Crop Prediction using machine learning

Pragnya Parmar, Henvi Patel, Akshita Patel, Hetvi Panchal, DR. Pooja Bhatt
Department of Computer Science And Engineering PIET, Vadodara Gujarat

Abstract—India is mostly an agricultural country. Agriculture is essential to both the survival of humans and the Indian economy. Agriculture employs a sizable portion of the workforce as well. 70% of people living in rural India depend on agriculture for their livelihood. Crop output forecasting is one of the most difficult and sought after responsibilities that any government can carry out. Each farmer wants to know how much crop production they might expect in the near future. Large datasets may be mined for accuracy and previously unknown patterns or information using machine learning techniques. In this study, machine learning (ML) was used to select the optimum crop by considering a range of inputs, including temperature, humidity, pH level, etc.

Keywords—crop, forecasting, temperature, humidity, ph level.

1. INTRODUCTION

In order to estimate crop yields or suggest appropriate crops for a certain location, crop prediction combining Python and machine learning uses historical data, environmental parameters, and machine learning algorithms. Planning crops, allocating resources, and increasing agricultural output are all goals of this process, which tries to support farmers, agricultural professionals, and policymakers in doing so.

An essential component of agriculture that aids farmers in making knowledgeable choices regarding their crops is crop yield prediction. Based on a variety of variables, including soil type, weather, and crop management techniques, it entails determining the quantity of crops that will be produced in a certain area.

2. METHODOLOGY

List of Algorithm used:
1) Logistic Regression
2) KNN (K-NEAREST NEIGHBORS)
3) Decision tree
4) Random forest

1. Logistic Regression.

Strategic relapse gauges the likelihood of an occasion happening, for example, casted a ballot or didn’t cast a ballot, in light of a given dataset of free factors. Since the result is a likelihood, the reliant variable is limited somewhere in the range of 0 and 1.

Advantages of utilizing Calculated Relapse Calculated relapse is a factual technique generally utilized for paired characterization errands, where the objective is to foresee the likelihood that an occasion has a place with a specific class. While it probably won’t be the best option for crop yield expectation, there are circumstances in which calculated relapse can offer specific advantages for crop-related forecast errands.

Logistic Regression For modelling of our data and data set.
1.1 Crop Production CVS

1.2 Probability of crops growing when temperature >30

Here are some predicted crops Probability of growing when temperature is less than 30 (mango holds 21.8 accuracy which is the highest), so it’s clear that which will suits best when temp is less than 30. Data set is taken from from kaggle which is used in this pie chart for prediction.

2: K-nearest neighbors:

We have used K-Nearest Neighbors (KNN) for the classification of data from a dataset. We calculated and classified the inputs obtained from the dataset, and the output obtained from this process is considered as a prediction. To begin, we converted the dataset into a pickle file. We used this data to perform calculations related to soil and environmental factors. For generating maps, we utilized Python code that involved analyzing the dataset using the Pandas library. We also employed the Random Forest classifier for classification tasks. Additionally, we used the Matplotlib library to generate graphs.

2.1 process diagram

2.0.1 KNN Classifiers

A K-Closest Neighbors (KNN) classifier is a straightforward and instinctive AI calculation utilized for characterization undertakings, including crop expectation. With regards to trim expectation, KNN orders a given area’s reasonableness for explicit yields in view of the closeness of that area to realized data of interest with named crop types. Here are the subtleties of carrying out a KNN classifier for crop expectation:

• Information Assortment: Accumulate information that incorporates data about the climate, soil, and authentic harvest yields. This dataset ought to incorporate elements like soil pH, temperature, precipitation, moistness, and some other pertinent natural variables.
• Information Preprocessing: Clean and preprocess the information. This incorporates taking care of missing qualities, scaling elements, and encoding unmitigated factors if vital. • Divide the Information: Separation the dataset into a preparation set and a testing set. The preparation set is utilized to prepare the KNN classifier, and the testing set is utilized to assess its presentation.
• Pick K Worth: Select a proper incentive for the K boundary in KNN. K addresses the quantity of closest neighbors to consider while making expectations. You can explore different avenues regarding various upsides of K to find the one that turns out best for your dataset through strategies like cross-approval. • Train the KNN Model: For every data of interest in the preparation set, register the distances to any remaining information focuses utilizing a distance metric. • Select the K closest neighbors of every information point in light of the figured distances.

• Decide the greater part class (crop type) among these K neighbors, and dole out that class to the data of interest

We use KNN To display final result based on parameters.

