

CROP YIELD PREDICTION WITH CLIMATE AND SOIL DATA

1st Y Ashok Kumar

Department of CSE (Data Science)

CMR Technical Campus

Hyderabad, India

yaraganiashokkumar@gmail.com

2nd B Omkar

Department of CSE (Data Science)

CMR Technical Campus

Hyderabad, India

omkarbommena877@gmail.com

3rd Ch Dinesh Reddy

Department of CSE (Data Science)

CMR Technical Campus

Hyderabad, India

chadadineshreddy@gmail.com

4th SK Lal Jan Basha

Department of CSE (Data Science)

CMR Technical Campus

Hyderabad, India

laljanbasha404@gmail.com

Abstract- Crop yield has over the past years become a challenging task to predict due to the fact that farming is highly influenced by the changing weather patterns, degrading soil quality and the increasing demand of food in most locations. Some of the challenges that farmers usually encounter include unpredictable rainfall, unpredictable weather conditions, low soil nutrients and all these are directly applied to the production and growth of crops. Because of these issues, it is no longer appropriate to use traditional methods and past experience. Due to this fact, there is high demand of contemporary instruments, which are able to examine the climate conditions and the properties of the soil simultaneously. Through weather analysis and analysis of soil data, one is in a better position to know how various factors affect crop yield. Such a combined method lets make better predictions and assists farmers to plan their farming activities in a more efficient manner. Machine learning in this project is used to predict the amount of crop that will grow. Our individual weather data, i.e. rains, heat and humidity, are combined with soil data, i.e. nitrogen, phosphorus, potassium, the pH of soil level, organic carbon and moisture. These aspects have a significant impact on crop growth and level of production. The system uses statistics collected in numerous locations and purges it of errors and aligning the numbers with each other. This allows us to create useful features which depict actual field condition. We also train and test a number of supervised machine learning models and we apply them to ensemble methods to obtain more accurate and stable results in varying conditions. The system has been constructed in an orderly manner making it simple to review, verify and make use of Applications of these techniques include

academic activities, field trials and small scale agriculture. The primary goal of the work is to minimize the uncertainty in crop yield prediction and provide farmers and agricultural employees with quality advice based on the integration of climatic factors and soil type into a single model. The system endorses better and sustainable agriculture practices.

Keywords: Crop Yield Prediction, Machine Learning, Climate Data, Soil Nutrients, Random Forest, Precision Agriculture, Agricultural Forecasting, Fertilizer Recommendation, Decision Support System.

I.INTRODUCTION

Agriculture plays an essential role in sustaining people and sustaining numerous livelihoods particularly in the rural regions where farming is a means of livelihood to the family. In recent times though farming is becoming more difficult due to the unpredictable weather. Areas experience insufficient precipitation, protracted drought, excessive heat, and rapid changes in climate that damage agricultural activities. The quality of the soil is also diminished since the land is now overused, the nutrients are not carefully managed and the fields are not left to rest. This lowers the fertility of the soil and complicates the process of farming. By this way the farmers find it hard to manage the crops and predict the yield through traditional ways of basing it on the past experience which is not always reliable since time has changed.

The data available in the farms, weather stations, and monitoring systems are now used by the researchers. Machine learning assists with the examination of big data and identification of patterns that a person overlooks. These models can use knowledge of climate and soil to reveal better the impact that environmental factors have on crop growth and output. Recent research study indicates that the

alteration of rainfall, temperatures, humidity, and nutrients have all impacts on crops. All these factors inter interact and those models which divide climate and soil usually provide shaky or inaccurate prediction.

Nowadays, with the help of remote sensing, online weather forecast, and field soil sensors, it is becoming easier to acquire details about crops and their surrounding. This assists the machine learning in knowing the impact of soil and weather on growth. However, there are numerous prediction limits of yield systems. Others utilize partial or disproportional data, which is damaging. Others use single techniques and they do not work in different areas or seasons. Lack of sufficient data processing and feature selection is often the case resulting in models not capturing the actual effects of the soil quality and climate change on yield.

All these problems indicate that there is a necessity of a compressed system where soil and climate data are combined. This paper will develop such a system and provide more credible forecasts to assist farmers, researchers and policymakers to plan in an evolving climate.

