KLE Society's
KLE Technological University, Hubballi.

A Minor Project Report

on

Analysing the effect of climatic change on the Maize crop yield

submitted in partial fulfillment of the requirement for the degree of

Bachelor of Engineering

in

Computer Science and Engineering

Submitted by

Neha Patil	01FE20BCS006
Ranjita Hegde	01FE20BCS009
Srishti Kadam	01FE20BCS010
Chaitra Hegde	01FE20BCS045

Under the guidance of
Prof. Karibasappa K.G

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

Hubballi – 580 031

2022 -2023

KLE Society's
KLE Technological University, Hubballi.

2022 - 2023

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that Minor Project titled Analysing the effect of climatic change on the crop yield is a bonafied work carried out by the student team comprising of Neha Patil (01FE20BCS006), Ranjita Hegde (01FE20BCS009), Srishti Kadam (01FE20BCS010), Chaitra Hegde(01FE20BCS045) for partial fulfillment of completion of sixth semester B.E. in Computer Science and Engineering during the academic year 2022-23.

Guide — Head, SoCSE

Prof. Karibasappa KG — Dr. Meena S. M

Viva -Voce:

Name of the Examiners — Signature with date

1.

2.

Acknowledgement

We would like to thank our faculty and management for their professional guidance towards the completion of the project work. We take this opportunity to thank Dr. Ashok Shettar, Vice-Chancellor, Dr. B.S Anami, Registrar, and Dr. P.G Tewari, Dean Academics, KLE Technological University, Hubballi, for their vision and support.

We also take this opportunity to thank Dr. Meena S. M, Professor and Head, SoCSE for having provided us direction and facilitated for enhancement of skills and academic growth.

We thank our guide Prof.Karibasappa, SoCSE for the constant guidance during interaction and reviews.

We extend our acknowledgement to the reviewers for critical suggestions and inputs. We also thank Project Co-ordinator Mr.Uday N.Kulkarni and Mr. Guruprasad Konnuramath for their support during the course of completion.

We express gratitude to our beloved parents for constant encouragement and support.

Neha Patil - 01FE20BCS006
Ranjita Hegde - 01FE20BCS009
Srishti Kadam- 01FE20BCS010
Chaitra Hegde - 01FE20BCS045

ABSTRACT

Climate change has the potential to significantly impact maize crop growth, as changes in temperature and precipitation patterns can alter growing conditions and lead to decreased yields. Maize crops are particularly vulnerable to changes in temperature and rainfall, as they are highly dependent on these factors for optimal growth and development.Temperature increases are one of the most important effects of climate change on maize crops. As temperatures rise, the growing season for maize can shift, resulting in changes in planting and harvesting times.Changes in precipitation patterns can also impact maize crop growth. Changes in rainfall amounts or distribution can result in drought or flooding, both of which can negatively affect crop yields.The impact of climatic variations on the growth of crops can be predicted using machine learning methods like Random Forest, Gradient Boosting and Decision Tree. These algorithms can be trained using historical climate and yield data, allowing them to identify patterns and relationships between climatic variables and maize yields.This project aims to analyze the impact of climate change on maize crop yield using machine learning methods.We will use Random Forest,Decision tree,Gradient Boosting and XG Boosting algorithms in particular to foresee how different meteorological factors would affect maize yield. We have used historical climate and yield data to train our models and evaluate their accuracy using cross-validation techniques. The accuracy of RF,Gradient Boosting,Decision tree and XG Boosting are 95%,87%,93%,86% respectively.This results will provide insights into climate change's potential impact on maize production, which could inform future agricultural policies and practices.

Keywords : *Maize crop growth,Temperature, Rainfall,Gradient boosting,XG boosting*

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Maize is an important crop worldwide, providing a staple food source for millions of people. However, its production is vulnerable to changes in climate factors, including temperature, rainfall, and pesticide use. Changes in these variables can significantly impact the crop's growth and yield, potentially reducing the quality of the produce and negatively affecting global food security.

The impact of climate change on maize production is causing worry, as extreme weather events such as droughts, floods, and heat waves become more frequent and severe. In response[7], researchers are using advanced statistical techniques like gradient boosting to address this issue. This algorithm uses a group of decision trees to predict results, making it a useful tool for analyzing complex data and identifying the critical factors that affect maize production in various areas.

By utilizing gradient boosting and other statistical techniques, the purpose of this study is[8] to investigate how climate change may affect maize output. Specifically, it will investigate how changes in temperature, rainfall, and pesticide use impact the growth and development of maize crops. Additionally, the study[9] will consider other factors that may influence maize production, such as soil type, fertilizer use, and irrigation practices.

The analysis will use gradient boosting and other statistical techniques to examine the effects of climatic change on maize production. The study will explore how changes in temperature, rainfall, and pesticides affect the growth and development of maize crops. The study will also consider other factors such as soil type, fertilizer use, and irrigation practices that may influence maize production. Overall, this analysis will provide valuable insights into the impact of climatic change on maize production and help policymakers and farmers to develop strategies to mitigate the effects of climate change on food security

1.1 Motivation

One of the most significant cereal crops in the world and a key component of global food security is maize. However[10], maize yield is highly sensitive to environmental factors such as temperature and precipitation, and the ongoing climate change is already having a significant impact on maize production. Therefore, the impact of climate change on maize yield is a major concern for global food security and agricultural sustainability. Extreme weather events brought on by climate change, including droughts and floods, can harm crops and lower yields, costing farmers a significant amount of money, especially in developing nations. This, in turn, can have far-reaching consequences for global food prices.

Analyzing the impact of climate change on maize yield using machine learning can provide valuable insights into the potential economic and societal consequences of yield reductions, particularly in developing countries. Machine learning techniques offer a powerful and flexible approach to analyzing complex datasets, [11]making it possible to identify patterns and predict future outcomes. Applying machine learning techniques to predict and analyze the effects of climate change on maize crops can offer more precise and dependable information for a range of stakeholders, including farmers, policymakers, and other decision-makers. Such information can enable these stakeholders to make better-informed choices regarding crop management, agricultural planning, and other measures aimed at mitigating the impacts of climate change on agriculture. Ultimately [12], the insights provided by machine learning models can support more effective adaptation strategies for farmers and other stakeholders, helping to ensure food security and agricultural sustainability in the face of a changing climate. Moreover, applying a machine learning model to analyze the impact of climate change on maize yield represents an opportunity for scientific advancement in several ways. It can help to deepen our understanding of the complex relationships between climate change and agricultural productivity by exploring large and complex datasets using machine learning techniques to uncover patterns and relationships that may not be readily apparent using more traditional statistical approaches. Additionally,[13] the application of machine learning techniques to the study of agriculture and climate change can help to open up new avenues of research and inquiry, paving the way for further scientific advancements in this critical area.

In conclusion, the insights from a machine learning model for analyzing the impact of climate change on maize yield can have significant implications for informing policy decisions, improving agricultural sustainability, and supporting global food security. By providing more accurate and reliable information[14]on the potential impacts of climate change on agriculture, such a model can help to identify strategies to build more resilient and sustainable food systems and ensure that food remains accessible and affordable for people around the world.

1.2 Literature Review / Survey

Renhai Zhong, Yue Zhu et al[1] proposed a method for accurately estimating crop yields at large spatial scales, which is crucial for global food security. However, the impact of extreme climate stress on crop yields is still not fully understood. To address this challenge, they developed a new approach that utilizes deep learning to not only predict yield patterns but also detect and attribute the effects of climatic extremes on crop yields. The study focused on the US Corn Belt and aimed to estimate variations in maize yield at the county level from 2006 to 2018, with a particular focus on the extreme yield loss experienced in 2012. The researchers utilized a multi-task learning framework that incorporated various data sources, including weather, soil, and remote sensing data. They first used an imperceptible clustering technique called k-means clustering to divide the US Corn Belt into a number of homogeneous midine production zones. They next created a multi-task learning framework based on deep neural networks to extract spatiotemporal patterns for yield estimation.Artificial neural networks were used to extract features from soil data, while LASSO, Random Forest, and LSTM were the three baseline models used. Data on key yields at the country level from 12 Corn Belt states in the US's central and northern regions were included in the study's dataset. The researchers were able to get insight into the spatiotemporal patterns of maize yield by dividing the Com Belt into six generally homogenous sections using the Mins Chistering method. However, the study had some limitations, including its inability to robustly demonstrate the impacts of climatic stresses.The study's findings indicate that developing heat-tolerant genotypes and adapting to minimise the effects of extreme heat stress during crucial growth stages, such as early planting, could be useful adaptation strategies to reduce the risk of extreme yield losses like those seen in 2012. Moving forward, the authors suggest that integrating expert knowledge and process-based crop models with data-driven deep-learning approaches could improve crop yield modeling and advance machine intelligence in crop modeling under changing climate conditions.

Martin Kuradusenge, Eric Hitimana et al [2] suggested machine learning techniques for predicting crop yield, concentrating on the primary crops in the Musanze district of Rwanda, which are maize and Irish potatoes. The dataset utilized in this study included harvest and meteorological parameters collected from various sources spanning 2006 to 2021.To analyse the gathered data, three models—Random Forest, Polynomial Regression, and Support Vector Regressor—were used. With a root mean square error of 510.8 and 129.9 for maize and potatoes, respectively, and an R2 of 0.875 and 0.817 for the same crop datasets, the study indicated that Random Forest was the most successful model. Nevertheless, limited data availability on crucial parameters such as air humidity, soil moisture, and solar radiation posed a challenge to the study.Despite this drawback, the study highlighted the need of early

yield information sharing to lower food insecurity, and the results will be used to create a crop production forecast system combining IoT and machine learning. The importance of determining each weather parameter's feature importance and studying the relationship between those factors and crop production,selecting the best machine learning model for predicting crop production underscores the need for more comprehensive data to improve the accuracy of crop yield prediction models. The adoption of IoT and machine learning for predicting crop yields could play a crucial role in mitigating the impact of climate change on agriculture and ensuring food security.

