Crop Yield Prediction

A Project Report

submitted in partial fulfillment of the requirements

of

Industrial Artificial Intelligence With Cloud Computing

by

Nikesh Jain, 220841102502

Ayushi Patel, 220841102503

Archi Borad, 210841102005

Prachi Pancholi, 210841102027

Under the Esteemed Guidance of

Abdul Aziz Md, Master Trainer, Edunet Foundation.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have contributed to the completion of this project. Without their support, guidance, and encouragement, this endeavor would not have been possible.

First and foremost, I extend my heartfelt thanks to Abdul Aziz Md, Master Trainer, Edunet Foundation, whose expertise and mentorship provided invaluable insights throughout the duration of this project. Their unwavering support and commitment played a pivotal role in shaping its outcome.

I am also immensely grateful to Vivek Joshi sir, and Foram Sukla Ma'am for their dedication and contributions at various stages of the project. Their collaborative efforts, feedback, and constructive criticism significantly enhanced the quality and effectiveness of our work.

I would also like to acknowledge the support of and friends, who provided encouragement, understanding, and patience throughout this journey. Their unwavering belief in me served as a constant source of motivation, and for that, I am truly grateful.

Thank you to everyone who contributed in any way, no matter how big or small. Your support has been invaluable, and I am deeply appreciative of the opportunity to work alongside such talented and dedicated individuals.

ABSTRACT

Predicting crop yields is crucial for resource management and agricultural planning. In this work, we create a prediction model to calculate crop yields according to several agricultural and environmental parameters. The dataset includes data on variables like temperature, crop type, geographic location, average rainfall, and pesticide use. First, actions related to data preprocessing such as feature engineering, data cleansing, and addressing missing values are carried out. Techniques for exploratory data analysis are used to learn more about how data is distributed across various variables. The preprocessed data is then used to train machine learning models, such as Decision Tree Regression, Lasso Regression, Ridge Regression, and Linear Regression. The predictive accuracy of the models is evaluated using performance evaluation metrics like Mean Absolute Error (MAE) and R-squared score. The model with the best performance among the models examined is the Decision Tree Regression model. Lastly, a function that predicts fresh data inputs is created, enabling users to calculate crop yields depending on predetermined criteria. The pickle library is used to serialize the learned model and preprocessing pipeline for later usage. This research offers a useful framework for predicting agricultural yield, which will help farmers make well-informed decisions.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1. Problem Statement:

Predicting crop yields is essential for managing and planning agriculture. Accurate forecasts are essential for farmers, policymakers, and stakeholders to make well-informed decisions regarding crop selection, resource allocation, and risk mitigation techniques. This research aims to create a prediction model that calculates crop production according to several agricultural and environmental parameters.

1.2. Problem Definition:

Data Acquisition: Compile extensive historical data on crop productivity, considering factors like temperature, crop type, typical rainfall, pesticide use, and geographic location. For improved prediction accuracy, include additional external data sources such as agricultural reports, satellite images, and soil quality indicators.

Feature Engineering: Determine the pertinent elements, such as past weather patterns, the properties of the soil, and agricultural methods, that have a major impact on crop output. To capture complex associations, extract significant features from the raw data, such as moving averages, interaction terms, and lag variables.

