International Journal of Modern Developments in Engineering and Science
Volume 2, Issue 1, January 2023
https://www.ijmdes.com | ISSN (Online): 2583-3138

Crop Yield Prediction for Cultivating Alternative Crops Based on Weather and Soil Conditions Using Machine Learning Algorithm

Akshara R. Nair[1*], Attla Anjali[2], Aditi Kashyap[3], K. Mohammed Afzal Ahmed[4], R. Suma[5]
[1,2,3,4]*Student, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, India*
[5]*Associate Professor, Department of Computer Science and Engineering, T. John Institute of Technology, Bengaluru, India*

***Abstract*: The impacts of climate change in India have severely impacted most crop production over the past two decades. Predicting crop yields before harvest helps policymakers and farmers take appropriate measures for marketing and storage. This project helps farmers to know the yield before planting crops in the field, which helps them make the right decisions. We are trying to solve this problem by building a prototype of an interactive prediction system. Implementation of such a system is done using an easy-to-use web-based graphical user interface and machine learning algorithms. Forecast results are provided to farmers. Therefore, there are various techniques or algorithms for this kind of data analysis in crop forecasting, with the help of which crop yield can be predicted. In this, we have made use of random forest algorithm. Even after analyzing all the issues of weather, temperature, humidity, precipitation, and humidity, there is no right solution or technology to deal with the situation we are facing. India has many opportunities to boost economic growth in the agricultural sector. Data mining can also help predict yields. In general, data mining is the process of analyzing data from different angles and summarizing them into key pieces of information. Random forest is the most popular and powerful supervised machine learning algorithm that can perform both classification and regression tasks. It works by building a large number of decision trees during training and producing an output for the class that is the mode of the class (classification) or the mean prediction of each tree (regression). The second part of the work builds a crop recommendation system that analyzes data including temperature, humidity, pressure, and soil parameters such as N, P, and K.**

***Keywords*: Agricultural, Crop recommendation, Crop yield prediction, Machine Learning, Random Forest, Weather.**

1. Introduction

In India, agricultural production is primarily dependent on weather conditions. In order for farmers to maximize crop production, timely crop productivity forecasting and analysis should be introduced. Yield forecasting is an important agricultural problem. Traditionally, farmers predicted yields based on their experience with the previous year's yields. Therefore, there are various techniques or algorithms for this kind of data analysis in crop forecasting, with the help of which crop yield can be predicted. The use of all these algorithms and the interrelationships between them have expanded the scope of applications and the role of big data analysis techniques in agriculture. Agriculture is gradually deteriorating as new innovative techniques are developed. Because of many of these inventions, people have focused on growing artificial and hybrid products that lead to unhealthy living. Modern people are not aware of growing plants in the right place at the right time. These agricultural techniques also change seasonal climatic conditions for soil, water, and air, leading to food shortages. There is no suitable solution or technology to solve the situation we face. In India, there are different ways to enhance economic growth in the agricultural sector. There are several ways to improve crop yield and quality. Data mining can also help predict crop yields. a) Use machine learning techniques to predict returns. b) Easy-to-use interface is given; and c) improve the accuracy of analysis of various climate parameters (cloudiness, precipitation, temperature).

A. Objectives

Project goals are what we want to achieve by the end of the project. This may include profits or assets or intangible goals such as increased productivity or motivation. Our project goals should be achievable, have a deadline, and be specific and measurable at the end of the project. Project objectives are an important part of project management. There is no way to communicate project goals succinctly before and during the project if there are no project goals. There is also no measurable way to measure the customer's success after the project is complete.

- Term definition has to be learned.
- Access to the latest approaches, methods, and theories.
- Discovering research topics based on existing research.

2. Literature Survey

The numerous studies and research did in the project's field of interest, as well as the results that have previously been published, are shown in a literature survey or literature review in a project report while also taking into account the project's varied constraints and scope.

