Crop Yield Prediction using MachineLearning Algorithm

G.Samyuktha
Department of Computer Science And Engineering
Panimalar Engineering College(Autonomous)
Chennai-05 Tamil Nadu
samyuganesh2004@gmail.com

S.Saima Nooreen
Department of Computer Science and Engineering,
Panimalar Engineering College(Autonomous)
Chennai-05 Tamil Nadu
saima18nooreen@gmail.com

Abstract --Agriculture is a cornerstone of global food security and economic growth. Accurate crop yield prediction is essential for optimizing resource allocation, improving agricultural planning, and addressing the challenges of global food security. Machine learning techniques, particularly the Decision Tree algorithm, have emerged as powerful tools for forecasting crop yields.By analyzing historical crop yield data in combination with meteorological and agricultural variables, these models enhance prediction accuracy. The Decision Tree algorithm is especially effective due to its ability to handle both categorical and continuous variables, making it suitable for uncovering complex relationships within the data. This integration of agriculture with data science offers a promising pathway for promoting sustainable farming practices and supporting informed decision-making on a global scale.

Keywords — Agriculture, Crop yield prediction, Machine learning, Decision Tree algorithm, Historical data Meteorological variables,Agricultural variables, Prediction accuracy,Categorical variables

I. INTRODUCTION

Agriculture is a vital component of India's economy, as it is essential for the survival of both humans and animals throughout the nation. In 2009, the global population was around 1.8 billion, with forecasts suggesting it could reach 4.9 billion by 2030, leading to a substantial rise in the demand for agricultural products. This anticipated increase will require effective management of farmland and improvements in crop yields. However, the effects of climate change often jeopardize crops due to extreme weather conditions. Factors such as soil fertility issues, fluctuations in climate, flooding, and limited groundwater can result in crop failures, which have a direct negative impact on farmers. In other regions, there is often encouragement for farmers to cultivate specific crops that are well-suited to local environmental conditions. With the population growing at an accelerated pace, precise estimation and monitoring of crop production is crucial. Thus, a suitable approach should be developed, taking into account the various factors that influence better crop selection in relation to seasonal changes.

The primary goal of estimating crop yield is to enhance agricultural production, and several established models are utilized to boost crop output. Currently, machine learning (ML) is gaining global traction for its effectiveness in various fields, including forecasting, fault detection, and pattern recognition. ML algorithms can also contribute to improving crop yield rates, particularly in adverse conditions. By applying ML techniques for crop selection, it is possible to minimize yield losses despite challenging environmental factors

II. LITERATURE SURVEY

[1] In this study, Aruvansh Nigam, Saksham Garg, Parul Agarwal, and Archit Agarwal concentrate on crop yield prediction through various machine learning algorithms. The results of this approach are evaluated using mean absolute error. Algorithms such as RNN and LSTM are utilized to ensure precise outcomes, helping farmers decide which crops are most advantageous to cultivate.

[2] The authors propose employing a deep reinforcement learning model for accurate crop yield estimation. Features that contribute to predictions are extracted using deep learning techniques, and the Q-learning method is applied, allowing the reinforcement learning algorithm to learn from user interactions and collaborations.

[3] This publication discusses the use of machine learning for predicting crop yields, applicable to almost any Indian crop. To meet its objectives, the research incorporates sophisticated algorithms like Kernel Ridge, Rasso, and Enet. Additionally, regression stacking is employed to improve prediction accuracy. The project monitors performance through root mean squared error, enabling users to input metric values for generating prediction results.

[4] In this research, Shivam Bang, Akshaya Kumar Dixit, Rajat Bishnoi, Indu Chawla, and Ankit Singh Chauhan explore crop yield prediction by analyzing rainfall and temperature as significant factors. They employ ARMA, SARIMA, and ARMAX models to improve prediction accuracy, followed by the application of fuzzy logic once the predictions are established. This approach relies on previous year's crop yield data to estimate yields for the upcoming year.

[5] Niketa Gandhi et al. present a system for estimating rice crop yield in India using the support vector machine (SVM) algorithm. They applied Sequential Minimal Optimization (SMO) classifiers within WEKA software to analyze the given datasets. The results from their analysis were integrated into the final conclusions, highlighting the

classifiers' effectiveness in accurately predicting rice production.

[6] In this research, the authors introduce a crop prediction method that accounts for soil and weather conditions to determine the crops best suited for a specific environment. They employ an artificial neural network (ANN) to achieve the intended results. The network learns from sample input and output data using a backpropagation technique, which is trained through supervised learning.

[7] Authors D. Ramesh and B. Vishnu Vardhan proposed a crop yield prediction technique using data mining processes, which are extensively applied to agricultural issues. This method is used for classifying and identifying patterns within large datasets. The model for predicting crop yields was developed using the Multiple Linear Regression (MLR) method.

[8] In this study, researchers B. Manjula Josephine et al. propose a system for predicting crop yields using machine learning. They input a crop yield dataset and apply data preprocessing techniques to transform the raw data into a more understandable format. Next, a random forest approach is employed, which involves data validation and forecasting to achieve accurate results. Additionally, dimensional reduction is conducted to reduce the number of random variables.