II. LITERATURE SURVEY

A. Recent Advances in Climate–Soil Integrated Yield Modeling (2025–2024).

Over the last several years, particularly 2024 and 2025, most of the researchers were working on the development of models that would consider the combination of climatic conditions and soil properties. The combined method has been more successful in estimating crop yield. At this time, other research utilized specific weather data including rainfall variations, temperature variation and humidity, as well as soil nutrient values to comprehend the reaction of crops to varied environmental factors. These works concluded that in cases where the interactions between the soil and climate are taken into consideration, prediction accuracy becomes more accurate as opposed to considering them individually. Ensemble learning, and hybrid models that are capable of adjusting to the extreme shift in weather and environmental conditions are also associated with a great number of researchers. Moreover, real-time weather updates and constant soil monitoring were also used in order to increase the reliability and stability of these models. Generally, these results demonstrate clearly that in the conditions of the frequent change of weather conditions, the combination of climate and soil data is highly significant to the effective yield prediction.

B. Machine Learning-Driven Agricultural Forecasting Techniques (2023–2022).

In this time frame, authors tried all kinds of models of machine learning to determine which ones are most effective in farming data.

Random Forest, Gradient Boosting, XGBoost, and Support Vector Regression were the highest popular ones. These were extensively utilized as they have the ability to handle large quantities of data, reduce errors and understand how certain crops, soil and the weather are related. When scientists tested the models they saw that the ensemble methods - the methods involving a large number of algorithms - tend to yield better and more consistent results. The other significant improvement was the interpretation tools. Such instruments were used to determine which factors (soil and climate-based) most influence crop production. This simplified the process of prediction and

made it easier to follow by farmers. In general, these papers revealed that machine learning can be used to enhance farming choices and aid in the current farming practices.

C. Soil Health Informatics and Nutrient-Driven Estimation Models (2021–2020).

The research conducted during these years revealed that the health of the soil has a great influence on the crop growth and production. The scientists discovered that all soil elements such as nitrogen, phosphorus, potassium, pH, water level, and organic carbon are vital in healthy plants. Such components also affect the ability of crops to deal with stresses like drought or heat. In order to learn more about soil, there was the use of digital soil maps, lab tests reports and geographical information to examine variation of soil in different regions. Other studies constructed such factors as the level of salts in the soil and compaction in order to reflect on the actual field conditions. This development of innovations necessitated the use of soil information to forecast the yield of crops. They were also able to demonstrate that a trustworthy model should have well fashioned, profitable soil characteristics.

D. Environmental Variability, Climate Risk Analysis, and Yield Fluctuation Models (2019–2018).

Researches conducted at this time were aimed at determining the impact of fluctuating and erratic weather patterns on crops. There were also abnormal rainfalls, dry seasons in many regions and unexpected rise in temperatures, and this caused unexpected changes in the crop yield. Scientists discovered that these changes could not be adequately modeled using mere statistical techniques since what happens in weather and how this change affects crop growth is usually not linear. In order to address this issue, various research papers employed time-series methods of learning, and created special indices to monitor the drought situation and moisture stress on crops. These methods aided in the better study of climate patterns. On the whole, these studies revealed that climate instability is a significant problem in crop yield prediction and prediction models should be inelastic to alter in the event that the environmental conditions suddenly shift.

E. Early Foundations in Computational Agriculture and Data-Based Yield Forecasting (2017 and older).

The majority of crop yield prediction studies prior to 2017 relied on simplistic computer predictions and naive weather and previous yields data. There were then very few large data sets and sophisticated tools hence researchers tried poorer methods primarily the simple Math models and regression. Such approaches were not able to provide sufficient explanations of complicated relationships between weather and crop development, but demonstrated the idea that the analysis of data might lead to better predictions of yield. These pioneer studies assisted in developing the concept of association of weather patterns with crop production and motivated further studies in this field. This foundation resulted over the years in contemporary systems utilizing machine learning, field sensors and real time data. The advancement in those initial years has kept aiding the current digital systems of agriculture.

