International Journal of Engineering Research and Science & Technology

IJERST

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed,Referred & Indexed Journal

Research Paper

CROP YIELD PRODUCTION PREDICTION USING MACHINE LEARNING

[1]D Sreedar, [2] R Ramya,[3] D Shiva , [4] M Karunakar
[1]AssistantProfessor, [234]Students
Department of Computer Science and Technology
Siddhartha Institute of Technology & Sciences, Narapally
sreedhargoud@siddhartha.org.in, 23TQ1A05J0@siddhartha.co.in, 23TQ1A05D0@siddhartha.co.in, 23TQ1A0518@siddhartha.co.in,

Abstract

Agriculture plays a vital role in the economic development of a nation, yet many agricultural sectors still face challenges due to limited adoption of advanced technologies and inefficient resource management. These issues often lead to reduced crop productivity, directly impacting the agricultural economy. Accurate prediction of crop yield can help farmers and policymakers make informed decisions to improve productivity and optimize resource utilization.

This project presents a machine learning-based approach for crop yield production prediction using historical agricultural data. The system analyzes various input parameters such as district name, season, year, and crop type to estimate crop yield. Data preprocessing techniques, including cleaning, normalization, and feature selection, are applied to enhance the quality of the dataset and improve model performance.

Multiple machine learning regression algorithms are implemented and compared to identify the most accurate model for yield prediction. Performance evaluation is carried out using metrics such as Mean Absolute Error (MAE) and R^2-score. The system is capable of predicting the yield for a specific crop based on given inputs, as well as estimating yields for multiple crops under different conditions.

I. Introduction

Agriculture is one of the primary sources of income in many developing countries and plays a crucial role in supporting livelihoods and economic growth. In recent years, the agricultural sector has undergone significant transformation due to advancements in technology, changing environmental conditions, and improved farming practices. However, farmers still face challenges in making informed decisions regarding crop selection and yield optimization due to the complexity of factors such as soil conditions, weather patterns, and seasonal variations.

With the increasing availability of agricultural data, the use of data mining and machine learning techniques has become essential for improving decision-making in farming. Data mining involves extracting meaningful and valuable information from large datasets, enabling better analysis and prediction. In agriculture, diverse datasets such as soil data, crop data, and weather data can be effectively analyzed using machine learning approaches to predict crop yield and enhance productivity.

Crop yield prediction using machine learning provides an efficient way to monitor and estimate agricultural output based on various influencing factors. By analyzing historical data and identifying patterns, machine learning models can assist farmers in selecting suitable crops and optimizing farming strategies. This approach reduces dependency on traditional methods and expert opinions, offering faster and more accurate predictions.

II. Literature Survey

Agricultural research has increasingly focused on improving crop yield prediction using advanced technologies such as remote sensing, data mining, and machine learning. Earlier studies emphasized the importance of soil properties and environmental factors in determining crop productivity. For instance, research on soil organic matter estimation using reflectance spectroscopy demonstrated that regression models such as Partial Least Squares Regression (PLSR) can effectively predict soil characteristics with good accuracy. These approaches provided a cost-effective and non-destructive method for analyzing soil conditions, which play a crucial role in crop growth.

Further studies explored the use of remote sensing and crop growth models for estimating soil nutrients and crop performance. The integration of time-series remote sensing data with models like WOFOST improved the accuracy and stability of soil nutrient simulation and crop growth prediction. These methods addressed the limitations of traditional statistical models by incorporating dynamic environmental factors. Similarly, research on heavy metal stress in crops used spectral indices and physiological data to assess crop health, highlighting the importance of environmental monitoring in agriculture.

Advancements in sensor technology also led to the development of innovative tools such as capacitive soil moisture sensors, which enable accurate measurement of soil water content. These technologies contribute to better decision-making in irrigation and crop management. In addition, studies on crop yield estimation using remote sensing data, such as MODIS-based models, demonstrated that regression-based approaches can accurately forecast crop production with minimal error, even before harvest. These methods are particularly useful in regions with limited ground data.

In recent years, machine learning techniques have gained significant attention in crop yield prediction. Research has shown that algorithms such as Random Forest, Extremely Randomized Trees, and Deep Learning models can effectively analyze large agricultural datasets, including climate data, soil properties, and vegetation indices. Comparative studies indicate that deep learning models often provide higher prediction accuracy, while ensemble methods offer robustness and reliability. These approaches have proven to be effective in handling complex and nonlinear relationships in agricultural data.