Florian Schierhorn, Max Hofmann et al[3] proposed a Random Forest machine learning model to assess the impact of climate and weather on winter wheat yields in Ukraine, which outperformed traditional regression-based methods in terms of explanatory power. The study collected daily measurements of temperature, precipitation, and snow cover from 190 meteorological stations in Ukraine from 1985 to 2018. Two approaches were used,Specifically, the fixed threshold approach and the percentile threshold approach.RF models with both intraseasonal climatic means and weather extremes performed better than models with only season-long variables. The models captured a significant portion of the variability in wheat yields, demonstrating the importance of considering both climatic means and weather extremes for predicting crop yields in Ukraine. The models captured 54% of the wheat yield variability country-wide, 58% in the Northwest, and 49% in the Southeast were calculated using the percentile threshold method, Weather extremes as well as climatic factors were considered. This study highlights the importance of considering both climatic means and weather extremes in predicting crop yields in Ukraine, which can help inform the development of early warning systems and climate-smart agricultural policies to mitigate the effects of extreme weather events on crop yields. However, one limitation of the study was the lack of comparison with other machine learning models. Nonetheless, the results demonstrate the potential of machine learning algorithms such as Random Forest to predict the impact of climate change on crop yields, which is critical for ensuring global food security. Overall, this study contributes to our understanding of the complex relationship between climate, weather, and crop yields, and how we can use machine learning to predict and mitigate the effects of climate change on agriculture.

Andrew Crane-Droesch et al[4] proposed a novel method for yield modeling that incorporates both parametric statistical models and deep neural networks. The approach is designed to capture the complex nonlinear relationships present in high-dimensional datasets and account for unobserved cross-sectional heterogeneity. The authors use maize yield data from the US Midwest to demonstrate the efficacy of this strategy, demonstrating that it is more accurate at predicting yields for next years than both fully nonparametric neural networks

and conventional statistical methods. To implement their method, the authors obtained data on corn yield from NASS quick stats and historic weather data from the Gridded Surface Meteorological Dataset. Their research revealed the adverse effects of the weather on maize yield, which were discovered to be less severe than anticipated using conventional statistical techniques. In addition to the accumulation of heat, they also found that when heat and moisture occur were significant predictors of maize yields. While the approach represents a significant improvement in yield modeling, the authors note that it has limitations. For instance, they encountered challenges in modeling deterministically crop models such as CO2 fertilization statistically. Despite this limitation, their findings offer valuable insights for farmers and policymakers seeking to optimize crop yields and mitigate the climate change's influence on agriculture.

Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni et al[5] proposed a machine learning-based system to forecast annual crop yields for six crops at the country level in West African countries. To help farmers and decision-makers anticipate annual crop yields, the system incorporated climate data, weather data, agricultural yields, and chemical data. To develop the system, the authors used three machine learning models: decision tree, multivariate logistic regression, and k-nearest neighbor. The authors achieved promising results with all three models and used hyper-parameter tuning techniques during cross-validation to avoid overfitting and obtain better models. The authors collected climate data, weather data, agricultural yields, annual rainfall data, and chemical data for the years 1990 to 2020 from Organisation for Food and Agriculture of the United Nations and the Climate Knowledge Portal World Bank. The authors found that the Ck-NN model outperformed the other two models - CDT and CMRL. However, they also noted that incorporating additional features such as soil data, wind data, and humidity could enhance the model's performance. Overall, the study demonstrated the potential for machine learning-based prediction systems to forecast annual crop yields, providing farmers and decision-makers with critical information that can inform their decisions and improve the standard of agricultural production in the face of climate change. Incorporating additional features, such as soil data, wind data, and humidity, could further enhance the model's performance and accuracy.

Kavita Jhajharia, Pratistha Mathur et al [6]proposed the using machine learning techniques like Random Forest, Lasso Regression, and Support Vector Machine (SVM), as well as deep learning models such as Gradient Descent and long short-term memory (LSTM), to predict crop yield. The study aimed to identify the most effective algorithm for crop yield prediction, and the outcomes demonstrated that Random Forest algorithm outperformed all other models, with an R2 value of 0.963, an RMSE value of 0.035, and an MAE value of 0.0251. To collect the data for the study, the researchers used various sources, including the official website of

the Rajasthan Government. The dataset covered the period from 1997 to 2019, except for the gap between 2002 and 2004 to 2010, and included information on climate, environment, and crop yields for each crop and year, providing a comprehensive set of variables for analysis. The study concluded that Deep learning models don't perform as well as machine learning model in predicting crop yield for the selected crops in Rajasthan. However, the researchers acknowledged the need for a larger dataset with more precise information on climate and environment for each crop year to make a more accurate comparison between deep learning and machine learning models. In summary, the study represents a significant step towards predicting crop yield in Rajasthan using machine learning algorithms. The Random Forest algorithm was being the most effective algorithm for crop yield forecast. The findings of the study provide valuable insights to farmers and policymakers for enhancing crop productivity. Nonetheless, the study's limitations present an opportunity for future research to explore ways to improve the accuracy of crop yield prediction.

S.k. Gudepu,v.k Burugari et al[7] proposed an Internet of Things (IoT) and machine learning approach for analyzing various climatic conditions, including temperature, soil moisture levels, and pH levels. The approach utilizes a Kalman-Filter algorithm, which enables continuous communication between sensors, ensuring reliable and accurate data for decision-making. The data collected is stored in cloud storage and made accessible to farmers via Google Assistant, which provides voice calls and vigilant messages in the regional language, updating them on changing weather conditions. The application of a support vector machine (SVM) algorithm helps to differentiate between plants and weeds, which is a common challenge in rice farming. This algorithm takes advantage of the physical similarities between plants and weeds to accurately differentiate between them, while Genetic Algorithm is used to analyze weather situations, which helps farmers to increase their yields and profits. The proposed approach achieved an accuracy of 98%. These advancements allow farmers to make informed decisions on the best farming practices to adopt, leading to increased yields and improved economic outcomes for the country.

Mamatha, J.C. Kavitha et al[8] proposed an automated hydroponic system as an alternative to traditional farming, offering the advantage of producing higher yields in a shorter time. However, hydroponic systems require continuous monitoring and maintenance and are vulnerable to power outages and waterborne diseases. To address these challenges, the proposed hydroponic system was automatized extensively, covering the entire greenhouse and producing different crops under varying climatic circumstances. The hydroponic system utilized an organic coconut coir medium for germination, and the NFT technique combined with the KNN algorithm obtained 93 percent accuracy in predicting crop growth rates, making it an effective commercial hydroponic system. The automation and intelligence provided by

this system have numerous benefits, including increased crop production while minimizing resource usage. organic coconut coir offers great water and oxygen retention, fostering plant growth and a strong root system. However, environmental conditions directly affect plant growth, and crop yields can decrease if grown in a section of a greenhouse. Despite the high setup cost and the need for constant maintenance, hydroponic systems offer an innovative solution for maximizing crop production while minimizing resource usage.

1.3 Problem Statement

Design the machine learning model for analyzing the impact of climatic change on the maize crop yield.

1.4 Applications

- Agricultural planning
  Machine learning models can offer valuable information to farmers and other stakeholders, allowing them to make better decisions related to crop management and agricultural planning. With precise predictions of climate change's impacts on maize yields, farmers can take proactive measures to safeguard their crops, minimize losses, and enhance their overall productivity.

- Better crop management
  The findings of the study can aid in the development of improved policies and strategies for addressing the impact of changes in climate on agriculture. This can involve the creation of new crop varieties that can better cope with changing weather conditions, the adoption of innovative soil management or irrigation techniques, or the provision of financial assistance to farmers to help them adapt to changing circumstances.

- Climate risk management
  Examining the climate change impacts on maize production can aid in managing climate risks by recognizing possible threats and weaknesses in the agricultural sector and creating plans to mitigate them. This can involve actions like enhancing water management, adopting novel techniques and methods, and introducing more robust crop types. By proactively managing climate-related risks, we can diminish the potential consequences of a changing climate on food security and agricultural output.

- Food security
  Analyzing the influence of climate change on maize yield can have important implications for global food security. As climate change continues to affect crop yields and agricultural

productivity, it is becoming increasingly important to develop sustainable and resilient food systems that can withstand these changes. By providing insights into climate change's potential effects on maize yields, machine learning models can support efforts to build more resilient and sustainable food systems, ensuring that food remains accessible and affordable for people around the world.

- Economic development
  Maize is a critical crop in many developing countries, and understanding climate change's potential effects on maize yield can help promote economic development and poverty reduction efforts. By supporting sustainable agriculture, machine learning models can help create new economic opportunities and improve livelihoods for farmers and rural communities.

- Environmental sustainability
  Studying the effects of climate change on maize yield can offer significant information on the environmental outcomes of changes in agricultural productivity. This knowledge can be used to support sustainable land use practices, minimize greenhouse gas emissions, and conserve biodiversity.

1.5 Objectives and Scope of the project

Objectives of Analysis of climatic change impact on maize typically outline the background, purpose, and overall goal. It may also describe the motivation for the project, such as the need to address the potential risks to food security and agricultural sustainability posed by climate change.

1.5.1 Objectives

- Exploring the models used to simulate the impact of climate change on maize crop growth and productivity.

- Develop the model to analyze the impact of Climate change on maize crop diseases.

- Compare the outcomes to the most recent models.

1.5.2 Scope of the project

Analyzing the impact of climatic change on crop yield, particularly focusing on maize, can be beneficial for a number of reasons. For millions of people worldwide, maize is one of the most significant crops and a primary source of sustenance. Therefore, understanding how climate

change may affect maize yield is crucial for maintaining food security in many regions. Climate change can impact maize crop yield in various ways, such as modifying the occurrence and severity of weather events, adjusting the climate's temperature and precipitation cycles[15], and accelerating the spread of pests and diseases. By analyzing the relationship between climatic variables and maize yield, researchers can develop strategies to assist secure a stable and enough food supply in the future and to help offset the negative consequences of climate change. Analyzing the effect of climate change on maize crop yield has become an increasingly important area of research in recent years. Agricultural effects of climate change can have significant implications for food security, particularly in regions where agriculture is a major source of income and food supply. By studying the relationship between climatic variables and maize crop yield, researchers and policymakers can gain valuable insights into the potential impacts of climate change on different crops, and develop strategies to mitigate these impacts. There are a number of different variables that can impact maize yield, including temperature, precipitation, and soil quality. By analyzing how these variables interact with each other and affect crop growth, researchers can gain a better understanding of how climate change may impact crop yields in different regions. In addition to examining the direct impacts of climate change on crop yield, researchers may also study[16] how climate change may impact other aspects of agricultural systems, such as pest and disease management, water use efficiency, and soil health. One important area of research within this field is the development of models to predict the impact of climate change on crop yield. Machine learning algorithms and other statistical models can be used to find patterns and connections between many variables using massive datasets. These models can be used to forecast how different climatic scenarios may impact crop yields in the future and can be a valuable tool for developing strategies to mitigate the climate change's effects on agriculture. Overall, there is a wide range of scope for analyzing the effect of climate change on crop yield, and this research is likely to become increasingly important in the coming years as climate change continues to impact agricultural systems around the world. By better understanding the relationships between different variables and developing effective strategies for adaptation and mitigation, it may be possible to minimize the adverse consequences of climate change on food security and ensure a sustainable and resilient agricultural system for future generations.

