engineering proceedings

Proceeding Paper

Agricultural Crop Yield Prediction using Advanced Data Analysis Techniques—Case Study

Guntha Karthik [1,*], B. Vaishnavi [2], N. Shashritha [2] and N. Anjali [2]

[1] Department of Electronics and communication, Stanley College of Engineering and Technology for Women (Autonomous), Abids, Hyderabad 500001, Telangana
[2] ECE Department, Stanley College of Engineering and Technology for Women (Autonomous), Hyderabad, Telangana, India; vaishuvnps@gmail.com (B.V.); shashritha7189@gmail.com (N.S.); anjalinarri@gmail.com (N.A.)
* Correspondence: gunthakarthik@rocketmail.com; Tel.: +91-9052611677/8596956466
† Presented at The 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: https://sciforum.net/event/ecsa-11.

Abstract: Agricultural crop yield prediction is crucial for enhancing food security, optimizing resource use, and ensuring sustainable agricultural practices. This project focuses on enhancing food security and sustainable agricultural practices by predicting crop yields using machine learning techniques. By integrating data from weather patterns, soil properties, and historical crop performance, the study aims to aid farmers and policymakers in decision-making. The research examines the correlation between crop yield and environmental factors such as nitrogen, phosphorus, potassium, rainfall, temperature, and fertilizer application. Multilinear Regression (MLR), Radial Basis Function (RBF), and Support Vector Machine (SVM) models are applied to predict yields, with SVM achieving the highest accuracy at 92.03%, followed by MLR at 88.56% and RBF at 75.36%. The data collection for this study includes nutrient levels in the soil, historical weather patterns, and fertilizer usage from the Peddapalli district, Telangana, India. MLR identifies linear relationships, RBF captures non-linear patterns, and SVM handles high-dimensional data to enhance prediction accuracy. The results indicate that while MLR and RBF provide valuable insights, SVM is the most robust tool for forecasting crop yields. This research holds significant potential for improving agricultural productivity and resource management, offering farmers crucial insights for better planning and allocation of resources.

Keywords: agricultural crop prediction; machine learning; multilinear regression; radial basis function; precision agriculture; environmental factors; crop yield forecasting

Citation: Karthik, G.; Vaishnavi, B.; Shashritha, N.; Anjali, N. Agricultural Crop Yield Prediction using Advanced Data Analysis Techniques—Case Study. *Eng. Proc.* **2024**, *6*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Published: 26 November 2024

1. Introduction

Machine learning is a valuable decision-making tool for predicting agricultural yields and deciding the type of crops to sow and things to do during the crop growing season. To aid crop prediction studies, several machine learning methods have been used. Machine learning techniques are utilized in various sectors, from evaluating customer behaviour in supermarkets to predicting customer phone usage. For some years, agriculture has been using machine learning techniques. Crop prediction is one of agriculture's complex challenges, and several models have been developed and proven so far. Because crop production is affected by many factors such as atmospheric conditions, type of fertilizer, soil, and seed, this challenge necessitates using several datasets [1–4]. This implies that predicting agricultural productivity is not a straightforward process; rather, it entails a series of complicated procedures. Crop yield prediction methods can now reasonably approximate the actual yield, although more excellent yield prediction performance is still desired. Agriculture, as the cornerstone of human civilization, faces unprecedented challenges in the 21st century due to factors such as climate change, population growth, and resource scarcity. To address these challenges, modern technologies, particularly data-driven approaches, are increasingly becoming integral to sustainable and efficient agricultural practices.

In the realm of agriculture, predicting crop yields is crucial for effective farming practices and resource management. This project, titled "Agricultural Crop Prediction Using Advanced Data Analysis Techniques—A Case Study", employs innovative machine learning methods to forecast crop yields based on numerous factors, including nitrogen, phosphorous, potassium, rainfall, temperature, and fertilizer application.