[9] This study recommends a combined approach to feature selection that integrates both filter and wrapper strategies. By merging soil, crop, and meteorological data, the study creates optimal features for a crop recommendation algorithm. Model performance is evaluated by comparing dataset features with those generated from the proposed method. The effectiveness of the feature selection technique is assessed using metrics like MSE, RMSE, MAE, and R^2.Using the selected features, artificial neural networks and decision trees are developed.

[10] This study proposes 22 crops by utilizing methods such as correlation analysis, distribution evaluation, ensemble strategies, and majority voting. It implements a three-tier architecture for crop recommendations that encompasses preprocessing, classification, and evaluation stages. Features are classified using correlation plots and density distributions, while the majority vote is used to determineoverall performance. The research incorporates ensemble learning with base models including decision trees, random forests, and Naive Bayes. Performance metrics are calculated based on majority voting, and the study provides outperforming the alternative method, which has an accuracy of 98.52%. Consequently, the Naive Bayes classifier is deemed more effective. Furthermore, the article highlights current challenges and suggests future research directions.

III. ARCHITECTURE DIGRAM

The diagram depicts a workflow for developing and deploying data-driven models. It starts with historical data, which undergoes feature engineering to improve the input quality. The data is then split into training and testing datasets. The training data is fed into a data science (DS) algorithm to build a predictive model, while the test data is used for model validation to ensure the model's accuracy on unseen information. After validation, the model processes new data to generate insights, leading to actionable results for decision-making.

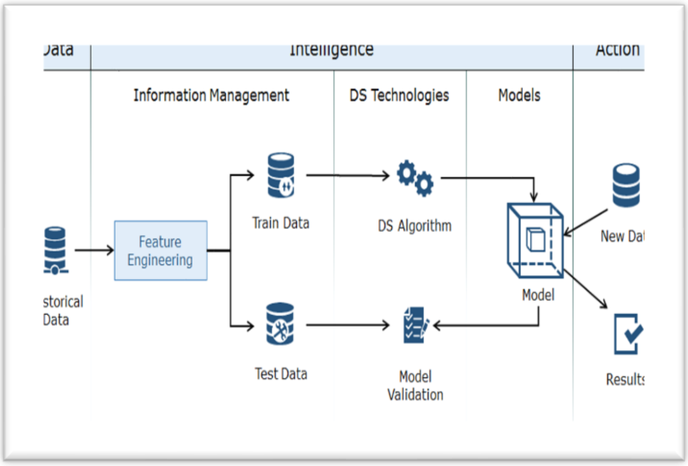


Fig.1 Architecture diagram

IV. EXISTING SYSTEM

A regression method has been applied in one of the existing systems for crop analysis. This system examined over 362 datasets, combining classification with decision tree analysis to provide solutions. For soil type prediction, the input data was divided into categories like natural, inorganic, and land. The results were consistent and reliable. Data management played a key role in collecting test data from another system during the study. Back Propagation Networks employed hidden layers to improve the accuracy of soil property predictions. Additionally, SVM and RVM were used as supervised learning techniques for predicting soil quality. Meteorological data was utilized to measure soil moisture through advanced remote sensing technologies, achieving 95% accuracy with a 15% error rate, though it hasn't yet been validated with the latest data. Basic engineering calculations were used to assess soil fertility and plant nutrient availability, leading to precise results and enhanced soil properties. While this approach yielded better outcomes than traditional methods, the system showed occasional inefficiencies. Essentially, this system uses soil factors to optimize crop yields. Currently, Naive Bayes classification is applied for soil categorization, achieving 77% accuracy. Additionally, Apriori algorithms are employed to identify crops with high yield potential based on soil type. It is proposed that the system integrates both crop yield data and environmental factors to predict crop development.

V. PROPOSED SYSTEM

The proposed system for crop yield prediction utilizes the Decision Tree Algorithm to analyze various factors influencing yield, such as soil properties, weather conditions, and crop type. The system will collect historical data on crop yields, soil nutrients, and meteorological information, followed by preprocessing steps to clean and normalize the data. The Decision Tree will then be trained on this processed data, allowing it to identify patterns and make numerical predictions about future crop yields based on new input variables. The output will provide a predicted yield value, enabling farmers to make informed decisions about crop management and resource allocation. This approach not only enhances prediction accuracy but also offers a clear and interpretable model that can aid in optimizing agricultural practices.

VI. METHODOLOGIES

A thorough evaluation and comparison of potential algorithms is essential before choosing the one that is most suitable for this dataset. Machine learning represents the most effective approach to addressing the challenges of crop production. Various machine learning techniques are utilized to predict agricultural yield.

