Crop Yield Prediction Based On Indian Agriculture

Dr . Ananda R kumar Mukkala
Assistant Professor, Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Chelamani Chandrika
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Jogu Harshitha
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Pullur Shankar
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

Nizampet Abhiram
Dept of CSE, Sreyas Institute of Engineering and Technology,
Telangana,India.

.

***Abstract*— Agriculture plays a significant role in the Indian economy, providing livelihood to a significant portion of the population and contributing to the country's GDP. It forms the backbone of rural India and plays a significant role in ensuring food security and economic growth. A country like India. In this study, machine learning techniques have been used to predict crop yield using various regression models such as linear regression, lasso regression, ridge regression and tree regression. This data covers many factors such as region, year, crop variety, season, rainfall, pesticide and insecticide usage. It is superior to other regression models in terms of accurately predicting crop yield. The ability of decision tree models to handle non-linear relationships and interactions among features helps them achieve high performance. This approach provides farmers and policymakers with powerful tools to make informed decisions, ultimately leading to profitable and sustainable agriculture. Using these forecasts, farmers can achieve higher yields, improve resource utilization, and therefore increase agricultural yields and economic benefits.**

Keywords—*Machine Learning, Crop Yield Prediction,Linear Regression,Lasso Regression,Ridge Regression.*

INTRODUCTION

Crop yield prediction is essential for effective agricultural planning and management. We can increase the prediction accuracy by using machine learning (ML) techniques. ML models analyze many factors like weather, crop type, and management. Regression models like linear regression, lasso regression, ridge regression, etc. Establish a relationship between input variables and crop yield. Classification algorithms like decision trees classify plants based on their characteristics to predict yield. Weather data includes temperature, precipitation, humidity, and other information. Soil information includes information such as soil type, pH, organic matter content, and moisture. Historical crop records, including past harvest records and management practices, are also important. Many things. This proves how effective the model is. Model evaluation is based on metrics such as mean squared error (MSE) and R-squared value (R2). Cross-validationEnsures the robustness and generalizability of models. Optimizes the use of resources such as water and fertilizer in precision agriculture based on predictive forecasts. Helps manage risks by identifying risks and developing strategies to reduce crop losses. Uses accurate forecasts to improve supply chain management, inventory, and distribution planning. ML provides insights into agricultural planning and management by analyzing large amounts of data and using advanced algorithms. On the technology front, the combination of machine learning with new technologies like the Internet of Things and blockchain will further transform agriculture by making it more profitable than ever before..

OBJECTIVES AND METHODOLOGY

The main goal of using machine learning for crop prediction is to build accurate models that predict yields based on various inputs such as crop type, region, year, season, rainfall, pesticides, and insecticides. This helps farmers make informed decisions, improve resource utilization, and increase productivity. The approach involves several basic steps: collecting historical agricultural data from reliable sources, preprocessing the data to clean and process it, and using selection techniques to identify the most important factors affecting the crops. Use machine learning algorithms such as linear regression, lasso regression, ridge regression, and decision trees to train the predictive models. Linear regression provides a simple way to predict crop yields as a function of inputs. Lasso regression introduces a continuous L1 time that helps in feature selection by controlling the variance in the model. On the other hand, ridge regression has an L2 constant term to address multicollinearity and increase model robustness. The decision tree algorithm subdivides the data into accessibility-based subsets, thus creating a tree-like structure that captures nonlinear relationships. High precision. Once developed, these models can accurately predict crop yields and provide recommendations to improve agricultural practices. These recommendations help farmers make data-driven decisions to improve the use of resources such as water, fertilizer, and pesticides, and ultimately increase yield and sustainability. With the use of these new technologies, agriculture can move towards a more data-driven and informed approach to crop management, making agriculture more profitable and rewarding.

I . LITERATURE SURVEY

A literature survey on machine learning for crop yield prediction reveals that various studies have focused on different algorithms and features to improve prediction accuracy. Commonly used algorithms include Linear Regression, Lasso Regression, Ridge Regression, and Decision Tree. Researchers have employed feature selection techniques such as correlation analysis and dimensionality reduction to identify the most relevant factors affecting crop yield. Some studies have also explored deep learning approaches like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, which can capture complex patterns in the data. Additionally, the integration of environmental factors, remote sensing technology, and vegetation indices has been found to enhance crop yield predictions. Overall, the literature highlights the potential of machine learning to provide valuable insights for optimizing agricultural practices and promoting sustainable farming.