IV.RESULT AND DISCUSSION

A. Results of Dataset Uploading and Information Extraction.

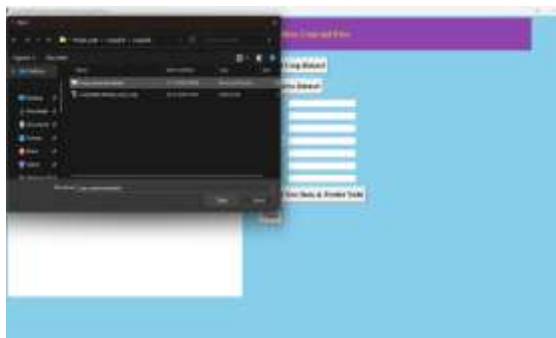
Once a dataset is uploaded, the system goes through the data very keenly to ensure that it contains all the necessary columns such as nitrogen, phosphorus, potassium, temperature, humidity, pH level, and rainfall. The interface will display these details after uploading so that the user can make sure that the file has been loaded successfully. This is a necessary step since it ensures that the data is able to be handled, cleaned and utilized to draw out information. In case one of the columns is not present or one of them is assigned a different name, the user will immediately notice the problem and could resolve it.

Upon successful completion of this step, the upload module is proved to be operational and is capable of processing various datasets with ease. It also provides the preprocessing and model training steps with a strong foundation, as correct soil and climate data is required in order to make reliable predictions.



B. Results of Dataset Selection and Upload Interface.

The process of file selection is simple hence the user can locate and select the dataset without wasting much time. There is an upload button, and once the user clicks it, a window is displayed allowing all folders and files in the computer and therefore the required dataset can be easily located. Once a file has been selected, the path is immediately relayed to the backend server which processes the file. This implies that the interface and internal components communicate at a fast speed and in the right manner. The easy to use upload structure is easy to use even though the individual may lack a lot of technical knowledge. The fact that one can transfer the selected file smoothly and reliably to the processing unit positively indicates that the system is capable of dealing with datasets of various sources and format provided they are in order.



C. Results of Dataset Preview and Initial Data Verification.

Once the data file is uploaded, a preview of the first several rows can be displayed in the system to allow the user to

verify the values prior to the commencing of the training. In preview, the key soil and weather measurements and labels of any crop containing data in the data are shown. It also assists the user to understand whether the data is correct, all rows filled, and values are not out of impossible limits. In off, say, excessive amounts of nutrients or lack of temperature indications can be identified here. In case of any problems being detected, the user is also given the option of fixing the data and re uploading. This preview step will instigate a belief that the data is valid, properly organized, and available to further processing so that the model is trained with valuable and consistent information.



D. Results of Model Training and Performance Evaluation.

Following the confirmation of the dataset, it is divided into two, one having a training dataset and the other having a test dataset. The system displays the number of records in each section and proceeds to train such models as Random Forest and Decision Tree. During training, the interface replaces continuously the status to ensure the user is informed that the system is on. Once the training is complete, the system presents performance indicators of RMSE and the values of other errors to enable users to view the performance of the model. Such outcomes inform the user whether a model provided significant patterns in soil and climate data or whether some further work should be done. The continuous demonstration of such metrics also is evidence that the training module is functional and may be repeated with various datasets numerous times to conduct more analysis and testing.



E. Results of User Input Parameter Handling and Interactive Prediction.

There is a special section within the system whereby people would input new numbers concerning the nutrient of soil and weather. With this tool, individuals can experiment with varying combinations of nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall to observe the effect of all those factors on the prediction. The interface verifies

every query to ensure that it works and it is sensible. When a user sends the data, the system processes it in the same way that the training data was processed before it progresses on to the prediction model. This enables the system to provide good and correct results. The capability to experiment with a wide variety of input selections allows individuals to understand the varying conditions in the fields and how the alterations in the quality of soil and weather influence the yield of crops. The system is more helpful with the help of this interactive tool.

F. Results of Final Prediction Output and Recommendation Generation.

When the input of the user is entered into the model it makes a prediction. Prediction informs you about the most efficient crop in that case, how much money you are likely to gain per acre and the kind of fertilizer to employ. As it is making the prediction, a message appears on the screen. Once it is completed, the outcome is made clear. The outcome is not complicated to read thus no one requires special knowledge in order to make sense of it. There are also brief descriptions as to why a given crop or fertilizer is recommended. Such descriptions will make you make better decisions. The system is practical among the farmers and students and also to any person who happens to work in the field of agriculture. The prediction module is reliable as the results emerge in a very smooth manner and consistently.