Chapter 2

REQUIREMENT ANALYSIS

Data design for a climate change impact analysis on maize yield involves identifying and gathering all the necessary components, data, and methods that are required to complete the project. The components and data may include climate data such as temperature, rainfall, and humidity, as well as maize yield data. The methods may include machine learning models for prediction and statistical analysis techniques for data exploration and visualization. To collect relevant climate data, it is important to identify appropriate sources of data and to determine the frequency and duration of data collection. This may involve accessing climate data from weather stations or remote sensing platforms. Once the data has been collected, it must be pre-processed and cleaned to ensure that it is in a usable format for analysis. This may involve dealing with missing data, outliers, and other data quality issues. A suitable machine learning model is selected to choose an appropriate machine learning model that can accurately predict the impact of climate change on maize yield.

2.1 Functional Requirements

Functional requirements are a set of specifications that describe the specific behavior and functionality that a system or product must have in order to meet the needs of its users. They are typically defined during the requirements-gathering phase of a project and serve as a foundation for design, development, testing, and implementation. The effect of climatic changes on maize crop yield requires a clear set of functional requirements to be successful. It must have access to a comprehensive dataset that covers various aspects of crop production, such as weather patterns, soil quality, Rainfall pesticides, and crop yields. The dataset must be sufficiently large to support machine learning algorithms such as Random Forest,Xg boosting,decision trees, and gradient boosting that can identify patterns and correlations between these variables. Appropriate machine learning algorithm is identified in which it is capable of handling the large dataset.

- The system shall be able to train the select model on preprocessed data.
- The system shall be able to evaluate the performance of models using accuracy and precision.
- The system shall be able to predict maize crop yield under different climatic conditions.

2.2 Non Functional Requirements

Non-functional requirements are the attributes that specify how it should behave in terms of characteristics such as usability, reliability, performance, security, and maintainability. The effect of climatic change on maize crop yield includes performance requirements such as the need for the system to be able to process large amounts of data in a reasonable amount of time, or reliability requirements such as the need for the system to be able to function without errors or crashes. It is essential for ensuring that the system meets the needs and expectations of its users in terms of its overall performance, usability, and reliability.

- The System should be able to process and analyze the large dataset efficiently and in a timely manner.
- The system should be able to produce accurate results with a high degree of precision.
- The system must minimize errors in the prediction and forecasting of maize crop yield.

2.3 Hardware Requirements

In machine learning, the computational resources required for the training and validation of models can vary depending on the algorithm used, the size of the dataset, and the complexity of the model. The minimum system requirements for training and validating machine learning models can depend on many factors. To efficiently train and validate machine learning models, a computer with a minimum of 4 cores is required. For large and complex datasets, more cores are recommended to speed up the process. Additionally, a minimum of 8 GB of RAM is necessary, but 16 GB or more is preferred to handle larger datasets and complex models. It is also important to have enough storage space to save the dataset and relevant libraries and packages. An SSD can enhance read and write speeds and overall system performance. For XGBoost, a GPU with CUDA support can greatly improve the training process. A GPU with at least 4 GB of VRAM is recommended to manage the computational load. Random Forest and Decision Tree algorithms, on the other hand, do not require a GPU for training.

2.4 Software Requirements

To perform climate change impact analysis on maize yield using Python, a Python 3.x distribution such as Anaconda must be installed. Anaconda provides the base for coding in Python and includes many pre-installed libraries and packages, making it easy to start working on data processing and analysis. In addition to the base Python distribution, several libraries are required for data processing, analysis, and visualization. These libraries include

NumPy, Pandas, Scikit-learn, and Matplotlib. Arrays and matrices are supported by the robust numerical computing toolkit NumPy. Panda is a library that provides support for data analysis and manipulation. Scikit-learn is a machine learning library that offers numerous methods for grouping, classification, and regression. Matplotlib is a library for data visualization that provides support for creating charts, plots, and graphs. The XGBoost library is also necessary for the XGBoost algorithm, which is a popular machine-learning algorithm used for predicting outcomes. The algorithm like XG Boost, Random Forest and Decision tree are particularly well-suited for handling large datasets and can provide accurate predictions with high efficiency. To code and analyze data, an integrated development environment (IDE) such as Jupyter Notebook or any other IDE can be used. These environments provide a user-friendly interface for coding and data analysis, and they can help streamline the workflow. It is also essential to keep the libraries and packages updated to the latest version to ensure compatibility with the algorithms used and to take advantage of the latest features and bug fixes.

Chapter 3

SYSTEM DESIGN

System design in analyzing the effect of climatic change on maize crop yield involves outlining the technical infrastructure required to collect and analyze data, model development and implementation, and decision support systems. It also involves identifying the different components of the system, including data sources, algorithms, hardware, and software requirements. The system design aims to develop a reliable and efficient framework for analyzing the impact of climate change on maize crop yield, generating accurate predictions, and providing useful insights for decision-making. A well-designed system can help to identify the different variables that affect crop yield, develop models for predicting yield, and integrate climate data and historical yield data to create a more accurate and reliable forecasting model. Ultimately, the system design supports the development of more effective adaptation strategies to ensure food security and agricultural sustainability in the face of a changing climate.

3.1 Architecture Design

The architectural design for analyzing the effect of climatic change on maize crop yield based on rainfall, and temperature pesticides can be divided into several steps, which include data preprocessing, hyperparameter tuning, combining datasets, model building, testing, and prediction analysis. The first step is combining datasets, where we merge the average temperature, pesticide, yield, and rainfall datasets to create a single dataset that will be used for analysis. The second step is data preprocessing, which involves cleaning and transforming the raw data to make it suitable for analysis. The third step is hyperparameter tuning, which involves selecting the best parameters for each model. This is done using techniques such as grid search or random search, which evaluate different combinations of hyperparameters and select the best-performing one. This may include removing duplicates, handling missing values, and converting data types. The output of this step is a cleaned and transformed dataset that is ready for analysis. The fourth step is model building, which involves building decision tree models using the best hyperparameters from the previous step. We also build other models such as Decision Trees, Extreme Gradient Boosting(XGboost), Gradient Boosting, and Random Forest to compare and select the best-performing model. The fifth step is testing, which involves evaluating the performance of the selected model on a separate testing dataset to measure its accuracy and identify any areas for improvement. Finally, prediction

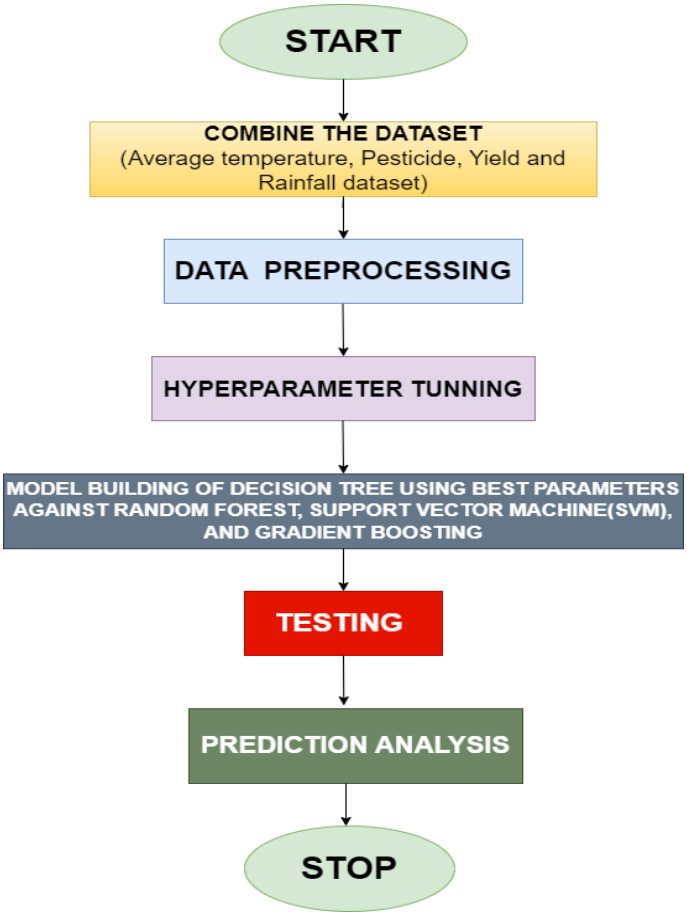

Figure 3.1: Flow Chart

analysis involves using the selected model to make predictions on new data to determine the effect of climatic change on crop yield based on rainfall temperature pesticides. Figure 3.1 illustrates the architectural design for analyzing the effect of climatic change on crop yield based on rainfall temperature pesticides:

3.2 Data Design

Data Design is a crucial aspect of analyzing the effect of climatic change on maize crop yield. The database tables should be constructed such that data may be stored and retrieved quickly for analysis. The data design should be flexible and scalable to handle changing requirements and data volumes.

3.2.1 Data Collection

The data collection process involved retrieving various datasets from different sources to examine the impact of climate change on crop yield. The crop yield dataset, obtained from the

FAO website, contained information on the yield of the ten most consumed crops worldwide, including country, item, year, and yield value. This dataset comprised 12 tuples and 56718 rows. The climatic dataset, consisting of rainfall and temperature data, was collected from the World Data Bank to study their effect on crop yield. The rainfall data is represented in 3 tuples and 6728 rows, while the temperature dataset contained 3 tuples and 71312 rows. Additionally, the pesticide dataset, containing information on pesticide usage in different countries and for various crops, was collected from the FAO database and had 6 tuples and 4350 rows. The purpose of collecting this data was to analyze the impact of climate change on agricultural productivity and develop potential strategies for mitigating its effects.

3.2.2 Data preprocessing

Combining datasets

In order to conduct an analysis, we will merge several datasets including the average temperature, pesticide usage, maize yield, and rainfall datasets. This will result in the creation of a consolidated dataset that will be used for analysis.

Table 3.1: Crop Yield Dataset

Year	**Yield**	**Rainfall**	**Pesticides**	**Temperature**
1990	36613	1485	121	16.37
1991	29068	1485	121	15.36
1992	24876	1485	121	16.06
...	...	...	...	...
1993	24185	1485	121	16.05
1994	25848	1485	201	16.96

The Table 3.1 contains the following attributes

- Year: The year in which the yield is measured.
- Maize Yield(hg/ha): The yield of the crop per hectare in hundredweight (hg/ha) units.
- Average rainfall per year: The average rainfall in "mm" received in that area during the year.
- Pesticides(tonnes): The number of pesticides used in that area during the year in tonnes.
- Average Temperature: Average temperature refers to the mean temperature in Celsius or Fahrenheit that was recorded in a specific area during a particular period.