Fig. 1: Agricutural Practices

II . PROPOSED SYSTEM

The proposed system for crop yield prediction utilizes machine learning to forecast yields based on various input parameters such as crop type, land size, year, season, rainfall, insecticides, and pesticides. By providing precise predictions and actionable recommendations, the system aims to help farmers optimize resource use, improve productivity, and promote sustainable farming practices.

A. User-Interference Design:

Design and develop a user-friendly interface that allows farmers to input their data easily. The interface should provide clear visualizations of predictions and recommendations. Features might include data entry forms, prediction results display, trend charts, and alerts for specific conditions.

Fig 2 : User Interface

B. Data Collection :

This involves gathering historical data on crop yield and the relevant input parameters such as crop type, land size, year, season, rainfall, insecticides, and pesticides. The data can be sourced from government agricultural departments, research institutions, weather stations, and farm records. It is crucial to ensure the data is comprehensive and covers a wide range of conditions and variables.

-

1	Crop	Crop_Year	Season	State	Area	Productio	Annual_R	Fertilizer	Pesticide	Yield
2	Arecanut	1997	Whole Ye	Assam	73814	56708	2051.4	7024878	22882.34	0.796087
3	Arhar/Tur	1997	Kharif	Assam	6637	4685	2051.4	631643.3	2057.47	0.710435
4	Castor see	1997	Kharif	Assam	796	22	2051.4	75755.32	246.76	0.238333
5	Coconut	1997	Whole Ye	Assam	19656	1.27E+08	2051.4	1870662	6093.36	5238.052
6	Cotton(lin	1997	Kharif	Assam	1739	794	2051.4	165500.6	539.09	0.420909
7	Dry chillie	1997	Whole Ye	Assam	13587	9073	2051.4	1293075	4211.97	0.643636
8	Gram	1997	Rabi	Assam	2979	1507	2051.4	283511.4	923.49	0.465455
9	Jute	1997	Kharif	Assam	94520	904095	2051.4	8995468	29301.2	9.919565
10	Linseed	1997	Rabi	Assam	10098	5158	2051.4	961026.7	3130.38	0.461364
11	Maize	1997	Kharif	Assam	19216	14721	2051.4	1828787	5956.96	0.615652

Fig 3: Data Collection

C. Data Preprocessing.

The collected data is often messy and may contain missing values, outliers, and inconsistencies. Preprocessing involves cleaning the data by handling missing values (e.g., using mean imputation), removing outliers, and normalizing the data to a common scale. This step ensures the data is suitable for machine learning models and improves their performance.

D. Feature Selection :

Not all input parameters may significantly impact crop yield. Feature selection techniques like correlation analysis, feature importance scores, and dimensionality reduction methods (e.g., Principal Component Analysis) are used to identify the most relevant factors. This step helps in simplifying the model and improving its interpretability and accuracy.

E. Model Selection :

Multiple machine learning algorithms are considered for the prediction task:

Linear Regression: Predicts crop yield as a linear combination of the input parameters.

Lasso Regression: Adds L1 regularization to linear regression, promoting sparsity and feature selection.

Ridge Regression: Adds L2 regularization to linear regression, handling multicollinearity and improving robustness.

Decision Tree: non-linear model that splits the data based on input parameters, forming a tree structure for predictions.

F. Model Training:

The data is split into training and testing sets to evaluate model performance. Cross-validation techniques, such as cross-validation, are used to ensure the model generalizes well to unseen data. This step involves fitting the selected models to the training data and fine-tuning their parameters.

G . Model Evaluation:

The trained models are evaluated using metrics like Mean Absolute Error (MAE) and R-squared (R^2). MAE measures the average magnitude of errors in predictions, while R^2 indicates the proportion of variance in the dependent variable explained by the model. These metrics help in comparing the performance of different models and selecting the best one.