In other terms, we aim to use the power of computers to analyse large sets of data related to farming conditions and develop smart predictions about how much crops can be expected. Think of it like having a virtual assistant for farmers that helps them anticipate how well their crops will grow based on specific elements like soil nutrients, weather patterns, and fertilizer usage. Our project utilizes two main machine learning techniques: Multilinear Regression and Radial Basis Function. These methods allow us to understand the relationships between distinct factors and crop yields. Multilinear Regression helps us with straightforward connections, while Radial Basis Function handles more complex, non-linear patterns in the data.

The motivation behind this research stems from the urgent need to empower farmers with reliable tools for decision-making. Accurate predictions of crop yields based on environmental factors are crucial for optimizing resource allocation, enhancing productivity, and mitigating the impact of unforeseen challenges.

By doing this research, we hope to empower farmers with better insights into their crop production. Imagine farmers having a tool that considers various aspects of their farming environment and gives them a heads-up on what to expect[5–8]. This not only helps farmers plan more effectively but also contributes to the broader goal of improving agriculture using advanced data analysis techniques.

Agricultural crop prediction is a crucial factor in maintaining crop yields and ensuring food security. This presentation will cover the purpose and importance of crop prediction, as well as a literature survey, and a proposed solution to the problem.

Crop Prediction

By analyzing the soil and atmosphere conditions at region to have more crop yield and the net crop yield can be predicated. Crop prediction, there are different techniques or algorithm, and with the help of those algorithms we can predict crop yield and we are using Multilinear regression and Radial basis function algorithms.

1.1. Objectives

1. To increase the accuracy of crop yield prediction.
2. To study and analyze the different parameters which influences the crop yield prediction such as rainfall, fertilizers, nitrogen, phosphorous, potassium and temperature.
3. To provide the recommendations such as amount of water, fertilizers, pesticides required for the crop yield, to the farmers on crop prediction.

1.2. Scope of the Project

The future scope of the "Agricultural Crop Yield Prediction using Advance Data Analysis Techniques" project is promising. It entails advancing predictive models for crop yield, disease detection, and optimal cultivation practices. Integration with emerging technologies like remote sensing, IoT, and AI-driven decision support systems will enhance precision farming. Collaboration with agricultural experts, government agencies, and industry stakeholders can lead to scalable solutions for sustainable agriculture. Continuous refinement of algorithms and data sources, coupled with user-friendly interfaces, will ensure widespread adoption. Ultimately, the project has the potential to revolutionize farming practices, increase productivity, and contribute to global food security.

1.3. Methodology

The flowchart consists of a series of steps that outline the process for analysing the relationship between weather variables and crop harvest, and then using that information to make predictions about crop yield. Here is a breakdown of each step:

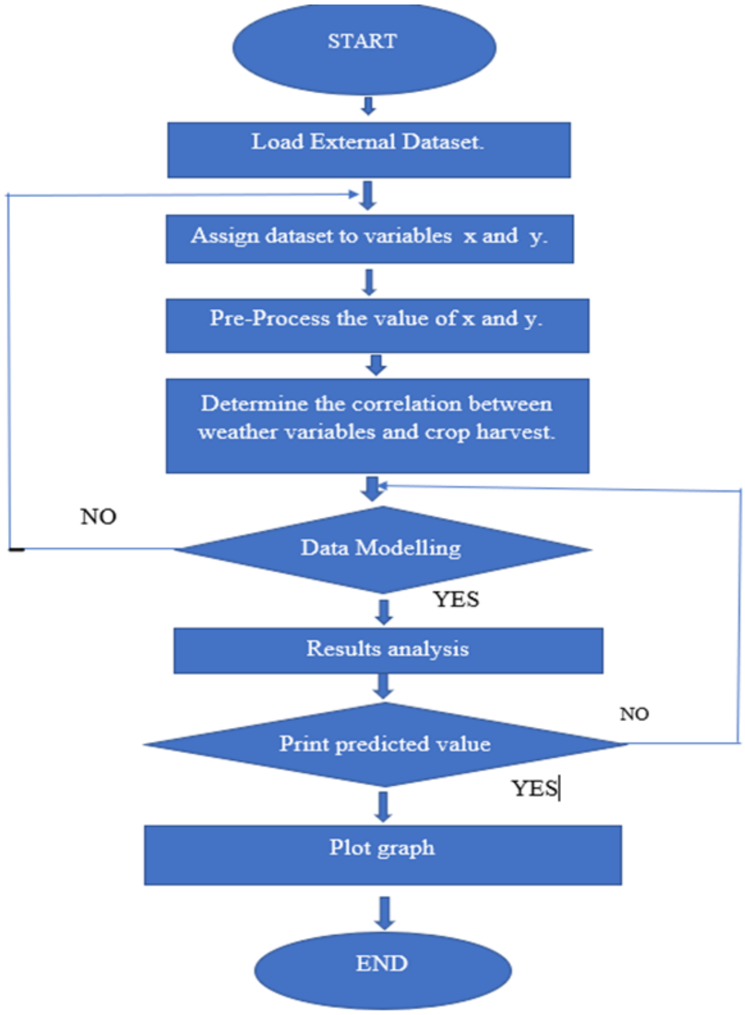


Figure 1. 1. Methodology.

Load External Dataset: The first step in the process is to load an external dataset that contains information about weather variables and crop harvest.

Assign dataset to variables x and y: This dataset will be assigned to variables x and y for further processing.

Pre-Process the value of x and y: Before determining the correlation between weather variables and crop harvest, the data needs to be pre-processed. This step may involve cleaning the data, handling missing values, and normalizing the data so that it can be more easily analysed.

Determine the correlation between weather variables and crop harvest: Once the data has been pre-processed, the next step is to determine the correlation between weather variables and crop harvest. This will help to identify which weather variables have the greatest impact on crop yield.

Data Modelling: After determining the correlation between weather variables and crop harvest, the next step is to create a data model that can be used to make predictions about crop yield based on weather data. If the data model is suitable for the dataset, then it proceeds to the next step otherwise again it starts from Assign dataset to variables x and y.

Results analysis: Once the data model has been created, the next step is to analyse the results to ensure that the model is accurate and reliable. This may involve testing the model with a separate dataset and evaluating its performance.

Print predicted value: After analysing the results, the next step is to print the predicted value for crop yield based on the data model. This value can be used to inform agricultural planning and decision-making. If the predicted value is correct it proceeds to the next step otherwise again it starts from Data modelling.

Plot graph: In addition to printing the predicted value for crop yield, the flowchart also indicates that a graph should be plotted to visualize the relationship between weather variables and crop harvest. This can help to identify trends and patterns in the data.

2. Methods

The flowchart consists of a series of steps that outline the process for analysing the relationship between weather variables and crop harvest, and then using that information to make predictions about crop yield. Here is a breakdown of each step:

Load External Dataset: The first step in the process is to load an external dataset that contains information about weather variables and crop harvest.

Assign dataset to variables x and y: This dataset will be assigned to variables x and y for further processing.

Pre-Process the value of x and y: Before determining the correlation between weather variables and crop harvest, the data needs to be pre-processed. This step may involve cleaning the data, handling missing values, and normalizing the data so that it can be more easily analysed.

Determine the correlation between weather variables and crop harvest: Once the data has been pre-processed, the next step is to determine the correlation between weather variables and crop harvest. This will help to identify which weather variables have the greatest impact on crop yield.

Data Modelling: After determining the correlation between weather variables and crop harvest, the next step is to create a data model that can be used to make predictions about crop yield based on weather data. If the data model is suitable for the dataset, then it proceeds to the next step otherwise again it starts from Assign dataset to variables x and y.

Results analysis: Once the data model has been created, the next step is to analyse the results to ensure that the model is accurate and reliable. This may involve testing the model with a separate dataset and evaluating its performance.

Print predicted value: After analysing the results, the next step is to print the predicted value for crop yield based on the data model. This value can be used to inform agricultural planning and decision-making. If the predicted value is correct it proceeds to the next step otherwise again it starts from Data modelling.

