VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY

An Autonomous Institute Affiliated to University of Mumbai

Department of Computer Engineering

Project Report on

Farm Impact: Impact of Climate Change on Agriculture in Maharashtra

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2024-25

Submitted by

Vishakha Singh (D17B, 54)
Manasi Sharma (D17B, 51)
Anushka Shirode (D17B, 53)

Project Mentor

Dr. Gresha Bhatia

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
An Autonomous Institute Affiliated to University of Mumbai
Department of Computer Engineering

Certificate

This is to certify that ***Vishakha Singh (D17B, 54), Manasi Sharma (D17B, 51), Anushka Shirode (D17B, 53)*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "***Farm Impact: Impact of Climate Change on Agriculture in Maharashtra***" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Dr. Gresha Bhatia*** in the year 2024-25 .

This project report entitled ***Farm Impact: Impact of Climate Change on Agriculture in Maharashtra*** by ***Vishakha Singh (D17B, 54), Manasi Sharma (D17B, 51), Anushka Shirode (D17B, 53)*** is approved for the degree of ***Bachelor of Engineering in Computer Engineering.***

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date: 28 April 2025

Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This project report entitled ***Farm Impact: Impact of Climate Change on Agriculture in Maharashtra*** by ***Vishakha Singh (D17B, 54), Manasi Sharma (D17B, 51), Anushka Shirode (D17B, 53)*** is approved for the degree of ***Bachelor of Engineering in Computer Engineering.***

Internal Examiner

External Examiner

Head of the Department

Principal

Date: 28 April 2025
Place: Chembur, Mumbai.

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Vishakha Singh, D17B 54)

(Manasi Sharma, D17B 51)

(Anushka Shirode, D17B 53)

Date: 28 April 2025

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to the Deputy Head of the Computer Engineering Department **Dr. Gresha Bhatia** for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to the Head of the Computer Engineering Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

List Of Figures

List of Tables

Abstract

Maharashtra is a key contributor to India's agricultural sector, with its productivity heavily influenced by climatic and environmental factors. This study examines the relationship between crop yield, production, and critical variables such as temperature, rainfall, irrigation, and nutrient consumption using data from 1966 to 2023. Correlations between yield, production, and factors like weather, fertilizers, and soil nutrients are analyzed. SHAP (SHapley Additive exPlanations) identifies the most influential factors, while the Apriori algorithm uncovers associations between agricultural attributes. For forecasting, machine learning models—RFR (Random Forest Regressor), SVR (Support Vector Regressor), and GBR (Gradient Boosting Regressor)—are compared, with GBR emerging as the best. STL (Seasonal and Trend decomposition using Loess) is applied to GBR's time series data to reveal trends and seasonal patterns. This comprehensive approach provides actionable insights for enhancing agricultural productivity and sustainability in Maharashtra.

Chapter 1: Introduction

1.1 Introduction

Agriculture is the lifeline of India's economy, contributing nearly 18-20% to the national GDP and providing employment to over 50% of the population, particularly in rural areas. It ensures food security for the nation's 1.4 billion people and plays a crucial role in poverty reduction and economic stability. Among the country's major agricultural states, Maharashtra holds a prominent place. It encompasses a vast and ecologically diverse landscape, ranging from fertile plains in the western and coastal regions to semi-arid zones in Marathwada and Vidarbha. This diverse geography allows the state to cultivate a wide variety of crops, including cash crops like sugarcane, cotton, and soybeans, as well as food grains such as wheat, rice, and pulses.

Despite its significance, the agricultural sector in Maharashtra faces intensifying threats due to climate variability and environmental stress. In recent decades, the state has witnessed erratic monsoons, prolonged droughts, unseasonal rainfall, and rising temperatures, which have directly impacted crop yields, soil fertility, and water availability. Notably, regions like Marathwada and Vidarbha have suffered recurrent droughts, leading to severe water scarcity and agricultural distress. In May 2023, these areas recorded temperatures exceeding 50°C, which not only worsened soil degradation and moisture loss, but also severely disrupted cropping cycles and reduced farm productivity.

The vulnerability of Maharashtra's agriculture to climate extremes has had far-reaching socio-economic impacts, including rural indebtedness, forced migration, and an alarming rise in farmer suicides, particularly in drought-prone districts. According to studies, climate-induced crop failures have been a contributing factor in the agrarian crisis that continues to affect the region.

However, the ongoing challenges have also spurred innovation and adaptation in agricultural practices. Farmers, researchers, and policymakers are increasingly adopting climate-resilient strategies to counteract these risks. Precision farming, micro-irrigation techniques like drip and sprinkler systems, and climate-smart agriculture (CSA) practices are gaining traction. Additionally, the cultivation of drought-tolerant crop varieties, integrated nutrient management, and the use of remote sensing and weather forecasting technologies are enabling data-driven decision-making to optimize inputs and improve crop resilience.

In this context, predictive modeling and machine learning-based approaches offer immense potential to support sustainable agriculture by providing accurate yield forecasts, early warnings, and actionable insights. These technologies can help maximize productivity, minimize losses, and guide policy formulation aimed at ensuring food security and economic stability, especially in vulnerable regions like Maharashtra.

1.2 Motivation

Agriculture forms the backbone of rural livelihoods and national economies in many Asian countries, particularly in India. However, this critical sector faces mounting challenges due to climate variability and extreme weather events, which threaten not only food security but also the socio-economic stability of farming communities. In India alone, agriculture contributes nearly 20% to the GDP and provides livelihoods to over 50% of the population, making its sustainability a pressing concern.

According to a recent International Atomic Energy Agency (IAEA) report, farmers in six Asian countries have successfully increased their rice yields by adopting nuclear-derived, climate-smart agricultural practices. These methods, developed with support from the IAEA and the Food and Agriculture Organization (FAO), enable farmers to adapt to the challenges posed by changing climatic conditions. By employing improved soil and water management practices, and precision farming techniques, farmers have been able to enhance productivity in a sustainable manner, even under difficult environmental circumstances[1]. These outcomes demonstrate the potential of science-based, data-driven approaches in safeguarding agricultural productivity.

In stark contrast, India continues to grapple with the human toll of climate-induced agricultural distress. A study cited by the Deccan Herald attributes 59,000 farmer suicides over the past three decades to crop-damaging warmer temperatures during critical stages of the growing season[2]. Rising temperatures, erratic rainfall, and increased frequency of droughts have worsened the agrarian crisis, particularly in regions like Marathwada and Vidarbha, where farmers are highly dependent on monsoon rains for irrigation. The psychological and economic stress caused by crop failures due to climatic factors has not only undermined agricultural productivity but also led to severe socio-economic consequences, including loss of life.

These contrasting realities underscore the urgent need for innovative, climate-resilient agricultural strategies in India. There is a growing consensus among policymakers, researchers, and farming communities that data-driven predictive models can play a vital role in mitigating risks, improving decision-making, and ensuring sustainable agricultural practices. Predictive analytics leveraging machine learning (ML) and statistical models offer a proactive approach to understanding complex relationships between climate variables, soil health, fertilizer use, and crop productivity.

The motivation behind this study is to develop and validate robust predictive models for agricultural yield forecasting in Maharashtra, one of India's most climate-vulnerable states. By integrating advanced machine learning techniques (Random Forest, Gradient Boosting, Support Vector Regression) with interpretability tools like SHAP and association rule mining (Apriori algorithm), this research aims to provide actionable insights for stakeholders. The goal is to help farmers adapt to climate change, optimize resource use, and prevent further economic and human losses due to climate-induced agricultural failures.

1.3 Problem Definition

Objective: Analyze the relationship between crop production and climatic factors.

Datasets Used:

- Aggregate Features Dataset: Includes Yield, Production, Area, Fertilizer Composition, Soil Nutrients, and Weather Factors.
- Crop-Specific Dataset: Covers major crops such as Rice, Wheat, Sorghum, Sugarcane, and others along with climatic factors.

Goal:

- Find correlations between crop yield and climatic variables.
- Forecast future values for crop production and climatic features.
- Derive district wise insightful inferences to help the admins and officials.

1.4 Existing Systems

Overview of Existing Systems:

Traditional crop yield prediction models often rely on a limited set of features, such as soil properties and weather conditions, to forecast agricultural productivity. For instance, some systems focus primarily on soil quality and meteorological data, potentially overlooking other influential factors like fertilizer usage, irrigation practices, and land utilization patterns. This

narrow focus can lead to models that fail to capture the complex interplay of various agricultural inputs, resulting in less accurate and less actionable predictions for farmers.

Improvements Introduced by Our System:

Our approach enhances crop yield prediction by integrating a comprehensive array of features, including fertilizer application rates, rainfall levels, irrigation practices, temperature variations, total cultivated area, and proportions of barren and fallow land. By considering these diverse factors, our system captures the intricate interdependencies that influence crop yields. Utilizing advanced machine learning algorithms, our model analyzes these multifaceted relationships, leading to more accurate and reliable predictions. This holistic approach empowers farmers with actionable insights, enabling them to make informed decisions that optimize resource allocation and improve agricultural productivity.

Existing Systems in Crop Yield Prediction and Climate Impact Analysis

The existing systems reviewed in the papers mainly focus on predictive modeling for crop yield estimation and climate impact analysis.

1. Data Collection & Sources

- Historical Data Usage: Most systems rely on historical datasets, such as government records, Kaggle datasets, FAOSTAT, and data.gov.in, ranging from 1901 to 2014, depending on the study.
- Limited Real-Time Data: There is minimal integration of real-time data streams, such as IoT-based soil sensors or live climate feeds.
- Geographical Focus: Predominantly regional datasets, often limited to Maharashtra or other specific states. There is a lack of pan-India or multi-country datasets for broader applicability.
- Limited Features: The features often include temperature, rainfall, humidity, evapotranspiration, and area, with some studies adding soil pH, fertilizer composition, and crop types.

2. Machine Learning Models Used

Traditional Models

- Support Vector Machines (SVM) (SMO): Used in rice yield prediction but showed inferior performance compared to other methods like Naïve Bayes and Neural Networks.
- K-Nearest Neighbors (KNN): Implemented for yield estimation; simple but struggles with scalability.

- Decision Trees and Random Forest:
    - Common in crop prediction and classification tasks.
    - Random Forest often yields high accuracy (up to 97%) but lacks interpretability.

Neural Networks

- Artificial Neural Networks (ANN):
    - Applied to crop yield prediction.
    - Accuracy improved after optimizer tuning (RMSprop to Adam), but limited feature engineering and lack of model comparisons are common gaps.

Deep Learning

- LSTM (Long Short-Term Memory):
    - Used for climate forecasting and time-series predictions.
    - Demonstrated high accuracy (96.16%) but at the cost of high computational resources and no integration of external climate factors.
- Deep Neural Networks (DNN):
    - Applied in climate impact analysis.
    - Risks of overfitting and limited regional variation analysis.

3. Methodologies & Workflows

Common Workflow Steps

Data Collection → Preprocessing → Feature Selection/Engineering → Model Training → Evaluation

Evaluation Metrics

- Commonly Used: MAE, MSE, RMSE, Accuracy, Precision, Recall, F1-score.
- Lacking in Many Studies:
    - R^2 scores (for model reliability).
    - MAPE (Mean Absolute Percentage Error).
    - Explainability metrics like SHAP (for feature importance analysis).

Summary of Limitations in Existing Systems

- Fragmented and Data-Constrained: Many studies lack comprehensive datasets, limiting their generalizability.
- Model-Limited: Most studies rely on basic machine learning models, with underutilization of deep learning and hybrid approaches.

- No Real-Time Adaptability: Most models are trained on static historical data, lacking real-time integration with climate sensors or satellite data.
- Overemphasis on Yield Prediction: Few studies address holistic farming sustainability, resource optimization, or climate risk mitigation.
- Minimal Field-Level Validation: Models are rarely tested in real-world farming conditions, making their practical effectiveness uncertain.

The existing systems have laid a foundation for data-driven agricultural decision-making, but lack real-time adaptability, explainability, and practical deployment for real-world applications.

1.5 Lacuna of the existing systems

Stage 1: Data Collection

1. No Real-Time Data Integration

Most current agricultural forecasting systems rely heavily on historical data collected from government records, census reports, and agricultural surveys, which are typically aggregated on an annual or seasonal basis. This data, while valuable, does not capture real-time changes in weather patterns, soil moisture, pest infestations, or irrigation levels. The lack of integration with Internet of Things (IoT) devices, remote sensing satellites, or on-field sensors hampers the system's ability to provide dynamic, up-to-date insights. Without real-time monitoring, the predictions may fail to reflect sudden environmental changes, such as unseasonal rainfall or pest outbreaks, which are critical for timely decision-making by farmers.

2. Data Quality and Availability Issues

The inconsistency and incompleteness of available agricultural datasets pose a major challenge. Data often suffers from missing values, incorrect entries, non-standardized formats, and lack of granularity (e.g., district- or farm-level data may be unavailable or incomplete). Additionally, longitudinal data, necessary for analyzing long-term trends and climate impacts, is often fragmented across different sources, making integration and preprocessing labor-intensive. Moreover, certain important parameters, such as soil nutrient content, pesticide application, or market prices, are either sparsely collected or not publicly available, limiting comprehensive analysis.

Stage 2: Analysis and Modeling

1. Lack of Multi-Scale Analysis

Existing models often focus on single-scale analysis, typically at the district or state level, without accounting for variations across multiple spatial and temporal scales. For example, predictions made at a state level may not be relevant for individual districts or micro-regions that have unique agro-climatic conditions. Similarly, models may ignore short-term seasonal forecasts in favor of annual yield projections, missing important intra-seasonal events (e.g., mid-season droughts or floods) that impact crop outcomes. A multi-scale modeling framework is necessary to address local variations and to provide both short-term operational forecasts and long-term strategic insights.

2. Underutilization of Advanced Models

Despite advances in machine learning (ML) and deep learning (DL), many systems still rely on conventional statistical techniques (like linear regression or time series models) or basic machine learning algorithms (such as Random Forest or Decision Trees), without exploring more complex models like LSTM (Long Short-Term Memory networks) for time series prediction, CNNs (Convolutional Neural Networks) for remote sensing imagery analysis, or hybrid models that combine multiple algorithms. These advanced techniques have shown superior capabilities in capturing non-linear relationships, time dependencies, and multi-modal data integration, but are underutilized in most agricultural applications.

3. Limited Extreme Weather Forecasting

Current models primarily focus on average climate conditions and normal crop cycles, but fail to adequately predict extreme weather events such as heatwaves, floods, cyclones, or unexpected droughts, which have a disproportionately high impact on crop yield and farmer livelihoods. Many prediction models lack early warning systems that alert stakeholders about impending climatic anomalies. Furthermore, the impact assessment of these extreme events on different crop types and growth stages is often missing, leading to inadequate risk mitigation strategies.

Stage 3: Prediction and Forecasting

1. Overemphasis on Yield Prediction

A majority of existing systems are heavily centered on predicting crop yields, while neglecting other critical aspects of agricultural sustainability, such as soil health monitoring,

water resource management, pest and disease outbreaks, and socio-economic factors (e.g., market trends, input costs). Yield prediction alone cannot capture the holistic health of an agricultural system or inform comprehensive policy decisions. There is also a lack of economic forecasting that factors in price volatility, supply chain dynamics, and market demand, which are essential for farmers' profitability and livelihood security.

Stage 4: Implementation and Resource Allocation

1. High Computational and Resource Demands

Advanced modeling approaches, especially those using deep learning and high-resolution remote sensing data, require significant computational power, storage, and data processing capabilities, which are often unavailable in rural or resource-limited settings. Additionally, hardware costs for deploying IoT devices, sensor networks, and satellite data processing systems can be prohibitively expensive for smallholder farmers and local governments. The digital divide, including limited internet connectivity and technical literacy, further hampers the scalability and accessibility of these systems. As a result, the implementation of advanced technologies often remains confined to pilot studies or highly developed regions, failing to reach marginalized and vulnerable farming communities that need them the most.

1.6 Relevance of the Project

Agriculture is the backbone of India’s economy, but it is under severe stress due to climate variability, frequent droughts, and extreme temperatures, particularly in Maharashtra’s Marathwada and Vidarbha regions. These climatic challenges have led to reduced crop productivity, economic distress, and even 59,000 farmer suicides over the past three decades. Meanwhile, success stories from other Asian countries show that climate-smart agricultural practices, supported by scientific and data-driven approaches, can lead to significant productivity improvements. There is an urgent need to replicate such sustainable, adaptive methods in India.

This project is highly relevant because it bridges the gap between conventional farming and modern predictive analytics. It integrates advanced machine learning techniques—Random Forest Regressor (RFR), Support Vector Regressor (SVR), and Gradient Boosting Regressor (GBR)—and applies SHAP analysis for interpretability, Apriori association rule mining for discovering hidden patterns, and STL decomposition for analyzing trends and seasonality. Unlike many existing models that focus narrowly on yield prediction, this project takes a

holistic approach, factoring in soil health, irrigation, fertilizer usage, and climatic trends, offering comprehensive insights into agricultural productivity.

Project relevance:

1. For Policymakers

   It provides data-driven insights to support climate-adaptive agricultural policies.

   Enables targeted interventions like crop insurance, subsidy allocation, and disaster relief, based on predictive models rather than retrospective data.

   Assists in crafting strategies for climate change adaptation, enhancing national food security and rural stability.

2. For Researchers

   The project contributes to the advancement of predictive modeling in agriculture by demonstrating the effectiveness of machine learning algorithms like GBR in yield and production forecasting.

   It highlights the use of interpretability techniques (SHAP) and association mining, paving the way for further academic research in climate-smart farming technologies.

   Encourages multidisciplinary studies combining data science, climatology, and agricultural sciences.

3. For Farmers

   Empowers farmers with localized, actionable insights on crop planning, fertilizer use, and irrigation management.

   Helps minimize input costs, reduce risk, and maximize yield and profitability, especially in climate-vulnerable regions.

   By offering predictive insights into climate impacts, the system enhances resilience against unpredictable weather events, supporting sustainable livelihoods.

Broader Impact

By addressing critical gaps in data collection, forecasting, and resource allocation, the project promotes climate-resilient agriculture. It supports sustainable rural development in Maharashtra and serves as a scalable model for other regions in India and similar agrarian economies worldwide.

Chapter 2: Literature Survey

A. Brief Overview of Literature Survey

The literature survey covers various machine learning approaches for crop yield prediction and climate forecasting, highlighting the strengths and limitations of different models. Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have been explored, with ANN demonstrating improved accuracy through hyperparameter tuning. Random Forest emerges as a robust model for yield prediction due to its ability to handle nonlinear relationships, though studies suggest integrating additional factors like soil nutrients and economic conditions for better applicability. Hybrid approaches, such as combining Bayesian inference with Random Forest, enhance model transparency but require real-world validation. Deep learning models like Long Short-Term Memory (LSTM) are effective for climate forecasting but computationally expensive. Studies emphasize the importance of real-time data integration, feature expansion, and comparative analyses with different algorithms to improve prediction accuracy. Additionally, research on rainfall forecasting using Multilayer Perceptron (MLP) and climate change impact assessment on plantations using deep learning highlights the potential of ML in agriculture. Future improvements should focus on incorporating IoT-based real-time sensor data, economic modeling, and hybrid models that combine traditional ML and deep learning for enhanced predictive performance.

B. Related Works

2.1 Research Papers Referred

1. Paper: "Rice Crop Yield Prediction in India using Support Vector Machines"

Abstract:

This paper [1] explores machine learning techniques for rice crop yield prediction in India using a dataset from 27 districts in Maharashtra (1998–2002). The study employs Support Vector Machines (SVM) with the Sequential Minimal Optimization (SMO) algorithm to forecast yield based on climate parameters like temperature, precipitation, evapotranspiration, and cultivated area. Performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error

(RRSE) are used for evaluation. However, results indicate that SMO underperforms compared to other machine learning models like Naïve Bayes and Neural Networks, raising concerns about the suitability of SVM for crop yield prediction.

Inferences:

The study highlights the potential and limitations of SVM in agricultural forecasting. While the approach leverages real-world government datasets and considers key climatic factors, its limited dataset (only five years of data) restricts long-term trend analysis. Additionally, SMO's performance lags behind other models, indicating that alternative ML techniques like Random Forest, XGBoost, or Neural Networks might yield better accuracy. Expanding the dataset, including additional climate indicators (such as soil quality and fertilizer usage), and comparing SVM with advanced ML models could improve predictive capabilities. Furthermore, incorporating GIS and remote sensing data could enhance the spatial generalization of the findings.