The following are some components of a literature review

*Corresponding author: akshararnair05@gmail.com

- Currently held, generally recognized theories about the subject.
- General and particular books that have been written on the subject.
- Field research is often conducted from the earliest to the latest findings.
- Any ongoing projects and challenges that are currently being faced. An analysis of the literature outlines the current project-related work. In addition to addressing the issue with the current system, it also provides users with clear instructions on how to handle any current issues.

[1] Using Machine Learning to Recommend Crops Based on Soil Analysis Anguraj Ka, Thiyaneswaran B, Megashree G, Preetha Shri J. G, Navya S, and Jayanthi J. Published: April 5, 2021. By transforming conventional farming to precision farming, new technology can be employed to increase agricultural output. Data analysis and the Internet of Things are examples of recent technologies that are utilized (IoT). The main problem that still has to be solved is when and how to cultivate specific crops. This is possible with the aid of machine learning algorithms, which are found to be a successful way of predicting the ideal crop. Using IOT, the soil metrics, including pH, temperature, humidity, and soil wetness, are collected from the sensors and provided to a graphical user interface (GUI). GUI receives inputs and makes suggestions for appropriate crops The IOT and ML technology significantly aids farmers in making wise decisions.

Benefits:

- The technique of gradient descent is employed. Increased prediction accuracy lowers the loss function.
- Makes use of the Nave Bayes classifier and Random Forest Among its drawbacks is that it doesn't offer yield prediction.
- The model is erroneously trained if the learning rate is high.

[2] Machine Learning-Based Crop Yield Prediction Mayank Champaneri, Darpan Chachpara, Chaitanya Chandvidkar, and Mansing Rathod are the authors. Published on: April 2020. The performance of the majority of India's agricultural crops has been negatively impacted by climate change during the past 20 years. Before harvest, forecasting crop yields would assist farmers and policymakers in making the right marketing and storage decisions. Before growing on the agricultural field, this project will assist the farmers in learning the yield of their crops, which will enable them to make the best choices. By creating a model of an interactive prediction system, it makes an effort to find a solution. the implementation of such a system with a user-friendly machine learning algorithm and web-based graphic user interface will be used. The farmer will have access to the prediction's outcomes. Therefore, there are various ways or algorithms for this type of data analytics in crop prediction, and we can anticipate crop production with the aid of those algorithms. It employs the random forest algorithm There are no suitable solutions or technologies to deal with the scenario we are in, despite the analysis of all these concerns and problems, including weather, temperature, humidity, rainfall, and moisture. In India, there are numerous approaches to boost agricultural economic development. Data mining can be used to forecast crop yield growth. Data mining, in general, is the act of examining data from many angles and putting the key points in a concise statement. The most well-known and effective supervised machine learning algorithm, known as random forest, can perform both classification and regression tasks. It works by building a large number of decision trees during the training phase and producing the output of the class that is the mean prediction (for regression) or mode of the classes (for classification) of the individual trees.

Benefits:

- It uses a classifier from the random forest
- Any user can utilize the user-friendly website created for crop yield prediction by giving the necessary location's climatic data.
- It does not offer crop recommendations, which is a drawback.
- Only 75% or more accuracy.

[3] Machine learning-based recommendation system for intelligent crops Aayush Kumar, Omen Rajendra Pooniwala, Swapnil Chakraborty, and Priyadharshini A. June 20, 2021. A role for agricultural crucial part in India's socioeconomic structure. For a nation where almost 58 percent of the people are engaged in farming, the failure of farmers to select the crop that is best suited for the land using conventional and non-scientific methods is a severe problem. Based on soil characteristics, sowing season, and geographic location, farmers occasionally failed to select the appropriate crops. As a result, people commit suicide, give up farming, and relocate to cities in search of work. This research effort has suggested a mechanism to help farmers choose crops by taking into account all the elements including sowing season, soil, and geographic location to address this problem. Additionally, precision agriculture is being used using contemporary agricultural technology and development in underdeveloped nations that focus on managing crops specifically for a certain site

The crops for the underlying soil series were suggested using five distinct algorithms.