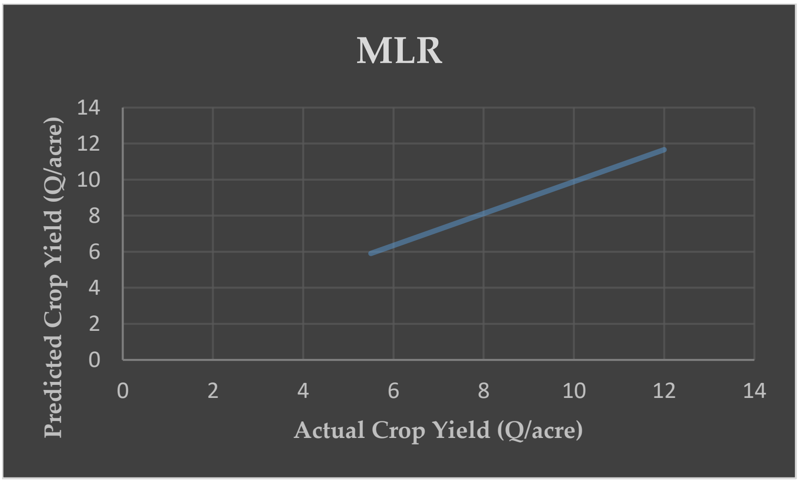
Plot graph: In addition to printing the predicted value for crop yield, the flowchart also indicates that a graph should be plotted to visualize the relationship between weather variables and crop harvest. This can help to identify trends and patterns in the data.

3. Results and Discussion

3.1. Data Interpretation Using Multilinear Regression

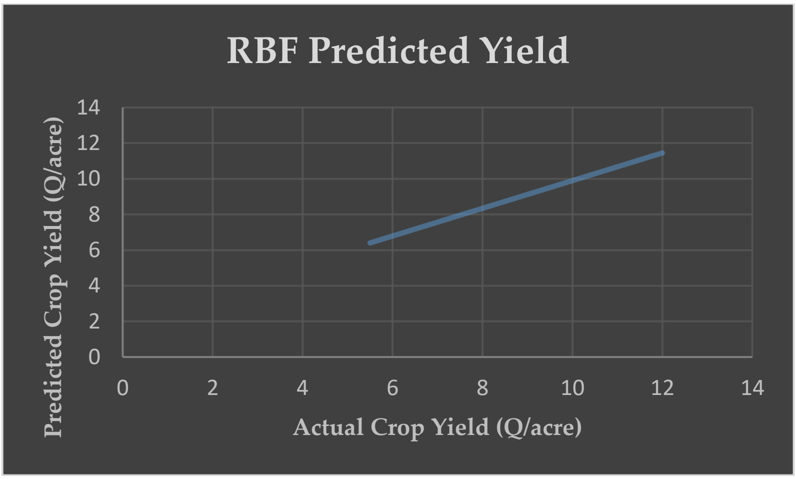
The Figure 3a represents the actual crop yield versus the predicted crop yield. The R-squared score for this model is 0.88, indicating a strong correlation between the actual and predicted crop yields. The graph displays two sets of data points, representing the actual crop yield (in Q/acre) and the predicted crop yield (in Q/acre) obtained from the model. The dots in the graph represent individual data points, with the x-axis showing the actual crop yield and the y-axis showing the predicted crop yield. A linear regression line (in blue) has been plotted to visualize the relationship between the actual and predicted crop yields. The R-squared value of 0.88 indicates that approximately 88% of the variation in the actual crop yield can be explained by the model's predictions. Ideally, the data points should be close to the regression line, indicating a strong fit between the actual and predicted values. In this graph, most of the data points are close to the regression line, suggesting that the model is a good fit for this case [9–12]. A high distance between the data points and the regression line would indicate a poor fit, suggesting that the model may not accurately predict crop yields.