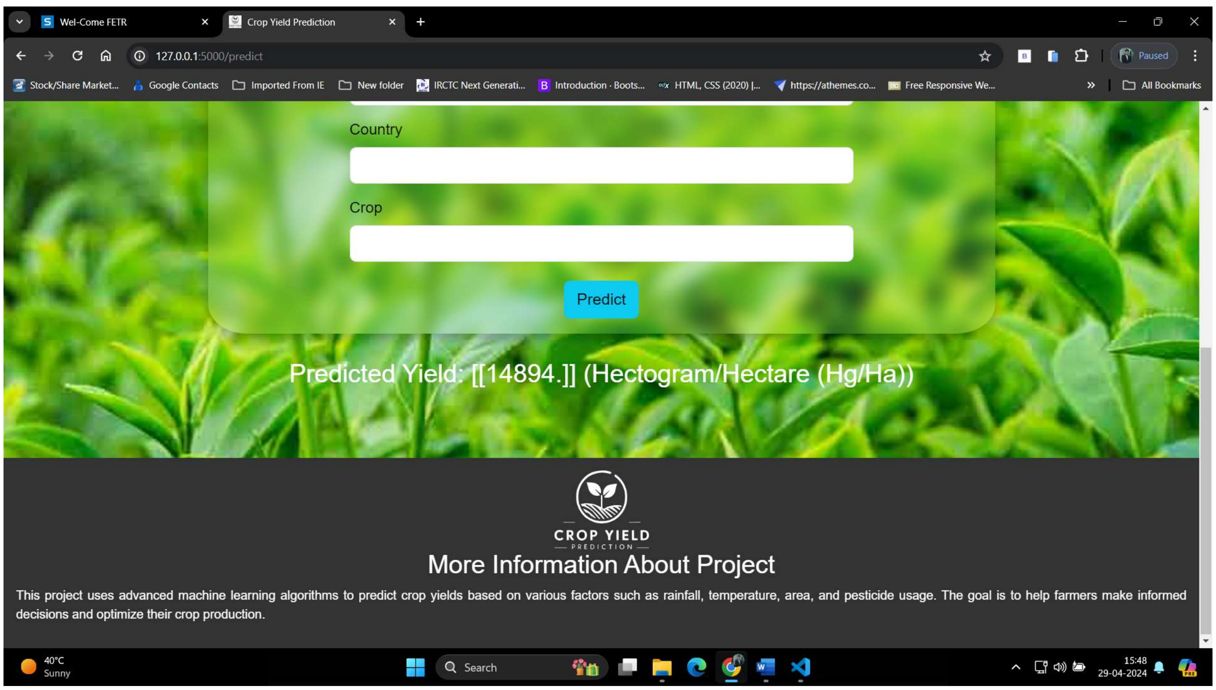
Model Development: Try out different machine learning techniques, including decision trees, Lasso, Ridge regression, linear regression, and random forests, that are appropriate for regression tasks. Utilizing methods like cross-validation to optimize hyperparameters and avoid overfitting, train models on the preprocessed dataset. To predict patterns in agricultural yield across a number of time periods, think about using time-series forecasting techniques.

Model Evaluation: Utilize measures such as coefficient of determination (R2), mean absolute error (MAE), and root mean square error (RMSE) to assess the effectiveness of trained models. To evaluate the resilience and generalizability of the model across several geographic locations and crop kinds, do cross-validation. To find the best model for predicting crop yield, compare the results of several algorithms.

Deployment and Integration: Provide crop production projections to stakeholders, including farmers, agricultural researchers, and policymakers, by integrating the trained model into an intuitive application or web interface. For real-time monitoring and assistance with decision-making, make sure there is a smooth interaction with any current agricultural systems or mobile applications. In order to clarify model predictions and draw attention to important aspects influencing crop yield, include interpretability features.

1.3. Expected Outcomes:

it aims to develop a robust prediction model leveraging machine learning techniques to accurately forecast crop yields based on various agricultural and environmental parameters. Through extensive data preprocessing, including feature engineering and addressing missing values, the model aims to enhance its predictive accuracy. The project also focuses on exploring different machine learning algorithms, such as Decision Tree Regression, Lasso Regression, Ridge Regression, and Linear Regression, to identify the most suitable model for crop yield prediction. Additionally, the project aims to deploy the developed model into a user-friendly interface, enabling stakeholders like farmers, policymakers, and agricultural researchers to access real-time crop yield projections. Overall, the project endeavors to provide a useful framework for predicting agricultural yield, empowering stakeholders to make informed decisions in resource management and agricultural planning.