- Naive Bayes, Adaboost, Bagged Tree, Support Vector Machine, and Artificial Neural Network are employed.

Cons:

- Not very accurate.
- Crop recommendations are only made after inspecting the soil, not after examining the weather.

[4] Using Machine Learning Techniques to Classify Soil and Crop Suggestion Author: A. Mythili and N. Saranya, February 2, 2020. The primary means of subsistence for the people of India is agriculture. The country's main economic engine is agricultural research. The quality of the soil is crucial to agriculture. India has a wide variety of soil. To foretell the sort of crop things can be grown in that specific soil type, we need to be aware of its features and characteristics. In this situation, machine learning techniques offer a flexible solution. It is very

helpful for farmers to be able to forecast which crops may be grown in a given type of soil by classifying the soil according to the soil nutrients. The use of data mining and machine learning in horticulture and agriculture is still in its infancy. In this essay, we present a system for categorizing soils based on their macro- and micronutrient contents and for forecasting the kinds of crops that can be grown there. Various machine learning methods are employed, including support vector machines (SVMs), bagged trees, and K-Nearest Neighbor (k-NN) Regression and Support Vector (SVM).

The sensors detect soil fertility and other in-soil minerals. • The current models work with old data that is currently accessible.

Cons:

- Farmers may find it challenging to comprehend.
- Based on the soil and weather, classification is the key issue.

[5] Machine learning agricultural yield prediction: Sanjay M. D., Rashmi K. T., Lohit V. K., and Vijayalakshmi Brunda, April 20, 2022. This paper makes yield predictions for all the different crops cultivated in India. This script is special since it predicts agricultural production for any year the user chooses using straightforward variables like state, district, season, and area. Regression methods including Kernel Ridge, Lasso, and ENet algorithms are used in this document.to forecast results, as well as to enhance algorithms through the use of regression stacking.

Advantages:

- It forecasts agricultural production for any year the user chooses using straightforward criteria like state, district, season, and region.
- To forecast output, the document uses regression techniques such as Kernel Ridge, Lasso, and ENet algorithms. It also employs the Regression Stacking concept to improve the algorithm.

Disadvantages:

- This model does not provide accurate statistics or suggest which season is best for which crop.

3. Methodology

A crucial component of every machine learning system is data. We choose to concentrate on the Indian state of Karnataka for implementing the system. Data collected at the district level were important since local climates vary. To put the system into place, historical information on the crops and climate of a certain area was required. This information was taken from many official websites. The information on the crops grown in each Maharashtra district was obtained from www.data.gov.in, and the information on the climate was obtained from www.imd.gov.in. Precipitation, temperature, cloud cover, vapor pressure, and the frequency of rainy days are the climatic factors that have the greatest impact on the crop. Therefore, information on these meteorological variables was acquired monthly. Collection of Datasets: During this stage, we gather data from multiple sources and create datasets. And analytics are being used with the provided dataset (descriptive and diagnostic). There are many online sources for abstracts, including Data.gov.in and indiastat.org. The annual abstracts of a crop will be used for at least ten years. These datasets typically permit time series with anarchic behavior. The primary and necessary abstracts were combined. Global and Regional Crop Yield Predictions Using Random Forests.

Partitioning data: The entire dataset is divided into two portions; for instance, let's say that 75% of the data is utilized to train the model and 25% is reserved for testing the model. To anticipate upcoming events Machine Learning Techniques

supervised education: Algorithms for supervised machine learning can Apply prior knowledge to fresh data by using tagged examples. The system may provide targets for any new input after sufficient training. The learning algorithm may also distinguish between its outputs and the correct, intended output and detect mistakes to adjust the model appropriately.

Contrarily, unsupervised machine learning techniques are employed when the data used to train is neither labeled nor categorized. Unsupervised learning examines how systems can extrapolate a function from unlabeled data to describe a hidden structure. The system does not determine the correct output in order to rise hidden structures from unlabeled data, but it studies the data and can make conclusions from datasets.