Overall, the literature highlights that combining remote sensing, sensor data, and machine learning techniques provides a powerful approach for accurate crop yield prediction. These studies form a strong foundation for developing intelligent systems that can assist farmers and policymakers in improving agricultural productivity and ensuring food security.

III. System Analysis

System analysis involves studying the current agricultural practices and identifying the need for an efficient crop yield prediction system. Traditional farming methods depend heavily on experience and environmental assumptions, which may not always provide accurate results. With the availability of large agricultural datasets, there is a need to analyze soil, weather, and crop-related data effectively. The proposed system uses machine learning techniques to process and analyze this data. It focuses on improving prediction accuracy by considering multiple influencing factors. The system ensures better decision-making support for farmers. It also reduces dependency on manual analysis. The analysis highlights the importance of automation in agriculture. Overall, it aims to develop a reliable and data-driven solution for crop yield prediction.

Existing System

The existing system for crop yield prediction mainly relies on traditional farming knowledge and basic statistical methods. Farmers often depend on past experiences and general climate conditions to estimate crop yield. Some systems use simple regression models, but they lack the ability to handle complex datasets. Government agencies provide reports based on limited data collected from specific regions. These systems do not use advanced technologies for prediction. They often fail to consider multiple factors like soil nutrients, weather conditions, and seasonal variations together. The prediction process is usually slow and less accurate. There is also limited use of real-time data. Overall, the existing system lacks efficiency and scalability.

Disadvantages of Existing System

- Low prediction accuracy
- Dependence on manual analysis and experience
- Inability to handle large and complex datasets
- Limited use of advanced technologies
- Lack of real-time data processing
- High chances of human error
- Poor scalability and adaptability
- Limited integration of multiple factors (soil, weather, crop data)
- Time-consuming process
- Inefficient decision-making support

Proposed System

The proposed system uses machine learning techniques to predict crop yield based on various input parameters. It collects agricultural data such as soil type, weather conditions, crop type, district, season, and year. The system preprocesses the data by cleaning, normalizing, and selecting relevant features. Machine learning algorithms such as Linear Regression, Random Forest, and Decision Trees are applied for prediction. The system can predict yield for a specific crop or multiple crops. It provides faster and more accurate results compared to traditional methods. The system is designed to handle large datasets efficiently. It can also be integrated with real-time data sources in the future. Overall, it offers a smart and automated solution for agriculture.

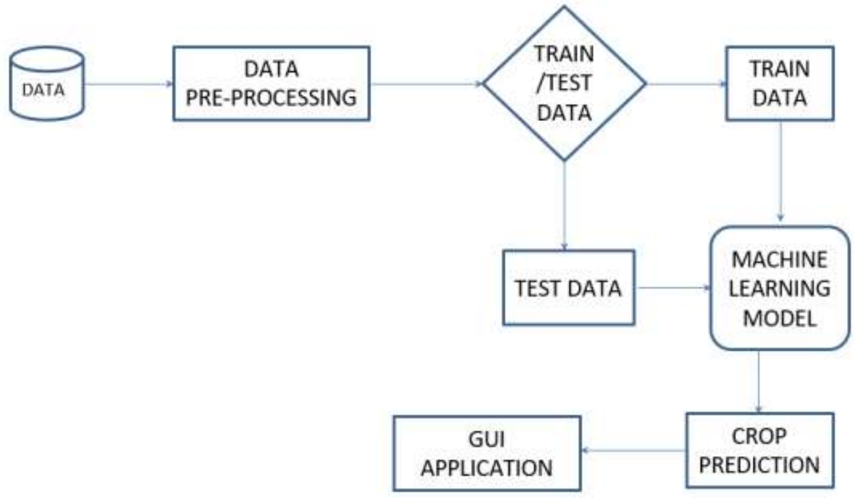
Advantages of Proposed System

- High prediction accuracy
- Faster processing and results
- Ability to handle large datasets
- Uses advanced machine learning techniques
- Reduces human effort and errors
- Supports data-driven decision making
- Scalable and flexible system
- Can be integrated with real-time data
- Helps farmers in crop planning
- Improves agricultural productivity