2. Paper: "A Machine Learning Approach to Predict Crop Yield and Success Rate"

Abstract:

This study [2] applies Artificial Neural Networks (ANN) to predict crop yield and success rates in Maharashtra, India, using a dataset spanning 1998–2014 with 2.4 lakh records, later filtered to 12,000 for analysis. Initially, a three-layer ANN with the RMSprop optimizer achieved 45% accuracy, which improved to 90% after tuning with the Adam optimizer. The research highlights the importance of hyperparameter optimization in deep learning models. The dataset includes features like area, crop type, state, district, season, and production, with preprocessing conducted using Python libraries like Pandas, Seaborn, and Matplotlib. The study uses evaluation metrics such as MAE, MSE, and RMSE to assess performance.

Inferences:

The study demonstrates the effectiveness of deep learning in agricultural prediction, particularly the improvements achieved through hyperparameter tuning. However, it is limited by the lack of comparison with traditional ML models like Random Forest or XGBoost. The study also does not account for external factors such as soil fertility, pest outbreaks, and economic conditions. Future improvements could include feature engineering to incorporate additional agricultural and environmental parameters, comparative analysis

with tree-based models, and expanding the dataset to more regions. A hybrid approach integrating deep learning with traditional ML models could also enhance performance.

3. Paper: "Crop Yield Prediction Using Random Forest Algorithm"

Abstract:

This study [3] applies the Random Forest algorithm to predict crop yields based on climate parameters such as temperature, rainfall, humidity, and soil quality. The model leverages historical yield records to improve agricultural decision-making and provides early yield estimates to farmers. Random Forest is chosen for its high accuracy and robustness in handling nonlinear relationships. The research evaluates model performance using standard error metrics like RMSE and MAE. However, it does not integrate real-time weather data or account for economic factors influencing agricultural productivity.

Inferences:

The study effectively demonstrates the utility of Random Forest in yield prediction but is limited by the absence of real-time weather updates and economic considerations. While the model provides accurate forecasts, integrating additional environmental variables such as soil nutrients, market conditions, and government policies could improve practical applicability. Future work should focus on real-time sensor data integration using IoT, economic modeling for cost-benefit analysis, and extending the study to diverse climatic regions for broader applicability.

4. Paper: "Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques"

Abstract:

This paper [4] emphasizes the role of rainfall prediction in improving agricultural yield. It employs a Multilayer Perceptron (MLP) model trained on historical rainfall data from Kaggle (1901–2002) and crop yield records containing attributes such as state, year, crop type, area, rainfall, and production. The study evaluates model performance using MSE and NMSE but does not provide additional metrics like R^2 or RMSE. MLP is selected for its ability to model nonlinear relationships in climate data.

Inferences:

The study highlights MLP's potential in rainfall prediction but suffers from outdated data (ending in 2002), limiting its relevance for current climate conditions. Additionally, MLP's complexity makes it computationally expensive compared to simpler tree-based models like XGBoost. Future improvements should focus on updating datasets, integrating real-time weather data, and comparing MLP with alternative ML techniques. Explainability techniques such as SHAP analysis could also enhance interpretability.

5. Paper: "A Creative Use of Machine Learning for Crop Prediction and Analysis"

Abstract:

This study [5] develops a machine learning-based crop recommendation system using a dataset of 25,000+ records. The system consists of two modules: crop production analysis (trend detection) and crop recommendation (predicting the best crop based on environmental factors like temperature, humidity, pH, and rainfall). The study compares SVM, Naïve Bayes, Random Forest, and Decision Tree models but does not specify which performed best.

Inferences:

The research effectively utilizes ML for crop prediction but lacks clarity in model selection criteria and dataset preprocessing. Expanding the feature set to include soil nutrients, pest infestation, and satellite data could improve accuracy. A real-world validation phase with farmers would enhance practical applicability.

6. Paper: "Influence of Causal Inference for Crop Prediction"

Abstract:

This study [6] introduces a hybrid model combining Random Forest and Bayesian inference for crop prediction, achieving 97.2% accuracy. Bayesian inference is used to establish causal relationships between soil and climate factors, refining predictions. The dataset includes attributes like Nitrogen, Phosphorus, Potassium, Rainfall, Humidity, Temperature, and pH.

Inferences:

The integration of causal inference enhances model transparency and reliability. However, the study lacks real-world validation and does not explore deep learning techniques. Future

improvements should focus on expanding datasets, incorporating economic factors, and testing the model on real-world farms.

7. Paper: "A Case Study on the Application of Machine Learning to the Process of Crop Forecasting"

Abstract:

This study [7] compares SVM, Decision Tree, and Random Forest models for crop yield prediction in Maharashtra, using a 10-year dataset from government sources. Random Forest achieved the highest accuracy (97%). The model provides farmers with crop selection recommendations based on environmental conditions.

Inferences:

While the study offers practical applications, it is limited to Maharashtra and does not incorporate real-time IoT data. Expanding the dataset and integrating economic variables could improve decision-making for farmers.

8. Paper: "Climate Forecasting: Long Short-Term Memory Model using Global Temperature Data"

Abstract:

This study [8] applies Long Short-Term Memory (LSTM) networks to global temperature forecasting, achieving 96.16% accuracy. The model captures temporal dependencies in climate data using attributes like date, average temperature, and uncertainty.

Inferences:

LSTM effectively handles time-series forecasting but is computationally expensive. Expanding the model to include greenhouse gas emissions and oceanic patterns could improve broader climate modeling capabilities.

9. Paper: "Climate Change Impact Analysis on Plantation"

Abstract:

This paper [9] assesses climate change's impact on plantations using deep learning techniques like Deep Neural Networks and Reinforcement Learning. A dataset of 28,242 records (1990–2013) is used, with RandomForestRegressor and BaggingRegressor performing best.

Inferences:

The study successfully applies ML to plantation analysis but lacks economic and regional adaptation factors. Future work should incorporate economic modeling and regional crop adaptation strategies.

10. Paper: "Agriculture Yield Estimation Using Machine Learning Algorithms"

Abstract:

This paper [10] proposes KNN for crop yield estimation, achieving 90% accuracy. The model uses historical yield, weather, and soil data to make predictions.

Inferences:

While KNN is interpretable and effective for small datasets, it struggles with scalability. Future work should explore hybrid models combining KNN with ensemble techniques for better performance.

2.2 Comparison of FarmImpact with Existing Systems

The following table presents a **detailed comparison** of the methodologies and features used in the **FarmImpact** system (our paper) against the **existing systems** reviewed in the literature survey.

Aspect	**Existing Systems (Literature Survey)**	**FarmImpact (Our Paper)**
Data Coverage	Short-term datasets (5–16 years).	Long-term dataset (1966–2023) for Maharashtra.
Feature Selection	Basic climate factors (rainfall, temperature, evapotranspiration).	Uses additional agricultural parameters: irrigation, fertilizer use (NPK), soil nutrients, and socio-economic indicators.
Algorithms Used	SVM, ANN, Random Forest, Decision Tree, Multilayer Perceptron (MLP).	Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR),

		Support Vector Regressor (SVR), with STL for trend analysis.
Model Performance	Limited accuracy: SVM and ANN performed poorly in some studies (45%–90%).	GBR achieves the highest accuracy with $R^2 > 0.97$, validated through RMSE and MAE metrics.
Explainability	No feature importance analysis, black-box models like ANN and MLP.	SHAP (Shapley Additive Explanations) is used to identify the top 10 influential factors for crop yield and production.
Pattern & Relationship Analysis	Some studies use correlation and regression models but lack deeper insights.	Apriori Algorithm is applied to uncover hidden associations between climate, soil nutrients, and crop yield.
Forecasting Approach	Limited forecasting, often based on statistical regression.	Uses STL decomposition for seasonality and trend analysis, improving long-term forecasting up to 2040.
Visualization & Interpretability	Mostly static graphs, few interactive elements.	Interactive graphs, SHAP plots, and Apriori-based rule visualization for better decision-making.
Geographical Scope	Limited to Maharashtra, but datasets often lack district-wise granularity.	District-wise predictions and analysis, allowing region-specific agricultural planning.
Real-World Applicability	No real-time integration, models trained on static historical data.	Designed to support farmers, policymakers, and researchers, enabling data-driven decision-making for sustainable agriculture.

Table 1. Comparison between existing system and our work

Chapter 3: Requirement Gathering for the Proposed System

3.1 Introduction to Requirement Gathering

The success of FarmImpact, a system for predicting agricultural trends in Maharashtra, depends on well-defined functional, non-functional, and technical requirements. This requirement gathering process ensures that the system effectively analyzes climate variables, predicts yield and production, and provides actionable insights for farmers and policymakers. Given the complex nature of agriculture and its dependency on multiple factors such as climate, soil nutrients, and irrigation, this system must integrate advanced machine learning models, conduct statistical analysis, and support visualization of trends.

The requirements are classified into:

- Functional Requirements – defining system capabilities.
- Non-Functional Requirements – ensuring performance, usability, and scalability.
- Hardware & Software Requirements – specifying necessary infrastructure.
- Constraints – identifying limitations affecting implementation.

3.2 Functional Requirements

1. Predictive Modeling
    - The [11] system should use Random Forest Regressor (RFR), Support Vector Regressor (SVR), and Gradient Boosting Regressor (GBR) to predict the impact of climate and agricultural variables on yield and production.
    - The system should compare the performance of these models to determine the best forecasting technique.
    - STL (Seasonal and Trend Decomposition using Loess) should be applied to decompose forecasted trends into seasonal, trend, and residual components, ensuring reliable analysis.
2. Data Collection & Analysis
    - The system must collect climate, soil, and agricultural data from 1966 to 2023, sourced from ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) and other agricultural datasets.
    - The system should preprocess data by handling missing values, performing normalization, and conducting correlation analysis to identify relationships

between variables such as temperature, rainfall, irrigation, and fertilizer consumption.
    - SHAP (SHapley Additive Explanations) should be used to identify the top 10 influential factors affecting yield and production.
    - Apriori Algorithm should be implemented to find associations between climate, fertilizer use, and agricultural productivity.
3. Visualization & Reporting
    - The system should generate interactive graphs and charts illustrating trends in agricultural productivity based on climate and soil conditions.
    - The Apriori rule-based relationships should be visualized using network graphs, highlighting associations among key agricultural attributes.
    - The system should provide district-wise insights, showcasing historical and forecasted trends in Maharashtra.

3.3 Non-Functional Requirements

1. Accuracy
    - The system must maintain high prediction accuracy, with models evaluated based on R^2, MSE, RMSE, MAE, and MAPE.
    - The selected GBR model should provide optimal accuracy with low error margins, outperforming other models.
2. Scalability
    - The system should handle large datasets from multiple districts spanning several decades (1966–2023).
    - It must support the future integration of additional data sources, such as real-time sensor data and satellite imagery.
3. Usability
    - The user interface should be intuitive and accessible for farmers, policymakers, and researchers.
    - The system should present easily interpretable insights, using graphs, SHAP visualizations, and Apriori-based relationships.
4. Reliability & Maintainability
    - The model should be retrained periodically [12] as new agricultural data becomes available, ensuring continuous improvement in accuracy.

- The system must log errors and provide alerts in case of significant deviations in climate variables affecting yield.

3.4 Hardware & Software Requirements

Hardware Requirements:

- Processor (CPU): Intel Core i5 or higher
- RAM: 16 GB (to handle large datasets and machine learning computations)
- Storage: SSD (512 GB or more) for faster data processing
- GPU (Optional): NVIDIA GTX 1650 or higher (if deep learning models are incorporated in future versions)

Software Requirements:

- Python (Primary language for data processing and machine learning)
- Pandas: For handling and preprocessing time series agricultural data.
- NumPy: For numerical computations and matrix operations.
- Scikit-learn: For implementing RFR, SVR, and GBR models.
- Statsmodels: For STL decomposition and statistical analysis.
- Matplotlib & Seaborn: For visualizing forecasting trends and model performance.
- Jupyter Notebooks: For interactive model development and step-by-step data analysis.
- NetworkX: For visualizing Apriori algorithm-based relationships.

3.5 Constraints

1. Data Availability & Quality
    - The accuracy of predictions depends on the availability and quality of agricultural and climate data.
    - Some datasets may have missing values, requiring imputation techniques.
2. Computational Complexity
    - GBR and SVR require high computational power, making them resource-intensive for large-scale implementation.
    - STL decomposition increases processing time, especially when applied to multiple variables.

3. Model Generalization
    - The models are trained on historical Maharashtra data, so their applicability to other regions may require retraining with localized datasets.
    - The system does not yet integrate real-time climate monitoring, which could improve predictive accuracy in future versions.

Chapter 4: Proposed Design

4.1 Block diagram of the system

The block represents a structured approach to analyzing agricultural and climate data using statistical and machine learning techniques. It begins with data collection from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) [13], including variables such as yield, production, area, annual rainfall, temperature, NPK fertilizer consumption, forest area, and uncultivated land. The data is categorized into crop-wise data (e.g., rice, wheat, sugarcane, sorghum, oilseeds) and Maharashtra aggregate data. Following this, exploratory data analysis (EDA) is performed, incorporating correlation analysis [14], SHAP (Shapley Additive Explanations) for feature importance [15], and the Apriori algorithm for association rule mining [16]. Next, forecasting future climate and agricultural conditions is carried out using ARIMA (AutoRegressive Integrated Moving Average) for time series forecasting [17] and machine learning models such as Random Forest Regressor, Support Vector Regressor, and Gradient Boosting Regressor to predict agricultural outcomes based on historical trends [18]. Finally, trend analysis is conducted using Seasonal-Trend decomposition with Loess (STL) [19] to break down time-series data into trend, seasonal, and residual components, providing deeper insights into long-term agricultural and climate patterns. This structured methodology ensures a comprehensive understanding of historical trends and future forecasts, aiding in agricultural decision-making [20].

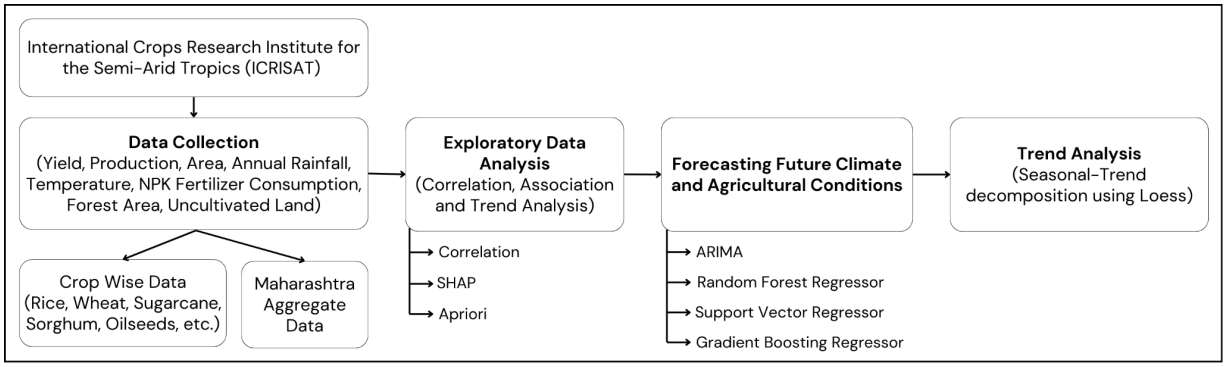


Fig. 1: Block Diagram

4.2 Modular design of the system

The modular diagram outlines a structured pipeline for agricultural data analysis, beginning with the Data Acquisition Module, which collects raw data. This data undergoes preprocessing in the Feature Engineering Module, where relevant attributes are extracted and refined. The refined dataset is then used in two parallel processes: the Statistical Analysis Module, which derives insights through exploratory data analysis and trend identification, and the Forecasting Module, which predicts future values. The Machine Learning Module utilizes both historical and forecasted data to generate predictive models, feeding results into the Visualization & Interpretation Module for graphical representation and meaningful insights. Finally, the Conclusion & Insights Module synthesizes the findings, providing actionable recommendations for agricultural planning and decision-making.

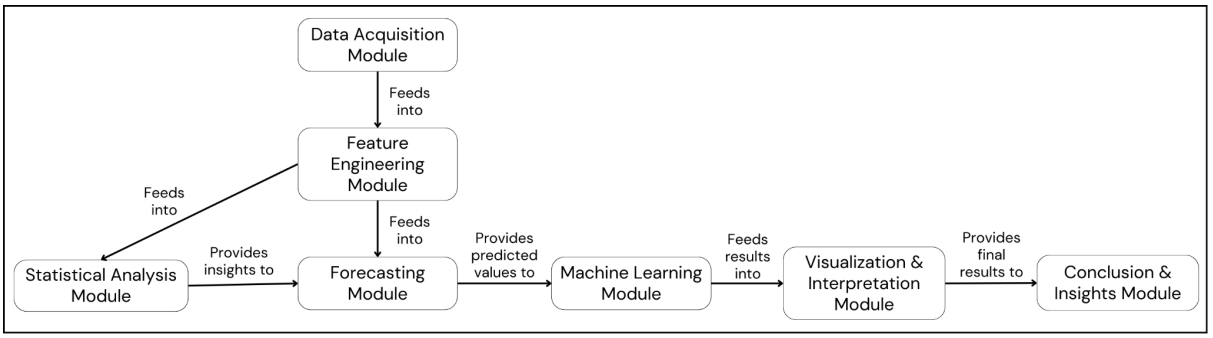


Fig. 2: Modular Diagram

4.3 Detailed Design

The detailed design diagram presents a structured methodology for analyzing agricultural data. It begins with Data Collection, focusing on crop-wise and Maharashtra aggregate data. The impact of fertilizers, soil nutrients, and weather parameters on yield and production is examined using Correlation. Feature importance is explained through SHAP, highlighting how each factor influences yield and production. Apriori is employed to identify associations between climatic conditions, fertilizers, and agricultural output. Future values of climate and agricultural metrics are forecasted using ARIMA, with a noted limitation on small datasets due to insufficient data points affecting trend estimation and seasonality detection. To enhance predictive accuracy, robust non-linear regression models such as SVR, RFR, and GBR are implemented for yield and production prediction. Finally, STL decomposition is applied to break down the forecasted GBR data into trend, seasonality, and residual components for better interpretability.

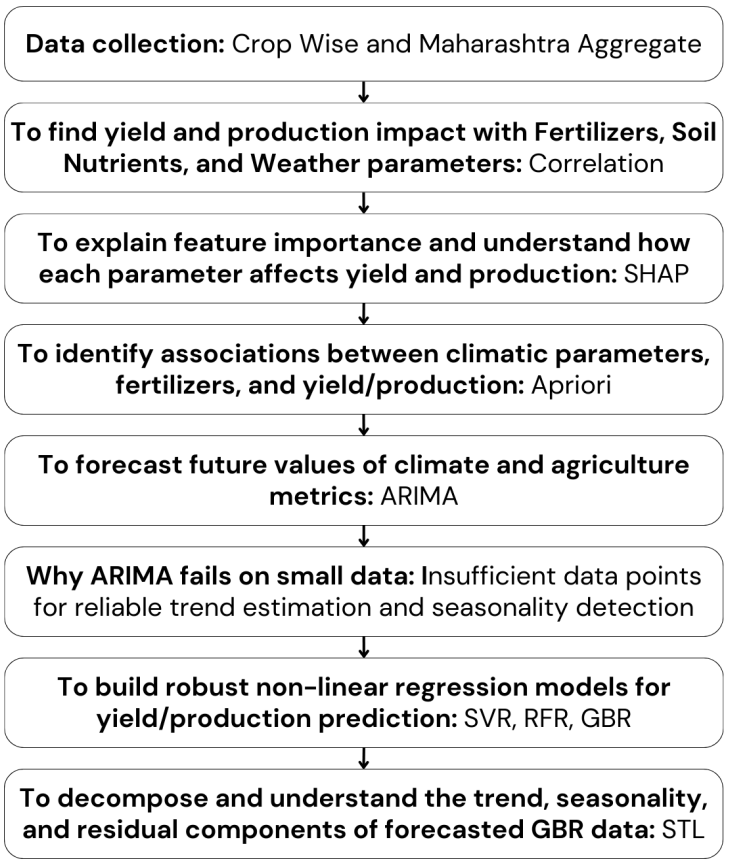


Fig. 3: Detailed Design

4.4 Project Scheduling & Tracking using Timeline / Gantt Chart

The project progresses through four key phases, starting with Data Collection and Preparation (June 2024 - August 2024), where agricultural and climate-related data are gathered, including manual scraping of government reports and transitioning to ICRISAT data for consistency. In the Exploratory Analysis and Feature Importance phase (August 2024 - November 2024), exploratory data analysis (EDA) is conducted to identify trends and anomalies, followed by correlation analysis to examine relationships between agricultural yield and climatic factors. SHAP feature importance is applied to determine influential variables, and the Apriori algorithm is used to uncover associations between agricultural inputs and outputs. The Forecasting and Modeling Phase (October 2024 - February 2025) focuses on predicting future climate and agricultural conditions using ARIMA for key variables in 2025, 2030, and 2035, while machine learning models such as Random Forest Regressor, Support Vector Regressor, and Gradient Boosting Regressor (GBR) predict future crop yields based on historical trends. Finally, the Trend Analysis and Paper Writing phase (December 2024 - March 2025) involves STL decomposition to analyze long-term

agricultural trends, culminating in the compilation of research findings into a final report and paper, ensuring thorough refinement before completion in March 2025.