r2_score = 0.88561



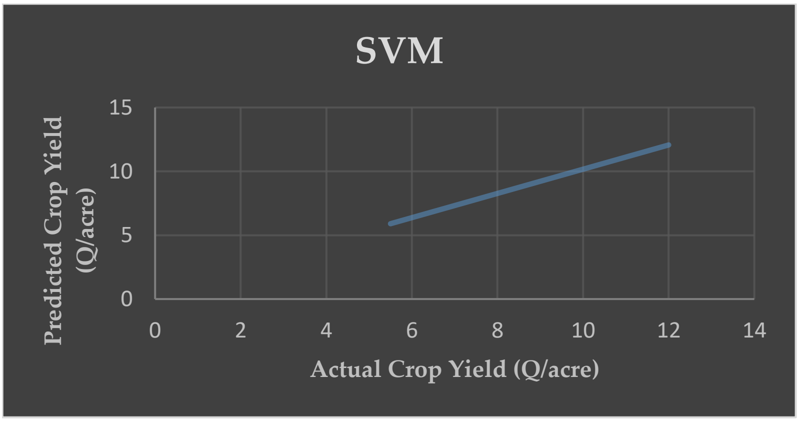
(**a**)

r2_score = 0.753604



(**a**)

r2_score = 0.920335



(**c**)

Figure 3. (**a**) Actual crop yield versus the predicted crop yield using MLR, (**b**) Actual crop yield versus the predicted crop yield using RBF, (**c**) Graph of SVM.

Based on the R-squared value and the proximity of data points to the regression line, the model appears to be a good fit for predicting crop yields in this agricultural case study. This graph can be used to demonstrate the model's performance and its ability to accurately predict crop yields.

3.2. Data Interpretation Using Radial Basis Function

The Figure 3b represents the actual crop yield versus the predicted crop yield for the project titled "Agricultural Crop yield prediction using Advanced Data analysis techniques-Case study." The R-squared score for this model is 0.753604, indicating a moderate to strong correlation between the actual and predicted crop yields. The graph displays two sets of data points, representing the actual crop yield (in Q/acre) and the predicted crop yield (in Q/acre) obtained from the model. The dots in the graph represent individual data points, with the x-axis showing the actual crop yield and the y-axis showing the predicted crop yield. A linear regression line (in blue) has been plotted to visualize the relationship between the actual and predicted crop yields. The R-squared value of 0.753604 indicates that approximately 75.36% of the variation in the actual crop yield can be explained by the model's predictions. Ideally, the data points should be close to the regression line, indicating a strong fit between the actual and predicted values. In this graph, most of the data points are relatively close to the regression line, suggesting that the model is a moderate to good fit for this case. However, there are some data points that are farther away from the regression line, indicating that the model may not accurately predict crop yields in certain cases. Based on the R-squared value and the proximity of most data points to the regression line, the model appears to be a moderate to good fit for predicting crop yields in this agricultural case study. However, there are some instances where the model may not accurately predict crop yields, as indicated by the data points that are farther away from the regression line. This graph can be used to demonstrate the model's performance and its ability to accurately predict crop yields, while also acknowledging the limitations of the model.

3.3. Data Interpretation Using Support Vector Machine

The Figure 3c represents the actual crop yield versus the predicted crop yield for the project titled "Agricultural Crop yield prediction using Advanced Data analysis techniques-Case study". The R-squared score for this model is 0.920335, indicating an extraordinarily strong correlation between the actual and predicted crop yields. The graph displays two sets of data points, representing the actual crop yield (in Q/acre) and the predicted crop yield (in Q/acre) obtained from the model. The dots in the graph represent individual data points, with the x-axis showing the actual crop yield and the y-axis showing the predicted crop yield. A linear regression line (in blue) has been plotted to visualize the relationship between the actual and predicted crop yields. The R-squared value of 0.920335 indicates that approximately 92.03% of the variation in the actual crop yield can be explained by the model's predictions. Ideally, the data points should be close to the regression line, indicating a strong fit between the actual and predicted values. In this graph, most of the data points are very close to the regression line, suggesting that the model is an excellent fit for this case. The proximity of the data points to the regression line indicates that the model is highly accurate in predicting crop yields.

Based on the R-squared value and the proximity of most data points to the regression line, the model is an excellent fit for predicting crop yields in this agricultural case study. The high R-squared value and the proximity of the data points to the regression line suggest that the model is highly accurate in predicting crop yields[12–14]. This graph can be used to demonstrate the model's performance and its ability to accurately predict crop yields.