Feature scaling

Scaling is an important data preprocessing step in machine learning that involves transforming the numerical features in a dataset to a common scale so that they have equal weight in the learning process. This is particularly important when the features in a dataset have different units or ranges, as this can cause some features to dominate the learning process and lead to suboptimal model performance. One common scaling technique is the min-max scaler, which scales the features to a specific range, typically between 0 and 1. To achieve this, the scaler subtracts the minimum value of each feature and divides it by the range (i.e., the difference between the maximum and minimum values). This ensures that the minimum value of each feature is scaled to 0 and the maximum value is scaled to 1, with all other values in between scaled proportionally. In summary, scaling is a crucial preprocessing step in machine learning that helps to ensure that all features are given equal weight in the learning process, regardless of their original magnitudes or units. The min-max scaler is one common technique for scaling features to a specific range and is particularly useful for datasets with features that have a bounded range.

Encoding categorical variables

One-hot encoding is a technique used in data preprocessing to convert categorical features into numerical features. In the dataset, the "Area" column contains categorical data in the form of country names. One-hot encoding has been applied to this column to create binary values representing the countries in the dataset. For instance, if the dataset contains three countries, such as the USA, Canada, and Mexico, the one-hot encoding process will create three binary columns representing these countries. For each row in the dataset, the value in the binary column for the corresponding country will be set to 1, and all other binary columns will be set to 0.This technique helps in the machine learning models by transforming the categorical data into numerical data that can be processed by the algorithms. In addition to creating binary columns for the country names, one-hot encoding has also been applied to the "Temperature", "Pesticides", and "Rainfall" columns. This process will create binary columns for the unique values in each of these columns, making the dataset suitable for machine learning algorithms that require numerical inputs. By applying one-hot encoding, we can avoid the issues of ordinality and biases that can arise when working with categorical data in machine learning models.

3.2.3 Data exploration

Heatmap

Heatmaps are a valuable tool that aids in the interpretation of a significant amount of data related to precipitation, temperature, and crop yield. They help analyze the impact of climate change on crop yield. By displaying data visually, heatmaps allow policymakers and researchers to easily recognize trends and patterns in the data. The heat map depicted in Figure 3.2 indicates that yield and temperature are negatively correlated, which means that an increase in temperature can lead to a decrease in yield. Conversely, rainfall and temperature are positively correlated, implying that as the temperature increases, the amount of rainfall also tends to increase. Heatmaps can also highlight other important relationships and correlations between different variables. Overall, using heatmaps as part of the data analysis process can provide valuable insights and inform decision-making related to agricultural practices and policies in the face of climate change.

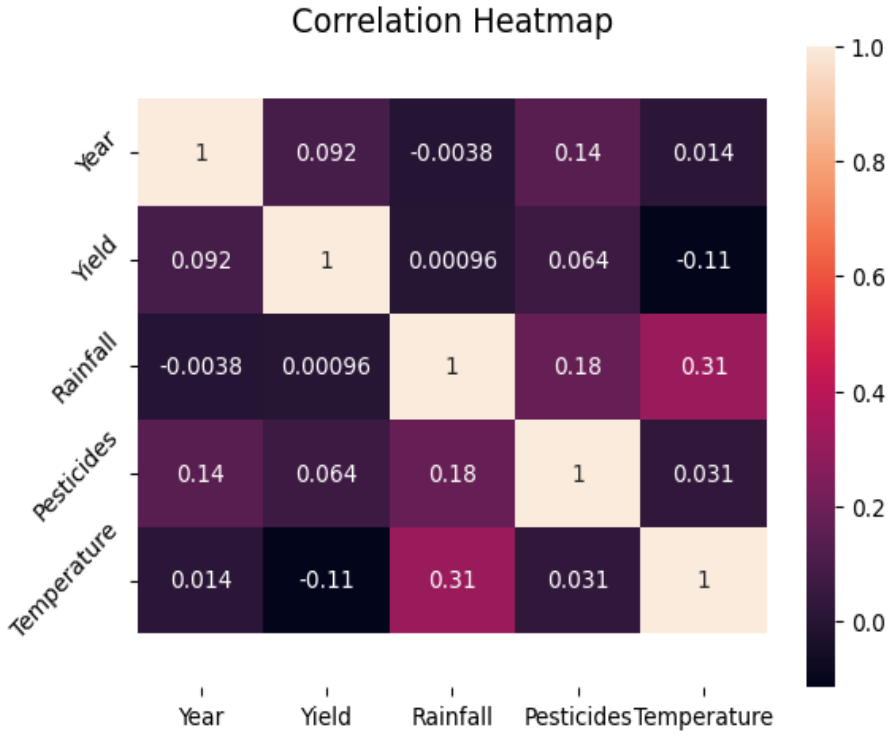

Figure 3.2: Correlation Heatmap

Data visualization

A plot for the year versus variable importance, such as rainfall, temperature, and pesticides, can provide an overview of how each variable impacts crop yield over time. The Figure 3.3

will typically have the year on the x-axis and the variable importance on the y-axis. The variable importance can be represented by a score or ranking that indicates how much of an impact each variable has on the crop yield. The plot can show trends and patterns in how each variable affects the crop yield over time. For example, it can show how the importance of rainfall varies from year to year and how this impacts crop yield. It can also show whether the importance of each variable is increasing or decreasing over time and whether there are any notable changes in the overall trend. Such a plot can be useful for identifying which variables are the most important in predicting crop yield and how this importance changes over time. It can help inform decisions about which variables to prioritize in crop management and help policymakers develop strategies for mitigating the impact of climate change on crop yields.

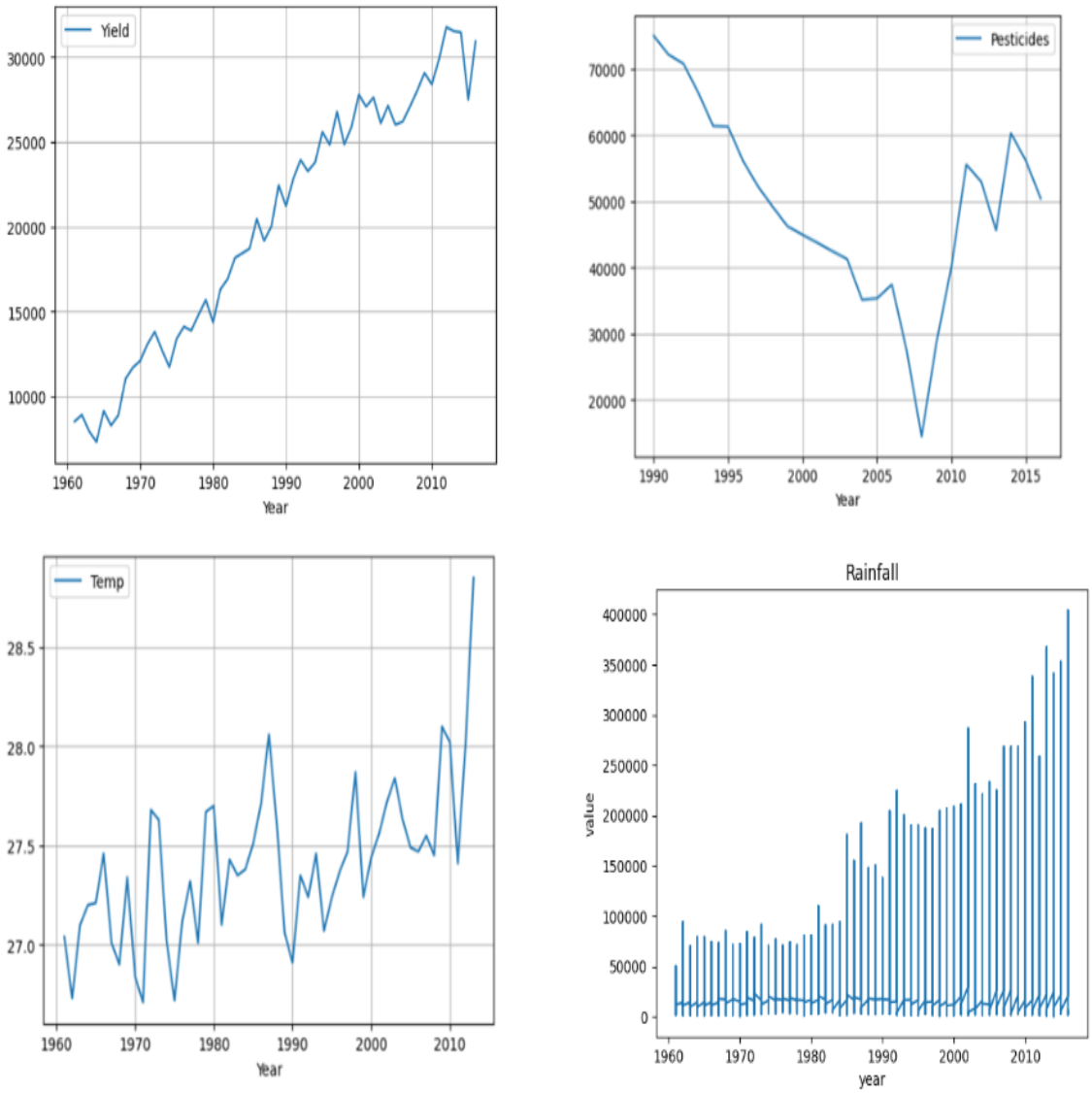

Figure 3.3: Year v/s Yield, Pesticides, Temperature, and Rainfall

3.2.4 Feature selection

Variable importance

Variable importance analysis is an important technique in analyzing crop yield as it helps in identifying the factors that have the greatest impact on the yield. In general, variable importance refers to the degree to which a variable contributes to the outcome of a model. By identifying the most important variables, policymakers and researchers can focus their efforts on addressing those factors that are most critical in improving crop yield. This can lead to more effective interventions and policies aimed at mitigating the negative effects of climate change and improving agricultural productivity.

In analyzing the effect of temperature, pesticides, and rainfall on crop yield, Figure 3.4 shows the variable importance.Variable importance can be determined through various methods such as feature selection, regression analysis, and machine learning algorithms. In machine learning algorithms such as decision trees and random forests, variable importance can be determined by calculating the decrease in impurity or variance when a variable is used for splitting the data. The variables with the highest decrease in impurity or variance are considered the most important variables. For example, a decision tree can be used to determine the impact of temperature, pesticides, and rainfall on crop yield. The variables with the highest decrease in impurity when used for splitting the data can indicate their importance in affecting crop yield.