3. Decision tree

1. Crop Yield Prediction: Decision trees can be used to predict crop yields based on various factors such as weather conditions (temperature, precipitation, humidity), soil properties (pH, nutrient levels), crop variety, and farming practices (irrigation, fertilization). By analyzing historical data, a decision tree model can provide insights into which factors most strongly influence crop yields.

2. Disease and Pest Detection: Decision trees can help identify the presence of diseases or pests in crops. By training a decision tree on data that includes information about crop health, symptoms, environmental conditions, and pest prevalence, the model can classify whether a crop is likely to be affected by a specific disease or pest.

3. Crop Recommendation: Decision trees can be used to recommend suitable crops for a given piece of land based on soil characteristics, climate, and other factors. The decision tree can guide farmers in selecting crops that are more likely to thrive in their specific conditions, optimizing yield and resource usage.

4. Irrigation Management: Decision trees can assist in determining when and how much irrigation is needed. By considering factors such as soil moisture levels, weather forecasts, and crop water requirements, a decision tree model can help farmers make informed decisions about irrigation scheduling.

In modelling and provide accuracy and take decision which crop is suitable.

4. Random forest

Irregular Backwoods, a strong AI device, assumes a significant part in reforming current horticulture. It investigates crop information with accuracy, improving result in light of the exactness of information data. This cycle starts with extensive information assortment, highlighting crop types, authentic yields, ecological information, and different impacting factors. After fastidious information preprocessing. Arbitrary Woods is prepared and enhanced, giving significant experiences into crop the board. It predicts yields, suggests appropriate harvest types, advances water system rehearses, supports bug and infection the executives, and recommends soil wellbeing enhancements. Standard updates guarantee its transformation to evolving conditions, setting its job as a central member in information driven decision-production for horticulture, eventually prompting further developed crop yields and reasonable cultivating rehearses.

Random forest used to analyse crop for better output based on accuracy and input provided to system.

3. SURVEY

Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta[1] describe The adverse impact of unnatural climatic changes on food production could lead to reduced agricultural yields, economic challenges for farmers, and increased uncertainty in crop forecasting, potentially threatening India's food security and the livelihoods of millions of farmers. Adaptation and mitigation measures are essential to address these concerns and ensure a resilient agricultural sector.

Ashwani kumar Kushwaha, Swetabhattachrya [2] describe The outcome of addressing these challenges in agriculture would likely lead to increased food security, sustainable farming practices, and improved livelihoods for farmers, ultimately contributing to economic development and environmental preservation.

Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K T[3] describe The outcome of this work is the development of a model that utilizes machine learning methods to classify different soil series based on chemical features and geographical attributes. This model not only helps in identifying specific soil characteristics but also suggests suitable crops for each soil series. The result is more informed and optimized crop planning, benefiting farmers and enhancing agricultural productivity.
Rahul Katarya, Ashutosh Raturi, Abhinav Mehndiratta, Abhinav Thapper [4] describe The adverse impact of unnatural climatic changes on food production could lead to reduced agricultural yields, economic challenges for farmers, and increased uncertainty in crop forecasting, potentially threatening India's food security and the livelihoods of millions of farmers. Adaptation and mitigation measures are essential to address these concerns and ensure a resilient agricultural sector.

Pijush Samui, Venkata Ravibabu Mandla, Arun Krishna and Tarun Teja [5] describe By implementing a variety of machine learning algorithms for crop selection and conducting a comparative analysis, this study can help identify the most effective method for optimizing crop choices. The outcome is likely to be a data-driven approach to crop selection that enhances agricultural productivity and resource utilization, ultimately benefiting farmers and the agricultural sector.

Shubham Prabhu, Prem Revandekar, Swami Shirdhankar, Sandip Paygude [6] describe Implementing this system for soil analysis and crop prediction can lead to sustainable agriculture, as it enables farmers to make informed decisions about crop selection, improving yield and resource efficiency. This approach contributes to increased food production and reduced environmental impact, ultimately benefiting both farmers and the agricultural sector.

Pavan Patil, Vireendra Panpatil, Prof. Shrikant Kokate [7] describe The outcome of this study indicates that machine learning techniques, particularly the random forest model, can be highly effective for crop yield prediction. This accuracy in prediction can empower farmers with valuable insights to make informed decisions and optimize their agricultural practices, ultimately leading to increased crop yields and resource management in Indian agriculture.