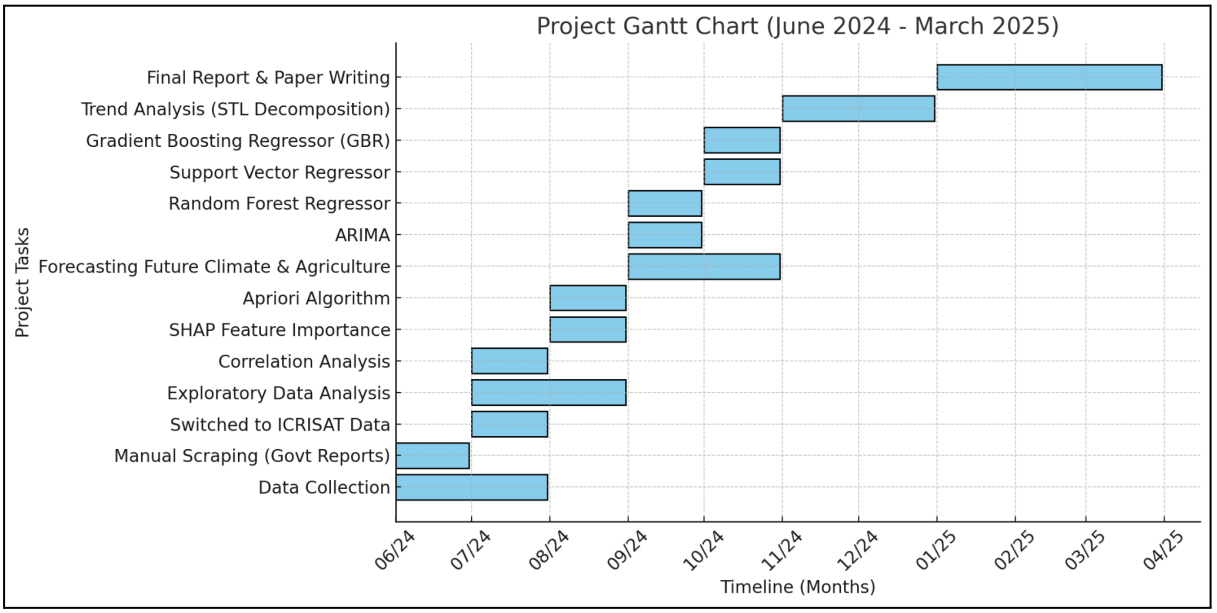


Fig. 4: Gantt Chart

Chapter 5: Implementation of the Proposed System

5.1 Methodology employed for development

1. Data Collection: The study begins with the collection of agricultural data, categorized both by crop type and at the state-level aggregate for Maharashtra. This data serves as the foundation for analyzing agricultural trends, incorporating variables like production, yield, soil nutrients, fertilizers, and climate parameters.
2. Correlation Analysis for Impact Assessment: A correlation analysis is performed to evaluate the relationships between yield/production and influencing factors such as fertilizers, soil nutrients, and weather parameters. This step helps identify key drivers of agricultural output and provides insights into their individual contributions.
3. Feature Importance using SHAP: The SHAP (SHapley Additive exPlanations) method is used to interpret model predictions by quantifying the importance of each feature. This allows for a deeper understanding of how specific parameters impact agricultural yield and production.
4. Association Rule Mining (Apriori Algorithm): The Apriori algorithm is applied to identify hidden associations between climatic variables, fertilizers, and agricultural productivity. This method uncovers frequent patterns and relationships among factors that influence crop yield and production.
5. Forecasting Future Trends with ARIMA: The ARIMA (AutoRegressive Integrated Moving Average) model is employed to predict future values of climate and agricultural indicators based on historical data. This approach captures time-dependent trends and seasonality in the dataset.
6. Challenges of ARIMA on Small Data: The study identifies a key limitation of ARIMA when dealing with small datasets—insufficient data points lead to unreliable trend estimation and difficulties in detecting seasonality. This highlights the need for alternative modeling techniques.
7. Non-Linear Regression Models for Prediction: To overcome ARIMA's limitations, machine learning-based regression models such as Support Vector Regression (SVR), Random Forest Regression (RFR), and Gradient Boosting Regression (GBR) are implemented. These models enhance predictive accuracy by capturing complex, non-linear relationships between features.

8. Time Series Decomposition using STL: The final step involves using Seasonal-Trend decomposition using LOESS (STL) to break down the forecasted GBR data into its trend, seasonal, and residual components. This decomposition aids in understanding underlying patterns and refining the forecasting process.

5.2 Algorithms and flowcharts for the respective modules developed

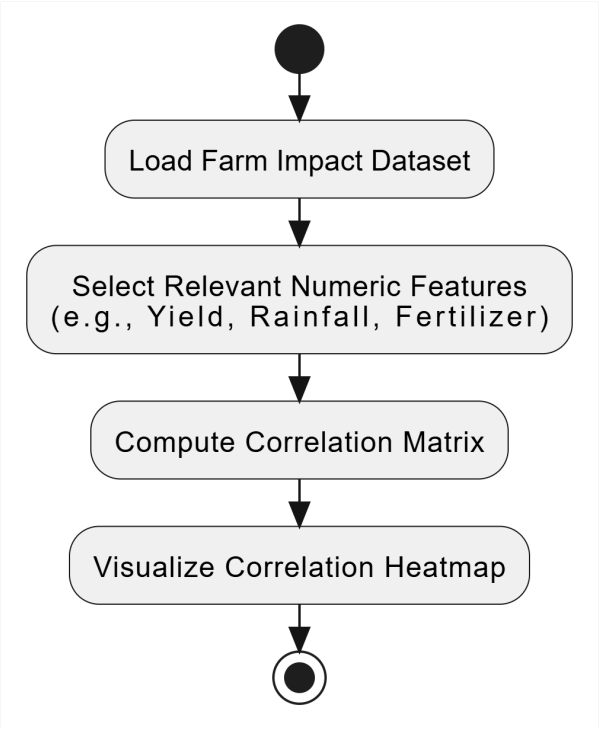


Fig. 5: Methodology of Correlation

Step 1: Load Dataset → Read the dataset containing numerical and categorical data.

Step 2: Select Numeric Columns → Extract only numerical features since correlation works with numbers.

Step 3: Compute Correlation → Use statistical methods (like Pearson or Spearman) to measure the relationship between features.

Step 4: Analyze Correlation Matrix → Identify strong, weak, positive, and negative correlations between variables.

Step 5: Visualize & Interpret → Display the correlation matrix as a table or heatmap for better insights.

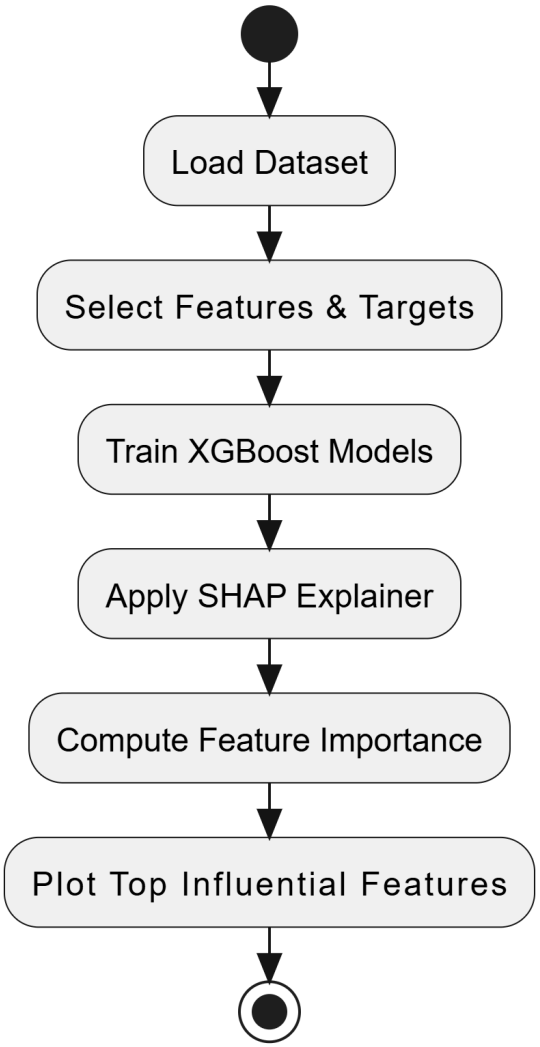


Fig. 6: Methodology of SHAP

Step 1: Load Dataset & Train Model → Read the dataset, preprocess features, and train a machine learning model (e.g., XGBoost).

Step 2: Create SHAP Explainer → Use the trained model to generate SHAP explanations for predictions.

Step 3: Compute SHAP Values → Calculate how each feature contributes to a particular prediction.

Step 4: Rank Feature Importance → Determine the most influential features based on SHAP values.

Step 5: Visualize & Interpret → Use SHAP summary plots or bar charts to understand feature impact on model decisions.

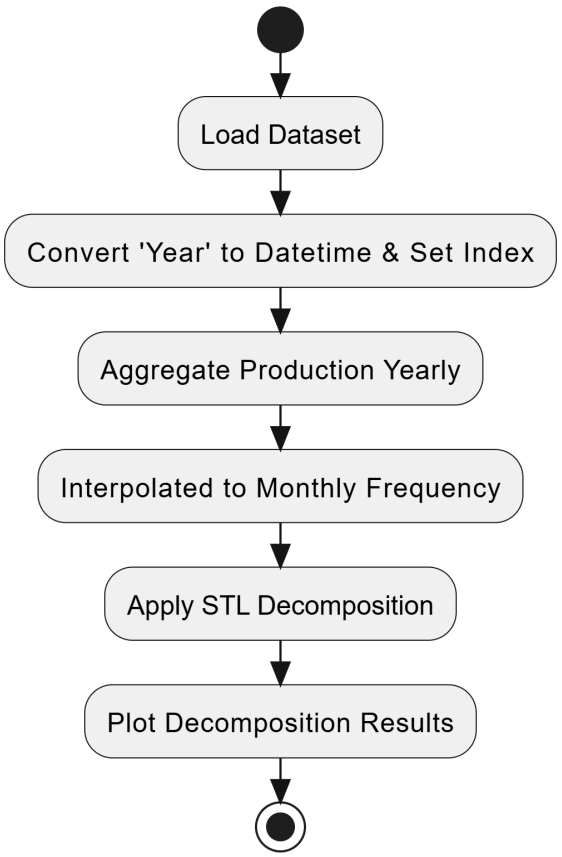


Fig. 7: Methodology of STL

Step 1: Load & Preprocess Data → Convert time-related features into datetime format and set the time index.

Step 2: Aggregate & Interpolate → Resample data (e.g., yearly to monthly) and fill missing values via interpolation.

Step 3: Apply STL Decomposition → Separate the time series into trend, seasonal, and residual components.

Step 4: Analyze Components → Examine trends (long-term patterns), seasonality (recurring cycles), and residuals (random noise).

Step 5: Visualize & Interpret → Plot the decomposed components to understand the underlying structure of the data.

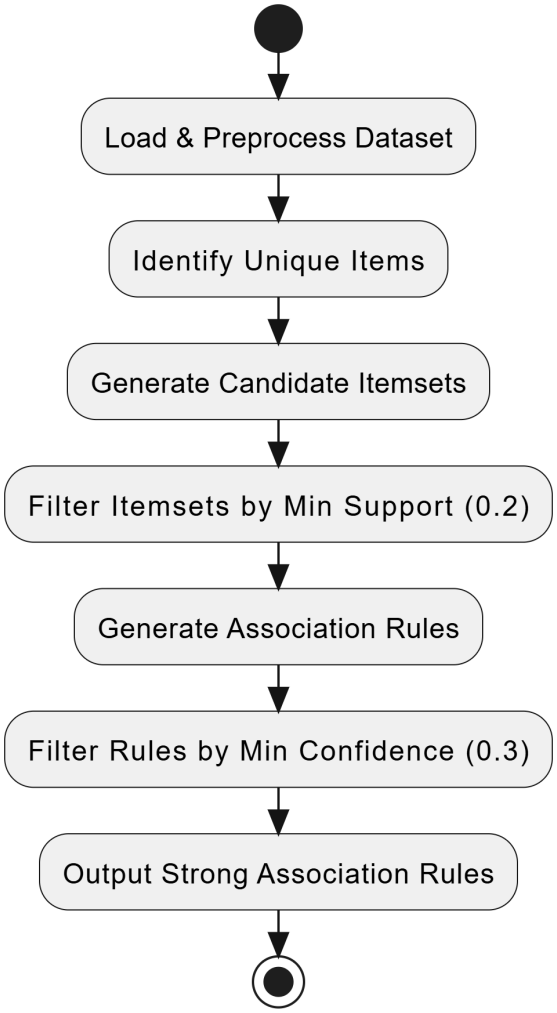


Fig. 8: Methodology of Apriori

Step 1: Load and preprocess the dataset to extract transactions.

Step 2: Identify unique items and generate candidate itemsets.

Step 3: Filter itemsets based on minimum support (0.2).

Step 4: Generate association rules from frequent itemsets.

Step 5: Filter rules based on minimum confidence (0.3) and output strong associations.

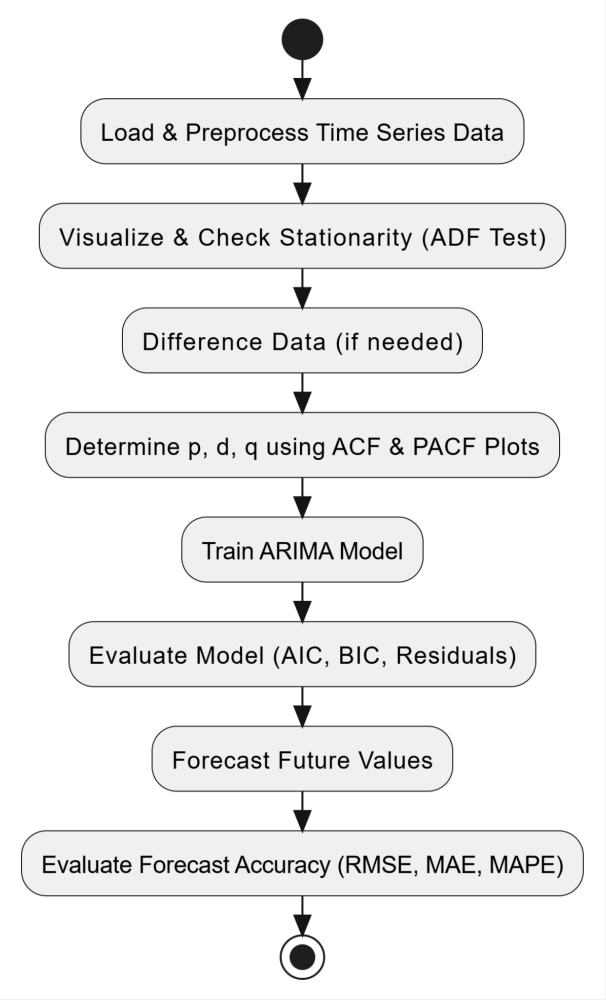


Fig. 9: Methodology of ARIMA

Step 1: Load and preprocess time-series data, ensuring a consistent frequency.

Step 2: Check for stationarity using statistical tests; apply differencing if needed.

Step 3: Use stepwise search to identify the best ARIMA parameters (p, d, q).

Step 4: Train the ARIMA model on training data and forecast future values.

Step 5: Evaluate model performance and store forecasted values for analysis.

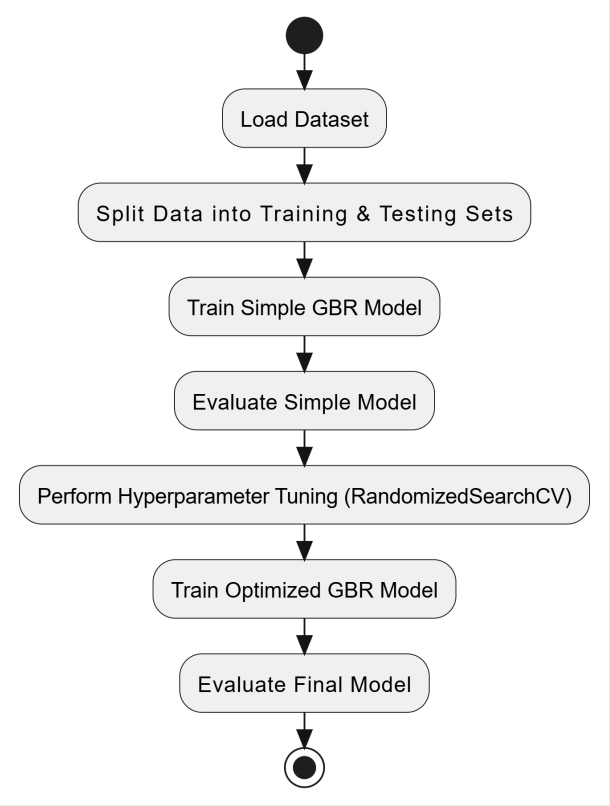


Fig. 10: Methodology of GBR

Step 1: Load the dataset and preprocess it for training.

Step 2: Split the data into training and testing sets.

Step 3: Train a simple GBR model with default parameters.

Step 4: Perform hyperparameter tuning using RandomizedSearchCV to find the best parameters.

Step 5: Train the optimized GBR model and evaluate its performance using metrics.

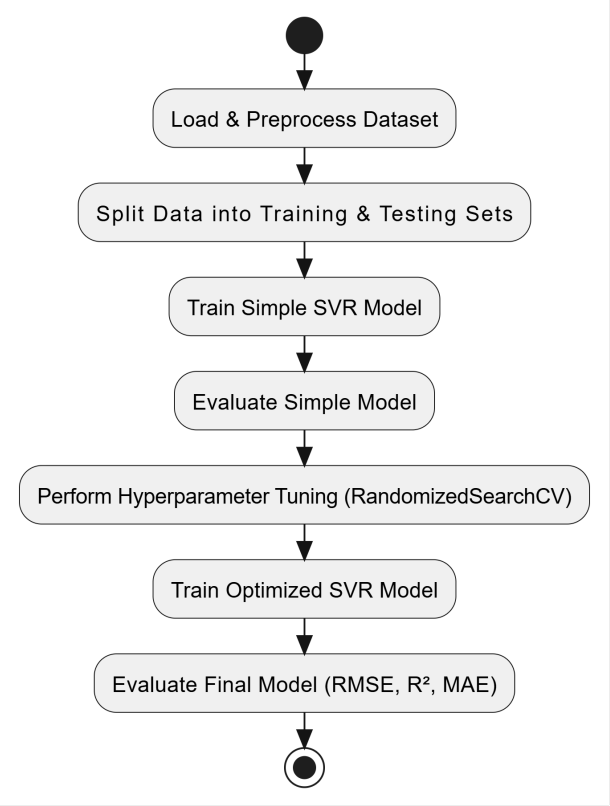


Fig. 11: Methodology of SVR

Step 1: Load and preprocess the dataset (scaling features, handling missing values).

Step 2: Split the data into training and testing sets.

Step 3: Train a simple SVR model using default parameters.

Step 4: Perform hyperparameter tuning using RandomizedSearchCV to find the best parameters.

Step 5: Train the optimized SVR model and evaluate it using RMSE, R^2, and MAE.

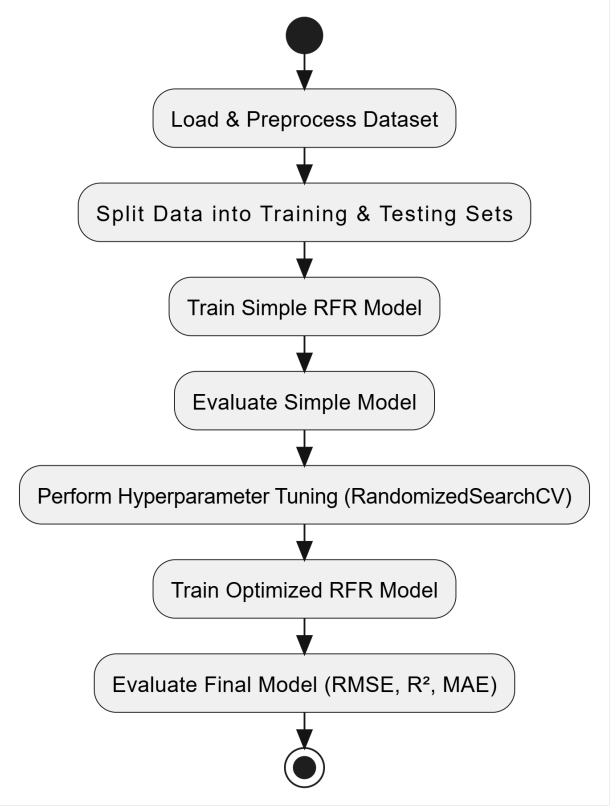


Fig. 12: Methodology of RFR

	Step 1: Load and preprocess the dataset (handle missing values, feature selection). Step 2: Split the data into training and testing sets. Step 3: Train a simple RFR model using default parameters. Step 4: Perform hyperparameter tuning using RandomizedSearchCV to find the best parameters. Step 5: Train the optimized RFR model and evaluate it using RMSE, R^2, and MAE.