4. Conclusions

In conclusion, the "Agricultural Crop Yield Prediction using Machine Learning" project stands as a pivotal tool in revolutionizing modern agriculture. The accurate predictions and insights provided by this project empower farmers with data-driven decision-making capabilities, enhancing efficiency, and overall productivity in the agricultural sector. The project has successfully demonstrated the potential of machine learning to predict agricultural crop production. By implementing advanced machine learning algorithms, the project has successfully identified patterns and relationships in the data that can be used to make informed predictions about future crop yields. Implementation of an algorithms using machine learning to predict the yield rate of the crops production. Based on the R-squared value and the proximity of most data points to the regression line, the model appears to be an excellent fit for predicting crop yields in this agricultural case study. The high R-squared value and the proximity of the data points to the regression line suggest that the model is highly accurate in predicting crop yields. According to our analysis using different machine learning algorithms, the suitable and best algorithm for agriculture crop yield prediction is **SUPPORT VECTOR MACHINE.** This project will help the farmers to know the yield of their crop before cultivating onto the agricultural field and help them to take the appropriate decisions. Ultimately, the project has the potential to revolutionize farming practices, increase productivity, and contribute to global food security.

Author Contributions:

Funding:

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Conflicts of Interest:

References

1. Joshua, V.; Priyadharson, S.M.; Kannadasan, R. Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamil Nādu. *Agronomy* **2021**, *11*, 2068.
2. Ansarifar, J.; Wang, L.; Archontoulis, S. An interaction regression model for crop yield prediction. *Nat. Portf. Sci. Rep.* **2021**, *11*, 17754.
3. Elbasi, E.; Zaki, C.; Topcu, A.E.; Abdelbaki, W.; Zreikat, A.I.; Cina, E.; Shdefat, A.; Saker, L. Crop Prediction Model Using Learning Algorithms. *Appl. Sci.* **2023**, *13*, 9288.
4. Benos, L.; Tagarakis, A.C.; Dolias, G.; Berruto, R.; Kateris, D.; Bochtis, D. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* **2021**, *21*, 3758.
5. Gupta, P.; Tyagi, N.; Kumar, V.; Kumar, A. Crop Yield Prediction using Machine Learning Techniques: A Review. *Sustainability* **2022**, *14*, 1222.
6. Karthik, G.; Jayanthu, S. Review on low-cost wireless communication systems for slope stability monitoring in opencast mines. *Int. J. Min. Mineral Eng.* **2018**, *9*, 21–31.
7. Jayanthu, S.; Karthik, G.; Shohood, A.P.M.G. Development of indigenous wireless tiltmeter for slope stability monitoring in opencast mines. 2016.
8. Yadav, D.K.; Karthik, G.; Jayanthu, S.; Das, S.K. Design of real-time slope monitoring system using time-domain reflectometry with wireless sensor network. *IEEE Sens. Lett.* **2019**, *3*, 2500304.
9. Karthik, G.; Jayanthu, S.; Rammohan, P. and Rahman, A. Utilisation of mobile communication in opencast mines. *Int. J. Comput. Sci. Mob. Comput.* **2014**, *3*, 373–378.
10. Guntha, K.; Jayanthu, S. Mathematical modelling of engineering problems. *J. Homepage* **2018**, *5*, 256–259. http://iieta.org/Journals/MMEP.
11. Kumar, V.; Gupta, P.; Kumar, A.; Tyagi, N. Prediction of Crop Yield using Machine Learning Techniques: A Comparative Analysis. *Sustainability* **2022**, *14*, 2828.
12. Mahajan, G.; Saini, A.; Kumar, V. A Review on Crop Yield Prediction using Machine Learning Techniques. In 2022 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 25–27 May 2022; pp. 1–5.

13. Prakash, A.; Kumar, V.; Gupta, P.; Kumar, A.; Tyagi, N. Crop Yield Prediction using Machine Learning Techniques: A Case Study of Wheat Crop in India. *Sustainability* **2022**, *14*, 1405.
14. Rathore, P.S.; Gupta, R.; Kumar, V.; Kumar, A.; Tyagi, N. Crop Yield Prediction using Machine Learning Techniques: A Case Study of Rice Crop in India. *Sustainability* **2022**, *14*, 1388.