(Fig 1.1)

1.4. Organization of the Report

1. **Introduction:** The issue statement summarizes: Using AI and ML technologies to predict crop productivity. For farmers, legislators, and other agricultural stakeholders, precise crop output forecasts are crucial report objectives and important analysis components.
2. **Data Acquisition and Preprocessing:** An explanation of the variables and sources included in the dataset used to forecast crop yield. Methods for cleaning data to address inconsistent formats, missing values, and duplication. Preprocessing procedures to get the data ready for modeling, include feature encoding, scaling, and normalization.
3. **Exploratory Data Analysis (EDA):** summaries and visualizations help comprehend trends, geographic variances, and agricultural yield distribution. Examination of the main variables affecting crop yield, such as temperature fluctuations, rainfall patterns, and farming methods.
4. **Feature Engineering:** Finding pertinent features and applying transformation methods to improve predictive modeling. To increase prediction accuracy, new features can be created or useful information can be taken out of variables that already exist.
5. **Model Development:** Using machine learning techniques, such as ensemble approaches, decision trees, and linear regression, to forecast crop productivity. Models are trained and evaluated using the proper criteria to determine their performance and capacity for generalization. Cross-validation methods and hyperparameter tweaking are used to maximize model performance.
6. **Results and Discussion:** Results of the model are presented, along with predicted insights, feature importance, and accuracy metrics. Key findings, their significance for agriculture, and possible directions for further research are discussed. Comparison of several models and how well suited they are to tasks involving the prediction of agricultural yield.
7. **Deployment and Integration:** Techniques for integrating the agricultural production forecast system into useful tools or decision assistance. For user accessibility, take into account connecting the predictive models with current agricultural platforms or systems. Recommendations that farmers, legislators, and other stakeholders should use and put into practice.
8. **Conclusion:** a summary of the major discoveries and advancements in agricultural yield prediction science. A consideration of how well AI and ML technology can solve problems in agriculture. Prospects for future study, invention, and agricultural predictive modeling use.
9. **References:** datasets, and methodologies used in the report.

CHAPTER 2

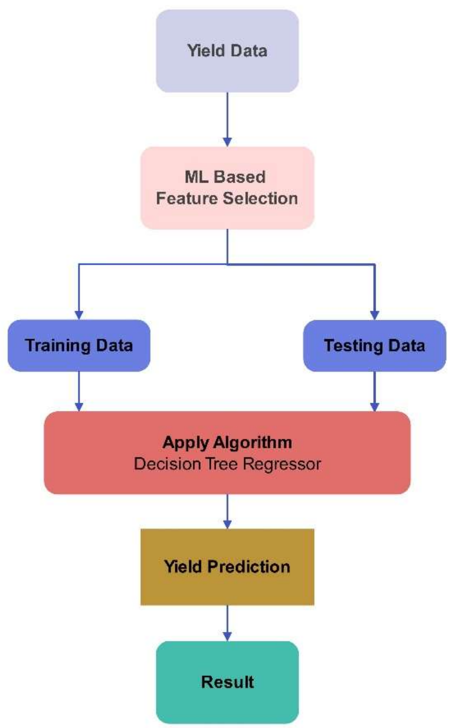
PROPOSED METHODOLOGY

CHAPTER 2

PROPOSED METHODOLOGY

2.1 Data Flow Diagram

A Data Flow Diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. DFDs can also be used for the visualization of data processing (structured design).



2.2Advantages

Ease of Use: Users can easily engage with the model using the user-friendly interfaces of web apps, even without requiring considerable technical knowledge.

Real-time Prediction: With the input of pertinent information like temperature, precipitation, year, and so forth, users can obtain real-time crop yield projections. They can use this to help them plan and manage their crops more intelligently.

Integration: Web applications can be integrated with other tools and services, such as weather APIs or agricultural databases, to enhance functionality and provide more comprehensive insights to users.

Scalability: Scaling up web applications is possible to accommodate higher processing demands or enormous user counts. The program can be readily expanded to handle changes in the user base or the addition of new features.

Continuous Improvement: The model can be updated and improved continuously thanks to web apps. Updates, bug fixes, and new features can be released by developers without any problems, guaranteeing that consumers always have access to the most recent version of the model.