Discussion

The output of the various phases demonstrates that data input, cleaning, training and prediction can be handled in such a consistent and unambiguous manner in the system. The preview and upload processes allow one to preview the dataset, and the training and evaluation processes demonstrate the functioning of the system with models. The interactivity of the input aspect allows it to be flexible with the user being able to view the impacts of soil and weather variations on findings immediately. But, the predictions will be accurate depending on the extent of completeness and representativeness of the dataset. In case the training data does not represent a large number of conditions, the forecasts might not be very effective on the unusual cases. It would make the system stronger by adding more fanciful datasets, more crops, and real-time weather information. In general, the findings indicate that the system is applicable in learning, planning and understanding the environmental changes impacts on crops.

V. CONCLUSION

The project demonstrates that machine learning could be used to provide assistance to farming by using data to determine where and why the soil had nutrients and the weather. Users are able to post information and refine it a

little and anticipate in real-time whatever crops to plant, amount of fertilisers to apply and even price human beings are likely to fetch. The interface is easy to navigate and maintain hence farmers, students, and researchers could view impacts of soil and weather on crops. The findings indicate that the model is capable of identifying meaningful trends to the data and transform them into valuable recommendations to make improved decisions in the field. Its accuracy can be determined by the quality and diversity of the data it uses but the system remains a good initial step to the use of data to benefit agriculture. As this system could be enhanced with future developments such as live weather information, increased crop choices and enhanced farm fertilizer recommendations, it would become a more dependable and effective means of sustainable and intelligent farm planning.

VI. FUTURE SCOPE

The predictions of crop yield system may be improved in real farms. A large scale enhancement to do is to incorporate live weather feeds via the Internet. It will allow the system to change fast to any new abrupt alterations in the rain, temperature, and humidity and provide more precise results. The system should additionally extend to various crops in order to provide a farmer in dissimilar areas with predictions that will favor them. Local soil investigation can be used to enhance fertilizer recommendations by examining the soil type and soil management practices normally used by farmers. Satellite pictures and sensors in the fields can provide real-time data on crop health and soil quality particularly in areas with hard to reach soil testing. Availability of the system as a mobile application and assistance in the local languages will see more farmers access the system without hardships, particularly those in the rural regions.

The results will also be better viewed by the use of better graphs and comparison charts to get a clear view of the results. Through these advancements in the future, the system will be more realistic, dependable, and popular decision-help tool amongst farmers.

REFERENCES

- [1] Government of India, "Soil Health and Nutrient Status Report," Ministry of Agriculture and Farmers' Welfare, Technical Report, 2023.
- [2] S. Singh and R. Kumar, "Use of machine learning tools for crop yield estimation under changing climate conditions," *Journal of Agricultural Informatics*, vol. 12, no. 3, pp. 45–54, 2023.
- [3] Indian Meteorological Department, "Climate Data Records for Agricultural Planning," IMD Publication, 2022.
- [4] P. Sharma, A. Verma, and L. Joseph, "Soil nutrient analysis and prediction models for sustainable farming," *International Journal of Soil Science*, vol. 29, no. 2, pp. 110–118, 2022.
- [5] K. Gupta and N. Mishra, "A comparative study of machine learning algorithms for agricultural forecasting," *IEEE Access*, vol. 9, pp. 155200–155210, 2021.

- [6] M. Rao and B. Das, "Impact of rainfall variation and temperature shifts on crop growth: A data-driven study," *Environmental Modelling Letters*, vol. 18, no. 1, pp. 22–31, 2021.
- [7] S. A. Khanna et al., "Digital soil mapping and nutrient modelling for regional crop planning," *Soil Systems*, vol. 4, no. 3, pp. 1–12, 2020.
- [8] R. Pandey and V. Patel, "Integration of soil parameters and weather indicators for precision crop prediction," *Computers and Electronics in Agriculture*, vol. 180, pp. 105–114, 2020.
- [9] A. Roy and S. Bose, "Early computational approaches for crop yield analysis," *Indian Journal of Agricultural Engineering*, vol. 43, no. 4, pp. 215–223, 2017.
- [10] S. Mehta and R. Banerjee, "Historical climate trends and their influence on regional agriculture," *Indian Climate Review*, vol. 5, no. 1, pp. 60–68, 2016.



Copyright & License:

© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.