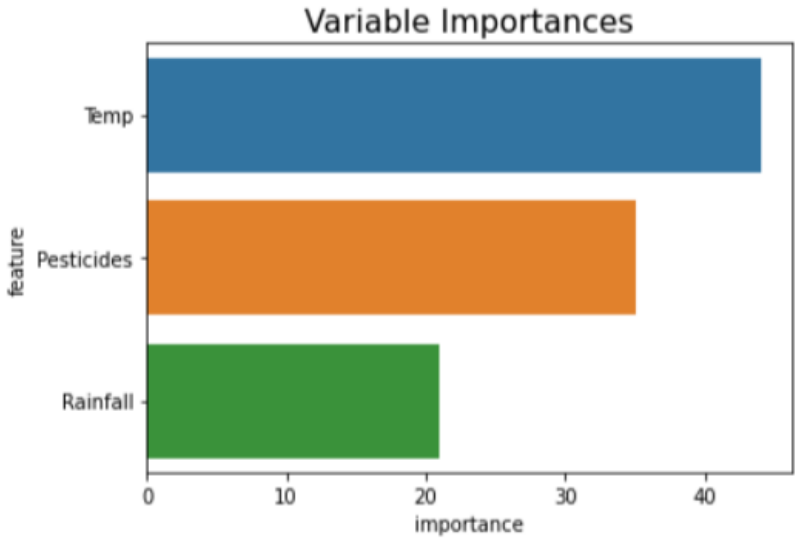

Figure 3.4: Variable importance

3.2.5 Model training

In order to estimate the production of the maize crop based on temperature, pesticide use, and rainfall data, algorithms such as Random Forest, XGBoost, Gradient Boosting, and Decision Trees are used to study the impact of climate change on agricultural output. An overview of the training methods for each algorithm:

1. Random Forest

Random Forest is used to predicting the effect of climate change on maize crop yield by training multiple decision trees on subsets of the training data and features. The algorithm can handle a large number of input features and is robust to noise and outliers. In the case of climate change analysis, the input features could include factors such as temperature, rainfall, and pesticide usage.

2. Decision

Decision tree algorithms recursively divide data based on selected attributes to anticipate how climate change would affect maize crop yield, aiming to maximize subset purity. These algorithms are straightforward, understandable, and can handle numerical and categorical input features. In climate change analysis, decision trees consider factors such as temperature range, rainfall intensity, and pesticide type as input features.

3. Gradient Boosting

Gradient boosting is utilized to anticipate the impact of climate change on maize crop production by iteratively incorporating decision trees into the model and training each tree to correct the shortcomings of the previous tree. This algorithm is especially beneficial when input features exhibit complex interactions and can handle missing values and outliers. In climate change analysis, input features for gradient boosting may involve average temperature, rainfall variability, pesticide application rate, and soil moisture content.

4. XGBoost

Decision trees are iteratively added to the model using XGBoost to forecast how climate change would affect maize crop yield. Each tree is trained to rectify the mistakes of the preceding tree. The algorithm is particularly useful when the input features are correlated and can handle missing values and outliers. In the case of climate change analysis, the input features could include factors such as maximum temperature, minimum temperature, rainfall amount, and pesticide concentration.

3.2.6 Model evaluation

We utilised R2, Mean Squared Error (MSE), and Mean Absolute Error (MAE) as the assessment metrics to assess the effectiveness of the machine learning model on analysing the effect of climate change on maize crop production based on rainfall, pesticides, and temperature information. The R2 score quantifies the percentage of variance in the target variable that is explained by the model, indicating how well the model fits the data. A higher R2 score indicates that the model provides a better fit to the data. It can range from 0 to 1, where 1 indicates a perfect fit. Therefore, a high R2 score indicates better model performance. The average squared difference between the expected and actual values is measured by the MSE. A lower MSE indicates that the model's predictions are more accurate, and it ranges from 0 to infinity. MSE is a popular metric for regression models, and it is particularly useful when the dataset contains outliers or errors that may skew the evaluation results.The average absolute difference between the expected and actual values is measured by the MAE. A lower MAE indicates that the model's predictions are more accurate, and it also ranges from 0 to infinity. MAE is another popular metric for regression models, and it provides a more robust evaluation of the model's performance as it is less sensitive to outliers. Overall, evaluating machine learning models using these metrics helps to identify the best-performing model for predicting the effect of climate change on maize crop yield. A model with a high R2 score, low MSE, and low MAE indicates that it provides a good fit to the data and accurate predictions.

3.3 User Interface Design

A user interface design for analyzing the effect of climatic change on crop yield would typically include several components. It would first provide a dashboard or landing page that would allow the user to input the necessary data, such as average temperature, pesticide, rainfall, and yield information. The design would then incorporate a data preprocessing step to clean and prepare the data for use in machine learning models. The user interface would then offer a selection of appropriate machine learning or statistical models for the user to choose from. The models would have relevant parameters, such as learning rates, regularization parameters, and model architectures, which the user could tune to their desired level of precision. Additionally, the user interface would include performance metrics for evaluating the accuracy of the models, such as mean squared error or coefficient of determination. Once the user has selected and trained a model, the interface would provide a visualization of the correlation between the different climatic variables and their effect on crop yield. This visualization would help the user understand how changes in climatic variables might affect crop yield, and also help them interpret the model's predictions. Finally, the user interface would allow the user to input a given combination of climatic variables and use the best-performing

model to predict crop yield. The predicted yield would then be visualized alongside the input climatic variables, providing the user with a clear understanding of how the different variables interact and affect crop yield.

The Table 3.2 represents the machine learning task involves analyzing the relationship between different climatic variables and crop yield. The input consists of the appropriate machine learning algorithm with relevant parameters and multiple datasets related to rainfall, pesticides, yield, and temperature. The output includes a visualization of the correlation between climatic variables and crop yield, which helps in understanding the impact of different factors on crop yield. Additionally, a list of the best-performing machine learning models with their respective hyperparameters and performance metrics is generated to select the most accurate model for predicting crop yield based on the climatic variables.

Table 3.2: User Interface Design

Content	**Details**
Corpus	Real-world data fetched from FAO and World Data Bank.
Input	(1)**TAAT**: The appropriate machine learning algorithm with relevant parameters, as well as performance metrics (2) **DAAT**: Multiple datasets such as rainfall, pesticides,yield, and temperature
Output	(1)**TAAT**: Visualization of the correlation between the different climatic variables and their effect on crop yield. (2)**DAAT** : Display a list of the best-performing models with their respective hyperparameters and performance metrics

Chapter 4

IMPLEMENTATION

The Implementation chapter includes sub-sections beginning from Data pre-processing to the different algorithms used on the maize crop yield dataset. This also leads to the comparison of 4 different algorithms with respect to efficiency.

This chapter begins with sub-section 4.1 as Data pre-processing followed by 4.2 - Hyperparameter Tuning, next with 4.3 Model-1 (Gradient Boosting) followed by 4.4 Model-2 (Random Forest Regressor), then 4.5 Model-3 (XG Boost) followed by 4.6 Model-4 Decision Tree, then concluding it with the sub-section 4.7 Comparison of the 4 algorithms.

4.1 Data Preprocessing

The dataset includes various crops such as Barley, Rice, Potato, Wheat, Soybeans, etc. Out of which we include particularly the maize crop and the climatic factors affecting the growth and yield of maize crops. Here, we have considered only the maize crop, as we have analyzed the detailing of the impact of climate change such as Temperature, Pesticides, and Rainfall on maize crop yield. In order to do this we have restricted the dataset to maize crop only. In addition to this, we have implemented one-hot encoding so as to the country name and applied it for the various effects such as Temperature, Pesticides, and Rainfall. This will create a binary value in the country(Area) column.

4.2 Hyperparameter Tuning

The performance of machine learning models can be improved by Hyper parameter tuning. The working of the hyperparameter tuning is given in the Figure 4.1, that are used to forecast crop yields based on several input variables, such as Rainfall, Temperature, and Pesticides, can be optimized by using hyperparameter tuning in the context of crop yield datasets. the collection of crop yield data for multiple farms over a number of years, together with various factors like Rainfall, temperature, Pesticides, etc. This dataset could be used to build a machine-learning model that would use these features to forecast crop yield. However, we must adjust the model's hyperparameters to get the greatest potential performance.

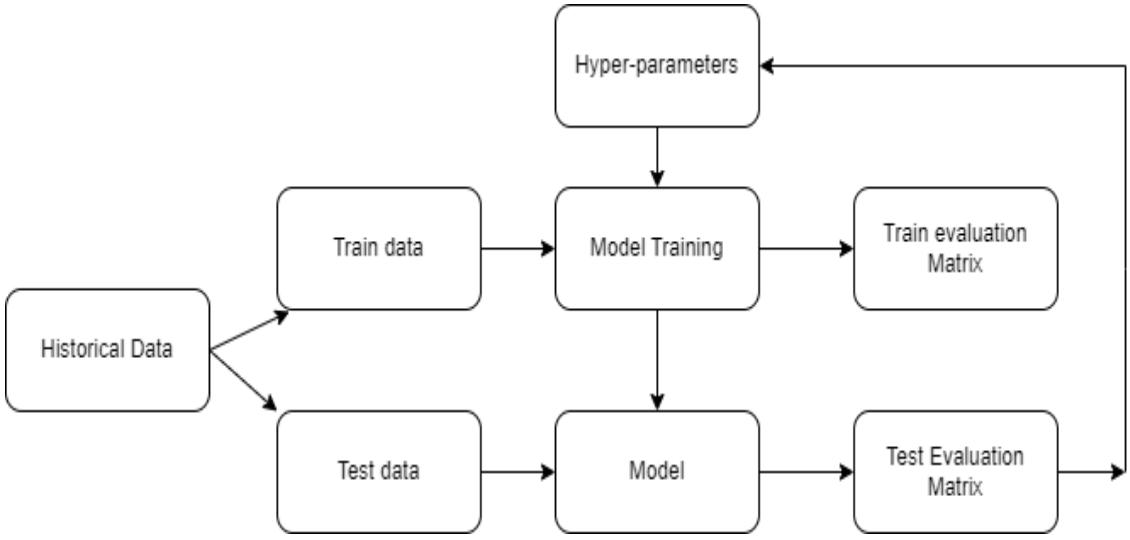

Figure 4.1: Hyperparameter Tuning Diagram

4.3 Model 1 - Gradient Boosting

First, we chose the Machine Learning model i.e. Gradient boosting algorithm. The approach operates by training decision trees iteratively using the residuals from the prior tree. The algorithm generates the loss function's negative gradient with regard to the preceding tree's prediction at each iteration. The goal variable for the next decision tree is then this negative gradient.