Random Forest is the most effective, well-known, and potent supervised machine learning algorithm that can perform both classification and regression tasks build a large number of decision trees during training, and produces outputs of the class that represent the mean prediction (for regression) or mode of the classes (for classification) of the individual trees. The prediction is more reliable the more trees there are in a forest.

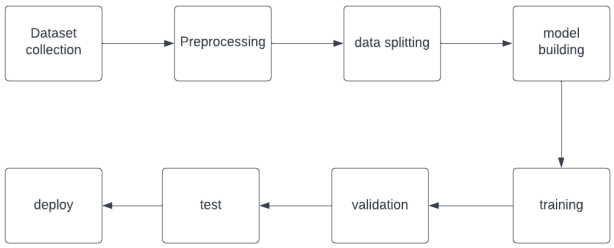

Fig. 1. Flow chart

4. Motivation

Farmers in India are facing so many problems and, in the end, they are going into loss. it is a social problem that we can solve or reduce the loss by using Technology. we trying to build a machine learning model to advise farmers to grow better crops in particular soil and who will inform what will be the yield from their field in the end. ML technology can solve this problem so we thought of working on this project.

5. Problem Statement

To accurately anticipate the crop output, we need to create a machine learning model based on features such as the state, district, soil type, N, P, and K, temperature, humidity, and pressure contrasting different machine learning models. We must create a system of recommendations that can suggest an

improved crop for a specific field that will provide the highest possible yield. The focus of the problem statement is on utilizing machine learning techniques to forecast agricultural yield. The project's objective is to assist users in selecting the best crop to produce to maximize output and, consequently, profit. The suggested system makes predictions by evaluating structured data to get around the limitations of current systems.

The approach we suggest is to create a system that takes into account the factors that have the greatest influence on how well a crop grows and to increase the variety of crops that may be cultivated throughout the season. This would make it easier for farmers to choose crops that will provide a high yield and hence optimize income, which in turn will lower the rate of suicide. The two major modules in this system are:

i. The yield prediction module gives the user two choices: they can either select a specific crop and get the yield for it or they may view the top 5 crops with the highest yield out of all the crops.
ii. The fertilizer module aids the user in determining if a specific period is appropriate for applying fertilizers.

A. Current System

Systems now in use have numerous restrictions. Few crop projections in the past have been more than 30% accurate. This simply provides the area and not the actual yield once the crop has been identified and the area has been calculated. The yields will vary based on several variables, such as the plant's health and the weather. Critical analysis, investigations into research gaps, recommendations, and future directions are lacking. The focus of current methods, which avoid other characteristics like soil nutrients, is on machine learning algorithms for predicting crop production based exclusively on meteorological parameters.

Most systems don't provide a critical assessment of the current machine learning techniques. Additionally, there is no investigation of the features that influence the crop yield forecast model.

Disadvantages:

- No thorough preparation of the data.
- No mathematics was used.
- In-depth graphical examination, no.

B. Suggested System

Numerous elements are included, such as meteorological factors, soil parameters, and topographical data. Together, we are constructing a crop yield and recommendation system. To gain useful insights into the data, we will analyze it using plotly, matplotlib, and seaborn. We'll use correlation to determine the chances of feature choice. In order to minimize overfitting, we remove unnecessary data using outlier identification using a z-score.

- Provides extremely accurate crop yield prediction models.
- Provides farmers with crop-growing recommendations depending on soil and weather conditions.
- Developing a crop recommendation system that would involve assessing data pertaining to temperature, humidity, pressure, and soil factors including N, P, and K.
- Including information about the soil in the system is beneficial because it influences the choice of crops.
- In order to anticipate the yield, it considers the state name, district name, year, season, crop, area, temperature, wind speed, pressure, and humidity.
- Recommending the ideal crop depending on the availability of nitrogen, phosphorus, and potassium as well as temperature, humidity, ph, and rainfall.