Nischitha K Assistant Professor [8] describe The outcome of this exploration of machine learning approaches for crop prediction is a better understanding of the diverse methods and technologies available. It highlights the potential for more accurate crop yield predictions and improved farming techniques, with each approach offering distinct advantages and considerations. Farmers can benefit from the evolving landscape of data-driven agriculture to enhance productivity and resource management.

m.champaneri [9] describe The outcome of this project is the creation of a valuable tool for farmers, enabling them to predict crop yields with greater accuracy. This empowers farmers to optimize resource allocation, crop management, and harvest planning, ultimately improving agricultural productivity and decision-making in the farming sector.

SaiNishant1, Pinapa Sai Venkat2, Bollu Lakshmi Avinash3, B. Jabber4 [10] describe The outcome of this research is the successful development of a machine learning-based approach for predicting crop yield in Indian agriculture. The Random Forest model's high R-squared value of 0.86 indicates its effectiveness in providing accurate predictions, which can significantly benefit crop management and planning, ultimately improving agricultural productivity.

J Kashyap1, Dr. Sivakumar V2, Rachel Rose Oommen3, Darshan A3, Bhoomika [11] describe The paper proposes a machine learning-based approach for crop yield prediction using weather, soil, and crop-related parameters. It evaluates multiple algorithms and finds that the random forest algorithm delivers the highest prediction accuracy among them.

Keerthi Yashwanth2, M Sridhar [12] describe The paper suggests a machine learning-based method for crop yield prediction using weather, soil, and crop parameters. After testing various algorithms, it concludes that the random forest algorithm exhibits superior prediction accuracy when applied to USDA data from 2012-2018...

Kuradusenge 1,* , Eric Hitimana 1, Damien Hanyurwimfura 1, PlacideRukundo 2, Kambombo mtone [13] describe The study explored machine learning models for predicting Irish potato and maize yields in Rwanda using weather and soil data. It found that Random Forest and XGBoost models performed best, indicating the potential of machine learning for accurate crop yield predictions, benefiting farmers' decision-making.

Venugopal, Aparna S, Jinsu Mani, Rima Mathew Proff [14] describe The study assesses three machine learning algorithms for predicting rice, wheat, and maize yields using weather, soil, and historical crop data. It finds that Random Forest exhibits the highest accuracy, with 93.5% for rice, 92.8% for wheat, and 89.6% for maize, highlighting its effectiveness in crop yield prediction.

Nermeen Gamal Rezk, Ezz El-Din Hemdan, Abdel-Fattah Attia, AymanEl-Sayed & Mohamed A [15] describe The paper introduces an IoT-based smart farming system and a machine learning method called WPART for predicting crop productivity and drought. The results demonstrate that WPART is highly accurate, achieving up to 98.15% accuracy for crop productivity and 98.04% for drought prediction, surpassing existing algorithms in various metrics, making it a robust tool for informed decision-making in agriculture.

Chowdhury1, Aamir feroze Siddiqui2, SaiTeja Nulla3, Ankit Panda4, Gali Sai Divakar Reddy5 [16] describe The paper examines the use of machine learning for crop prediction, focusing on its benefits, drawbacks, and outcomes. The specific findings and results are not provided in the given text.
K.J. Somaiya [17] describe the paper introduces a machine learning-based approach for crop yield prediction, demonstrating its superiority over traditional regression methods with an accuracy of up to 87% when applied to a real-world wheat yield dataset. This approach has the potential to aid farmers, policymakers, and researchers in making informed decisions about crop management.

Miss Vaishali Patil Pro. Sachin [18] describe the paper presents a Crop Prediction System using Machine Learning for Indian agriculture. It accurately predicts crop yield and recommends suitable crops based on historical climate and soil data, offering valuable decision support for crop selection in specific regions.

Karan Chaudhary Prof. Farhana Kausar Student [19] describe the paper investigates the application of machine learning for crop yield prediction in agriculture and assesses its pros and cons. However, specific findings or outcomes are not provided in the given text.

Ishwarya R*1, Nagapooja BN*2, Raghavi R*3, Soundarya K*4, Prof. Chitra [20] describe the paper introduces a machine learning-based method for crop yield prediction using the Random Forest algorithm, achieving an average prediction accuracy of 85%. This approach holds promise for assisting farmers in making informed decisions to enhance productivity.
4. Analysis & Design

4.1 Flow Chart

Analysis is the process of examining, evaluating, and interpreting data, information, or a complex subject to gain a deeper understanding of its components, structure, or underlying principles. It often involves breaking down the subject into its constituent parts, studying their relationships, and drawing conclusions or making informed judgments based on the findings. Analysis can be applied to various fields, including science, mathematics, economics, literature, and more, to uncover patterns, insights, or insights that may not be immediately apparent. It is a fundamental method for making sense of complex information and making informed decisions.