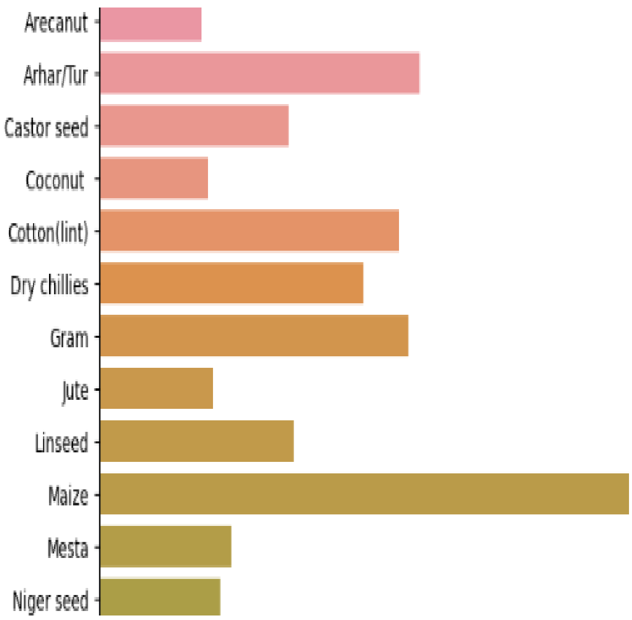

Fig 4 : Evaluation of model

H . Model Optimization:

Hyperparameter tuning is performed to further improve the model's accuracy. Techniques like grid search and Bayesian optimization are used to find the optimal values for model parameters. This step ensures the model is fine-tuned for the best possible performance.

I . Prediction:

The optimized models are used to predict crop yields for new data inputs. This involves feeding the input parameters into the trained model and obtaining the predicted yield values. The predictions can be used to make informed decisions about crop management.

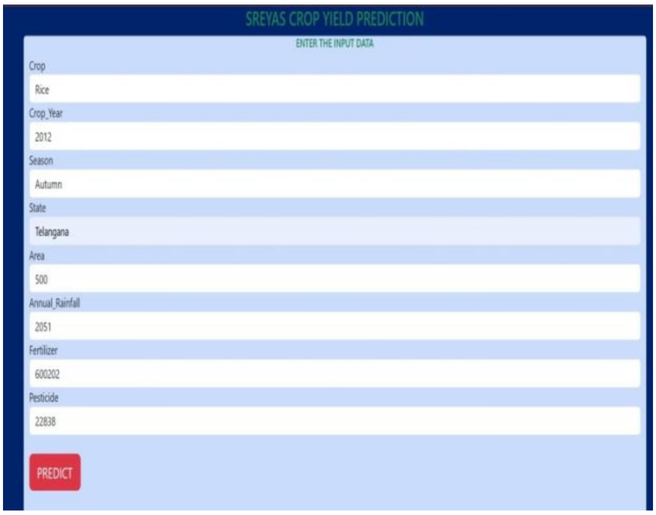

Fig 5 : Data Inputs

J. Validation:

The model's predictions are validated against actual crop yield data to ensure reliability. This step involves comparing the predicted values with the observed values and assessing the model's accuracy. Validation helps in identifying any discrepancies and making necessary adjustments to the model.

K . Deployment:

Final step involves deploying the trained and validated model in a real-world agricultural setting. This can be done through a user-friendly interface or an application that provides farmers with actionable insights and recommendations based on the model's predictions. The deployed system helps farmers optimize resource use, improve productivity, and promote sustainable farming practices.

Fig 6 : Predicted Yield

III .MPLEMENTATION

The implementation of the proposed system for crop yield prediction involves several critical steps. First, historical agricultural data is collected from reliable sources and preprocessed to clean and normalize it, ensuring consistency and accuracy. Feature selection methods are then applied to identify the most relevant factors influencing crop yield. Machine learning algorithms such as Linear Regression, Lasso Regression, Ridge Regression, and Decision Tree are used to train predictive models. These models are evaluated and fine-tuned using metrics like Mean Absolute Error (MAE) and R-squared (RÂ²) to ensure high accuracy. The optimized models are then used to predict crop yields for new data inputs. To make the system accessible and user-friendly, a user interface is developed, allowing farmers to input their data and receive insights. The interface includes features like data entry forms, prediction results display, trend charts, and alerts for specific conditions. The final step is deploying the trained model and user interface in a real-world agricultural setting, providing farmers with actionable insights and recommendations to optimize their practices. This comprehensive approach aims to enhance resource utilization, improve productivity, and promote sustainable farming practices.