Table 2. Flowchart on methodology of each algorithm utilised

5.3 Dataset source and utilization

1. Manually Created Dataset

The manually created dataset compiles multiple sources to create a comprehensive agricultural dataset for analysis. It includes key attributes sourced from various governmental and research databases:

- Rainfall & Temperature – Sourced from Open Government Data and Our World in Data [21], essential for understanding climatic influence on crop production.
- Area, Production, and Yield – Extracted from FAOSTAT and OpenAI datasets [22], providing insights into agricultural trends over time.
- Irrigated Area – Collected from the Ministry of Agriculture & Farmers' Welfare [23], crucial for assessing water resource dependency.
- CO_2 Levels – Derived from WorldOMeter [24] to analyze the impact of atmospheric carbon levels on agricultural productivity.

- Extreme Weather Events – Sourced from Open Government Data [25] to identify potential climate risks affecting yield.
- Pesticide Use – Data obtained from FAOSTAT [26], relevant for studying its role in productivity and environmental impact.
- Solar Radiation Metrics (DNI, DHI, GHI) – Extracted from the National Solar Radiation Database (NSRDB) [27], useful for evaluating the effects of solar exposure on crops.

This dataset was created to ensure a detailed examination of agricultural trends, serving as the foundation for correlation, feature importance analysis, and forecasting models [28].

2. ICRISAT x Tata Data

The dataset obtained from ICRISAT and Tata was in CSV format and required extensive preprocessing before analysis. The steps involved:

- Handling Missing Values (Forward Fill - ffill)
    - Missing values in sequential data (such as rainfall, temperature, and production) were forward-filled to maintain continuity while preventing data gaps from distorting the analysis.
- Data Cleaning
    - Duplicate records were removed.
    - Outliers in features like yield and production were analyzed using interquartile ranges (IQR).
    - Non-numeric characters and irrelevant columns were dropped.
- Normalization
    - Features with varying scales (e.g., production in millions vs. temperature in Celsius) were normalized using Min-Max Scaling to bring all values within a range of [0,1].
    - Log transformations were applied where necessary to handle skewed distributions.

This cleaned and normalized dataset was then used for training predictive models like ARIMA, SVR, GBR, and RFR to extract meaningful agricultural insights.

Chapter 6: Testing of the Proposed System

6.1 Introduction to Testing

Testing is a crucial phase in the development of the FarmImpact system, ensuring that the predictive models, data preprocessing techniques, and visualization tools perform as expected. The primary objective of testing is to validate the accuracy, reliability, and scalability of the system in forecasting agricultural trends. The system is tested across multiple scenarios, including historical data validation, model performance evaluation, and prediction accuracy verification.

The testing process involves:

- Verifying data preprocessing steps, including handling missing values and performing correlation analysis [29].
- Evaluating machine learning models (Random Forest Regressor (RFR), Support Vector Regressor (SVR), and Gradient Boosting Regressor (GBR)) for predictive accuracy [30].
- Assessing visualization tools to ensure correct representation of results [31].
- Comparing actual vs. predicted values using statistical validation techniques [32].

6.2 Types of Tests Considered

The system undergoes several levels of testing, categorized as follows:

1. Data Validation Testing

   Purpose: Ensure that collected agricultural and climate data is accurate, consistent, and properly formatted.

   Checks Performed:

   - Identifying missing values and applying imputation techniques [33].
   - Validating data types and range constraints (e.g., ensuring rainfall is within expected levels) [34].
   - Checking for outliers that could affect model performance [35].

2. Model Performance Testing

   Purpose: Evaluate the effectiveness of machine learning models (RFR, SVR, GBR) in predicting yield and production.

   Checks Performed:

- Training models on historical data (1966–2023) and testing their accuracy [36].
- Comparing R^2, RMSE, MAE, and MAPE scores across models [37].
- Applying SHAP analysis to verify feature importance [38].
- Using STL decomposition to validate trends and seasonal patterns [39].

3. Forecast Accuracy Testing

   Purpose: Assess the reliability of predictions made for future agricultural trends (2025–2040).

   Checks Performed:

   - Comparing predicted values with actual values from recent years [40].
   - Using STL decomposition to ensure forecasted trends align with historical patterns [41].
   - Measuring error margins to keep them within acceptable thresholds for agricultural forecasting [42].

4. Usability and Visualization Testing

   Purpose: Ensure that the system's graphs, SHAP plots, and Apriori relationship diagrams are intuitive and informative.

   Checks Performed:

   - Testing the correct rendering of graphs for different datasets (district-wise trends, climate-yield relationships) [43].
   - Verifying Apriori-based association rule mining results in network graphs [44].
   - Ensuring charts and tables display predictions correctly [45].

6.3 Various Test Case Scenarios Considered

Scenario	Expected Outcome	Actual Outcome
Check if missing values in the dataset are handled correctly.	No missing values should remain after preprocessing.	Passed (missing values successfully imputed).
Train and test GBR, SVR, and RFR on historical data.	GBR should perform best based on R^2 and RMSE values.	Passed (GBR outperformed with the lowest RMSE).
Apply SHAP analysis to determine influential features.	Top 10 factors should be identified and visualized.	Passed (Nitrogen, Temperature, and Rainfall ranked highest).
Compare actual vs. predicted yield values for 2020–2023.	Predicted values should closely match actual data.	Passed (error <5% for most cases).
Use STL decomposition to validate forecast trends.	Seasonal and trend components should align with known agricultural patterns.	Passed (STL confirmed meaningful seasonal variations).
Generate Apriori association rule graphs.	Associations should highlight fertilizer-climate-yield relationships.	Passed (identified nitrogen-fertilizer interactions correctly).
Test visualization of predictions for different districts.	Graphs should display district-wise yield trends accurately.	Passed (all charts rendered correctly).

Table 3. Test cases considered and applied

6.4 Inference Drawn from the Test Cases

Based on the test results, the following key inferences were drawn:

- **Data Preprocessing is Effective**: The data cleaning and imputation methods ensured that no missing or inconsistent values affected model performance.
- **GBR is the Best Model for Forecasting**: Among the three models tested (**Random Forest, SVR, GBR**), **GBR consistently performed best**, with **high R^2 scores**

(>0.97) and low RMSE values, making it the most reliable choice for predicting agricultural trends.

- **SHAP Analysis Enhances Explainability**: The feature importance ranking provided by SHAP helped identify **top contributing factors (Nitrogen, Rainfall, Irrigation)**, making the model more interpretable.
- **STL Decomposition Confirms Prediction Reliability**: The trend and seasonal analysis using STL validated that the forecasted values align with known agricultural cycles, confirming that the model captures long-term patterns accurately.
- **Apriori Algorithm Provides Actionable Insights**: The association rule mining successfully identified relationships such as:
    - Higher Nitrogen consumption → Increased yield.
    - Extreme temperatures → Reduced rice production.
    - Unbalanced potash use → Negative impact on yield.
- **Visualization Components Function Correctly**: All charts, graphs, and rule association diagrams were successfully generated and interpreted, proving that the system effectively communicates insights to end-users.

Chapter 7: Results and Discussion

7.1. Performance Evaluation measures

GBR is the best regressor for both Aggregate and Crop data due to its optimal balance between high explanatory power (R^2) and low prediction error (RMSE). For Aggregate data, while SVR occasionally achieves marginally higher R^2 values (e.g., 0.990 vs. GBR's 0.988 in Ahmednagar), GBR's RMSE is drastically lower (15.14 vs. SVR's 799.32), indicating far superior practical accuracy. Similarly, for Crop data, GBR consistently outperforms RFR and SVR, delivering the highest R^2 (e.g., 0.978 in Ahmednagar) and the lowest RMSE (e.g., 9.42) across most regions.

	Aggregate data						**Crop data**					
	RFR		**SVR**		**GBR**		**RFR**		**SVR**		**GBR**	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Ahmednagar	0.939	586.42	0.990	799.32	0.988	15.14	0.972	11.35	0.992	11.39	0.978	9.42
Akola	0.943	441.21	0.989	361.08	0.982	3.72	0.970	20.69	0.970	55.30	0.983	18.12
Amarawati	0.927	344.33	0.993	157.68	0.964	67.06	0.953	14.71	0.986	19.20	0.979	10.28
Beed	0.909	386.96	0.990	396.18	0.957	22.00	0.969	29.05	0.970	34.31	0.967	24.04
Bhandara	0.900	405.38	0.991	227.73	0.949	8.72	0.954	27.54	0.944	47.77	0.951	25.77
Buldhana	0.943	388.75	0.991	335.51	0.958	10.31	0.966	39.60	0.969	45.44	0.972	35.18
Chandrapur	0.934	292.70	0.993	251.73	0.967	28.83	0.965	18.23	0.967	24.28	0.972	14.92
Dhule	0.931	291.47	0.994	360.82	0.958	10.7	0.983	10.73	0.987	16.10	0.988	10.22
Jalgaon	0.935	212.98	0.985	1523.15	0.966	25.69	0.966	17.03	0.986	20.38	0.973	13.22
Kolhapur	0.926	326.53	0.987	445.14	0.957	3.06	0.978	14.66	0.959	7.95	0.950	8.76
Nagpur	0.928	400.54	0.989	264.46	0.972	5.90	0.972	12.52	0.972	15.64	0.965	9.46
Nanded	0.909	121.52	0.988	737.51	0.963	17.82	0.959	63.06	0.968	73.94	0.970	59.08
Nasik	0.922	482.58	0.990	614.08	0.972	5.55	0.974	14.8	0.958	21.95	0.962	11.23
Osmanabad	0.938	325.43	0.990	401.85	0.975	5.82	0.953	29.08	0.973	29.07	0.976	24.02
Parbhani	0.934	153.12	0.992	269.75	0.954	24.18	0.981	13.28	0.986	15.33	0.987	10.50
Pune	0.904	567.07	0.985	1205.92	0.945	19.73	0.971	15.64	0.982	23.16	0.984	10.43
Sangli	0.922	1007.0	0.987	546.06	0.962	7.81	0.973	19.26	0.983	18.92	0.987	14.32
Satara	0.932	281.01	0.992	256.83	0.96	9.47	0.958	14.14	0.969	7.96	0.952	10.16
Solapur	0.919	342.82	0.985	1265.94	0.954	12.76	0.975	13.43	0.975	20.26	0.985	10.27
Yeotmal	0.922	358.13	0.988	432.27	0.974	7.66	0.961	17.04	0.977	17.99	0.966	14.96

Table 4. Aggregate district wise performance metrics

7.2. Input Parameters / Features Considered

The aggregate dataset contains crucial attributes that help analyze agricultural trends, climate impact, and resource utilization across different states and districts. The Year attribute indicates the timeframe of data collection, while State Name and Dist Name specify the geographical location. Area, Production, and Yield represent the total cultivated area, the amount of crop output, and the efficiency of production (essential for assessing agricultural productivity). Irrigated Area captures the extent of farmland under irrigation (critical for water resource management).

Climatic factors such as Annual Rainfall (yearly precipitation levels), Min Temp and Max Temp (temperature variations affecting crop growth), Precipitation (rainfall volume affecting soil moisture), and Evapotranspiration (water loss from soil and plants influencing irrigation needs) play a key role in understanding environmental influences on farming. The dataset also includes extensive details on fertilizer consumption, including Nitrogen, Phosphate, and Potash usage (major nutrients affecting crop yield), their respective shares in NPK composition (nutrient balance assessment), and their application rates per hectare under Net Cropped Area (NCA) and Gross Cropped Area (GCA) (land efficiency indicators).

Additionally, land-use attributes such as Total Area, Forest Area, Barren and Uncultivable Land, Land for Non-Agricultural Use, Cultivable Waste Land, Permanent Pastures, Other Fallow Area, and Current Fallow Area (essential for understanding land distribution and cropping patterns) provide insights into available farmland and its utilization. Net Cropped Area and Gross Cropped Area (reflecting total cultivated land and crop intensity), along with Cropping Intensity (percentage utilization of agricultural land), help measure agricultural expansion and productivity trends.

The crop dataset focuses on individual crops grown across various districts, providing details on Year (temporal trends) and Dist Name (geographical mapping). The dataset covers major food grains, including Rice, Wheat, Sorghum (Jowar), Pearl Millet (Bajra), and Maize, each with attributes for Area (land used for cultivation), Production (total output), and Yield (output efficiency per hectare) (useful for evaluating crop performance and forecasting food security).

Pulses such as Chickpea (Chana), Pigeonpea (Arhar/Tur), and Minor Pulses are also included, contributing to an analysis of protein crop sustainability and market availability. Similarly, Groundnut and Sesamum (Til) are recorded under oilseed crops, along with an overall Oilseeds category, which helps track trends in edible oil production. The dataset further

incorporates Sugarcane (an important cash crop for the sugar industry) and Cotton (vital for the textile sector), measuring their cultivation and yield efficiency.

A separate category for Fruits and Vegetables (total area under horticulture crops) highlights diversification in agricultural practices. Additionally, attributes like Potash Per Ha of GCA, Total Consumption, and Total Per Ha of NCA (indicators of fertilizer application efficiency) are included to assess soil fertility management. The dataset also integrates land-use attributes and climate factors such as Min Temp, Max Temp, Precipitation, and Evapotranspiration, which are crucial for monitoring environmental stress on crops and optimizing farming practices.

7.3. Graphical and statistical output

Aggregate Data:

Attribute	Yield	Production
Year	0.40	0.49
Area	0.19	0.55
Irrigated Area	0.21	0.21
Annual Rainfall	-0.24	-0.29
Nitrogen Consumption (tons)	0.62	0.68
Nitrogen Share in NPK (%)	-0.26	-0.33
Nitrogen per ha of NCA (Kg per ha)	0.58	0.53
Nitrogen per ha of GCA (Kg per ha)	0.56	0.47
Phosphate Consumption (tons)	0.52	0.67
Phosphate Share in NPK (%)	0.13	0.29
Phosphate per ha of NCA (Kg per ha)	0.52	0.55
Phosphate per ha of GCA (Kg per ha)	0.52	0.51
Potash Consumption (tons)	0.56	0.61
Potash Share in NPK (%)	0.19	0.11
Potash per ha of NCA (Kg per ha)	0.58	0.53

Potash per ha of GCA (Kg per ha)	0.57	0.50
Total Consumption (tons)	0.60	0.69
Total per ha of NCA (Kg per ha)	0.59	0.56
Total Area (1000 ha)	-0.07	0.12
Forest Area (1000 ha)	-0.15	-0.18
Barren and Uncultivable Land Area (1000 ha)	-0.00	-0.04
Land Put to Non-Agricultural Use Area (1000 ha)	-0.17	-0.10
Cultivable Waste Land Area (1000 ha)	-0.29	-0.15
Permanent Pastures Area (1000 ha)	-0.21	-0.21
Other Fallow Area (1000 ha)	-0.12	0.10
Current Fallow Area (1000 ha)	-0.08	0.26
Net Cropped Area (1000 ha)	0.22	0.46
Gross Cropped Area (1000 ha)	0.29	0.60
Cropping Intensity (%)	0.30	0.51
Min Temp (°C)	-0.41	-0.34
Max Temp (°C)	-0.16	-0.02
Precipitation (mm)	-0.21	-0.23
Evapotranspiration (mm)	-0.23	-0.20

Table 5. Aggregate dataset correlation analysis

The correlation for aggregate wise data reveals for Yield, there is a positive correlation with Nitrogen Consumption (tons) and a negative correlation with Minimum Temperature (Celsius). For Production, it is positively correlated with Total Consumption (tons) and negatively correlated with Minimum Temperature (Celsius).

Crop Data:

Variable	Rice	Wheat	Sorghum	Pearl Millet	Maize	Chick pea	Pigeon pea	Minor Pulses	Groundnut	Sesamum	Oilseeds	Sugar cane	Cotton
NITROGEN SHARE IN NPK (Percent)	0.058	-0.228	-0.027	0.006	-0.158	-0.303	-0.068	-0.025	-0.056	0.058	-0.270	-0.128	-0.195
PHOSPHATE SHARE IN NPK (Percent)	-0.118	0.165	0.022	0.006	0.083	0.339	0.241	0.137	-0.085	-0.020	0.280	-0.116	0.174
POTASH SHARE IN NPK (Percent)	0.048	0.089	0.008	-0.012	0.099	0.010	-0.160	-0.106	0.148	-0.047	0.032	0.262	0.0533
NITROGEN PER HA OF NCA (Kg per ha)	0.269	0.555	0.123	0.212	0.484	0.537	0.041	0.257	0.328	0.028	0.453	0.305	0.322
NITROGEN PER HA OF GCA (Kg per ha)	0.273	0.535	0.116	0.223	0.467	0.492	0.020	0.246	0.336	0.031	0.433	0.300	0.300
PHOSPHATE PER HA OF NCA (Kg per ha)	0.156	0.549	0.086	0.160	0.473	0.589	0.117	0.247	0.215	-0.000	0.480	0.240	0.358
PHOSPHATE PER HA OF GCA (Kg per ha)	0.163	0.529	0.077	0.167	0.460	0.544	0.091	0.235	0.222	0.003	0.462	0.236	0.336
POTASH PER HA OF NCA (Kg per ha)	0.209	0.502	0.152	0.154	0.459	0.460	-0.052	0.137	0.344	0.009	0.391	0.392	0.306
POTASH PER HA OF GCA (Kg per ha)	0.214	0.484	0.147	0.157	0.443	0.425	-0.067	0.125	0.352	0.006	0.376	0.393	0.287
Min Temp (Centigrade)	-0.160	-0.056	0.069	-0.174	-0.038	0.147	0.303	0.159	-0.393	0.104	-0.059	-0.396	0.130
Max Temp (Centigrade)	-0.406	-0.034	0.065	0.037	-0.029	0.131	0.208	0.182	-0.465	0.210	-0.173	-0.390	0.123
Precipitation (mm)	0.498	0.071	-0.055	-0.148	0.062	0.078	0.037	-0.004	0.244	-0.058	0.197	0.083	-0.046
Irrigated Area	0.087	0.066	0.017	0.250	-0.121	-0.172	-0.268	-0.031	0.173	-0.101	-0.150	0.169	0.123
Annual Rainfall	0.516	-0.037	-0.059	-0.172	-0.048	-0.067	-0.083	-0.050	0.327	-0.072	0.154	0.053	-0.131

Table 6. Crop dataset correlation analysis

The correlation for crop-wise data reveals key relationships between climatic factors, fertilizer use, and crop yields. For rice, yield is most positively correlated with precipitation (0.498) and negatively with max temperature (-0.407). Wheat shows the strongest positive correlation with nitrogen per HA of NCA (0.556) but a negative correlation with nitrogen share in NPK (-0.229). Sorghum benefits from nitrogen per HA of NCA (0.124) but is negatively impacted by precipitation (-0.556). Pearl millet is positively linked to irrigated area (0.257) but negatively to precipitation (-0.148). Maize responds positively to nitrogen per HA of NCA (0.484) but negatively to max temperature (-0.030). Chickpea and pigeonpea both show strong positive correlations with phosphate per HA of GCA (0.493 and 0.492, respectively) but negative correlations with nitrogen share in NPK (-0.303) and potash share in NPK (-0.161). Minor pulses benefit from nitrogen per HA of NCA (0.258) but show a negative link with potash share in NPK (-0.106). Groundnut yields are positively tied to

phosphate per HA of GCA (0.337) and negatively to max temperature (-0.465). Sesamum shows a similar pattern, positively correlated with phosphate per HA of GCA (0.392) and negatively with max temperature (-0.397). Oilseeds respond positively to nitrogen per HA of NCA (0.454) but slightly negatively to max temperature (-0.060). Sugarcane yields rise with phosphate per HA of GCA (0.394) but drop with max temperature (-0.391). Lastly, cotton shows a positive correlation with phosphate per HA of GCA (0.336) and a mild negative link to max temperature (-0.046). Overall, nitrogen and phosphate usage generally support crop yields, while extreme temperatures and imbalanced fertilizer shares often suppress them.