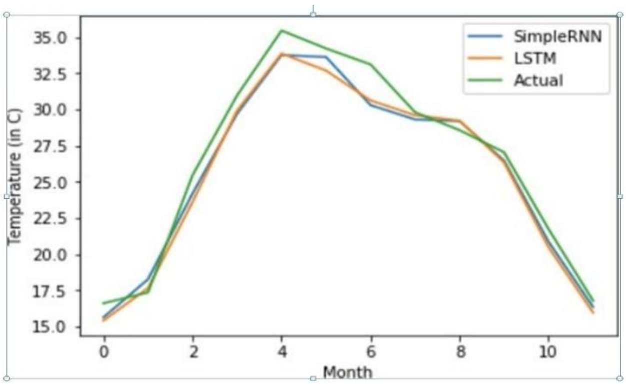
Supervised Learning encompasses one of the most widely used machine learning algorithms: logistic regression. This technique is applied to predict a categorical dependent variable based on a set of independent variables. Logistic regression serves as a classification method in supervised learning, used to estimate the probability of a target variable. Since the dependent variable is binary, there are only two possible classes. The logistic regression model achieves an accuracy of approximately 87.8% when tested on our dataset.

VII. CONCLUSION

The Decision Tree Algorithm to accurately predict crop yield based on various environmental and agricultural factors. With its transparent decision-making process and adaptability to diverse datasets, this system has the potential to assist farmers and agricultural planners in optimizing crop production and resource allocation.

VIII. RESULT

Predicting and forecasting crop yield can significantly boost agricultural production. Implementing crop rotation enhances soil fertility. This system offers user-friendly fertilization recommendations. The accuracy of this approach is around 92%. The proposed solution gathers input parameters from users to recommend the most effective fertilizer, thereby improving production efficiency.



1) Temperature and Rainfall: Models are created district wise and time series algorithms are trained by taking a look back of 10 years.

Fig 1. Deviation of predicted temperature with actual values.

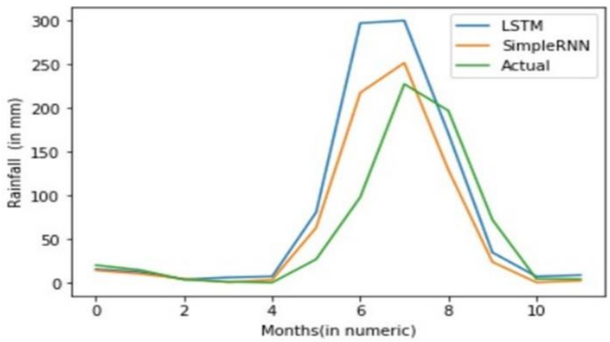


Fig 2. Deviation of predicted rainfall with actual values.

VIII. FUTURE ENHANCEMENTS

Future enhancements in crop yield prediction are likely to leverage advanced technologies such as artificial intelligence, machine learning, and remote sensing. By integrating data from various sources—including satellite imagery, weather patterns, soil health, and historical yield records—predictive models can become more accurate and adaptable. Additionally, real-time data collection through IoT devices and drones will enable farmers to monitor conditions closely and make informed decisions. These advancements aim to improve precision agriculture, optimize resource allocation, and ultimately increase food security in the face of climate change and population growth.

IX. REFERENCES

[1] Shivam Bang, Akshaya Kumar Dixit Rajat Bishnoi, Indu Chawla,Ankit Singh Chauhan, "Predicting Crop Yield using Fuzzy Logic using ARMA, SARIMA, and ARMAX Model Predicted Temperature and Rainfall Parameters, IEEE-2019

[2] Aruvansh Nigam, Saksham Garg, Parul Agarwal, Archit Agarwal.“Crop yield prediction in Indian agriculture using machine learning”,
[3] Potnuru Sai Nishant, Pinapa Sai Venkat, and Bollu Lakshmi Avinash B., “Crop Yield Prediction based on Indian Agriculture using Machine Learning"
[4] Dhivya Elavarasan and P. M. Durairaj Vincent., “Crop yield prediction using a deep reinforcement learning model for environmentally responsible agricultural applications”, IEEE.
[5] Niketa Gandhi, Owaiz Petkar, Leisa J. Armstrong, and Amiya Kumar Tripathy "Rice Crop Yield Prediction in India UsingSupport Vector Machines,"
[6] Snehal S. Dahikar and Sandeep V. Rode's Article "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach"published in the international journal of creative research
[7] D. Ramesh and B. Vishnu Vardhan, Analysis of crop yield prediction utilising data mining techniques, International Journal of Research in Engineering and Technology
[8] B. Manjula Josephine; K. Ruth Ramya; K.V.S.N. Rama Rao;Swarna Kuchibhotla; P.Venkata Bala Kishore; S. Rahamathulla;“Crop yield prediction using machine learning”; international journal of scientific & technology research
[9] M. J. and I. M., "Hybrid filter and wrapper approaches-based feature selection for crop recommendation." 2020 International Conference on Electronic Systems and Intelligent Computing(ICESIC),
[10] R. K. Ray, S. K. Das, and S. Chakravarty, “Smart Crop Recommender System-A Machine Learning Approach”, 12th International Conference on Cloud Computing, Data Science, and Engineering (Confluence), 2022.