IV. Methodology

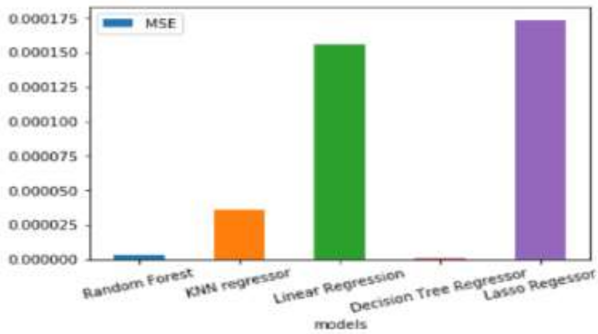
The methodology begins with data collection from agricultural datasets containing soil, weather, and crop information. The collected data is preprocessed by handling missing values, removing noise, and normalizing features. Feature selection is performed to identify important parameters influencing crop yield. The dataset is then divided into training and testing sets. Machine learning algorithms such as Linear Regression, Random Forest, and Decision Tree are applied. The models are trained using the training dataset. After training, the model is tested for accuracy using evaluation metrics like MAE and R^2-score. The best-performing model is selected for prediction. The system then predicts crop yield based on input parameters. Finally, the results are visualized for better understanding.

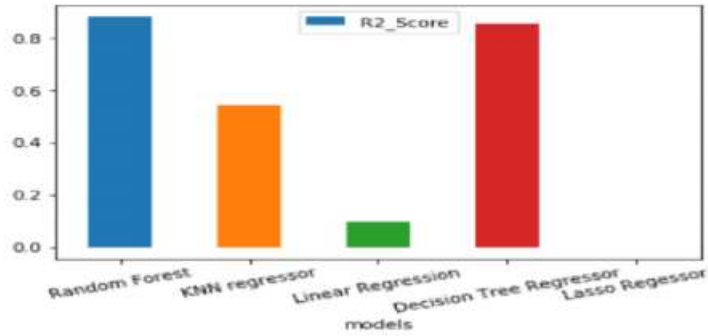
System Architecture

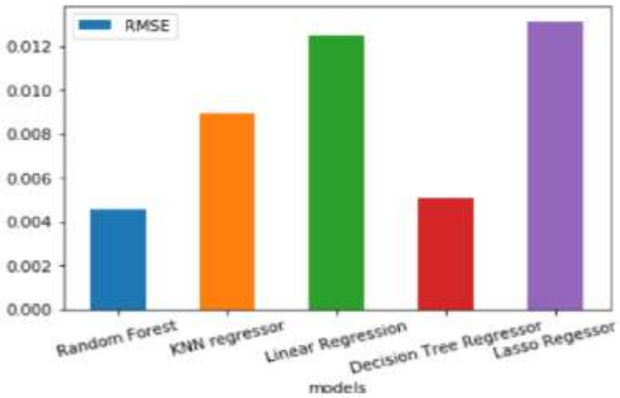


The system architecture consists of multiple layers that work together for crop yield prediction. The first layer is the data collection layer, which gathers data from agricultural datasets. The next layer is data preprocessing, where the data is cleaned and normalized. The processed data is passed to the machine learning layer for model training and prediction. In this layer, algorithms analyze the relationship between input features and crop yield. The prediction results are generated in the output layer. A visualization layer is used to display results in the form of graphs and charts. The system also includes a storage layer to store datasets and results. The architecture ensures smooth data flow between components. It is scalable and efficient for handling large datasets.

V. Result and Output







VI. Conclusion

The process of crop yield prediction in this project began with data cleaning, handling missing values, and performing exploratory data analysis, followed by model building and evaluation using machine learning algorithms. By applying different algorithms, the system is able to generate accurate predictions of crop yield based on various input parameters. The results demonstrate that machine learning techniques can effectively analyze agricultural data and provide reliable insights for better decision-making.

The proposed system covers a wide range of crops, enabling farmers to explore different cultivation options, including crops that may not have been previously considered. This helps farmers make informed choices about which crops to grow based on predicted yield and conditions. Additionally, the system considers historical production data, allowing farmers to understand trends in crop demand and market conditions.

Overall, the system provides a practical and intelligent solution for crop prediction, supporting farmers in improving productivity and planning. By reducing uncertainty and enhancing decision-making, it contributes to the advancement of modern, data-driven agriculture. Future improvements can include real-time data integration, advanced models, and deployment through user-friendly platforms to further enhance its usability and effectiveness.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, “An advanced movie recommender using collaborative filtering and sentiment analysis,” International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, “Automatic crop recommendation system using LightGBM and decision tree machine learning models,” Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, “Smart agriculture through IoT and machine learning for analyzing carbon footprints,” in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.