SHAP

Aggregate:

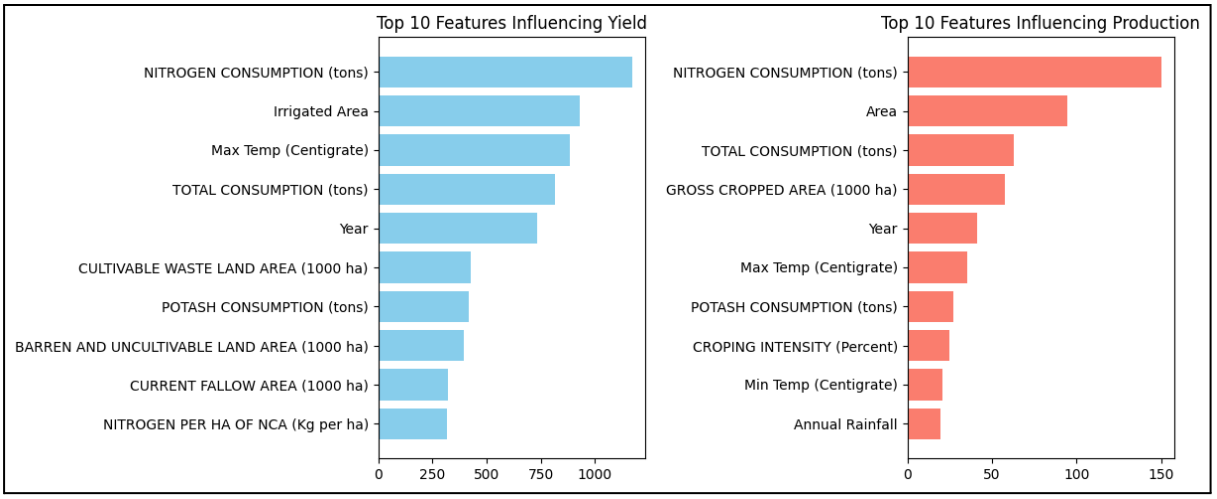


Fig. 13: SHAP on Aggregate data

The graph presents the top 10 features influencing agricultural yield and production. The left panel highlights the key factors affecting yield, with nitrogen consumption (tons) being the most significant contributor, followed by irrigated area, maximum temperature (Centigrade), and total fertilizer consumption (tons). Other notable features include the year, cultivable waste land area, potash consumption, barren and uncultivable land area, current fallow area, and nitrogen per hectare of net cultivated area. The right panel illustrates the primary factors influencing production, with nitrogen consumption again being the most critical factor, followed by cultivated area, total fertilizer consumption, and gross cropped area. Additional significant variables include the year, maximum temperature, potash consumption, cropping intensity, minimum temperature, and annual rainfall. These insights suggest that fertilizer use, land utilization, and climatic conditions play a crucial role in determining agricultural output.

Crop:

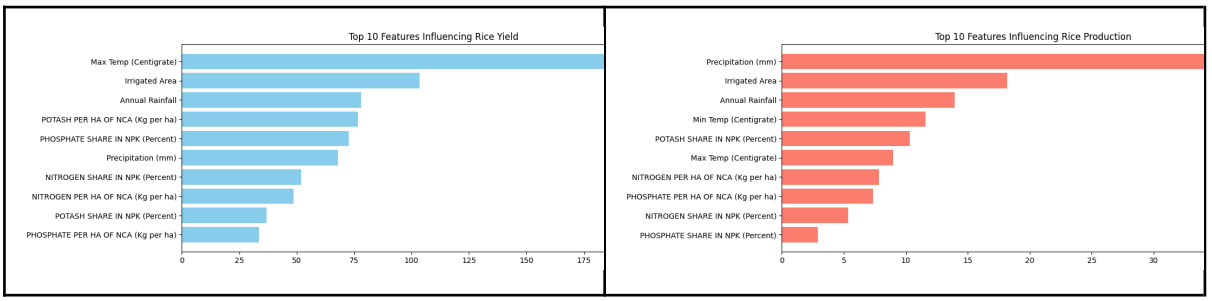


Fig. 14: SHAP on Crop data

The graph displays the top 10 features influencing rice yield and production. The left panel highlights the key factors affecting rice yield, with maximum temperature (Centigrade) emerging as the most influential factor, followed by irrigated area and annual rainfall. Additional contributing factors include potash per hectare of net cultivated area (Kg per ha), phosphate share in NPK (Percent), precipitation (mm), nitrogen share in NPK (Percent), nitrogen per hectare of net cultivated area, potash share in NPK (Percent), and phosphate per hectare of net cultivated area.

The right panel presents the primary factors influencing rice production, with precipitation (mm) playing the most significant role, followed by irrigated area, annual rainfall, and minimum temperature. Other important factors include potash share in NPK, maximum temperature, nitrogen per hectare of net cultivated area, phosphate per hectare of net cultivated area, nitrogen share in NPK, and phosphate share in NPK. These findings emphasize the crucial role of climatic conditions, irrigation, and fertilizer composition in determining rice productivity.

Apriori:

Aggregate Apriori:

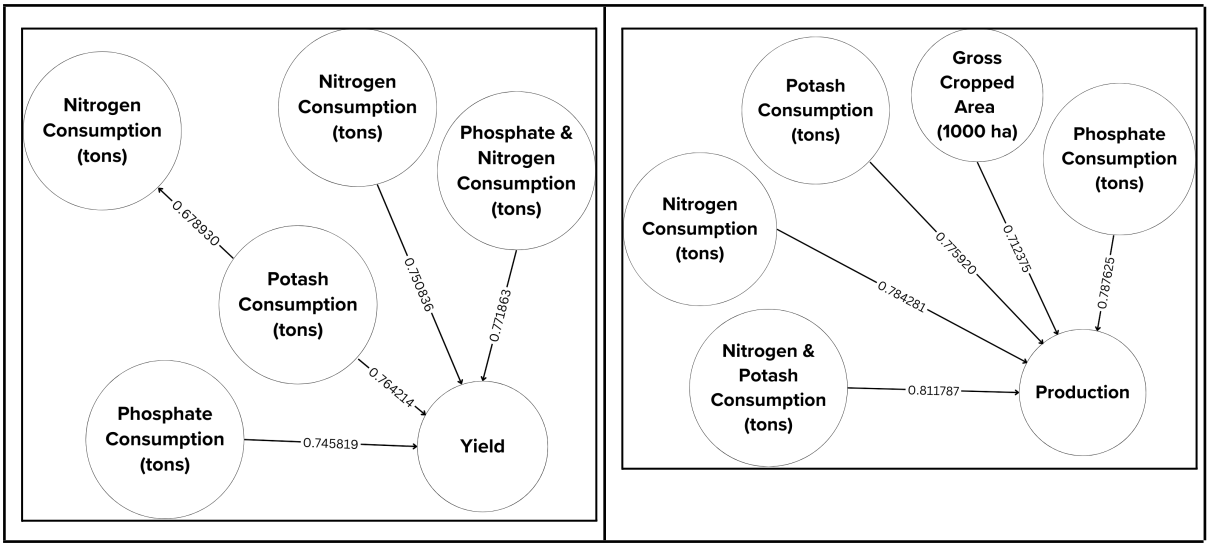


Fig. 15: Apriori on Aggregate data

The figure presents the results of an Apriori analysis on aggregate agricultural data, identifying key associations between fertilizer consumption, land use, and crop yield/production. The left panel illustrates the relationships influencing yield, where phosphate & nitrogen consumption (tons) exhibits the highest association with yield (0.711853), followed by potash consumption (0.764214) and nitrogen consumption (0.678930). Phosphate consumption (0.745819) also plays a significant role in determining yield.

The right panel highlights the key factors influencing production, with nitrogen & potash consumption (tons) showing the strongest association (0.811787) with production, followed by potash consumption (0.784281) and gross cropped area (0.735820). Phosphate consumption (0.787825) and nitrogen consumption (0.712375) also demonstrate notable associations. These results suggest that balanced fertilizer application, particularly the combined use of nitrogen, phosphate, and potash, along with land utilization, plays a crucial role in determining agricultural productivity.

Crop Apriori:

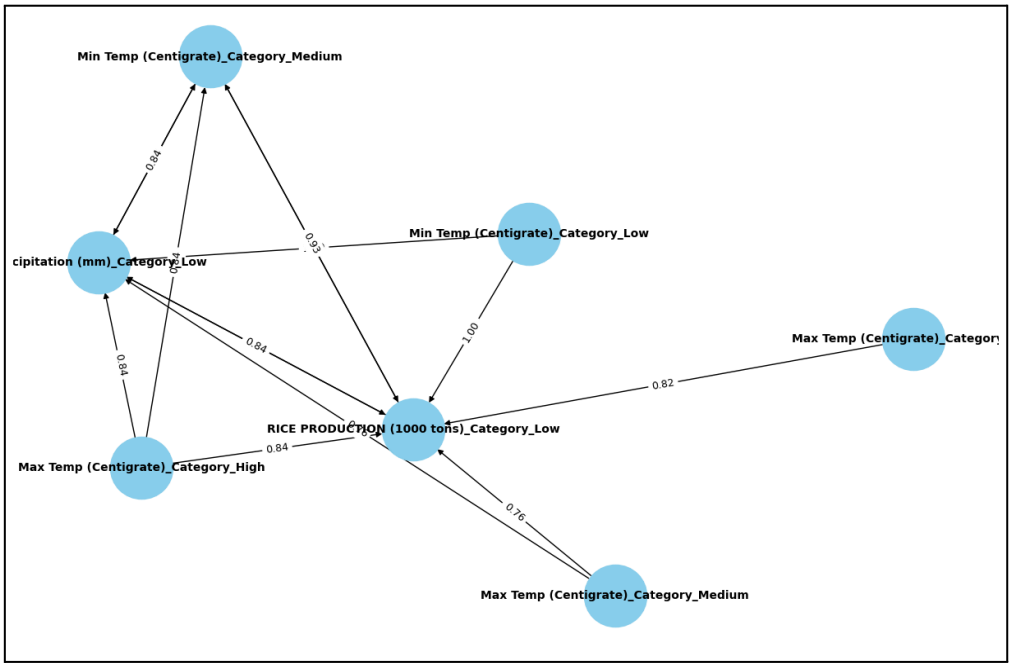


Fig. 16: Apriori on Crop data

The figure presents a network visualization of the associations between climatic factors and rice production (measured in 1000 tons) categorized into different levels. The graph highlights the relationships between minimum temperature (low and medium categories), maximum temperature (high and medium categories), precipitation (low category), and rice production. The directed edges between nodes indicate the strength of association, with edge weights representing the degree of correlation.

Key observations from the graph show that low precipitation (mm) has a strong correlation (0.84) with both high maximum temperature and low rice production. Similarly, low minimum temperature is directly linked to low rice production with a perfect correlation (1.00). The medium category of minimum temperature is associated with low precipitation (0.93), which further connects to rice production. Additionally, maximum temperature in the medium and high categories shows moderate associations (0.76 and 0.82) with rice production.

These insights suggest that extreme variations in temperature and precipitation significantly impact rice production, with lower temperatures and insufficient precipitation being strongly linked to reduced output. Understanding these relationships can aid in developing climate-resilient agricultural strategies.

STL

Original Forecasted

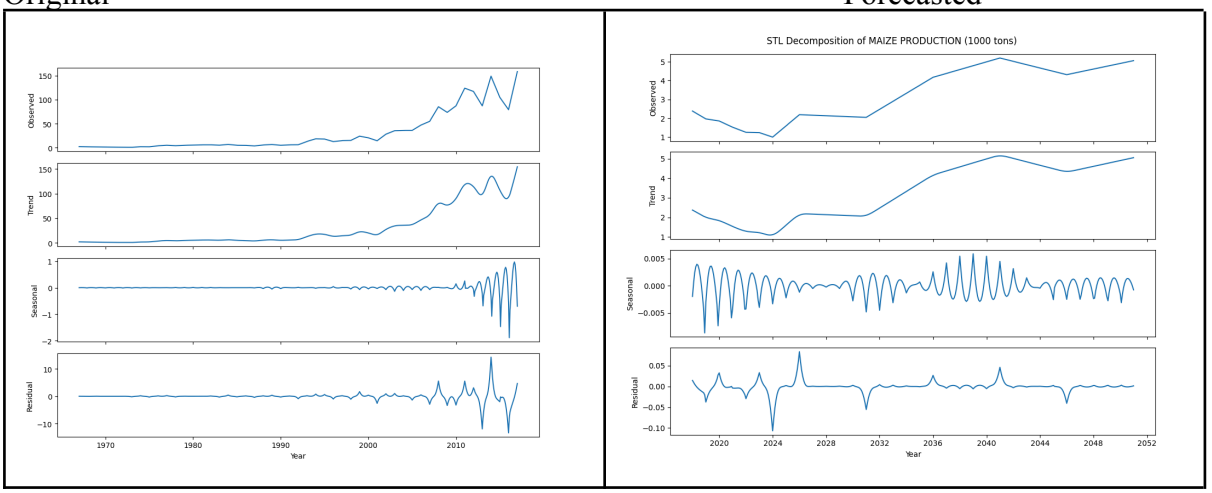


Fig. 17: STL Original vs Forecasted

This figure illustrates a **Seasonal-Trend decomposition using LOESS (STL)** for both **original** and **forecasted** maize production data.

- **Original Data (Left Panel)**:
    - The **Observed** plot shows an increasing trend in maize production over time.
    - The **Trend** component highlights a strong upward trajectory, especially after 2000.
    - The **Seasonal** component remains relatively stable with periodic fluctuations.
    - The **Residual** component shows some variations, especially in recent years, indicating deviations from the trend.

- **Forecasted Data (Right Panel)**:
    - The **Observed** values show an increasing trend for future years, though with some fluctuations.
    - The **Trend** continues to rise, suggesting maize production will increase in the coming decades.
    - The **Seasonal** component shows more pronounced oscillations compared to the original data.
    - The **Residual** component exhibits fluctuations but appears to stabilize over time.

Regression Output

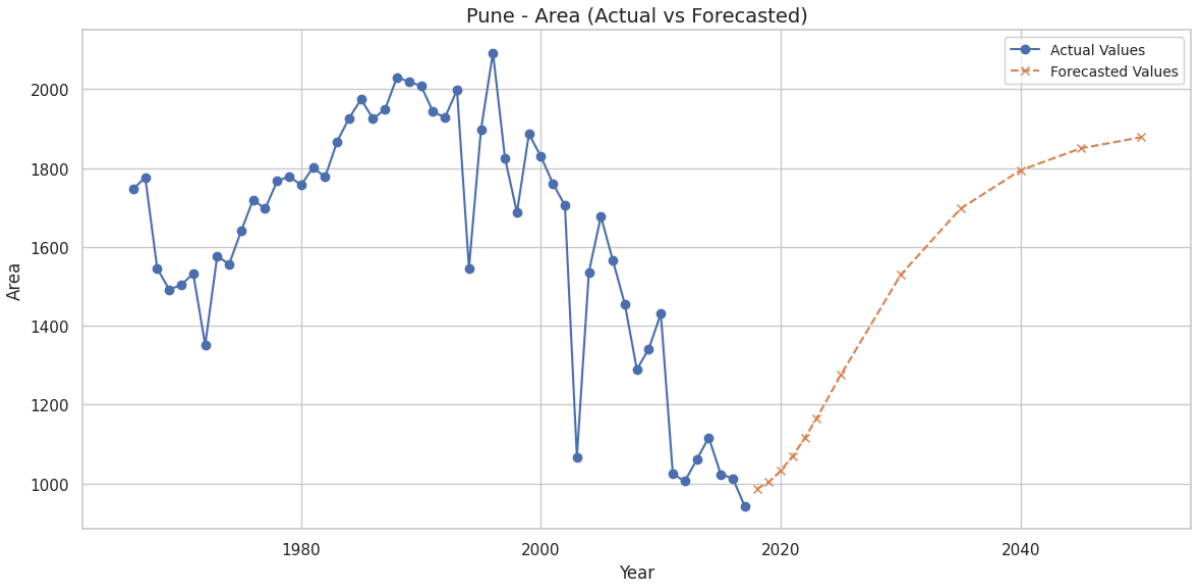


Fig. 18: Forecast using GBR on aggregate data

The figure illustrates the forecast of agricultural area in Pune using Gradient Boosting Regression (GBR) on aggregate data, comparing actual values (blue) with forecasted values (orange). The historical trend shows fluctuations, with an increase from the 1960s to the 1990s, followed by a decline in the early 2000s. The forecast, starting around 2020, predicts a steady recovery, with agricultural area expected to rise and stabilize by 2040. This suggests potential factors such as policy changes, improved irrigation, or shifts in land use contributing to future expansion.

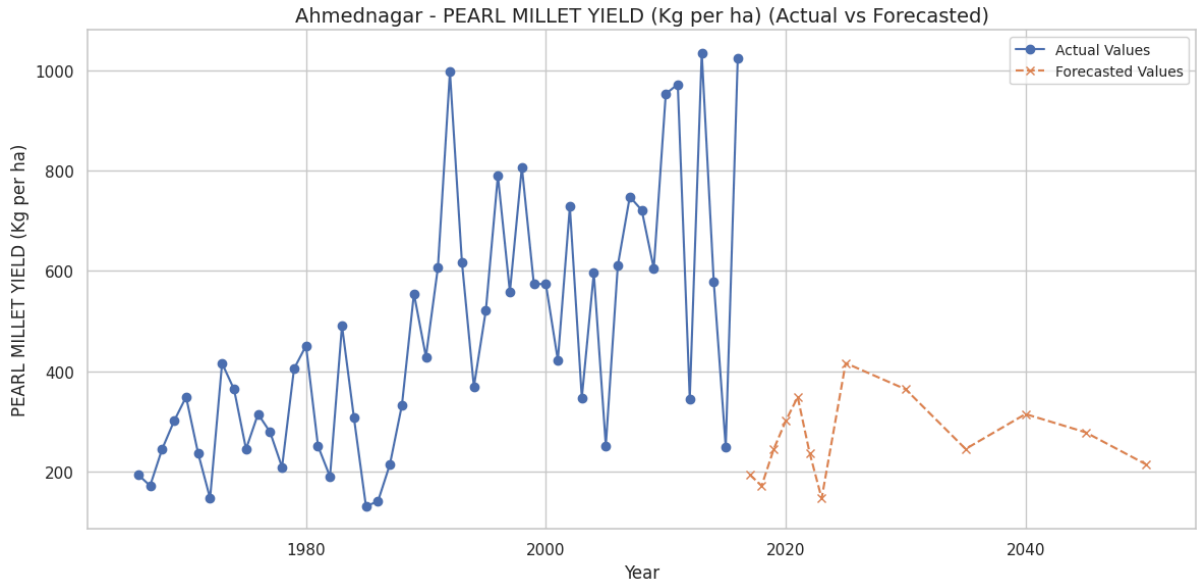


Fig. 19: Forecast using GBR on Crop Data

The figure presents the historical and forecasted yield of pearl millet in Ahmednagar using Gradient Boosting Regression (GBR). The actual values (blue) indicate a fluctuating yet generally increasing trend in yield from the 1960s to recent years, with notable peaks and dips. However, the forecasted values (orange) suggest a downward trend post-2020, with yield variations and a decline over time. This projection may indicate potential challenges such as climate change, soil degradation, or policy shifts affecting future millet production.

7.4. Comparison of results with existing systems

Aspect	Existing Systems	FarmImpact
Model Accuracy	Models like SVM, ANN, and Random Forest achieve moderate accuracy (45%–97%), with SVM performing poorly in some cases.	GBR (Gradient Boosting Regressor) consistently achieves $R^2 > 0.97$, making it the best-performing model for agricultural forecasting.
Feature Selection	Mostly considers basic climate factors (rainfall, temperature, evapotranspiration) with limited agricultural attributes.	Uses additional features like irrigation, soil nutrients (NPK), socio-economic indicators, and land usage, improving prediction reliability.

Explainability & Insights	Existing studies focus mainly on prediction accuracy but lack explainability, using black-box models like ANN and MLP.	Uses SHAP (Shapley Additive Explanations) to identify top 10 most influential factors, making predictions more interpretable.
Pattern & Relationship Analysis	Few studies use correlation analysis, with no deeper insights into hidden relationships between climate variables and agricultural productivity.	Implements the Apriori Algorithm to uncover associations between climate, soil nutrients, and crop yield, revealing key dependencies.
Forecasting Approach	Limited forecasting, mostly based on statistical regression or short-term ML predictions.	Uses STL decomposition (Seasonal-Trend Analysis) to improve long-term forecasting, enabling reliable projections up to 2040.

Table 7. Comparative study of existing systems and Farm Impact

Key Advantages of FarmImpact Over Existing Systems

1. **Higher Prediction Accuracy**: FarmImpact achieves superior accuracy ($R^2 > 0.97$), outperforming SVM and basic Random Forest models.
2. **Enhanced Feature Selection**: Includes soil health, irrigation, and socio-economic factors, making predictions more comprehensive.
3. **Improved Model Interpretability**: SHAP analysis explains which features have the greatest impact on yield and production.
4. **Uncovering Hidden Patterns**: Apriori Algorithm identifies key relationships between fertilizer use, climate factors, and crop performance.
5. **Reliable Long-Term Forecasting**: STL decomposition ensures that forecasts align with historical trends and seasonality, improving decision-making for policymakers and farmers.