5. Implementation

| Front-End  | HTML, CSS, JavaScript |
|------------|-----------------------|
| Back-End   | Python                |
| Algorithm  | Machine Learning      |
| Framework  | Flask                 |
| IDE        | Visual Studio Code    |
| Presentation | MS Power Point         |
| Diagram    | MS Visio-2013         |
| Platform   | Windows-10 or Above   |
5.2 Probability of crops growing when temperature > 30

Here are some predicted crops Probability of growing when temperature is less than 30 (mango holds 21.8 accuracy which is the highest) so, it’s clear that which will suits best when temp is less than 30. Data set is taken from from kaggle which is used in this pie chart for prediction.
5.1 Login Page

Here are some predicted crops Probability of growing when rainfall is less than 150mm jute and rice holds 23.5% accuracy which is the highest so, it’s clear that which crop will be best to grow when rainfall is less than 150mm. Data set is taken from kaggle which is used in this pie chart for prediction.

5.3 Probability of crops growing when rainfall>150

5.4 Probability of crops when soil pH>7.5 (Alkaline Nature)

Here are some predicted crops Probability of growing when soil pH is less than 7.5 (chickpea and mothbeans holds 23.8% and 24.1% accuracy which is the highest) so, it’s clear that which crop will suits well when soil pH is less than 7.5 Data set is taken from kaggle which is used in this pie chart for prediction.

5.5 Input

5.6 Final Result

6. Result Discussion

| Nitrogen Level | Temperature (°C) | Phosphorus | Humidity (% Relative) | PH Level | Rainfall Level | Accuracy | Predicted Crop |
|----------------|------------------|------------|-----------------------|----------|----------------|----------|----------------|
| 90             | 3                | 10         | 4                     | 10       | 5              | 100      | Coffee         |
| 20             | 4                | 10         | 5                     | 10       | 5              | 100      | Chickpea       |
| 80             | 6                | 10         | 8                     | 10       | 5              | 100      | Wheat          |
| 100            | 8                | 10         | 10                    | 10       | 5              | 100      | Soybean        |
| 150            | 10               | 10         | 10                    | 10       | 5              | 100      | Tomato         |
| 200            | 12               | 10         | 10                    | 10       | 5              | 100      | Apple          |
| 250            | 14               | 10         | 10                    | 10       | 5              | 100      | Orange         |
6.1 Predicted Outcome by System

Based on nitrogen, Temperature (°C), phosphorus level, humidity, potassium, pH and rainfall level, our system will give a 97.7273% accurate result. Like what crop to grow based on your given parameters of nitrogen, Temperature (°C), phosphorus level, humidity, potassium, pH and rainfall level. Here’s some outcome.

6.2 Average of Accuracy by Predicted Crop

Average of accuracy by predicted crop. Which show us which crop will be suitable grown on soil as per all parameters.

6.3 Sum of Humidity by Predicted Crop

If humidity is 50% mothbeans is suitable crop for soil. Under 50% humidity apple, banana, chickpea, etc crop will be grown.

6.4 Sum of Phosphorus by Predicted Crop

If phosphorous level is close to 150 then chickpea, coffee and kidney beans these crops will be suitable crops to grow. Based on predictions.
6.5 sum of nitrogen level by predicted crop

If nitrogen level is 199 then best suitable crop is coffee and if it silty over 100 the pomegranate is best. Under 100 banana, chickpea, mango will be grown.

6.6 sum of potassium predicted crop

If potassium level is up 400 then suitable crop for soil is grapes and under 200 apple, banana, chickpea, etc. will be grown.

7. Conclusion

In our development and implementation journey so far, we have gained expertise in Machine Learning, Python, and the Flask framework. Our current system operates with a predefined dataset and methods. However, for future work, we plan to empower our system by adding new features. These additions will enhance its functionality and provide additional capabilities to users. The specific details of these features will depend on project requirements. By incorporating real-time data integration, predictive modeling, interactive visualizations, or advanced analytics techniques, our system will become more versatile and adaptable. Our focus is on continuous improvement, innovation, and creating a robust and user-friendly system and we try our best to achieve all this.