2.3 Requirement Specification

2.5.1 Hardware Requirements:

- **Processor:** i5 Intel 11250H or Ryzen 5
- **Hard Disk:** 500 GB
- **Ram:** 8 GB

2.5.2 Software Requirements:

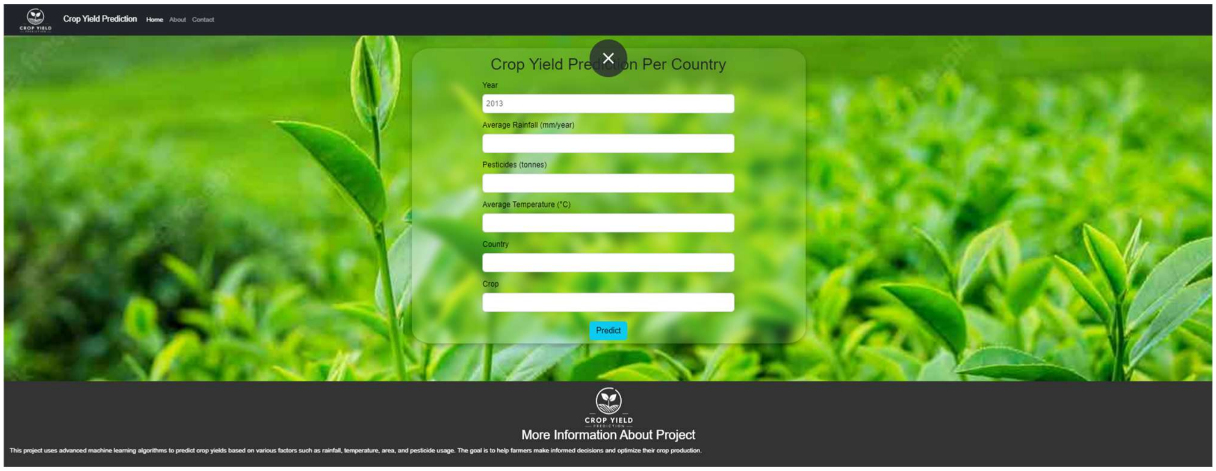
- **Operating System:** Windows 11
- **Language:** Python
- **Libraries:** NumPy, Pandas, Flask, Scikit-learn (Sklearn), Pickle
- **IDE:** Visual Studio Code, PyCharm, Anaconda

