knowledge about the harvest's historical context, executives cannot prevent horticulture threats. Consequently, this work paves the way for the prediction of agricultural output. Before starting work in the field, the reaper will check the yield of collecting against the harvested area.

Literature Review

Agriculture has a significant role in India's economy. The agricultural sector contributes significantly to the income of 70% of Indian households. From 2015 to 2016, the agricultural sector was responsible for 17% of India's GDP. However, agricultural support for GVA has been steadily dwindling. The food that farmers provide is vital to our existence, and we are deeply grateful to them. Here is the corresponding correlation: A study on crops was carried out in [1] using a Regression method. More than 362 datasets are analyzed using classification and decision tree computing. This helpful index has been divided into three parts: "natural," "inorganic," and "land"—to help distinguish between different types of soil. The documented setups of the system are reliable and correct. The data needed to be entered into the testing database was handled by the quiz in [2]. To improve the accuracy of soil property predictions, Back Propagation Networks include a hidden layer. This method outperforms its predecessors in terms of efficiency and quality of results. However, the system encounters slowdowns and unexpected output on rare instances. Two AI approaches that deal with relapses are used in the study [3]: Using SVM or RVM to forecast soil quality is a piece of cake. To measure the amount of moisture in the soil, meteorological data and a clever remote invention are used. With the remote device, you may expect an accuracy level of 95% and a margin of error of 15%. Whatever the situation may be, there is no evidence from the data that this has been attempted. Soil fertility and plant nutrition analysis using back-engineering computation is part of the research [4].

The reliable results pave the way for soil enhancement. When compared to more traditional approaches, the results are superior. However, structure is unequal and quite inefficient. Scientists in the agricultural field are increasingly trusting soil components to boost crop yields. At this time, a 77% accuracy rate for dirt classification may be achieved via an innocent Bayes computation. Estimates based on apiorism also link healthy soil to abundant harvests. We show that JRIP, Naive Bayes, and J48 all contribute to accurate grouping [5]. Similarly, Rub et al. published findings on relapse models for crop yield prediction. Recognition Model, Regreen,

Support Vector Machine, and Radial Basis Function Network are some of the contemporary computer approaches. Due to their design, SVM is clearly the better model [6]. In their paper on agricultural production forecasting, Sujatha et al. J48 explains the logic behind various arrangement ideas. There is no shortage of data mining techniques; some examples are Naive Bayes, support vector machines, arbitrary woodlands, and fake neural networks. Meteorological data and yield parameters are included into an assessment model for crop growth [7]. Using climate and disease-related informative indicators, Kushwaha et al. provides the suitability in terms of yield for certain climatic traits and possible consequences, allowing you to improve the quality of your harvests. In [8], they provide a calculation for predictions, an order, and an inquiry to help construct a reliable emotional support system with more accuracy. The Hadoop document store forms the basis of its structure.

PROPOSED SYSTEM

Soil pH, precipitation, humidity, and temperature are just a few of the meteorological factors that the proposed system would use to determine which crops will thrive.

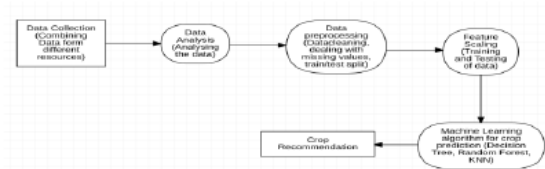


Figure 1. Proposed System Architecture

As shown in the flow diagram, we began by gathering data from a variety of sources, including public records, kaggle, and private sources. Then, we analyzed the data to determine the importance of each parameter in crop prediction. In the third step, known as preprocessing, we dealt with issues like data missingness and cleaned up our main dataset. This allowed us to divide it into two tables: one with the final crops and their unique identifiers, and the other with all of the other parameters. The feature scaling section followed, in which we split our main dataset into two datasets: one to train our model and the other to test its accuracy. Finally, we applied a machine learning algorithm to forecast the crops. The following system is proposed:

3.1. Data Collection: Data collecting plays a crucial role in machine learning. A big and varied dataset is necessary for training an efficient machine learning model. Poor data quality or data that does not accurately reflect the issue the model is attempting to address might hinder the model's ability to provide accurate predictions and perform suitable actions. So We gather information from a variety of

between two variables is the job of pair plots. For the purpose of building simple classification models, it is helpful to first generate some basic linear separations or lines in our dataset. The graphic depicts the diagonal distribution of two qualities in all conceivable combinations, which is an essential concept! Visualizing the differences between classes in a real-world context is very useful.

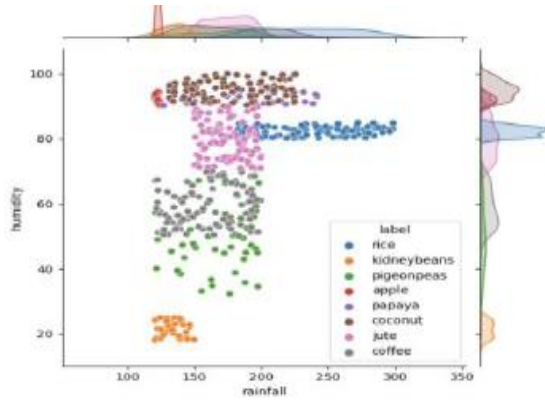


Figure 3. Jointplot of humidity vs rainfall

From this plot, we can deduce that only 8 out of 21 crops need cold weather and heavy rainfall, as shown by the requirements of crops with rainfall over 120 mm and temperatures below 30°C. This indicates a significant quantity of water and cold weather.