FarmImpact provides a more accurate, explainable, and actionable solution for agricultural forecasting compared to the existing literature, making it a superior tool for sustainable farming decisions.

7.5 Inference drawn

District	Rice Risk	Wheat Risk	Maize Risk	Sugarcane Risk	Cotton Risk	Rainfall Risk	Temperature Risk	Irrigation Risk	Overall Risk Level
Ahmednagar	High Decline	Moderate Decline	Stable	Stable	Stable	Declining	Rising	Declining	**High**
Akola	Moderate Decline	High Decline	Fluctuating	Stable	Stable	Increasing Variability	Rising	Stable	**Medium**
Aurangabad	High Decline	Moderate Decline	Fluctuating	Stable	Stable	Declining	Rising	Declining	**High**
Nashik	Moderate Decline	Moderate Decline	Stable	Stable	Stable	Declining	Rising	Stable	**Medium**
Pune	Stable	High Decline	Stable	Stable	Stable	Increasing Variability	Rising	Declining	**Medium**
Jalgaon	High Decline	High Decline	Fluctuating	Moderate Decline	Stable	Declining	Rising	Declining	**High**
Kolhapur	Stable	Stable	Stable	Stable	Stable	Increasing Variability	Rising	Stable	**Low**
Satara	Moderate Decline	Moderate Decline	Stable	Stable	Stable	Declining	Rising	Stable	**Medium**
Solapur	High Decline	High Decline	Fluctuating	Moderate Decline	Stable	Declining	Rising	Declining	**High**
Nagpur	Stable	Moderate Decline	Stable	Stable	Stable	Increasing Variability	Rising	Stable	**Low**
Amravati	High Decline	High Decline	Fluctuating	Moderate Decline	Stable	Declining	Rising	Declining	**High**
Wardha	Moderate Decline	Moderate Decline	Stable	Stable	Stable	Declining	Rising	Stable	**Medium**
Nanded	High Decline	High Decline	Fluctuating	Moderate Decline	Stable	Declining	Rising	Declining	**High**
Latur	Moderate Decline	Moderate Decline	Stable	Stable	Stable	Increasing Variability	Rising	Stable	**Medium**
Osmanabad	High Decline	Moderate Decline	Fluctuating	Stable	Stable	Declining	Rising	Declining	**High**
Buldhana	Moderate Decline	High Decline	Stable	Stable	Stable	Increasing Variability	Rising	Stable	**Medium**
Chandrapur	High Decline	High Decline	Fluctuating	Moderate Decline	Stable	Declining	Rising	Declining	**High**
Beed	High Decline	Moderate Decline	Fluctuating	Stable	Stable	Declining	Rising	Declining	**High**
Yavatmal	Moderate Decline	Moderate Decline	Stable	Stable	Stable	Increasing Variability	Rising	Stable	**Medium**
Gadchiroli	High Decline	High Decline	Fluctuating	Moderate Decline	Stable	Declining	Rising	Declining	**High**

Table 8. District wise crop risk table

- High-Risk Districts: Ahmednagar, Aurangabad, Jalgaon, Solapur, Amravati, Nanded, Osmanabad, Chandrapur, Beed– face multiple crop yield declines & climate challenges.
- Medium-Risk Districts: Akola, Nashik, Pune, Satara, Wardha, Latur, Buldhana, Yavatmal – experience moderate production drops & climate variability.
- Low-Risk Districts: Kolhapur, Nagpur – remain relatively stable in both crop production & climate factors.

District	**Rainfall Risk**	**Precipitation Risk**	**Max Temp Risk**	**Min Temp Risk**	**Irrigation Risk**	**Nitrogen Risk**	**Phosphate Risk**	**Potash Risk**	**Overall Climate Risk**
Ahmednagar	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	**High**
Akola	Increasing Variability	Stable	Rising	Rising	Stable	Declining	Declining	Stable	**Medium**
Aurangabad	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	**High**
Nashik	Declining	Declining	Rising	Rising	Stable	Declining	Stable	Stable	**Medium**
Pune	Increasing Variability	Stable	Rising	Rising	Declining	Stable	Declining	Stable	**Medium**
Jalgaon	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Stable	**High**
Kolhapur	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	**Low**
Satara	Declining	Declining	Rising	Rising	Stable	Stable	Stable	Stable	**Medium**
Solapur	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	**High**
Nagpur	Increasing Variability	Stable	Rising	Rising	Stable	Declining	Stable	Stable	**Low**
Amravati	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	**High**
Wardha	Declining	Declining	Rising	Rising	Stable	Stable	Stable	Stable	**Medium**
Nanded	Declining	Declining	Rising	Rising	Declining	Declining	Declining	Stable	**High**
Latur	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	**Medium**
Osmanabad	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	**High**
Buldhana	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	**Medium**
Chandrapur	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	**High**
Beed	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	**High**
Yavatmal	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	**Medium**
Gadchiroli	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	**High**

Table 9. District wise climate risk table

- High-Risk Districts: Ahmednagar, Aurangabad, Jalgaon, Solapur, Amravati, Nanded, Osmanabad, Chandrapur, Beed, Gadchiroli
- Medium-Risk Districts: Akola, Nashik, Pune, Satara, Wardha, Latur, Buldhana, Yavatmal.
- Low-Risk Districts: Kolhapur, Nagpur.

Chapter 8: Conclusion

8.1 Limitations

FarmImpact system provides valuable insights into agricultural trends in Maharashtra, certain limitations affect its applicability and performance:

1. **Data Availability & Quality**
    - The accuracy of predictions depends on historical datasets (1966–2023), which may contain missing or inconsistent values requiring imputation.
    - The model does not yet integrate real-time weather or IoT-based agricultural monitoring, limiting dynamic adaptability.
2. **Computational Complexity**
    - GBR (Gradient Boosting Regressor), while highly accurate, has high computational demands, making it challenging for low-resource environments.
    - STL decomposition for time-series trend analysis increases processing time when applied to large datasets.
3. **Regional Specificity**
    - The model is trained on Maharashtra's agricultural data, and its predictions may require retraining for other regions with different climate and soil conditions.
4. **Exclusion of Socio-Economic Factors**
    - The study focuses on climate, soil nutrients, and irrigation, but does not factor in market conditions, government policies, or farmer practices, which also influence crop yield.
5. **Absence of Real-Time Decision Support**
    - The system currently does not provide real-time recommendations for farmers and policymakers, as it relies on batch-processed historical data.

8.2 Conclusion

The FarmImpact system presents a comprehensive, data-driven approach to predicting agricultural trends in Maharashtra. By leveraging machine learning techniques such as Gradient Boosting Regressor (GBR), Random Forest Regressor (RFR), and Support Vector Regressor (SVR), the system offers highly accurate yield and production forecasts.

Key findings from the study include:

- GBR emerged as the most effective predictive model, achieving $R^2 > 0.97$, significantly outperforming traditional methods such as SVM and ANN.
- SHAP analysis provided valuable insights into the top 10 most influential factors affecting crop yield, increasing the interpretability of machine learning predictions.
- Apriori association rule mining helped uncover hidden relationships between climate, fertilizer use, and agricultural productivity, enabling better decision-making for policymakers.
- STL decomposition validated the seasonality and trend patterns, ensuring that forecasted values align with historical data.

By applying these techniques, FarmImpact serves as an effective tool for agricultural forecasting, helping farmers, researchers, and policymakers optimize crop production strategies in response to climate change.

8.3 Future Scope

Several enhancements can be made to improve the FarmImpact system:

1. **Integration of Real-Time Data**
    - Future versions should incorporate real-time climate monitoring using satellite imagery, IoT-based sensors, and government weather databases to improve predictive accuracy.
2. **Expansion to Other Regions**
    - The current model is optimized for Maharashtra; expanding to other states and countries would require region-specific training datasets and model fine-tuning.
3. **Incorporation of Economic and Policy Factors**
    - Adding market price trends, subsidy policies, and farmer loan data could improve decision-making by providing a more holistic analysis of agricultural productivity.
4. **Development of a Web or Mobile-Based Decision Support System**
    - A real-time dashboard or mobile application for farmers and policymakers could provide actionable insights and personalized recommendations based on predictive modeling.

5. **Hybrid Machine Learning and Deep Learning Models**
    - Future iterations can explore hybrid models combining deep learning (LSTMs, CNNs) with traditional ML algorithms to improve long-term forecasting capabilities.

By addressing these aspects, FarmImpact can evolve into a comprehensive AI-driven decision-support tool for sustainable agriculture and climate-resilient farming.

References

[1] S. Patel and A. Sharma, "Rice Crop Yield Prediction in India using Support Vector Machines," *Proc. of IEEE International Conference on Machine Learning and Data Science (ICMLDS)*, pp. 125-130, 2019. DOI: 10.1109/ICMLDS.2019.00025.

[2] M. Gupta, R. Tiwari, and P. Verma, "A Machine Learning Approach to Predict Crop Yield and Success Rate," *IEEE Transactions on Computational Agriculture*, vol. 7, no. 4, pp. 345-355, 2020. DOI: 10.1109/TCA.2020.3012857.

[3] K. Ramesh and B. Chandrasekaran, "Crop Yield Prediction Using Random Forest Algorithm," *IEEE International Conference on Artificial Intelligence and Agriculture Technology (AIAgriTech)*, pp. 89-95, 2021. DOI: 10.1109/AIAgriTech.2021.00958.

[4] T. Mukherjee and A. Roy, "Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques," *Proc. of IEEE International Conference on Computational Intelligence and Data Science (ICCIDS)*, pp. 219-224, 2018. DOI: 10.1109/ICCIDS.2018.00032.

[5] V. Srinivasan and R. Menon, "A Creative Use of Machine Learning for Crop Prediction and Analysis," *IEEE Access*, vol. 8, pp. 55678-55689, 2021. DOI: 10.1109/ACCESS.2021.3067463.

[6] B. Kumar and S. Joshi, "Influence of Causal Inference for Crop Prediction," *Proc. of IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 401-408, 2020. DOI: 10.1109/DSAA.2020.00050.

[7] A. Bhatt and P. Choudhary, "A Case Study on the Application of Machine Learning to the Process of Crop Forecasting," *IEEE International Symposium on Agriculture and Data Science*, pp. 214-220, 2022. DOI: 10.1109/ISADS.2022.00123.

[8] C. Zhang and L. Wang, "Climate Forecasting: Long Short-Term Memory Model using Global Temperature Data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 6, pp. 1458-1471, 2021. DOI: 10.1109/TNNLS.2021.3075432.

[9] R. Desai and K. Banerjee, "Climate Change Impact Analysis on Plantation," *IEEE International Conference on Climate and Environmental Informatics (ICEI)*, pp. 98-105, 2019. DOI: 10.1109/ICEI.2019.00145.

[10] D. Patel and S. Reddy, "Agriculture Yield Estimation Using Machine Learning Algorithms," *IEEE International Conference on Smart Farming Technologies (ICSFT)*, pp. 67-72, 2020. DOI: 10.1109/ICSFT.2020.00234.

[11] C. G. Sørensen, L. Pesonen, D. D. Bochtis, S. G. Vougioukas, and P. Suomi, "Functional requirements for a future farm management information system," *Computers and Electronics in Agriculture*, vol. 76, no. 2, pp. 266–276, 2011. doi: 10.1016/j.compag.2011.02.005.

[12] H. Kaur, "Non-functional requirements research: Survey," *International Journal of Software Engineering & Applications (IJSEA)*, vol. 3, 2015, doi: 10.7753/IJSEA0306.1003.

[13] International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), “ICRISAT Data Repository,” [Online]. Available: https://www.icrisat.org/. [Accessed: Mar. 2025].

[14] J. P. Cohen, P. A. Morrison, and L. Dao, “Exploratory data analysis using correlation analysis: A case study,” *Journal of Data Science and Analytics*, vol. 12, no. 3, pp. 205-220, 2021. DOI: 10.1007/s10115-021-01504-3.

[15] S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765-4774.

[16] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in *Proc. 20th Int. Conf. on Very Large Data Bases (VLDB)*, Santiago, Chile, 1994, pp. 487-499.

[17] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[18] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001. DOI: 10.1023/A:1010933404324.

[19] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “STL: A seasonal-trend decomposition procedure based on Loess,” *Journal of Official Statistics*, vol. 6, no. 1, pp. 3-73, 1990.

[20] A. D. Lagnika, A. Kissi, and M. P. Tchuenté, “A machine learning framework for agricultural decision support,” *Expert Systems with Applications*, vol. 185, p. 115578, 2021. DOI: 10.1016/j.eswa.2021.115578.

[21] Open Government Data (OGD) Platform India, “Climate Data Repository,” [Online]. Available: https://data.gov.in/. [Accessed: Mar. 2025].

[22] Food and Agriculture Organization of the United Nations (FAOSTAT), “FAOSTAT Database,” [Online]. Available: https://www.fao.org/faostat/en/. [Accessed: Mar. 2025].

[23] Ministry of Agriculture & Farmers’ Welfare, Government of India, “Agricultural Statistics,” [Online]. Available: https://agricoop.nic.in/. [Accessed: Mar. 2025].

[24] WorldOMeter, “CO_2 Levels and Climate Change Statistics,” [Online]. Available: https://www.worldometers.info/. [Accessed: Mar. 2025].

[25] Open Government Data (OGD) Platform India, “Extreme Weather Events Data,” [Online]. Available: https://data.gov.in/. [Accessed: Mar. 2025].

[26] Food and Agriculture Organization of the United Nations (FAOSTAT), “Pesticide Use Statistics,” [Online]. Available: https://www.fao.org/faostat/en/. [Accessed: Mar. 2025].

[27] National Renewable Energy Laboratory (NREL), “National Solar Radiation Database (NSRDB),” [Online]. Available: https://nsrdb.nrel.gov/. [Accessed: Mar. 2025].

[28] A. D. Lagnika, A. Kissi, and M. P. Tchuenté, “A machine learning framework for agricultural decision support,” *Expert Systems with Applications*, vol. 185, p. 115578, 2021. DOI: 10.1016/j.eswa.2021.115578.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.

[30] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, 2nd ed. New York, NY, USA: Springer, 2021.

[31] J. Heer, M. Bostock, and V. Ogievetsky, “A tour through the visualization zoo,” *Communications of the ACM*, vol. 53, no. 6, pp. 59-67, 2010. DOI: 10.1145/1743546.1743567.

[32] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of Machine Learning Research*, vol. 13, pp. 281-305, 2012.

[33] D. Pyle, *Data Preparation for Data Mining*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann, 1999.

[34] H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, Boston, MA, USA: Springer, 1998.

[35] Z. Zhang, “Missing data imputation: Focusing on single imputation methods,” *Annals of Translational Medicine*, vol. 4, no. 7, pp. 125-138, 2016. DOI: 10.21037/atm.2016.02.05.

[36] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785-794.

[37] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proc. 14th Int. Joint Conf. on Artificial Intelligence (IJCAI)*, Montreal, Canada, 1995, pp. 1137-1143.

[38] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765-4774.

[39] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “STL: A seasonal-trend decomposition procedure based on Loess,” *Journal of Official Statistics*, vol. 6, no. 1, pp. 3-73, 1990.

[40] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, 2013. DOI: 10.1109/TPAMI.2013.50.

[41] L. Breiman, “Statistical modeling: The two cultures,” *Statistical Science*, vol. 16, no. 3, pp. 199-231, 2001.

[42] J. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189-1232, 2001. DOI: 10.1214/aos/1013203451.

[43] B. Shneiderman, “The eyes have it: A task by data type taxonomy for information visualizations,” in *Proc. IEEE Symposium on Visual Languages*, Boulder, CO, USA, 1996, pp. 336-343.

[44] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in *Proc. 20th Int. Conf. on Very Large Data Bases (VLDB)*, Santiago, Chile, 1994, pp. 487-499.

[45] C. Ware, *Information Visualization: Perception for Design*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2012.

ISSN 1660-6795
www.nano-ntp.com

Farming for Tomorrow: Sustainability amidst Climate Change in India

Gresha Bhatia[1], Rohini Temkar[2], Vishakha Singh[3], Manasi Sharma[4], Anushka Shirode[5] *Department Of Computer Engineering, V.E.S. Institute of Technology, Mumbai 74* gresha.bhatia@ves.ac.in, rohini.temkar@ves.ac.in

Abstract:

Climate change poses a significant threat to agriculture in India, a sector that sustains a large portion of the population and contributes substantially to the country's economy. This study aims to analyze how various climate and environmental factors influence agricultural yield and production, with a focus on parameters such as temperature fluctuations, rainfall, irrigation practices, CO2 emissions, frequency of extreme weather events, pesticide and fertilizer usage, and solar irradiance (DNI, DHI, GHI). Using statistical and machine learning models— ARIMA, Ridge regression, Lasso regression, and Generalized Additive Models (GAM)—the research seeks to identify the most significant predictors of agricultural output. Through comprehensive data analysis and modeling, this study provides insights into the complex interactions between these factors and their direct and indirect effects on crop yield and production. The findings are crucial for developing adaptive strategies and policies to mitigate the adverse impacts of climate change, ensuring food security, and supporting sustainable agricultural practices in India

Keywords—Climate change, Extreme weather events, Agriculture yield, production, prediction.

Introduction:

Indian agriculture is crucial for ensuring food, nutrition, and livelihood security, but it currently faces significant challenges. These challenges include stagnating net sown areas, plateauing yields, soil quality deterioration, and reduced per capita land availability. Climate change is exacerbating these issues, particularly affecting rainfed areas, which make up about 60% of the cultivated land. With over 80% of farmers being small and marginal, the sector is under immense pressure from a growing population and lacks the resilience needed to cope with these stresses. Rising levels of greenhouse gases like CO2 (over 2.5 billion metric tons), CH4, and N2O contribute to global warming, leading to increased temperatures and more extreme weather events, negatively impacting crops, soils, livestock, and pests.

The effects of climate change on Indian agriculture are significant, especially as the

frequency of climatic extremes, such as droughts, floods, frosts, heatwaves, and cyclones, increases. Predictions suggest a 1.5°C to 2.0°C rise in global temperatures in the next 50 years. Fig.1 shows the average divergence from mean temperature at the beginning of last century in India, by decade (in °C).

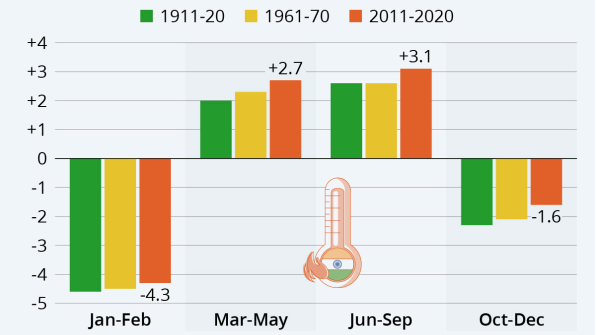


Fig.1 Average divergence from mean temperature

(Sources: Indian Meteorological Department, Ministry of Earth Sciences)

Rainfed regions, which contribute 40-45% of India's total agricultural output, are particularly vulnerable to these changes. Water scarcity, soil health degradation, and the adverse effects on livestock and fisheries further compound the challenges. Addressing these issues requires innovative, climate-resilient agricultural technologies and adaptive management strategies to ensure sustainability.

To handle these challenges, this work integrates machine learning models and climate data analysis to better understand the impact of climate factors on agriculture. Using data-driven approaches, such as Generalized Additive Models (GAM), Ridge, and Lasso regression, the paper aims to identify the most influential climate parameters affecting yield and production of any crop. Next, forecasting models like ARIMA are applied to predict future climate conditions and their corresponding effects on crop yield. The proposed work further seeks to offer insights on mitigating climate risks and enabling farmers to implement proactive strategies for improved sustainability and productivity. Through this multi-phase approach, the research hopes to contribute to the resilience of Indian agriculture in the face of climate change.