CHAPTER 4

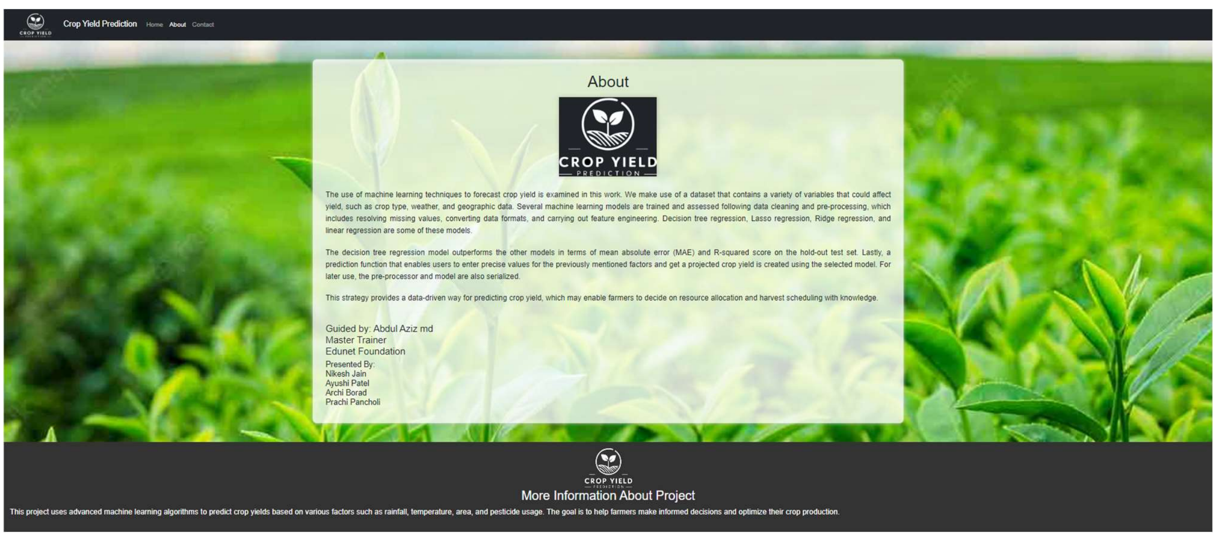
Implementation and Result

CHAPTER 4

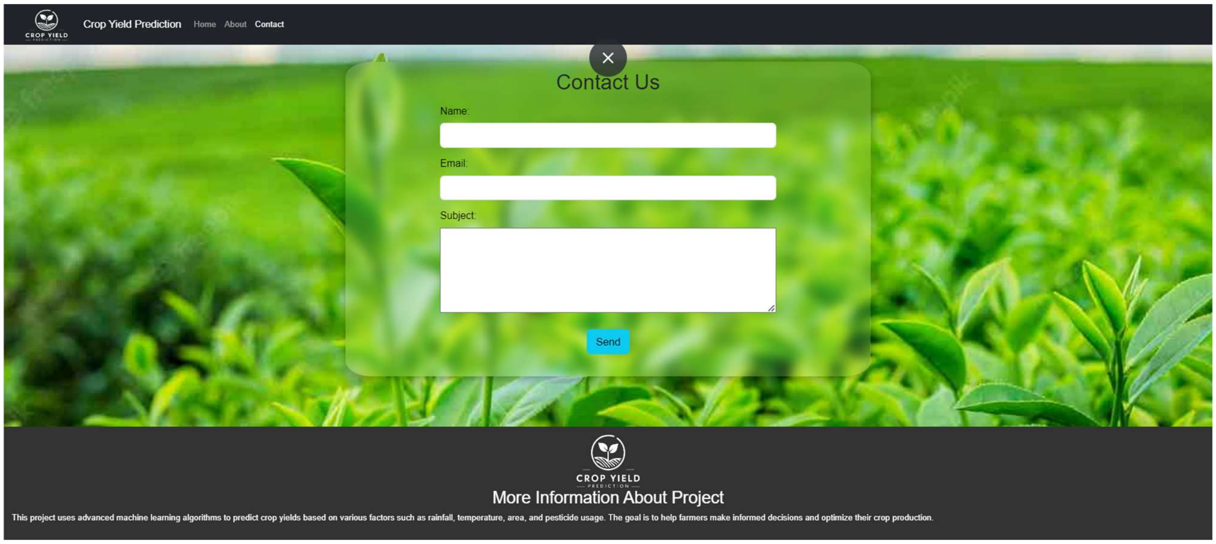