6. Conclusion

This strategy is suggested as a solution to the rising number of farmer suicides and to aid in their financial development. The Crop Recommender system aids farmers in making decisions regarding which crop to cultivate as well as yield predictions for certain crops. Additionally, it instructs the user when to apply the fertilizer. Machine learning tools were used to gather, analyze, and train relevant datasets. Based on the user's location, the system retrieves the necessary data from the backend. As a result, only a few details, such as region and soil type are required from the user. The current study demonstrated the possible use of data mining approaches in agricultural production prediction based on meteorological input factors. The user-friendly website that is being developed will provide predictions that are more accurate than 90% for all the crops and study districts that were chosen. Any user can utilize the user-friendly website designed for crop yield prediction by submitting climate data for their preferred crop. Additionally, we will be including a suggestion for improved crop forecasts based on local soil and weather conditions.

References

[1] P. Priya and U. Muthiah, M. Balamurugan, "Prediction of crop yields using the machine learning method," International Journal of Engineering Science Research Technology.
[2] J. Resop, J. Jeong, N. Mueller, “Global and regional crop yield prediction Using Random forests,” PLoS ONE Journal.
[3] Govindarajan Muthukumarasamy and Narayanan Balkrishnan, “Crop production Ensemble Using Machine Learning model for prediction,” International Journal of Computer Science and Software Engineering.
[4] S. Veenadhari, Bharat Misra, and C. D. Singh, “Forecast crop yields based on climate factors Using machine learning,” International Conference on Informatics and Computer Communication.
[5] Prajakta V. Tawale and Shweta K. Shahane, “Crop cultivation prediction,” International Journal of Advanced Research in Computer Science and Electronics Engineering, vol. 5, no. 10, Oct. 2016.
[6] D. Ramesh and B. Vishnu Vardhan, “Utilizing Data Mining Techniques and analyse crop yield prediction,” International Journal of Research in Engineering and Technology.
[7] Gour Hari Santra, Subhadra Mishra, and Debahuti Mishra, “Applications of Machine Learning in the Production of Agricultural Crops,” Indian Journal of Science and Technology, vol. 9, no. 38, Oct. 2016.
[8] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson ID and Dionysis Bochtis, “Agriculture using machine learning,” Lincoln Institute for Agri-Food Technology (LIAT), University of Lincoln’s, Brayford Way, Brayford Pool, Lincoln, UK.
[9] Baisali Ghosh, “An investigation of crop insurance yield utilising precision farming from an aerial platform,” Symbiosis Institute of Geoinformatics Symbiosis International University.
[10] Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler,Soo-Hyung Kim, “Global and

Regional Crop Yield Predictions Using Random Forests," United States of America, University of Minnesota, Institute on the Environment.
[11] Ecochem Online, (2009). Crop yields and soil health. http://ecochem.com/healthysoil.html
[12] Food and Agriculture Organization (2006). The state of Agricultural Commodity Markets, 37–39.
[13] Aditya Shastry, H. A. Sanjay, and E. Bhanushree, "Crop yields Prediction using Regression Technique," International Journal of Computing, vol. 12, no. 2, pp. 96–102, 2017.
[14] E. Manjula and S. Djodiltachoumy, "A Model for Prediction of Crop Yield," International Journal of Computational Intelligence and Informatics, vol. 6, no. 4, March 2017.
[15] S. Nishanthini, D. Santhiya, and K. R. Sri Preetha, "Crop Yield Prediction," International Journal of Engineering, Technology, and Sciences, vol. 3, no. 3, March 2016.
[16] Jharna Majumdar, Sneha Naraseeyappa, and Shilpa Ankalaki, "Analysis of farm data using data mining techniques: use of big data," J. Big Data, 4:20, 2017.
[17] D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015.
[18] N. G. Yethiraj, "Applying data mining techniques in the field of Agriculture and allied sciences", 2012.
[19] Zelu Zia (2009), "An Expert System Based on Spatial Data Mining using Decision Tree for Agriculture Land Grading," Second International Conference on Intelligent Computation Technology and Automation, Oct. 10, China.