8. Future Work

In our current project, we have completed the necessary work and successfully implemented tasks such as uploading the dataset and applying various methods of exploratory data analysis. We have also employed important data cleaning techniques to ensure the quality of the dataset. The results we obtained are based on a predefined dataset. Looking ahead to our next future milestone, we aim to enhance our system by introducing functionality that allows end users to dynamically insert inputs. By incorporating this dynamic aspect, users will be able to obtain accurate results in the form of visualizations. This enhanced system will enable more precise and informed decision-making processes. By empowering end users to interact with the system and provide inputs, we anticipate improved accuracy and relevance in the results generated. This increased flexibility and adaptability will greatly benefit decision-making processes based on the insights derived from the system.
9. Reference

1. Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta “Crop Prediction System using Machine Learning”.

2. Girish L, Gangadhar S, Bharath T R, Balaji K S, Abhishek K T “Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning”.

3. Rahul Katarya, Ashutosh Raturi, Abhinav Mehdiratta, Abhinav Thapper “Impact of Machine Learning Techniques in Precision Agriculture”.

4. Pijush Samui, Venkata Ravibabu Mandla, Arun Krishna and Tarun Teja “Prediction of Rainfall Using Support Vector Machine and Relevance Vector Machine”.

5. Himani Sharma, Sunil Kumar “A Survey on Decision Tree Algorithms of Classification in Data Mining”.

6. Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate “Crop Prediction System using Machine Learning Algorithms”.

7. Crop Prediction using Machine Learning Approaches Nischitha K Assistant Professor, Department of E&C Engineering P.E.S College of Engineering, Mandya India Dhanush Vishwakarma Department of E&C Engineering P.E.S College of Engineering, Mandya India.

8. Crop Yield Prediction based on Indian Agriculture using Machine Learning Potnuru Sai Nishant1, Pinapa Sai Venkat2, Bolu Lakshmi Avinash3, B. Jabber4 1,2,3,4 Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India. 1 sai.nishant8@gmail.com, 2 pinapasaivenkat98@gmail.com, 3 avichowdary999@gmail.com, 4 jabberhi5@gmail.com

9. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize Martin Kuradusenge1,* , Eric Hitimana1, Damien Hanyurwimfura1, Placide Rukundo2, Kambombo Mtonga3, Angeline Mukasine1, Claudette Uwitonze1, Jackson Ngabonziza1 and Angeline Uwamahoro 1
14. Published by, www.ijert.orgNCREIS - 2021 Conference ProceedingsCrop Yield Prediction using Machine Learning
Algorithms
Anakha Venugopal, Aparna S, Jinsu Mani, Rima Mathew, Prof. Vinu Williams
Department of Computer Science and Engineering College of Engineering, Kidangoor
Kottayam, India

15. Published: 04 September 2020 An efficient IoT based smart farming system using machine learning algorithms Neroen Gamal
Rezk, Ezz El-Din Hemdan, Abdel-Fattah
Atta, Ayman El-Sayed & Mohamed A. El-Rashidy
Multimedia Tools and Applications volume 80, pages 773–797 (2021)
Cite this article 2121 Accesses 33 Citations

16. Crop Prediction using Machine Learning Ayanesh
Chowdhury1, Aamir feroze Siddiqui2, Sai Teja Nulla3,
Ankit Panda4, Gali Sai Divakar Reddy5

17. CROP YIELD PREDICTION USING MACHINE LEARNING Mayank Champaneri
Department of Information technology, K.J. Somaiya Institute of Engineering and Information Technology, University of Mumbai, India.

18. Crop Prediction System using Machine Learning Miss Vaishali Patil Pro. Sachin Deshpande
Department of Computer Engineering Vidyalankar Institute of Technology, Mumbai, India
Department of Computer Engineering Vidyalankar Institute of Technology, Mumbai, India

19. PREDICTION OF CROP YIELD USING MACHINE LEARNING Karan Chaudhary Prof. Farhana Kausar
Student, Atria Institute of Technology, Assistant Professor, Atria Institute of Technology Visvesvaraya Technological University, Bangalore Visvesvaraya Technological University

20. CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM Ishwarya R*1, Nagapooja BN*2, Raghavi R*3, Soundarya K*4, Prof. Chitra C*5
*1,2,3,4,5 Department of Information Science And Engineering, Maharaja Institute Of Technology, Mysore, Karnataka, India.