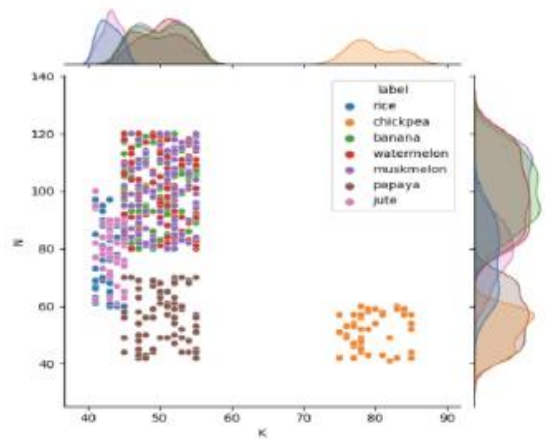


Figure 4. Jointplot of Nitrogen vs Potassium

This figure shows the crops where the nitrogen and potassium levels are above 40 ppm. Out of the 21 crops tested, only 7 had soil concentrations that are too high for their growth.

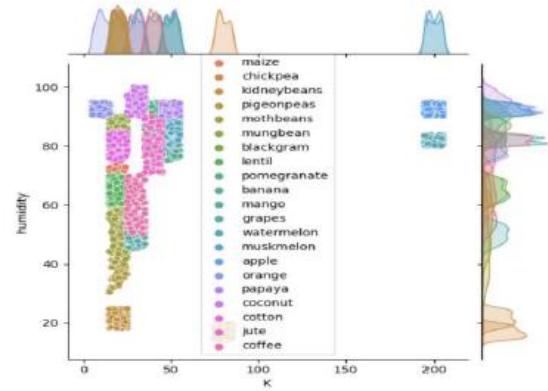


Figure 5. Jointplot of Humidity vs Potassium

As this graph shows, the majority of crops only needed 0-50 ppm of potassium, while a small number of crops—apples and grapes—required over 200 ppm.

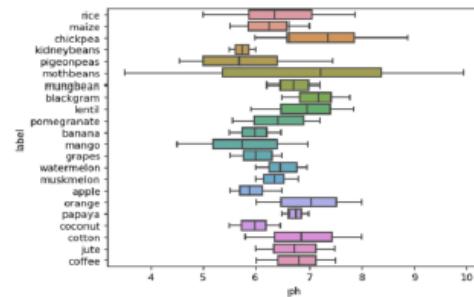


Figure 6. Boxplot of all the crops to the PH value of the soil

This box plot shows that most soils have a pH of 6-7, which means that a slightly acidic soil is required for high producing crops.

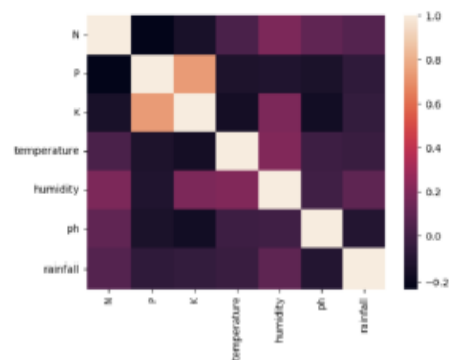


Figure 7. Heat map for every parameter except the label parameter

This heat map clearly shows that, compared to the other parameters, the correlation between potassium and phosphorus is rather strong.

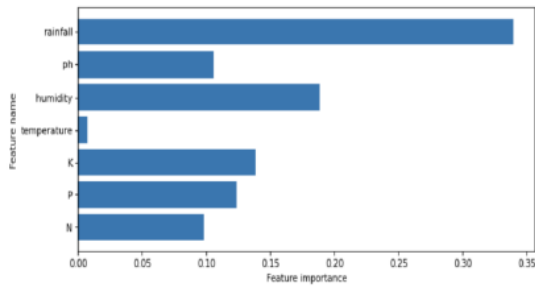


Figure 8. Barh for Decision Tree Prediction

We display Barh, a horizontal bargraph, here because it reveals the model with the most accurate forecast. With rainfall having the greatest value in this model, humidity is the second most important variable, and so on down the list until temperature, the least dependant variable, which is the final one. This indicates how important the feature is.

CONCLUSION AND FUTURE SCOPE

Applying machine learning to crop prediction may greatly enhance the efficiency and profitability of agricultural operations. A more accurate prediction of a crop's performance in a given location can be achieved by developing a model that accounts for a number of factors that influence crop development, such as soil pH, temperature, humidity, nitrogen, and potassium levels, as well as other environmental factors. By using machine learning methods, the model may learn from previous data and improve its prediction abilities over time. Better yields and higher profits will come from farmers being able to plan when and how to sow, irrigate, and harvest their crops. So that farmers don't have to input rainfall data, we may update our model to collect it from another source. To further simplify farmers' job, we can develop an app that is both simple to use and available in several languages 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.