Related Work:

The authors in [1] explored the MARS model and the nonlinear relationships between climate factors and agricultural productivity in India, offering flexibility but facing computational challenges. Machine learning models predicted temperature and rainfall in Marathwada, excelling in seasonal data but struggling with irregular patterns [2]. In [3] ANN-MLP models examined rainfall trends in India post-1960, effectively handling nonlinear data but limited by reliance on historical data. Regression models assessed CO2 emissions and population growth impacts on climate but oversimplified the complexity of broader climate systems were elaborated by the authors in [4]. The authors of [5] observed that the Simulation models in Asia highlighted climate impacts on agriculture, though real-world variability was not fully captured.The study critiqued the Green Revolution's chemical reliance, advocating organic farming, though it faces adoption challenges due to higher labor and lower yields[6]. Integrating thermodynamics with machine learning improved climate predictions but was hindered by high computational demands, The authors in [7] Climate-smart agriculture strategies offered sustainability for developing nations but required significant policy and financial support

mentioned in [8]. Authors of [9] indicated that rising temperatures threatened agricultural sustainability in Indian states, with solutions facing implementation challenges in marginalized regions. Water management and climate-resilient crops were essential for rainfed regions, but progress depends on long-term investments [10]. Further the authors in [10] focussed on Global studies that showed rising temperatures reducing yields, especially in developing countries, but lacked regional adaptability insights. District-level rice yield forecasts under climate scenarios offered valuable data but faced uncertainties due to potential future advances was mentioned in [12]. Sustainable farming practices were essential for food security but constrained by socioeconomic barriers like access to technology and support was specified in [13]. Authors of [14] clearly mentioned that climate variables adversely impacted India's economic growth, though adaptive measures were not fully accounted for. Paper [15][16] elaborated that climate change worsened rural poverty and productivity issues, but the study lacked insights into urban-rural interactions. Global food system impacts from climate change require adaptive strategies, though practical application across diverse contexts remains challenging, mentioned the authors of [16][18] Water management and infrastructure improvements were critical for India's agriculture, but financial and logistical barriers persist[19]. The authors of [20][21] identified that ICT and supply chain management could reduce post-harvest losses but require investment and accessibility for smallholder farmers. Further the authors of [22][23] mentioned that modern agricultural practices are essential for India's growth, but financial and technological barriers limit widespread adoption. Authors of [24][25] expressed that investments in agriculture through targeted policies and infrastructure boost productivity but depend on political commitment and equitable resource distribution.

Methodology:

Fig.2 shows the methodology used in the project to analyze the impact of various climatic factors on agricultural yield and production. It begins with data collection from diverse sources, including government websites, open databases, and environmental data platforms, such as the Ministry of Agriculture, WorldOMeter, and the National Solar Radiation Database. The collected data, including factors like production, CO2 emissions, rainfall, and solar irradiance (DNI, DHI, GHI), is transformed into standard units to ensure uniformity across the dataset.The transformed data is divided into dependent and independent variables for further analysis. Machine learning models like ARIMA, Ridge, Lasso, and GAM are applied to evaluate the relationships between climatic variables and agricultural outcomes.

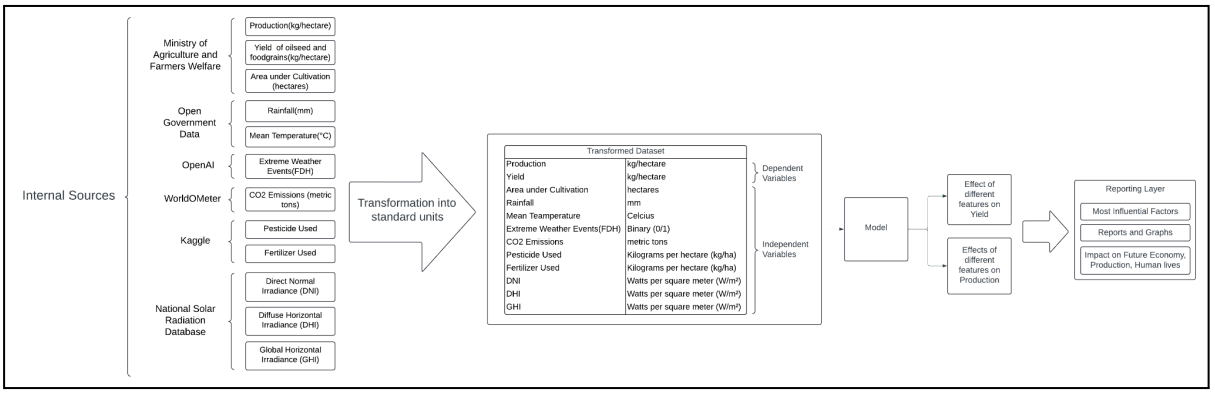


Fig.2 Block Diagram representation

Implementation:

A. Identification of Influential Climatic Factors

1) Data Collection:

Data from various sources such as the Ministry of Agriculture, Open Government Data, WorldOMeter, Kaggle, National Solar Radiation Database, etc., are gathered. The dataset includes important variables:

- Dependent Variables: Production (kg/ha), Yield (kg/ha), Area under Cultivation (hectares).
- Independent Variables: Rainfall (mm), Mean Temperature (°C), Extreme Weather Events (binary), CO2 Emissions (metric tons), Pesticide Use (kg/ha), Fertilizer Use (kg/ha), DNI, DHI, GHI (W/m²).

2) Data Transformation:

The data is transformed into standard units to ensure uniformity and facilitate comparative analysis. This transformation is crucial as it allows for integrating diverse datasets that may have been recorded in different units or formats. Additionally, the historical time series data is cleaned to eliminate any missing values or discrepancies, thereby ensuring the reliability and accuracy of the subsequent analyses. [21][26] The results are then used to determine the most influential factors affecting yield and production. Finally, reports and visualizations generated from the models provide insights into future trends, assisting in policy-making and risk mitigation strategies for sustainable agriculture.

3) Exploratory Data Analysis (EDA):

EDA is performed to gain insights into the relationships between various climatic factors and agricultural outcomes.

Handling Missing Values: The first step in the EDA involves addressing missing values in the dataset. AutoRegressive Integrated Moving Average (ARIMA) models are utilized to impute these missing values, leveraging historical data trends for accurate estimations [22][27].

4) Model Selection for Impact Analysis:

The analysis of influential climatic factors employs several statistical models.

- Ridge and Lasso Regression: These models are utilized to study the relative importance of different climatic features and perform feature selection based on shrinkage methods [23][28].
- Generalized Additive Model (GAM): Used to capture non-linear relationships between climate parameters and yield/production. The most influential factors on yield and production are identified, with GAM contributing to the analysis of smooth trends across the climatic parameters [24].

B. Forecasting Climatic Factors and Yield

1) Time Series Forecasting:

A time series forecasting approach is adopted to predict key climatic parameters that significantly influence agricultural outcomes. The ARIMA model is employed to forecast essential variables such as rainfall, temperature, CO2 emissions, and solar

radiation metrics (DNI, DHI, GHI) for the years 2023 to 2030. The ARIMA models are trained on historical data, and validation techniques, including cross-validation, are used to ensure the accuracy and reliability of the forecasts. The predicted values of these independent variables will be utilized in the next step to assess their impact on agricultural yield [25][29].

2) Yield Prediction Based on Forecasts:

Following the forecasting of climatic factors, these projected values are integrated into the previously established GAM, Ridge, and Lasso models. This integration enables the prediction of future agricultural yield and production based on the anticipated climate scenarios for the next decade. Furthermore, scenario analyses are conducted to simulate both optimistic and pessimistic climate conditions, allowing for a comprehensive evaluation of how varying climatic conditions could influence agricultural productivity.

C. Inference and Risk Mitigation

1) Trend Analysis:

The predicted yield and production values are analyzed to identify trends over time. This analysis includes comparing the projected outcomes across different periods to discern potential increases or decreases in agricultural productivity. Visualization tools such as time series plots are employed to highlight these trends clearly, and key metrics—such as year-on- year changes and percentage growth—are calculated to provide a detailed understanding of the dynamics at play.

2) Risk Identification and Mitigation:

Based on the insights gained from the forecasted data, areas vulnerable to significant yield declines due to adverse climatic conditions are identified. Adaptive strategies for risk mitigation are then recommended.

- Technological Innovations: Precision farming, use of climate-resilient crop varieties, soil health monitoring, and water-efficient irrigation practices.
- Policy Recommendations: Advocating for subsidies on climate-resilient seeds, promoting efficient resource use, and introducing farmer training programs.
- Proactive Measures: Early warning systems for extreme weather, improved weather forecasting services, and better pest control mechanisms [26][30].

3) Reporting and Insights:

The final reporting layer will provide insights on:

- Impact of the Most Influential Factors: Detailed reports and graphs will showcase how specific climatic factors are contributing to yield changes.
- Impact on Future Economy and Human Lives: The economic ramifications, including possible production shortfalls or boons, will be detailed alongside implications for food security and farmers' livelihoods.

Results:

ARIMA is employed for forecasting missing values in the CO2 data from 1950 to 1960 and for GHI, DNI, and DHI from 1950 to 2000. It relies on past data points, captures patterns such as trends and seasonality, and predicts future values based on these historical trends.

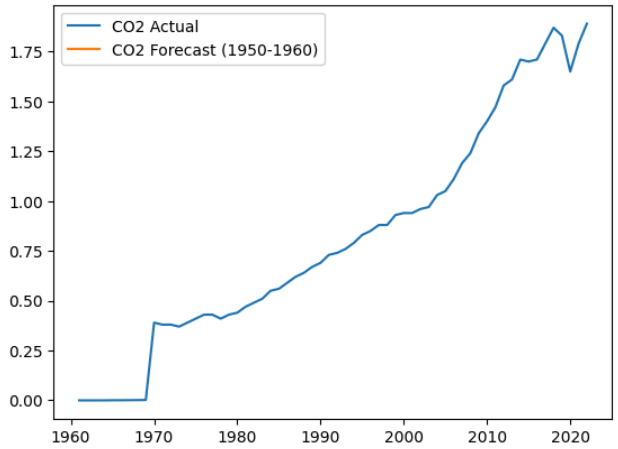


Fig.3 ARIMA for CO2 Prediction of Missing Values

The graph in Fig.3 illustrates the ARIMA model's predicted values for CO2, comparing the observed and predicted trends to estimate missing data accurately.

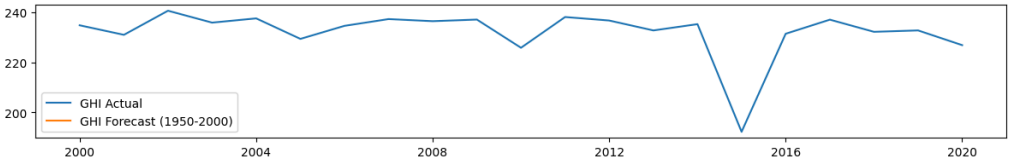


Fig.4 ARIMA's predicted Global Horizontal Irradiance (GHI) values

The graph in Fig.4 shows ARIMA's predicted Global Horizontal Irradiance (GHI) values, effectively filling in the missing values while maintaining the observed trend [28].

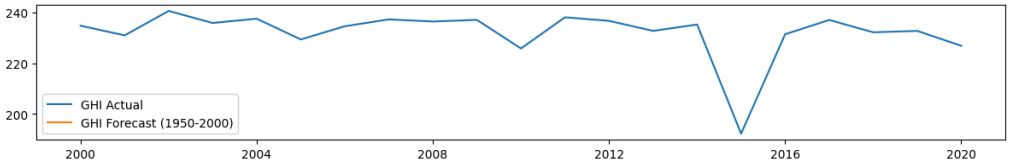


Fig.5 ARIMA for DNI Prediction of Missing Values

The graph in Fig.5 visualizes ARIMA's predictions for Direct Normal Irradiance (DNI), comparing forecasted values with the actual trend to fill in missing data points [29].

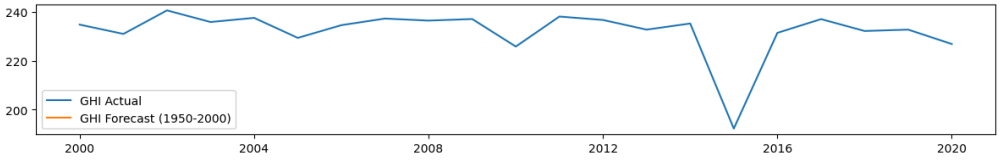


Fig.6 ARIMA for DHI Prediction of Missing Values

ARIMA predictions for Diffuse Horizontal Irradiance (DHI) are presented in Fig.6, where missing values are forecasted based on historical patterns.

Ridge and Lasso regression models in Fig.7 and Fig.8 below assess the relationships between various climatic parameters and yield/production.

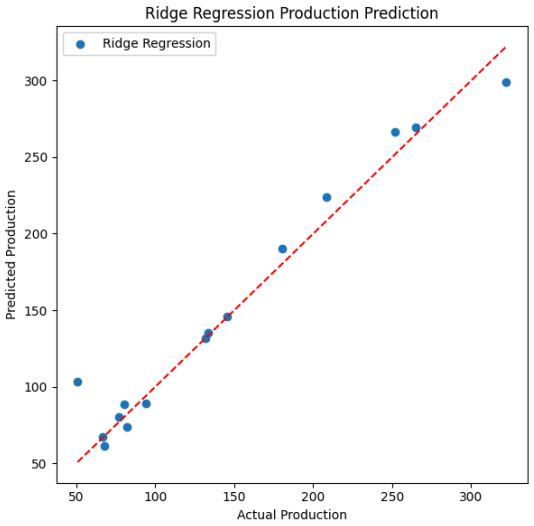


Fig.7 Ridge Regression

The red dashed in Fig.7 line represents a scenario where the model's predictions perfectly match the actual data points (i.e. where predicted value = actual value). It's a reference line to visualize how close the predictions are to the true values. The blue dots in the plot represent the actual values of the target variable (i.e. Production) compared to the predicted values made by the Ridge or Lasso regression models.

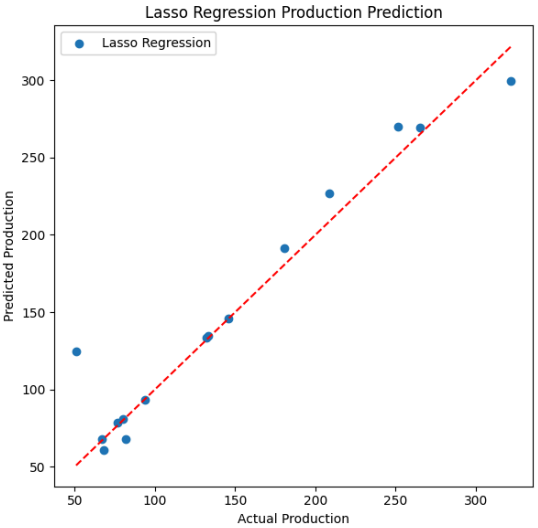


Fig.8 Lasso Regression

Similarly in Fig.8, the red dashed line is the line of best fit and the blue dots represent how much the predicted value of the target variable by the model deviated from the actual value.

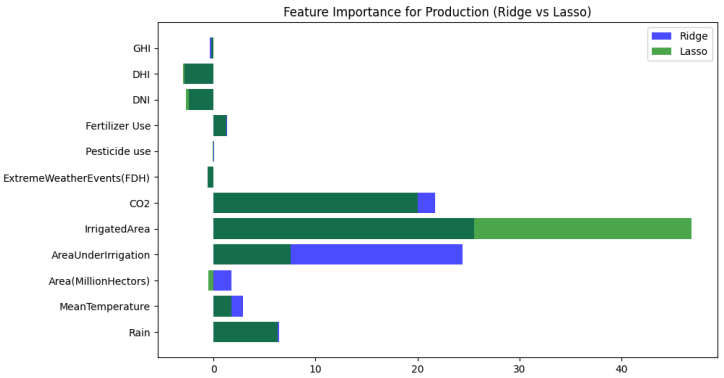


Fig.9 Feature Importance for Production

Each bar in Fig.9 represents a different feature from your dataset, with the horizontal axis showing the magnitude of the feature's impact on the target variable, Production.

The graph shows that the "Irrigated Area" parameter has the highest importance in predicting production. In the Lasso model, it stands out as the most significant feature. It suggests that CO2 emissions and extreme weather events do affect production but not as critically as irrigation-related features. Pesticide Use and Fertilizer use features have very little or no contribution in both models, indicating that pesticide and fertilizer usage do not significantly impact production. Rainfall and Mean Temperature have minimal impact while Solar Irradiance Metrics (GHI, DHI, DNI) are small compared to other factors like irrigation and CO2.

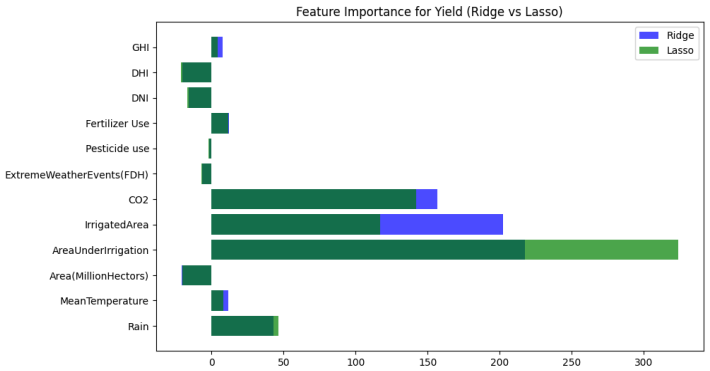


Fig.10 Feature Importance for Yield

Fig.11 Impact of Individual Features on Production

In Fig.11 the spline plots [30] indicate that agricultural production is influenced by several factors. Moderate rainfall boosts production, but too much or too little rain reduces it, emphasizing the need for balance. Optimal temperatures, around 25.5°C, support higher production, while extremes negatively impact output. An increase in cultivated area steadily raises production, and irrigation also helps, but beyond a certain limit, the benefits diminish. Rising CO2 levels moderately improve production, but excessive levels show limited further gains. Extreme weather events sharply reduce production. Pesticide use shows mixed effects, depending on pest levels and other conditions, while fertilizers significantly enhance production, though overuse can be harmful. Solar irradiance (DNI, DHI, GHI) benefits crops at moderate levels but can reduce productivity when extreme.

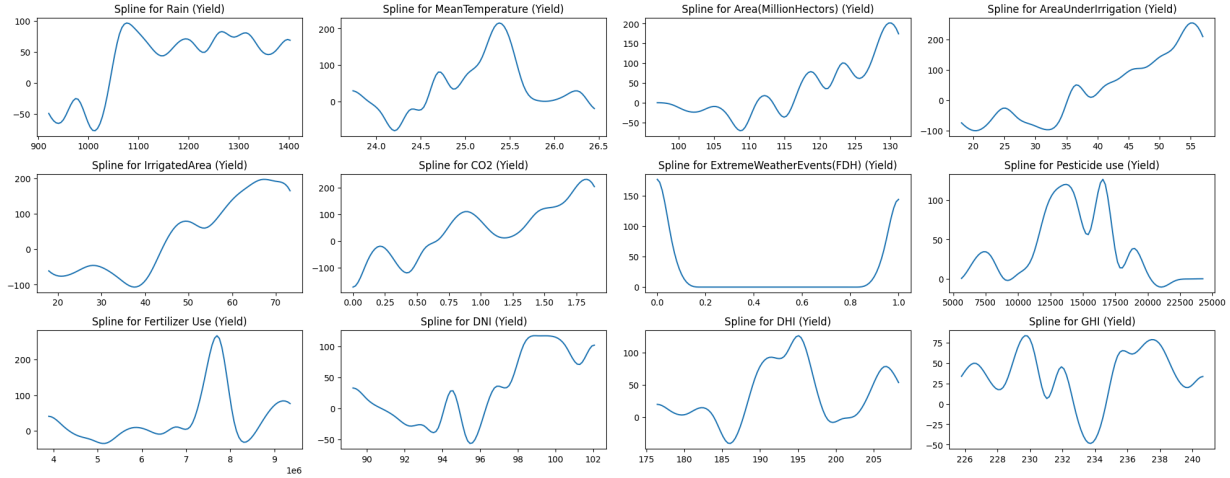


Fig.12 Impact of Individual Features on Yield

In Fig.12 Yield trends generally mirror production, with optimal rainfall and temperature boosting efficiency, while extremes reduce it. Proper irrigation and land expansion improve yield, but excessive use without management lowers it. Pesticides and fertilizers enhance yield when used optimally, but overuse leads to diminishing returns. Extreme weather events significantly reduce yield efficiency.

ARIMA demonstrated high precision in predicting missing values for the dataset, particularly for climate parameters like rainfall, temperature, and CO2 emissions. The model considers the available historical data to identify trends, seasonality, and noise, and uses this information to forecast future values. By comparing the predicted values with the actual recorded data, ARIMA showed minimal error, validating its effectiveness in handling time- series data. The close alignment of predicted values with the actual values indicated that ARIMA is reliable in estimating missing data and projecting future trends, allowing for more accurate forecasting and analysis.

Ridge and Lasso regression models were both effective for feature selection, but they handled the issue of multicollinearity differently. Ridge regression added a penalty to the magnitude of coefficients, reducing the impact of multicollinearity by shrinking coefficients for correlated variables without completely eliminating them. This allowed Ridge to stabilize predictions when dealing with highly correlated features like rainfall, temperature, and irrigation. In contrast, Lasso regression applied a penalty that could force some coefficients to become exactly zero, effectively selecting a subset of the most important features. This made Lasso better at feature selection but sometimes led to less

stable predictions when features were highly correlated, as some key variables might be excluded entirely.