A. Architecture Diagram:

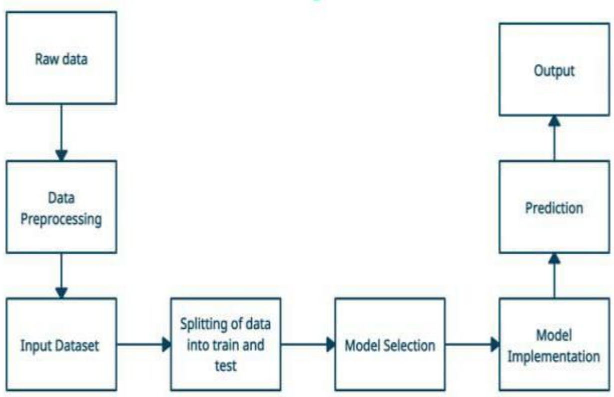

Fig 7 : Architecture diagram

B. Test Cases:

Test cases are essential for ensuring the functionality and accuracy of the crop yield prediction system. They help validate that the system handles data correctly, performs accurate predictions, and interacts seamlessly with the user interface. Data validation test cases ensure the system properly handles missing values, inconsistencies, and outliers. Feature selection test cases verify the accurate identification of relevant factors impacting crop yield. Model training test cases assess the performance of machine learning algorithms, ensuring they do not overfit or underfit the data. Model evaluation test cases check the accuracy of the models using metrics like Mean Absolute Error (MAE) and R-squared (R^2). User interface test cases validate the ease of data entry, clarity of visualizations, and responsiveness of the interface. Integration test cases ensure that all components work together seamlessly, providing a smooth end-to-end experience for farmers. By implementing these test cases, the system can deliver reliable, accurate, and user-friendly crop yield predictions that empower farmers to make data-driven decisions and improve agricultural productivity.

Description	Preconditions	Test Steps	Expected Result	Actual Result	Status
Predict yield with valid data	Model is trained with valid data	1. Input valid weather and crop data 2. Run prediction model	Accurate yield prediction	Accurate yield prediction	Success
Predict yield with missing data	Model is trained with missing data	1. Input weather and crop data with missing values 2. Run prediction model	Model handles missing data gracefully	Model handles missing data gracefully	Success
Predict yield with outlier data	Model is trained with outlier data	1. Input weather and crop data with outliers 2. 2. Run prediction model	Model handles outliers effectively	Model handles outliers effectively	Success
Predict yield with new crop type	Model is trained with existing crop data	1. Input weather and new crop type data 2. Run prediction model	Model predicts yield for new crop type	Model predicts yield for new crop type	Success

Fig 8 : Test Cases

IV . DISCUSSION

A . Usability and User Experience:

The usability and user experience (UX) of the crop yield prediction system are critical to its success. Usability ensures that farmers can easily input their data and access predictions through a straightforward and intuitive interface. Key factors include simplicity, accessibility across devices, consistency in design, immediate feedback for user actions, and readily available help and support resources. UX, on the other hand, focuses on the overall satisfaction and engagement of the user. A visually appealing and clean design, interactive features like trend charts, and personalized recommendations based on user preferences contribute to a positive experience. The system should be easy to learn with minimal training required and evoke positive emotions by providing accurate, reliable predictions and handling errors gracefully. By prioritizing usability and UX, the crop yield prediction system can become an indispensable tool for farmers, helping them make data-driven decisions, optimize resources, and improve productivity with ease and confidence.

B . Applications in Real World Scenarios:

Machine learning-based crop yield prediction systems have numerous real-world applications that can greatly benefit the agricultural sector. These systems enable precision agriculture by providing farmers with precise recommendations for planting, fertilization, and irrigation based on data analysis, which helps in optimizing resource use, reducing waste, and increasing yields. They also assist in risk management by offering early warnings and insights into potential yield losses due to weather variability, pests, and diseases, allowing farmers to take proactive measures. Accurate yield predictions improve supply chain optimization by enabling better planning and coordination within the agricultural supply chain, reducing food wastage, and managing inventory effectively. Financial planning for farmers is enhanced as they can use yield predictions to secure loans, purchase insurance, and plan investments, while financial institutions can assess farmers' creditworthiness. Policymakers can utilize these systems to design and implement effective agricultural policies, ensuring food security and supporting farmers. Researchers can study the impact of different farming practices and environmental conditions on crop yields, driving innovation and the development of resilient crop varieties. Additionally, these systems facilitate efficient resource management by optimizing the use of water, fertilizers, and pesticides, thus reducing environmental impact and production costs. Lastly, they help farmers adapt to climate change by providing insights into how different climatic conditions affect crop production, enabling informed decisions about crop selection and management practices. These applications demonstrate the potential of machine learning-based crop yield prediction systems to transform agriculture, improve productivity, and promote sustainable farming practices**.**