The Algorithm 1 calculates the difference between the known correct target value and the present prediction. Gradient Boosting is a boosting algorithm that relies on the idea of the stagewise addition procedure, in which numerous weak learners are taught and as a result, we obtain strong learners.The function fit is particularly versatile because it provides a variety of hyperparameter tinkering possibilities and can optimise different loss functions. In gradient boosting, a new weak model is trained at each iteration to anticipate the "error" of the prevailing strong model (also known as the pseudo response). The discrepancy between the forecast and a regressive label is the "error".

Algorithm 1 Gradient Boosting

For $F_0(x) = argmin_\rho \sum_{i=1}^{N} L(y_i, \rho)$
for m = 1 to M do: **do**
Compute the negative gradient

$$y_i = -\left(\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right)$$

Fit a model

$$\alpha_m = argmin_\alpha, \beta \sum_{i=1}^{N} [y - \beta h(x_i; \alpha_m)]^2$$

Choose a gradient descent step size

$$\rho_m = argmin_p \sum_{i=1}^{N} L(y_i, F_m - 1(x_i) + \rho h(x_i, \alpha))$$

Estimate the estimation of F(x)

$$F_m(x) = F_m - 1(x) + \rho_m h(x_i; \alpha_m)$$

end for
Output the final regression function F_m(x)

4.4 Model 2 - Random Forest Regressor

When using the Random Forest Regressor method,it starts by randomly choosing a subset of the training data and a subset of the input features. The decision tree is then trained using the conventional CART algorithm on this subset of data and features, with the split being based on the feature that offers the maximum information gain. The procedure of randomly selecting subsets of data and features and training the decision tree is done numerous times to produce a forest of decision trees. A lot of decision trees are constructed during training, and after that, each tree's mean/mode of prediction is shown. The final stage is to combine all of the trees' predictions to get a single prediction. The projected values' mean or median is used. The model's accuracy and stability are enhanced by the aggregate. The bootstrapping method, which is used by the Random Forest Regressor algorithm which is shown in Algorithm 2, includes randomly selecting the data with replacement to produce several datasets. This helps to lessen overfitting by generating new datasets with slightly different distributions. As the number of explanatory variables rises, random forest performs better.

Algorithm 2 Random Forest Regressor

$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} f_i(x)$
where:
$\hat{y}$ is the predicted output (regression value)
N is the number of decision trees in the forest
$f_i(x)$ is the prediction of the i-th decision tree on input x

4.5 Model 3 - XG Boost

The data will be prepared by XGBoost by being divided into training and testing sets and properly encoding any categorical features. A single decision tree representing the full dataset must be created at initialization. Next, using the prediction from the present model, the gradient of the loss function is determined for each observation in the training set. By iteratively dividing the dataset into smaller subsets that minimise the loss function. Purity and coverage are combined to determine the split, and overfitting can be prevented by adjusting hyperparameters such minimum samples per leaf and tree depth. After the decision tree has been constructed, the current model prediction is combined with its forecast before the process is repeated.

The algorithm 3 represents the algorithm for XGBoost. XGBoost makes use of sophisticated regularisation (L1 & L2), which enhances the capacity for model generalization. The method used by XGBoost is based on trees, and each tree is constructed in turn. Each consecutive tree in the algorithm is built to fix the mistakes caused by the preceding tree, starting with a single tree. Using a greedy algorithm, it creates trees by selecting the split that maximizes a specified objective function. The quality of the splits produced by each tree is assessed by XGBoost using a particular objective function. A loss function that calculates the discrepancy between anticipated and actual values plus a regularisation function that penalizes model complexity to prevent overfitting makes up the objective function.

Algorithm 3 XG Boost

1: **Input:** The image feature, the loss function, the total number of sub-tree 'M';
2: **Output:** The estimated probability of image feature $feat_z$ Repeat
3: Initialize the m-th tree $f_m(x_i)$
4: Compute $g_i = \partial_{y_i}(m-1)Loss_{XGBoost}(y_i, y_i^{(m-1)})$
5: Compute $h_i = \partial^2_{y_i}(m-1)Loss_{XGBoost}(y_i, y_i^{(m-1)})$
6: Use the statistics to greedily grow a new tree $f_m(x_i)$:
7: $\sum_{i=1}^{n} \frac{1}{i}$
$obj^{(m)} = -\frac{1}{2}\sum_{j=1}^{M} \frac{G_j^2}{H_j+\lambda} + \gamma M$
8: As shown above , Add the best tree $f_m(x_i)$ into the current model
9: Until all M sub-trees are processed
10: Obtain a strong regression tree based on all weak regression sub-trees
11: Output the estimated probability based on the strong regression tree
12: Output the final regression function F_m(x)

4.6 Model 4 - Decision Tree

A non-parametric supervised learning approach used in machine learning and data mining is the decision tree. The decisions and potential outcomes are represented by a hierarchical model made up of nodes and edges. With each division represented by a node in the tree, the algorithm 4 iteratively divides the input data into subsets depending on the values of input features using an information gain or Gini index-based criterion. The objective is to develop a model that can identify a new input's class or label based on its attributes. Decision trees are simple to comprehend and analyze and can handle both category and numerical data. They may, however, become overfit and sensitive to slight alterations in the data. These problems can be solved using a variety of strategies, including pruning, regularisation, and ensemble methods like Random Forests.

Algorithm 4 Decision Tree

1: $\hat{y}_i = \sum_{j=1}^{J} c_j \cdot \mathbb{I}(x_i \in R_j)$
2: $\hat{y}_i$ is the predicted output value for the i-th observation.
3: J is the total number of leaves or terminal nodes in the tree.
4: c_j is the constant value assigned to the j-th leaf or terminal node.
5: $I(x_i \in R_j)$ is the indicator function that takes the value 1 if the i-th observation belongs to the j-th leaf or terminal node, and 0 otherwise.
6: R_j is the region or subset of the feature space defined by the j-th leaf or terminal node.

4.7 Comparison of algorithms

The four algorithms - Gradient Boosting, Random Forest Regressor, XG Boost, and Decision Tree - were used to analyze the dataset for maize crop yield and compared their performance. A description of each approach is provided in sections 4.3 to 4.6. While comparing them we come across the efficiency of each algorithm on the Maize crop yield dataset, before moving to the conclusion, let us compare the working of each algorithm. Recursively dividing the data into subsets based on the features that best divide the data, Decision Trees do this until the subgroups are sufficiently pure or a maximum depth is achieved. Decision Trees will try to divide the data in the crop yield dataset based on characteristics like temperature, rainfall, pesticides, etc. to estimate the yield.

The Random Forest Regressor is an ensemble technique that generates predictions by building numerous decision trees and averaging them. To prevent overfitting, each decision tree is trained using a random subset of the features and the data. Similar to Random Forest Regressor, Gradient Boosting is an ensemble method that builds numerous decision trees,

but it takes a different approach to train the trees. Each decision tree attempts to enhance the predictions produced by the preceding trees as it is trained successively. XGBoost is a type of Gradient Boosting algorithm that trains decision trees using a more streamlined and regularised method. It also includes some extra features, such as how to deal with missing values and how to apply L1 and L2 regularisation to lessen overfitting.

Chapter 5

RESULTS AND DISCUSSIONS

Analyzing climate change's impact on maize crop yield enables farmers and decision-makers to better comprehend the possible climate change effects on crop production. In the course of this study, we aimed to predict the impact of climatic variables, such as temperature, precipitation, and pesticide use, on maize yield using machine learning models. We used four different models to analyze the data and make predictions. The study's findings can assist farmers and decision-makers in making sensible choices about yield management practices and policies in the face of changing climate conditions.

The purpose of the plot 5.1 is to assess a machine-learning model's performance that predicts the effects of climate factors on maize crop yield. The plots show a scatter plot that compares the actual and predicted values of the model, with the observed values shown on the x-axis and the predicted values shown on the y-axis. Additionally, a dashed black line is included to show the regressor line, which compares the observed and predicted values. The plot for actual vs predicted values using the Random Forest algorithm shows that the data points are tightly clustered around the diagonal line, indicating that the actual and anticipated values have a strong linear relationship. This implies that the model's predicted values closely match the actual values and the model is accurate. Contrarily, the plots for other algorithms show that the data points are more widely spread out, indicating a weaker linear relationship and suggesting that the predictions of these models may be less accurate.

The GradientBoostingRegressor has an accuracy score of 87%, RandomForestRegressor has an accuracy score of 97%, DecisionTreeRegressor has an accuracy score of 96%, and XGboosting has an accuracy score of 86%. In addition to accuracy, Mean Squared Error (MSE) and Mean Absolute Error (MAE) are often used as performance metrics for regression problems, including examining the impact of climate change on agricultural productivity. After analyzing all of the models, the Random Forest model has the greatest accuracy score of 97%, making it the best model for predicting the influence of climate change on maize yield.