IMPLEMENTATION and RESULT



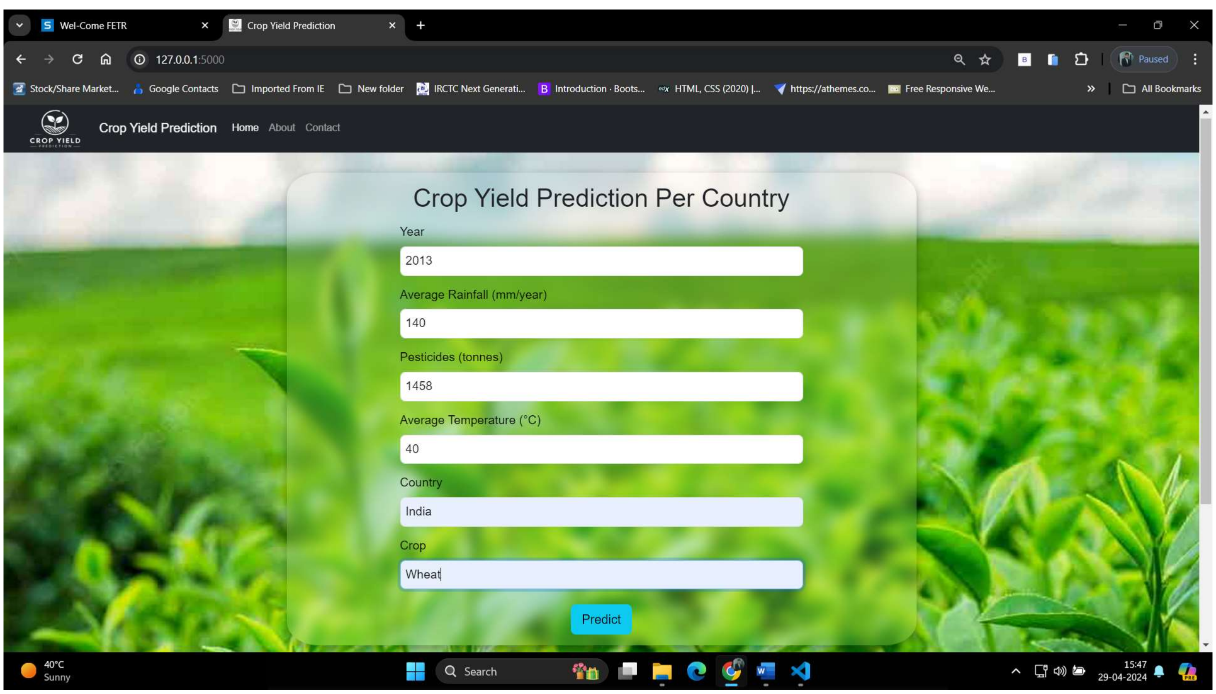
(Fig 4.1)



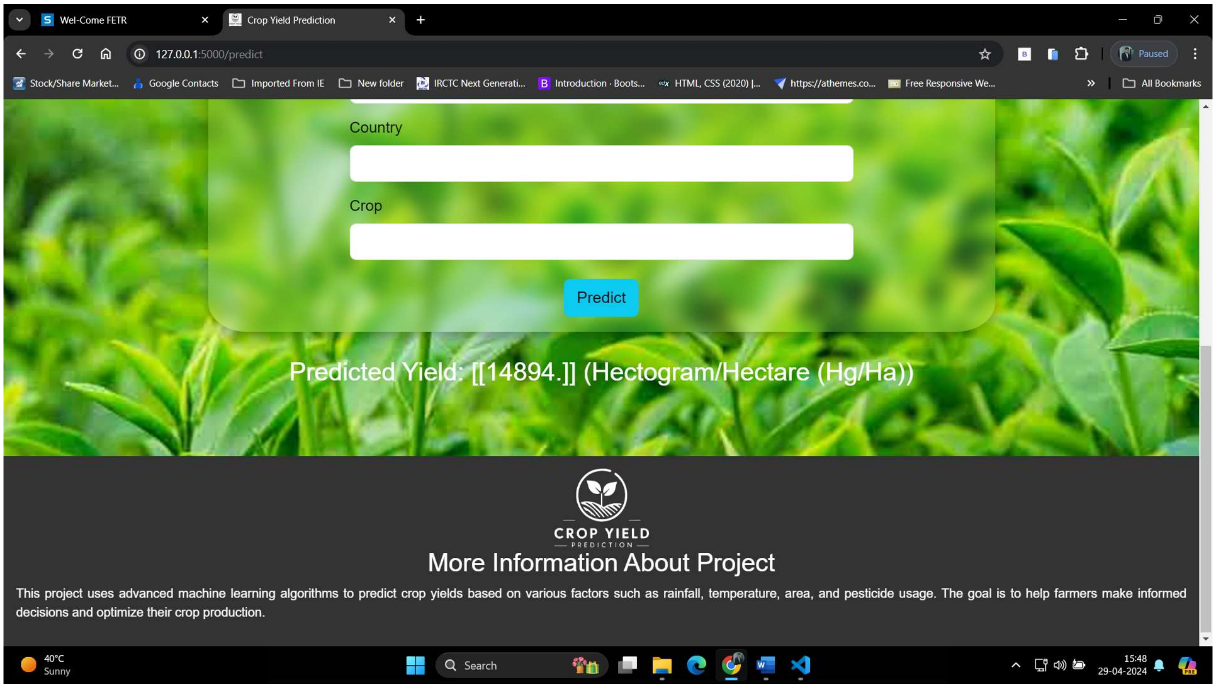
(Fig 4.2)