V . CONCLUSION AND FUTURESCOPE

The future of crop production is based on predictive machine learning and risk. As technology continues to evolve and agriculture continues to face new challenges, opportunities to enhance and expand the capabilities of these machines continue to emerge. Key areas for future development include integrating IoT devices for real-time data collection to increase accuracy, using satellite imagery and analytics to monitor crop health and the environment, and changing methods to predict outcomes under various climate change scenarios. While automated data collection using drones and other tools can increase efficiency, precision agriculture can be tailored to specific regions or land areas to deliver benefits and resource utilization. Expanding the approach to predict yields of multiple crops simultaneously and integrating with existing agricultural management systems can improve the use of data. Advanced machine learning techniques like deep learning can improve forecast accuracy. Additionally, creating a multilingual and user-friendly interface will make the system accessible to farmers around the world, providing instant updates and notifications when needed. Changing the environment will improve decision-making. Integrating sustainability metrics to assess environmental impacts and linking crop forecasts to economic models can help farmers make better financial decisions. Increasing the mobility of mobile devices, creating collaborative research, and integrating pest and disease prediction models can further benefit farmers. Providing insight into how best to use resources and incorporating user feedback will continue to improve the system for real-world use. The system integrates multiple components, including data collection, pre-processing, model training, and deployment. Data such as state, year, crop type, season, rainfall, soil, pesticides, and herbicides are first entered by farmers and then validated and processed to ensure accuracy. Machine learning models trained using historical data can predict high-population crop yields. The system involves interaction between users, project managers, and agricultural scientists to ensure accuracy and continuous improvement of data processing. Being able to predict enables farmers to make informed decisions. The user-friendly interface, including a web and mobile app, provides easy access to data and forecasts. Leveraging the power of machine learning and data analytics, the system represents a breakthrough in agricultural technology, empowering farmers and encouraging action for sustainable agriculture.

VI . REFERENCES

[1] P. Priya, U. Muthaiah M. Balamurugan. Predicting yield of the crop using machine learning algorithm. International Journal of Engineering Science Research Technology.

[2] J. Jeong, J. Resop, N. Mueller and team. Random forests for global and regional crop yield prediction. PLoS ONE Journal.

[3] Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy. Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4] S. Veenadhari, Dr. Bharat Misra, Dr. CD Singh. Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5] Shweta K Shahane, Prajakta V Tawale. Prediction On Crop Cultivation. IInternational Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

[6] D Ramesh, B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology.

[7] Subhadra Mishra, Debahuti Mishra, Gour Hari Santra.Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science

and Technology, Vol 9(38), DOI:10. 17485/ijst/2016/v9i38/95032, October 2016.

[8] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson ID, Dionysis Bochtis. Machine Learning in Agriculture. Lincoln Institute for Agri-food Technology (LIAT), University of Lincoln, Brayford Way, Brayford Pool, Lincoln LN6 7TS, UK.

[9] Baisali Ghosh. A Study to Determine Yield for Crop Insurance using Precision Agriculture on an Aerial Platform. Symbiosis Institute of Geoinformatics Symbiosis International University 5th & 6th Floor, Atur Centre, Gokhale Cross Road, Model Colony, Pune – 411016.

[10] Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.

[11] Ecochem Online. (2009). Soil Health and Crop yields. Last modified January 28th 2009. Retrieved on March 4th 2009 from http://ecochem. com/healthy_soil. html Food and Agricultural Organization. (2006). The state of Agricultural Commodity Markets. 37-39.

[12] Aditya Shastry, H. A Sanjay And E. Bhanushree, “Prediction of crop yield using Regression Technique”, International Journal of computing r12 (2):96- 102 2017, ISSN:1816-914] E. 14]E. Manjula, S.

[13]Djodiltachoumy, “A Model for Prediction of Crop Yield”, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.

[14] Mrs. K. R. Sri Preethaa, S. Nishanthini, D. SanthiyaK. Vani Shree, “CropYield Prediction”, International Journal On Engineering Technology and Sciences – IJETS™ISSN(P): 2349-3968, ISSN (O):2349-3976 Volume III, Issue III, March- 2016.

[15] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, “Analysis of agriculture data using data mining techniques: application of big data” Majumdar et al. J Big Data (2017) 4:20 DOI 10. 1186/s40537-017- 0077-4

[16] 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.

[17] Yethiraj N G, ” Applying data mining techniques in the field of Agriculture and allied sciences”, Vol 01, Issue 02, December 2012.

[18] Zelu Zia (2009). An Expert System Based on Spatial Data Mining used Decision Tree for Agriculture Land Grading. Second International Conference on Intelligent Computation Technology and Automation. Oct10-11, China.