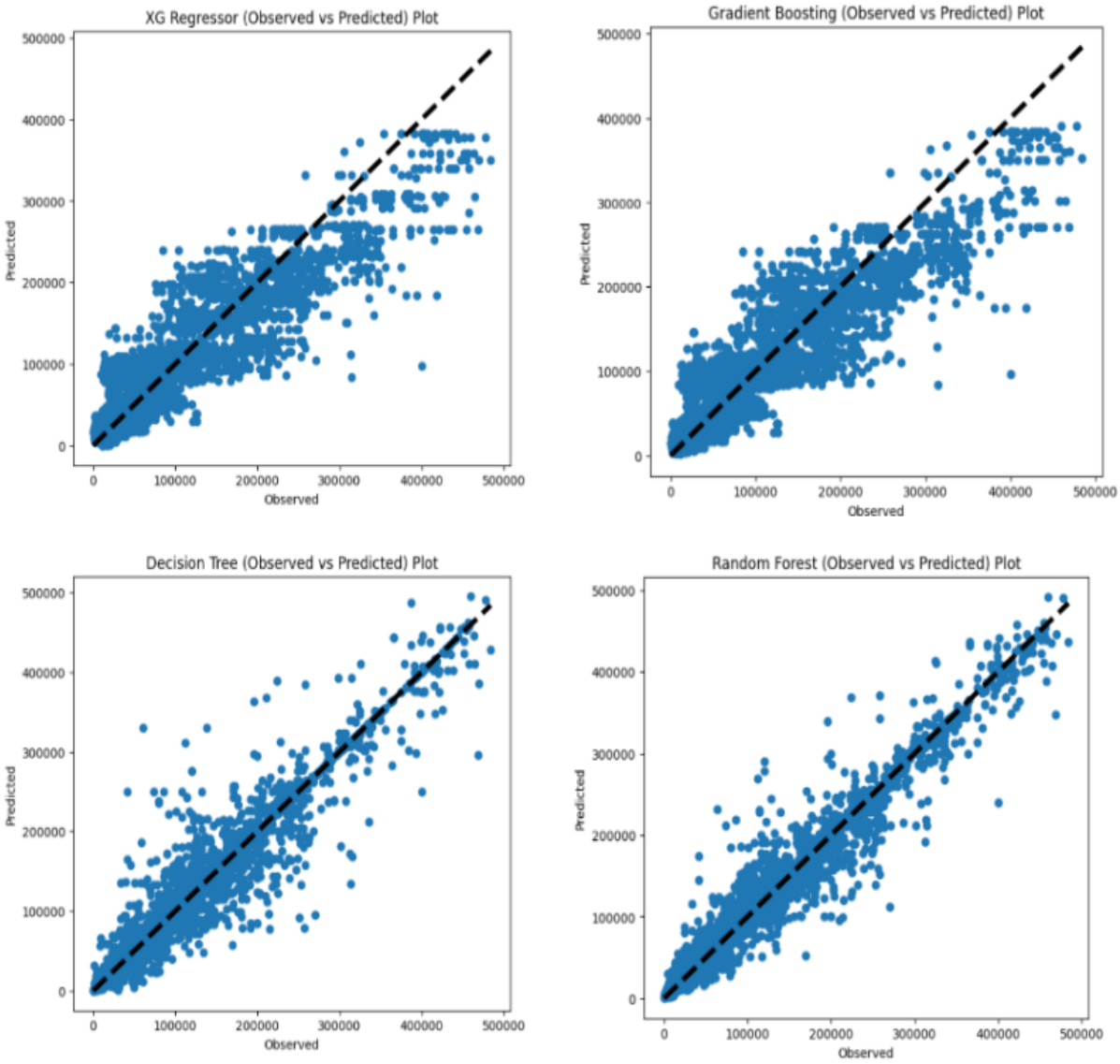

Figure 5.1: Actual Vs Predicted yield

Chapter 6

CONCLUSION AND FUTURE SCOPE

Analyzing the impact of climatic change on maize crop yield using XGBoost, gradient boosting, random forest, and decision tree algorithms can provide predictions of future maize yields under different climatic scenarios. The analysis 6.1 can assist in pinpointing the primary climatic variables that influence maize yields and in formulating suitable adaptation plans to lessen the negative effects of climate change on maize production. The random forest model is promising because to its great accuracy when compared to other models. But the accuracy may vary depending on the dataset and the climatic conditions. Therefore, further research is needed to validate the results and identify the best model for different regions and climate scenarios

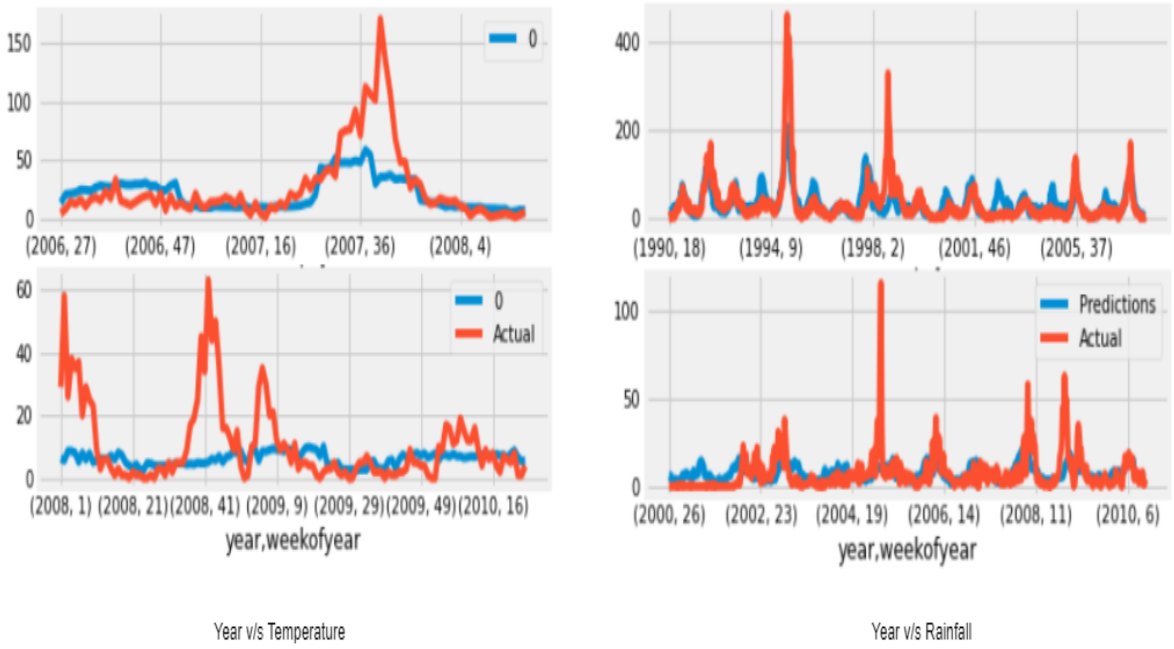

Figure 6.1: Actual Vs Predicted yield

There can be a lot of advancements in the AI and Machine learning field which can further lead to the Crop yield prediction models combined with the use of satellite data to deliver precise and current information on climatic elements like rainfall, temperature, soil moisture content, and pesticides. This can assist farmers in making knowledgeable choices on crop

management techniques. The optimized Crop yield prediction models can be used to create strategies for adjusting to changing conditions as climate change continues to have an impact on agriculture. In order to reduce the effects of droughts or floods, planting timings may need to be optimised or crops with greater resistance to shifting weather patterns may need to be identified.

IoT devices like soil moisture sensors, temperature sensors, and weather stations can be integrated to offer real-time information about the field's circumstances. Machine learning algorithms that can more accurately estimate agricultural yield can be trained using this data. For several crops, crop yield prediction models can be created based on the unique growth circumstances of each crop. This can aid farmers in making the best crop choices and maximising yield. Big data analytics can be used to find patterns and trends in crop yield data, which can then be utilised to create prediction models that are more precise. Deep learning algorithms, for example, can be used to process massive amounts of data and produce precise predictions.

ANALYSING

ORIGINALITY REPORT

16%	8%	12%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to B.V. B College of Engineering and Technology, Hubli Student Paper	3%
2	Z Han, B Zhang, G Hoogenboom, X Li, C He. "Climate change impacts and adaptation strategies on rainfed and irrigated maize in the agro-pastoral ecotone of Northwestern China", Climate Research, 2021 Publication	2%
3	"Quantification of Climate Variability, Adaptation and Mitigation for Agricultural Sustainability", Springer Science and Business Media LLC, 2017 Publication	1%
4	Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni, Rubby Aworka, Jérémie Thouakesseh Zoueu et al. "Crops yield prediction based on machine learning models: Case of West African countries", Smart Agricultural Technology, 2022 Publication	1%

5	www.mdpi.com Internet Source	1 %
6	Qaisar Saddique, Muhammad Imran Khan, Muhammad Habib ur Rahman, Xu Jiatun et al. "Effects of Elevated Air Temperature and CO2 on Maize Production and Water Use Efficiency under Future Climate Change Scenarios in Shaanxi Province, China", Atmosphere, 2020 Publication	1 %
7	Florian Schierhorn, Max Hofmann, Taras Gagalyuk, Igor Ostapchuk, Daniel Müller. "Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages", Climatic Change, 2021 Publication	1 %
8	Renhai Zhong, Yue Zhu, Xuhui Wang, Haifeng Li et al. "Detect and attribute the extreme maize yield losses based on spatio-temporal deep learning", Fundamental Research, 2022 Publication	<1 %
9	Submitted to The Robert Gordon University Student Paper	<1 %
10	Shuxiang Song, Chen Fei, Haiying Xia. "Lithium-Ion Battery SOH Estimation Based on XGBoost Algorithm with Accuracy Correction", Energies, 2020 Publication	<1 %

11	etd.aau.edu.et Internet Source	<1%
12	Submitted to Coventry University Student Paper	<1%
13	Kindie Tesfaye, Gideon Kruseman, Jill E. Cairns, Mainassara Zaman-Allah et al. "Potential benefits of drought and heat tolerance for adapting maize to climate change in tropical environments", Climate Risk Management, 2018 Publication	<1%
14	core.ac.uk Internet Source	<1%
15	nottingham-repository.worktribe.com Internet Source	<1%
16	R. K. Mall, Ranjeet Singh, Akhilesh Gupta, G. Srinivasan, L. S. Rathore. "Impact of Climate Change on Indian Agriculture: A Review", Climatic Change, 2006 Publication	<1%
17	hdl.handle.net Internet Source	<1%
18	Coster, A. S., and A. I. Adeoti. "Economic Effects of Climate Change on Maize Production and Farmers' Adaptation	<1%

Strategies in Nigeria: A Ricardian Approach", Journal of Agricultural Science, 2015.
Publication

19	"Emerging Research in Electronics, Computer Science and Technology", Springer Science and Business Media LLC, 2019 Publication	<1 %
20	M. V. K. Sivakumar, H. P. Das, O. Brunini. "Impacts of Present and Future Climate Variability and Change on Agriculture and Forestry in the Arid and Semi-Arid Tropics", Climatic Change, 2005 Publication	<1 %
21	Submitted to University of Wales Institute, Cardiff Student Paper	<1 %
22	akworkblog.wordpress.com Internet Source	<1 %
23	Submitted to WorldQuant University Student Paper	<1 %
24	www.slideshare.net Internet Source	<1 %
25	Submitted to Bahrain Polytechnic Student Paper	<1 %
26	Huang, Jun-Jie, and Wan-Chi Siu. "Practical application of random forests for super-	<1 %

resolution imaging", 2015 IEEE International Symposium on Circuits and Systems (ISCAS), 2015.
Publication

27	Submitted to University of Technology, Sydney Student Paper	<1 %
28	Submitted to University of the Pacific Student Paper	<1 %
29	Walid Kamal Abdelbasset, Shereen H. Elsayed, Sameer Alshehri, Bader Huwaimel et al. "Development of GBRT Model as a Novel and Robust Mathematical Model to Predict and Optimize the Solubility of Decitabine as an Anti-Cancer Drug", Molecules, 2022 Publication	<1 %
30	Submitted to Ahmedabad University Student Paper	<1 %
31	Submitted to Canterbury Christ Church University Student Paper	<1 %
32	Submitted to Sim University Student Paper	<1 %
33	Andrew Crane-Droesch. "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture", Environmental Research Letters, 2018	<1 %