Crop Yield Prediction
Home
About
Contact
Contact Us
Name:
Email:
Subject:
Send
CROP YIELD
More Information About Project
This project uses advanced machine learning algorithms to predict crop yields based on various factors such as rainfall, temperature, area, and pesticide usage. The goal is to help farmers make informed decisions and optimize their crop production.

(Fig 4.3)


Wel-Come FETR
Crop Yield Prediction
127.0.0.1:5000
Crop Yield Prediction
Home
About
Contact
Crop Yield Prediction Per Country
Year
2013
Average Rainfall (mm/year)
140
Pesticides (tonnes)
1458
Average Temperature (°C)
40
Country
India
Crop
Wheat
Predict

(Fig 4.4)


Wel-Come FETR
Crop Yield Prediction
127.0.0.1:5000/predict
Country
Crop
Predict
Predicted Yield: [[14894.]] (Hectogram/Hectare (Hg/Ha))
CROP YIELD
PREDICTION
More Information About Project
This project uses advanced machine learning algorithms to predict crop yields based on various factors such as rainfall, temperature, area, and pesticide usage. The goal is to help farmers make informed decisions and optimize their crop production.

(Fig 4.5)

CHAPTER 5

CONCLUSION

CHAPTER 5

CONCLUSION

In summary, crop yield prediction is extremely important to modern agriculture because it provides insightful information about production trends and helps researchers, farmers, and policymakers make well-informed decisions. By employing sophisticated data analytics methods, machine learning algorithms, and agronomy-specific domain expertise, we can leverage the potential of data to predict crop yields more precisely and consistently. Crop yield prediction makes proactive management measures easier to implement, giving farmers the ability to better allocate resources, customize cultivation techniques, and reduce risks related to pests, diseases, and the environment. Predictive models can produce projections that help with crop planning, market analysis, and food security by utilizing historical yield data in conjunction with a wide range of environmental and agronomic variables, including weather patterns, soil characteristics, crop type, and management approaches.

SCOPE:

Enhanced Predictive Models: More advancements in artificial intelligence, statistical modeling, and machine learning research and development may result in more reliable and accurate crop yield prediction models. Model performance may be further enhanced by integrating cutting-edge approaches like reinforcement learning, deep learning, and ensemble methods.

Big Data Analytics: The agricultural industry has witnessed a surge in data collection technologies, leading to an abundance of information on various aspects affecting crop yields. These elements include weather patterns, soil composition, agricultural methods, and genetic data. Large and heterogeneous datasets can be usefully analyzed using advanced analytics approaches like data fusion and big data analytics to generate actionable insights that improve predictive power and comprehension of intricate correlations between variables.

Predictive Analytics for Climate Change Adaptation: With variations in temperature, precipitation patterns, and extreme weather events affecting crop yields, climate change is predicted to have a substantial influence on agricultural output. Predictive models that take climate change projections into account can assist farmers and policymakers in anticipating and preparing for future issues, including altered agricultural patterns, altered dynamics of pests and diseases, and altered management of water resources.

REFERENCES

[1]. https://chat.openai.com/.
[2]. https://youtu.be/vGOvUoqn_M8?si=wfypITp56JYVHT8c
[3]. https://github.com/611noorsaeed/Crop-Yield-Prediction-Using-Machin-Learning-Python.git
[4]. https://flask.palletsprojects.com/en/3.0.x/
[5]. https://streamlit.io/
[6]. https://getbootstrap.com/docs/5.0/getting-started/introduction/
[7]. https://www.canva.com/
[8]. https://www.smartdraw.com/
[9]. https://gemini.google.com/app