IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, July 2024

# Crop Yield Prediction System Using Machine Learning

**H K Charan[1] and Usha M[2]**

Student, Department of MCA, Bangalore Institute of Technology, Bangalore, India[1]

Professor, Department of MCA, Bangalore Institute of Technology, Bangalore, India[2]

**Abstract:** *Agriculture is the field which plays an important role in improving our countries economy. Agriculture is the one which gave birth to civilization. India is an agrarian country and its economy largely based upon crop productivity. Hence we can say that agriculture can be backbone of all business in our country. Selecting of every crop is very important in the agriculture planning. The selection of crops will depend upon the different parameters such as market price, production rate and the different government policies. Many changes are required in the agriculture field to improve changes in our Indian economy. We can improve agriculture by using machine learning techniques which are applied easily on farming sector. Along with all advances in the machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. The concept of this project is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.*

**Keywords:** Crop Yield, Prediction, Water Usage, UV Radiation, Fertilizers, Pesticides, Soil Quality

## I. INTRODUCTION

Agriculture is the pillar of India's economy, playing a crucial role in sustaining the livelihood of millions of people. Despite its importance, the sector faces numerous challenges that affect crop yield, including unpredictable weather patterns, varying soil conditions, and the quality of seeds used. Traditional farming techniques often fall short in addressing these complexities, resulting in inconsistent and often disappointing yields. In recent years, the domain of information mining has emerged as a robust tool obtaining useful knowledge from large datasets, making it possible to predict crop yields with greater accuracy.

Data mining involves the mechanism of discovering patterns and relationships within large sets of data, a practice that spans various disciplines such as statistics, machine learning, and artificial intelligence. The primary goal of data mining in the field of agriculture is to analyze historical data to forecast future crop yields, thereby enabling farmers to optimize their farming practices. This study focuses on using machine learning techniques to forecast the results of crops like wheat and maize. By incorporating things like the type of oil, weather circumstances, and crop-specific parameters, the study aims to develop a robust predictive model that can assist farmers in making informed decisions.

The significance of accurate crop yield prediction cannot be overstated. It not only helps farmers in planning and resource allocation but also aids policymakers in ensuring food security and managing supply chains effectively. This study employs various regression algorithms to analyze the data and evaluate their performance based on key metrics like RMSE, $R^2$, and MAPE. By comparing these models, the research seeks to identify the most effective approach for predicting crop yields.

In summary, this introduction sets the stage for an in-depth exploration of how data mining and machine learning can revolutionize agricultural practices in India. By utilizing the strength of advanced data analysis techniques, this research aims to provide practical solutions to the longstanding challenges faced by farmers, ultimately contributing to the expansion and durability of the agricultural sector.

## II. EXISTING SYSTEM

The existing systems for crop yield prediction in India primarily rely on traditional statistical methods and empirical approaches. These methods often involve the use of basic linear regression models or historical yield averages to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, July 2024

forecast future yields. While these approaches provide some level of insight, they are limited in their ability to capture the complex and dynamic nature of agricultural data.

Traditional methods typically do not account for the numerous factors that influence crop yield, such as soil properties, weather conditions, and crop management practices. As a result, the predictions generated by these methods are often inaccurate and unreliable. Additionally, these systems tend to be static, lacking the ability to adapt to changing environmental conditions and advancements in agricultural practices.

Furthermore, existing systems usually do not leverage the vast amounts of data available from modern agricultural technologies, such as remote sensing, IoT devices, and precision farming tools. This data, if properly analyzed, could provide valuable insights into crop health, soil quality, and other critical parameters.

In summary, the limitations of existing systems underscore the need for more advanced and dynamic approaches to crop yield prediction. This research aims to address these gaps by applying machine learning algorithms that can effectively analyze and interpret complex agricultural data, leading to more accurate and actionable predictions

## III. PROPOSED SYSTEM

The proposed system aims to leverage the power of machine learning algorithms to develop a more accurate and reliable predictive model for crop yield. Unlike traditional methods, this system will utilize advanced regression techniques to analyze large datasets encompassing various factors such as soil properties, weather conditions, and crop-specific parameters.