Publication

34	Submitted to RMIT University Student Paper	<1 %
35	Submitted to University of Suffolk Student Paper	<1 %
36	utahrivers.org Internet Source	<1 %
37	Chen-Sen Ouyang, Rei-Cheng Yang, Rong-Ching Wu, Ching-Tai Chiang, Lung-Chang Lin. "Determination of Antiepileptic Drugs Withdrawal Through EEG Hjorth Parameter Analysis", International Journal of Neural Systems, 2020 Publication	<1 %
38	Submitted to Higher Education Commission Pakistan Student Paper	<1 %
39	Submitted to University College London Student Paper	<1 %
40	Yuning Yang. "Further Results for Perron–Frobenius Theorem for Nonnegative Tensors", SIAM Journal on Matrix Analysis and Applications, 2010 Publication	<1 %
41	Submitted to Midlands State University Student Paper	<1 %

42	Submitted to Swinburne University of Technology Student Paper	<1 %
43	G. Kapetanaki, C. Rosenzweig. "Impact of climate change on maize yield in central and northern Greece: A simulation study with CERES-Maize", Mitigation and Adaptation Strategies for Global Change, 1997 Publication	<1 %
44	www.frontiersin.org Internet Source	<1 %
45	Stehfest, E.. "Simulation of global crop production with the ecosystem model DayCent", Ecological Modelling, 20071216 Publication	<1 %
46	V.A Alexandrov, G Hoogenboom. "The impact of climate variability and change on crop yield in Bulgaria", Agricultural and Forest Meteorology, 2000 Publication	<1 %
47	globalrust.org Internet Source	<1 %
48	www.iiss.org Internet Source	<1 %
49	Kattarkandi Byjesh, Soora Naresh Kumar, Pramod Kumar Aggarwal. "Simulating	<1 %

impacts, potential adaptation and vulnerability of maize to climate change in India", Mitigation and Adaptation Strategies for Global Change, 2010
Publication

---

50 Marwa G. M. Ali, Mukhtar Ahmed, Mahmoud M. Ibrahim, Ahmed A El Baroudy et al. "Optimizing sowing window, cultivar choice, and plant density to boost maize yield under RCP8.5 climate scenario of CMIP5", International Journal of Biometeorology, 2022 <1%
Publication

---

51 Sajjad Rahimi-Moghaddam, Jafar Kambouzia, Reza Deihimfard. "Optimal genotype×environment×management as a strategy to increase grain maize productivity and water use efficiency in water-limited environments and rising temperature", Ecological Indicators, 2019 <1%
Publication

---

52 Martin Kuradusenge, Eric Hitimana, Damien Hanyurwimfura, Placide Rukundo et al. "Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize", Agriculture, 2023 <1%
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

REFERENCES

[1] Renhai Zhong, Yue Zhu, Xuhui Wang, Haifeng Li, Bin Wang, Fengqi You, Luis F. Rodríguez, Jingfeng Huang, K.C. Ting, Yibin Ying, Tao Lin,Detect and attribute the extreme maize yield losses based on spatio-temporal deep learning,Fundamental Research,2022.

[2] Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.; Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. Agriculture 2023, 13, 225. https://doi.org/10.3390/agriculture13010225.

[3] Schierhorn, Florian & Hofmann, Max & Gagalyuk, Taras & Ostapchuk, Igor & Müller, Daniel. (2021). Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages. Climatic Change.169.

[4] Andrew Crane-Droesch 2018 Environ. Res. Lett. 13 114003DOI 10.1088/1748-9326/aae159

[5] Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni, Rubby Aworka, Jérémie Thouakesseh Zoueu, Franck Kalala Mutombo, Moez Krichen, Charles LebonMberi Kimpolo,Crops yield prediction based on machine learning models:Case of West African countries,Smart Agricultural Technology,Volume 2,2022.

[6] Kavita Jhajharia, Pratistha Mathur, Sanchit Jain, Sukriti Nijhawan,Crop Yield Prediction using Machine Learning and Deep Learning Techniques,Procedia Computer Science,Volume 218,2023.

[7] S. k. Gudepu and V. K. Burugari, "Weather Prediction using Support Vector based Genetic Algorithm in Rice Farming," 2021 International Conference on Computing, Communication and Green Engineering (CCGE), Pune, India, 2021, pp. 1-8, doi: 10.1109/CCGE50943.2021.9776357

[8] Mamatha, J.C. Kavitha,Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques,Measurement: Sensors,Volume 25,023,

[9] Guntukula, R., & Goyari, P. (2020). Climate Change Effects on the Crop Yield and Its Variability in Telangana, India. Studies in Microeconomics, 8(1), 119–148. https://doi.org/10.1177/2321022220923197

[10] Raza A, Razzaq A, Mehmood SS, Zou X, Zhang X, Lv Y, Xu J. Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. Plants (Basel). 2019 Jan 30;8(2):34. doi: 10.3390/plants8020034. PMID: 30704089; PMCID: PMC6409995.

[11] Rettie FM, Gayler S, K D Weber T, Tesfaye K, Streck T. Climate change impact on wheat and maize growth in Ethiopia: A multi-model uncertainty analysis. PLoS One. 2022 Jan 21;17(1):e0262951. doi: 10.1371/journal.pone.0262951. PMID: 35061854; PMCID: PMC8782302.

[12] Champaneri, Mayank & Chachpara, Darpan & Chandvidkar, Chaitanya & Rathod, Mansing. (2020). CROP YIELD PREDICTION USING MACHINE LEARNING. International Journal of Science and Research (IJSR). 9. 2.

[13] Ishwarya , Nagapooja BN, Raghavi R, Soundarya K, Prof. Chitra C, "CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHM" e-ISSN: 2582-5208 | Volume:04/Issue:07/July-2022

[14] Aksheya Suresh, K. Monisha, R. Pavithra,B. Marish Hariswamy, "Crop Selection and it's Yield Prediction", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

[15] Dr. T N Anitha., 2Anushka Aman., 3Gaurav V Salgaonkar., 4Jyotsna A Patel., 5Kunal Tapse, "Crop Yield Prediction Using Machine Learning", © 2023 JETIR May 2023, Volume 10, Issue 5

[16] Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. Remote Sens. 2020, 12, 236. https://doi.org/10.3390/rs12020236

[17] Strobl, C.; Boulesteix, A.L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinform. 2007, 8, 25. [Google Scholar] [CrossRef][Green Version]

[18] Benali, A., Carvalho, A. C., Nunes, J. P., Carvalhais, N., and Santos, A. (2012). "Estimating air surface temperature in Portugal using MODIS LST data." Remote Sensing of Environment, Elsevier Inc., 124, 108–121

[19] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., and Kudlur, M. (2016). "TensorFlow: A System for Large-Scale Machine Learning." In 12th USENIX symposium on operating systems design and implementation (OSDI 16), 101(C), 265–283

[20] Chen, Z., Zhang, B., Stojanovic, V., Zhang, Y., and Zhang, Z. (2020). "Event-based fuzzy control for T-S fuzzy networked systems with various data missing." Neurocomputing, Elsevier B.V., 417, 322–332

[21] Elboushaki, A., Hannane, R., Afdel, K., and Koutti, L. (2020). "MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences." Expert Systems with Applications, Elsevier Ltd, 139, 1128

[22] Gallego, J., Carfagna, E., and Baruth, B. (2010). "Accuracy, Objectivity and Efficiency of Remote Sensing for Agricultural Statistics." Agricultural Survey Methods, John Wiley & Sons, Ltd, Chichester, UK, 193–211.

[23] Hochreiter, S., and Schmidhuber, J. (1997). "Long Short-Term Memory." Neural Computation, MIT Press Journals, 9(8), 1735–1780.

[24] Khaki, S., Wang, L., and Archontoulis, S. V. (2020). "A CNN-RNN Framework for Crop Yield Prediction." Frontiers in Plant Science, 10(January), 1–14.

Appendix A

A.1 Scipy

Scipy is a crucial Python utility for scientific computing. For data analysis, machine learning, and scientific research, it is frequently used in conjunction with other libraries like NumPy, Pandas, and Matplotlib. Its potent mathematical functions and algorithms make it an essential library for tackling challenging issues in scientific computing. Scipy's optimisation procedures, which are used to determine the ideal parameter values for a particular model or system, are one of its main advantages. Scipy provides a number of optimisation methods, including global, least-squares, and unconstrained as well as constrained optimisation. These techniques can be applied in a variety of situations, including parameter estimation, parameter tuning, and model fitting to data. For resolving difficult integration issues, Scipy additionally offers a selection of numerical integration techniques, including both adaptive and non-adaptive integration. These operations can be used to compute probabilities in statistical applications, simulate dynamic systems, and solve differential equations. Scipy's capabilities for signal processing, which include Fourier transforms, wavelet transforms, digital filtering, and spectrum analysis, are another crucial aspect of the software. Applications for these tools include communication systems, audio analysis, and image processing. Additionally, Scipy provides a number of linear algebra operations, such as matrix decomposition, eigenvalue issues, and the resolution of linear equation systems. Such complicated issues in scientific computing as the solution of partial differential equations or the optimisation of large-scale systems depend on these functions. Scipy offers statistical functions for data analysis in addition to its mathematical capabilities, including probability distributions, hypothesis testing, and regression analysis. Numerous disciplines, such as biology, finance, and the social sciences, use these methods. Scipy is a crucial Python library for scientific computing since it provides strong algorithms and mathematical functions for resolving challenging issues in a variety of applications. It is a useful tool for data science, machine learning, and scientific research workflows and combines well with other scientific Python modules. Scipy additionally offers tools for interpolation, spatial data analysis, and image processing. These tools can be used to model geographical data, smooth and resample data, and analyse and alter photographs. The image processing module of Scipy includes tools for tasks including feature recognition, morphology, and picture filtering. These processes can be applied to a variety of image processing tasks,

such as computer vision, remote sensing, and medical imaging.Scipy's interpolation module provides a number of interpolation techniques, such as kriging, radial basis function interpolation, and spline interpolation. These techniques can be used to resample data, smooth noisy data, and complete data gaps. In Scipy, spatial data analysis entails looking at information about places or spatial coordinates. In many disciplines, including environmental science, geography, and urban planning, Scipy offers functions for spatial interpolation, spatial grouping, and spatial statistics. A complete and potent library, Scipy offers a wide variety of tools for scientific computing. Its capabilities can be utilised to find solutions to a range of issues in disciplines like physics, engineering, biology, and economics. It is a useful tool for researchers, data analysts, and developers alike due to its flexibility and compatibility with other scientific Python modules.

- To install the XGBoost library, we can use the pip package installer by executing the appropriate command in terminal or command prompt. For example, enter the command "pip install xgboost" to install the library
- SciPy: can accomplish this by executing the following command at a command prompt or terminal:install numpy scipy with pip