The key components of the proposed system include:

- **Data Collection and Preprocessing**: Extensive datasets related to soil, weather, and crop parameters will be collected from reliable sources. Preprocessing steps will ensure the data is clean, consistent, and suitable for analysis.
- **Machine Learning Algorithms**: Various regression algorithms, including linear regression, Support Vector Regression (SVR), and Random Forest Regression, will be implemented to develop predictive models. These algorithms will be trained using the preprocessed data to ensure accurate yield predictions.
- **Model Evaluation**: The performance of the predictive models will be evaluated using metrics such as RMSE, $R^2$, and MAPE. This evaluation will help identify the most effective algorithm for crop yield prediction.
- **User Interface**: A user-friendly interface will be developed to present the prediction results to farmers in an easily interpretable format. This interface will provide actionable insights that farmers can use to optimize their agricultural practices.

## IV. METHADOLOGY

### Data Collection

The variables in the dataset for this research are water usage, UV radiation, fertilizer application data, regional climate data, and pesticide usage data. The data was gathered from the agricultural databases, meteorological stations, and field surveys. Other sources of data include the satellite imagery and soils health records.

### Data Preprocessing

Data preprocessing involved cleaning, normalization, and feature selection. Handling missing values was done with interpolation techniques, while outliers were detected and removed. Feature selection was carried out using correlation analysis to select variables of influence.

### Model Selection

Several machine learning algorithms were evaluated, which included the following.

- Decision Trees: Simple and interpretable but prone to overfitting.
- Random Forest: An ensemble method that mitigates overfitting and improves accuracy.
- Support Vector Machines: High dimensional spaces, effective in high-dimensional spaces, computationally intensive.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, July 2024

- Neural Networks: Complex relationships can be captured, but large datasets and high computational power are required. Model Training and Evaluation

The dataset was split 80% for training and 20% for testing. The models were trained on the training dataset, and their performance was evaluated using the test dataset by accuracy, precision, recall, and the F1 score. Cross-validation techniques such as k-fold cross-validation are taken into account in order to make the model more robust. Hyperparameter tuning has been done using grid search for optimum performance of the model.

## V. IMPLEMENTATION

Crop Yield Prediction System was implemented in Python, which is one of the most flexible and popular programming languages available. It gains much of its popularity from the broad variety of libraries supporting different applications, especially in scientific computing and data analysis. In this project, extensive use has been made of the following libraries: Pandas for data manipulation, Scikit-learn, and Matplotlib. Pandas contributed a great deal with its data structures, DataFrames, which are highly useful in efficiently handling and processing large volumes of data. It has supported tasks such as data cleaning, handling missing values via interpolation, removing duplicates, filtering outliers, and making feature engineering to improve the performance of models.

Scikit-learn was used for the implementation of machine learning, looking forward to applying various algorithms like Decision Trees, Random Forests, Support Vector Machines, and Neural Networks. This library provided very impressive coverage of tools on model training, cross-validation, hyperparameter tuning, and model performance evaluation. This was greatly beneficial in using Scikit-learn's pipeline functionality to homogenize data processing and model train sequences for reproducibility and ease of maintenance.

Data visualization was done using Matplotlib with the addition of Seaborn, which assists in creating plots that become aids in understanding data distribution, feature importance, and model performance. Using these libraries, it performed EDA through histograms, box plots, scatter plots, and heatmaps. All of these were quite instrumental in the exploration of a dataset for getting patterns. In addition, model performance was visualized using ROC curves, precision-recall curves, and feature importance plots that gave insight into model performance.

The user interface was developed using Flask, a Python lightweight web application framework. Flask helped create a web interface where the user could feed environmental and regional data in order to get crop yield predictions on the fly. HTML and CSS enabled building the interface in a manner that is user-friendly and could be visually striking. It was through this rigorous implementation that the Crop Yield Prediction System became not only robust and accurate but also user-friendly and accessible.

## VI. RESULT

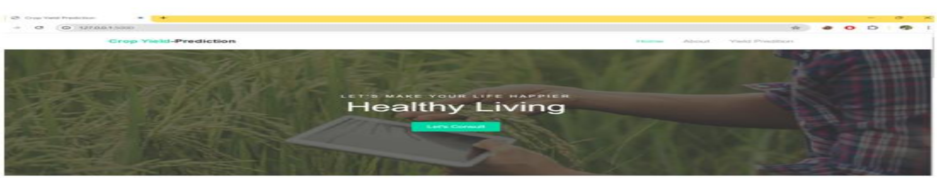


Fig 6.1 Home page


Fig 6.2 About page


DOI: 10.48175/568

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 2, July 2024

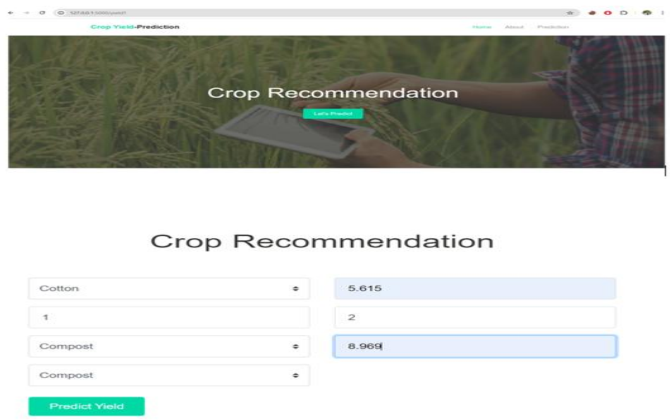


Fig 6.3 Prediction page


## VII. CONCLUSION

The Crop Yield Prediction System is a comprehensive solution designed to address the critical need for accurate yield predictions in agriculture. By leveraging advanced machine learning algorithms and integrating diverse data sources, this system provides valuable insights that can help farmers and agricultural stakeholders make informed decisions. The implementation of this system on a local server ensures reliability and accessibility, even in areas with limited internet connectivity.

**Key Achievements:**

- **Accurate Predictions:** The system employs robust machine learning models, such as linear regression, SVR, and random forest, to deliver accurate crop yield predictions.
- **Data Integration:** It successfully integrates data from soil sensors, weather stations, and historical databases, providing a holistic view of the factors affecting crop yields.
- **User-Friendly Interface:** The system features an intuitive user interface that allows users to easily input data and view predictions and insights.
- **Rigorous Testing:** Extensive testing, including unit, integration, system, and user acceptance testing, ensures that the system is reliable, accurate, and user-friendly.

## VIII. FUTURE ENHANCEMENTS

**Integration of Additional Data Sources**

To enhance the accuracy and reliability of predictions, the system can integrate additional data sources such as:

- **Satellite Imagery:** Incorporating remote sensing data can provide high-resolution insights into crop health and growth patterns.
- **IoT Devices:** Using IoT devices for real-time monitoring of soil moisture, temperature, and other environmental factors can enhance data accuracy and timeliness.
- **Market Data:** Including market prices and trends can help predict the economic viability of different crops.
- **Advanced Machine Learning Techniques**
- Exploring advanced machine learning techniques can improve prediction accuracy and model robustness:
- **Deep Learning Models:** Implementing deep learning architectures such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can capture complex patterns in the data.
- **Ensemble Methods:** Using ensemble methods that combine multiple models can enhance prediction accuracy and reduce the risk of overfitting.

## REFERENCES

**[1].** Oikonomidis, A., Catal, C., &Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction. *Journal of Applied Artificial Intelligence, 36*(1). Retrieved fromSpringerLink

**[2].** Ali, M., et al. (2022). Coupled online sequential extreme learning machine model with ant colony optimization algorithm for wheat yield prediction. *Scientific Reports, 12*(1), 5488. Retrieved fromSpringerLink

**[3].** Dubey, S.K., & Sharma, D. (2018). Assessment of climate change impact on yield of major crops in the Banas River Basin, India. *Science of The Total Environment, 635*, 10-19. Retrieved fromSpringerLink

**[4].** Kamilaris, A., &Prenafeta-Boldú, F.X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70-90. Retrieved fromScienceDirect

**[5].** Khan, S.N. (2022). A geographically weighted random forest approach to predict corn yield in the US corn belt. *Remote Sensing, 14*(12), 2843. Retrieved from MDPI

**[6].** Fei, S., Li, L., Han, Z., Chen, Z., & Xiao, Y. (2022). Combining novel feature selection strategy and hyperspectral vegetation indices to predict crop yield. *Plant Methods, 18*(1), 1-13. Retrieved from SpringerLink

**[7].** Bian, C. (2022). Prediction of field-scale wheat yield using machine learning method and multi-spectral UAV data. *Remote Sensing, 14*(6), 1474. Retrieved from MDPI

**[8].** Tripathi, A., Tiwari, R.K., & Tiwari, S.P. (2022). A deep learning multi-layer perceptron and remote sensing approach for soil health based crop yield estimation. *International Journal of Applied Earth Observation and Geoinformation, 113*, 102959. Retrieved fromScienceDirect

**[9].** Ragam, P., &Nimaje, D.S. (2018). Evaluation and prediction of blast-induced peak particle velocity using artificial neural network: A case study. *Noise and Vibration Worldwide, 49*(3). Retrieved fromSpringerLink