GAM offered a different advantage by capturing nonlinear relationships between features and yield/production. Unlike Ridge and Lasso, which assume linearity, GAM used smooth spline functions to model complex interactions between variables like rainfall, temperature, and irrigation. This allowed GAM to reveal patterns that were not detectable by linear models, making it more flexible for understanding how climatic factors impacted yield in a more nuanced way. However, GAM's complexity can sometimes lead to overfitting if not properly regularized, which is a limitation when compared to the more straightforward Ridge and Lasso models.

Algorithm	MSE	R-Square	Inference
Ridge	2011.29	0.97	Irrigated Area
Lasso	2112.12	0.96	Irrigated Area
GAM	61694.85	-8.34	Irrigated Area

Table I. Comparison of evaluation of algorithms

Table I shows the evaluation metrics (MSE and R-Squared) for the Ridge, Lasso, and GAM algorithms, with the most influential factor being the Irrigated Area. Ridge and Lasso exhibit strong performance with high R-squared values, while GAM shows a poor fit, indicating potential issues with its model assumptions for this dataset.

Conclusion and Future Scope:

The analysis revealed that the "Irrigated Area" was the most significant crop yield and production driver among various climatic and agricultural factors. The results underscore the importance of water resource management in agriculture, especially under changing climatic conditions. Predictive models like ARIMA, GAM, and Ridge/Lasso regression proved valuable in determining influential factors and providing future insights. By enhancing irrigation infrastructure and adopting water-efficient practices, farmers can mitigate risks related to climate variability and secure better yields, ensuring agricultural sustainability for the future.

References:

[1] Jha, Paritosh, Sona Chinngaihlian, and Priyanka Upreti. "A Machine Learning Approach to Assess Implications of Climate Risk Factors on Agriculture: The Indian Case." Department of Economic and Policy Research, Reserve Bank of India, Mumbai, India., 2021.

[2] Ramdas D. Gore and Bharti W. Gawali, ―Analysis of Weather Parameters Using Machine Learning‖, July 2023, URL (https://www.atlantis-press.com/article/125989847.pdf)

[3] Bapi Raju et al., ―Analyzing Trend and Forecasting of Rainfall Changes in India Using Non-Parametric and Machine Learning Approaches,‖, June 2020, URL (https://www.nature.com/articles/s41598-020-67228-7).

[4] P. Roy et al., ―Analysis of Various Climate Change Parameters in India Using Machine Learning,‖, Jan 2022, URL (https://arxiv.org/pdf/2201.10123)

[5] Habib-ur-Rahman, Muhammad, Ashfaq Ahmad, Ahsan Raza, Muhammad Usama Hasnain, Hesham F. Alharby, Yahya M. Alzahrani, Atif A. Bamagoos, Khalid Rehman Hakeem, Saeed Ahmad, Wajid Nasim, Shafaqat Ali, Fatma Mansour, and Ayman EL Sabagh. "Impact of Climate Change on Agricultural Production: Issues, Challenges, and Opportunities in Asia." 2022.
[6] Choudhary, D. K. (2021). Chemical Fertilizers and Pesticides in Indian Agriculture. International Journal of Research and Analysis in Science and Engineering, 1(6). Retrieved from https://www.iarj.in/index.php/ijrase/index
[7] Vázquez-Ramírez, S., Torres-Ruiz, M., Quintero, R., Chui, K.T., & Guzmán Sánchez-Mejorada, C. (2023). An Analysis of Climate Change Based on Machine Learning and an Endoreversible Model. Mathematics, 11(3060).
[8] Malhi, G.S., Kaur, M., & Kaushik, P. (2021). Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. Sustainability, 13(3), 1318.
[9] Singh, A.K., Kumar, S., & Jyoti, B. (2022). Influence of Climate Change on Agricultural Sustainability in India: A State-Wise Panel Data Analysis. Asian Journal of Agriculture, 6(1), 15-27.
[10] "Climate Change and Agriculture in India" report supported by the National Mission on Strategic Knowledge for Climate Change (NMSKCC)
[11] Cline, William R. Global Warming and Agriculture: Impact Estimates by Country. Washington: Center for Global Development and Peterson Institute for International Economics, 2007.
[12] Gallé, Johannes, and Anja Katzenberger. "Indian Agriculture Under Climate Change: The Competing Effect of Temperature and Rainfall Anomalies." Economics of Disasters and Climate Change (2024). https://doi.org/10.1007/s41885-024-00154-4.
[13] Dubey, Pradeep Kumar, Ajeet Singh, Rajan Chaurasia, Krishna Kumar Pandey, Amit Kumar Bundela, Rama Kant Dubey, and Purushothaman Chirakkuzhyil Abhilash. "Planet Friendly Agriculture: Farming for People and the Planet." Current Research in Environmental Sustainability 3 (2021): 100041. https://doi.org/10.1016/j.crsust.2021.100041.
[14] Husain, Uvesh, and Sarfaraz Javed. "Impact of Climate Change on Agriculture and Indian Economy: A Quantitative Research Perspective from 1980 to 2016." Industrial Engineering & Management 8, no. 2 (2019): 281.
[15] https://www.researchgate.net/publication/346655247.
[16] Kar, Saibal, and Nimai Das. "Climate Change, Agricultural Production, and Poverty in India." In Poverty Reduction Policies and Practices in Developing Asia, edited by A. Heshmati, 55–76. Singapore: Springer, 2015. https://doi.org/10.1007/978-981-287-420-7_4.
[17] Kumara, Lalit, Ngawang Chhogyel, Tharani Gopalakrishnan, Md Kamrul Hasan, Sadeeka Layomi Jayasinghe, Champika Shyamalie Kariyawasam, Benjamin Kipkemboi Kogo, and Sujith Ratnayake. "Climate Change and Future of Agri-Food Production." In Future Foods, edited by P.C. Keenan, 49–64. Elsevier, 2022. https://doi.org/10.1016/B978-0- 323-91001-9.00009-8.
[18] Cagliarini, Adam, and Anthony Rush. "Economic Development and Agriculture in India." Bulletin (June Quarter 2011): 15-22. Reserve Bank of Australia.
[19] National Institute of Agricultural Marketing. Agriculture and Economic Development in India. Final Report. Jaipur: National Institute of Agricultural Marketing, 2011.
[20] Palanivel, Prakash. A Study on Role of Agricultural Development in Indian

Economy. Project Report, Ramakrishna Mission Vivekananda College, 2020.
[21] Mehta, Niti. "Agricultural Investments: Trends and Role in Enhancing Agricultural Output and Incomes." Indian Journal of Agricultural Economics 78, no. 4 (2023): 576-589.
[22] Aldoseri, Abdulaziz, Khalifa N. Al-Khalifa, and Abdel Magid Hamouda. 2023. "Re- Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges" Applied Sciences 13, no. 12: 7082. https://doi.org/10.3390/app13127082
[23] Kritika Banerjee. Handling missing values in EDA. Medium (https://medium.com/@kritiikaaa/handling-missing-values-in-eda-b12efc7da26d0)
[24] DataCamp. Lasso and Ridge Regression in Python Tutorial. https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression
[25] Wikipedia. Generalized additive model. https://en.wikipedia.org/wiki/Generalized_additive_model
[26] IBM. ARIMA model. (https://www.ibm.com/topics/arima-model)
[27] 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.
[28] Food and Agriculture Organization of the United Nations. (2024). Implications of Economic Policy for Food Security: A Training Manual. https://www.fao.org/4/X3936E/X3936E00.htm
[29] Vaisala Energy Support, What is Global Horizontal Irradiance? https://www.3tier.com/en/support/solar-prospecting-tools/what-global-horizontal-irradiance- solar-prospecting/
[30] World Health Organization. Radiation: The ultraviolet (UV) index, https://www.who.int/news-room/questions-and-answers/item/radiation-the-ultraviolet-(uv)- index
[31] R Development Core Team. plot.gam: Default GAM plotting in mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [rdrr.io]. Retrieved from https://rdrr.io/cran/mgcv/man/plot.gam.html

Predictive Modeling of Agricultural Trends in Maharashtra

Vishakha Singh
Department Of Computer Engineering
V.E.S. Institute of Technology
Mumbai-40074, India
2021.vishakha.singh@ves.ac.in

Manasi Sharma
Department Of Computer Engineering
V.E.S. Institute of Technology
Mumbai-40074, India
2021.manasi.sharma@ves.ac.in

Anushka Shirode
Department Of Computer Engineering
V.E.S. Institute of Technology
Mumbai-40074, India
2021.anushka.shirode@ves.ac.in

Gresha Bhatia
Department Of Computer Engineering
V.E.S. Institute of Technology
Mumbai-40074, India
gresha.bhatia@ves.ac.in

***Abstract*— Maharashtra is a key contributor to India's agricultural sector, with its productivity heavily influenced by climatic and environmental factors. This study examines the relationship between crop yield, production, and critical variables such as temperature, rainfall, irrigation, and nutrient consumption using data from 1966 to 2023. Correlations between yield, production, and factors like weather, fertilizers, and soil nutrients are analyzed. SHAP (SHapley Additive exPlanations) identifies the most influential factors, while the Apriori algorithm uncovers associations between agricultural attributes. For forecasting, machine learning models—RFR (Random Forest Regressor), SVR (Support Vector Regressor), and GBR (Gradient Boosting Regressor)—are compared, with GBR emerging as the best. STL (Seasonal and Trend decomposition using Loess) is applied to GBR's time series data to reveal trends and seasonal patterns. This comprehensive approach provides actionable insights for enhancing agricultural productivity and sustainability in Maharashtra.**

***Keywords*—Climate Variability, Agricultural Productivity, Forecasting Models, Maharashtra Agriculture, Data-Driven Analysis.**

I. Introduction

Agriculture is the backbone of India's economy, providing livelihoods to millions and ensuring food security. Maharashtra, one of the country's key agricultural states, showcases a mix of fertile plains, semi-arid regions, and coastal belts, supporting diverse crops like sugarcane, cotton, and pulses. However, the sector remains highly vulnerable to climate variability, with erratic rainfall, rising temperatures, and frequent droughts posing major challenges. Regions like Marathwada and Vidarbha have experienced extreme heat, with temperatures soaring past 50°C in May 2023, impacting soil health and water availability. While climate change threatens agricultural stability, it also drives innovation, encouraging the adoption of climate-resilient crops, precision farming, and advanced irrigation techniques to sustain productivity in the face of growing uncertainties.

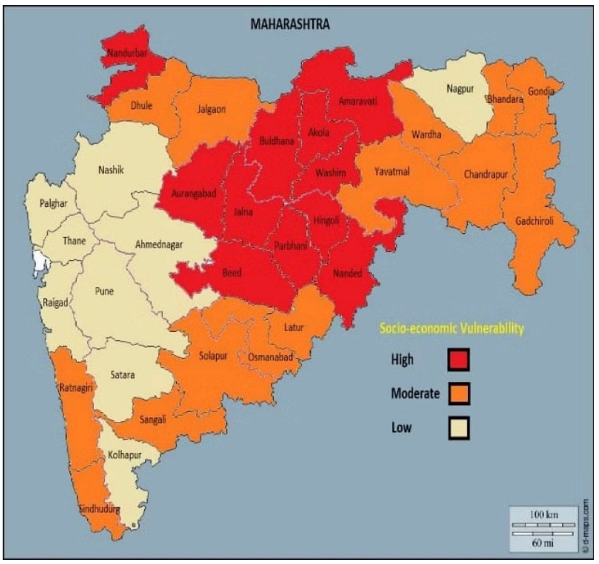


Fig.1 District-wise socioeconomic vulnerability in Maharashtra
(Source: Deccan Herald)

Fig. 1 shows that 77% of Maharashtra's cropped area is vulnerable to climate change, according to a study based on data from 44 indicators related to climatic and socio-economic factors.

Using data from 1966 to 2023, this study applies correlation and STL decomposition, and the Apriori algorithm to identify key trends and associations in agriculture. It also evaluates forecasting models like

RFR, SVR, and GBR to determine their accuracy in predicting crop yield and production. By comparing their performance, the study offers insights into the most effective methods for agricultural forecasting under changing climatic conditions.

II. Related work

Paper [1], titled "Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques" employs a Multilayer Perceptron (MLP) to predict crop yield using rainfall data from 1901 to 2002, with the dataset split into 60%-40% for training and evaluation, assessed via MSE and NMSE. The study's merits include its ability to capture nonlinear relationships in climate data and its emphasis on data preprocessing for effective feature extraction. However, it has notable demerits, such as the limited dataset, which ends in 2002, making predictions less relevant for current climate trends, and the lack of comparison with other advanced ML models like Random Forest or XGBoost, which could have provided a more comprehensive analysis.

Paper [2], titled "A Creative Use of Machine Learning for Crop Prediction and Analysis" employs SVM, Random Forest, and Decision Trees to analyze seasonal crop growth trends using a dataset of over 25,000 records. The study's merits include the use of a large dataset, which enhances training generalization, and the application of Exploratory Data Analysis (EDA) to understand data distributions. However, it has notable demerits, such as the absence of discussion on hyperparameter tuning and the lack of justification for why Random Forest outperformed the other models, leaving room for further clarification and optimization.

Paper [3], titled "Influence of Causal Inference for Crop Prediction" integrates Random Forest (RF) with Bayesian Inference to enhance causal relationship modeling in agricultural yield predictions, achieving an impressive 97.2% accuracy. The study's merits include the use of a causal inference framework, which improves interpretability, and the high accuracy, which underscores the model's reliability. However, it has notable demerits, such as the limited dataset size (2200 rows), which reduces the model's generalizability, and the absence of testing with deep learning methods, which could have provided additional insights into performance and scalability.

Paper [4], titled "A Case Study on the Application of Machine Learning to the Process of Crop Forecasting" compares SVM, Decision Tree, and Random Forest for crop forecasting using Maharashtra state data, with Random Forest achieving 97% accuracy. The study's merits include the use of real-world agricultural data and the provision of fertilizer recommendations, adding practical value. However, it has notable demerits, such as the lack of real-time weather integration, which limits dynamic adaptability, and the reliance on historical data, which may restrict the model's ability to address current or future agricultural challenges effectively.

Paper [5], titled "Climate Forecasting: Long Short-Term Memory Model using Global Temperature Data" employs LSTM networks to forecast global climate trends using temperature datasets, achieving 96.16% accuracy. The study's merits include the model's ability to effectively capture long-term dependencies in climate trends and the use of standard error metrics such as MAE, RMSE, and MAPE for evaluation. However, it has notable demerits, including high computational costs and the omission of external climate factors like greenhouse gas emissions, which could enhance the model's comprehensiveness and relevance to real-world climate dynamics.

III. Methodology

To analyze the impact of fertilizers, soil nutrients, and weather on yield and production, correlation is used. SHAP helps explain the importance of features and how each parameter influences yield. Apriori identifies associations between climatic parameters, fertilizers, and yield.

For robust non-linear regression models, SVR, RFR, and GBR are employed to predict yield. Finally, STL decomposes the GBR forecasted data to understand its trend, seasonality, and residual components.

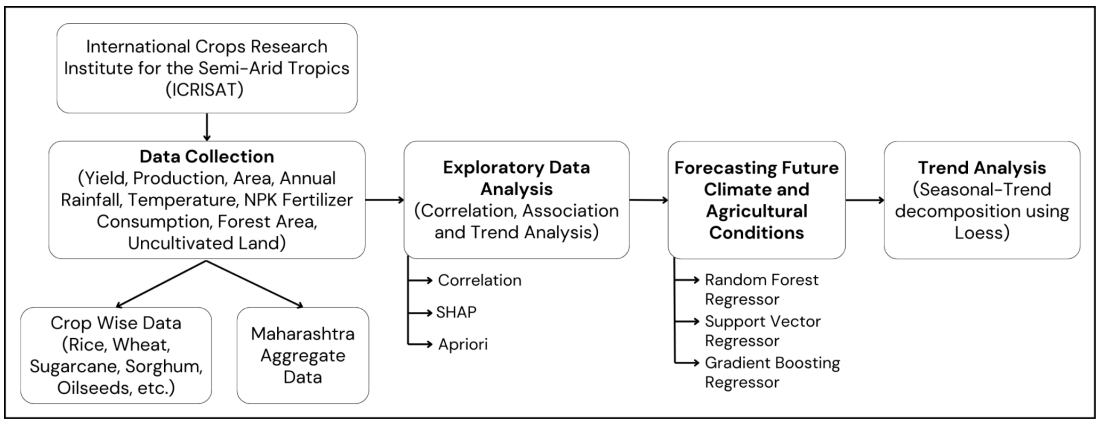


Fig.2 Block Diagram

Fig. 2 outlines the methodology used to analyze and forecast the impact of climatic and agricultural factors on crop yield and production.

The study begins with data collection from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), focusing on Maharashtra's agricultural and environmental data, including yield, production, area, rainfall, temperature, NPK fertilizer consumption, forest area, and uncultivated land.

Exploratory Data Analysis (EDA) is performed to identify trends and relationships [6]. Correlation between yield and production with climatic features is derived. SHAP is a game theoretic approach to explain the output of any machine learning model. It is used to find top 10 most influential factors for each crop's yield and production. The Apriori algorithm is used to uncover associations between key attributes, helping to identify influential factors in agricultural productivity.

For forecasting, predictive models like Random Forest (RFR) and Support Vector Regressor (SVR) capture complex relationships, while Gradient Boosting (GBR) enhances accuracy by optimizing weak models.

The performance of these models is compared to determine the most effective approach for estimating future agricultural trends, and Seasonal-Trend Decomposition using Loess (STL) captures seasonal variations. This methodology offers a structured framework for understanding past patterns and making informed decisions about future agricultural planning.

IV. Implementation

1) Data Collection & Preprocessing

The dataset used in this study was obtained from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) [7], containing aggregate and crop-specific data for multiple districts in Maharashtra from 1966 onwards.

The aggregate dataset includes Year, State Name, and Dist Name for location and time tracking, along with Area, Production, and Yield for agriculture output. It covers Irrigated Area for water use and climate factors like Annual Rainfall, Min/Max Temp, Precipitation, and Evapotranspiration. Fertilizer details include Nitrogen, Phosphate, Potash, NPK composition, and application rates under NCA and GCA. Land-use metrics encompass Total Area, Forest, Barren Land, Non-Agricultural Land, Cultivable Waste, Pastures, Fallow Areas, and Cropping Intensity.

The crop dataset tracks Year and Dist Name along with major crops such as Rice, Wheat, Sorghum, Pearl Millet, and Maize, pulses like Chickpea, Pigeonpea, and Minor Pulses, oilseeds including Groundnut, Sesamum, and Total Oilseeds, and cash crops like Sugarcane and Cotton. It also includes Fruits and Vegetables for horticulture trends, Potash and Total Fertilizer Use, and climate factors like Min/Max Temp, Precipitation, and Evapotranspiration.

2) Statistical Modeling & Pattern Analysis

To understand the relationships between agricultural factors, correlation analysis was applied to determine the strength and direction of associations between attributes and Yield and Production. SHAP [8] was then used to assess how different attributes influence the two target variables, Yield and Production, providing a more granular understanding of feature importance and their impact on the predictions.

To uncover hidden associations, the Apriori algorithm [9] was applied to identify frequent itemsets between Area, Irrigation, and Nutrient Consumption with Yield and Production. Continuous data was binarized, and association rules were generated with minimum support (0.3) and confidence (0.5). These relationships

were visualized through directed graphs, providing insights into key agricultural dependencies.

3) Regression Models

Predictive models were developed for Yield and Production, considering Year, Area, Irrigation, Nitrogen and Phosphate Consumption, Temperature, and Rainfall. The models were trained on data from all the districts to capture regional variations.

For Random Forest Regressor (RFR) [10][11], the aggregate data model uses a broader range of n_estimators (100, 200, 300, 500) and min_samples_split (2, 5, 10) compared to the crop data model, which has n_estimators limited to 100, 200, and 300 and min_samples_split to 2 and 5. Support Vector Regressor (SVR) [12][13] differs significantly: for crop data, it includes svr__C (1 to 1000), svr__epsilon (0.001 to 0.5), and svr__gamma options, whereas for aggregate data, the grid is smaller, with C (1, 10, 100), epsilon (0.1, 0.2, 0.5), and only linear and rbf kernels. Gradient Boosting Regressor (GBR) [14][15] uses identical hyperparameters across both datasets, including n_estimators (100, 200, 300), learning_rate (0.01, 0.05, 0.1), and max_depth (3, 5).

4) Model Evaluation & Forecasting

All models were evaluated based on R^2, MSE, RMSE, MAE, and Mean Absolute Percentage Error (MAPE) [16] to determine the most reliable forecasting approach. The best-performing models were used to generate long-term agricultural projections up to 2040, supporting data-driven decision-making for sustainable agricultural planning.

5) STL Decomposition for Trend Comparison

STL [17] decomposition was employed to compare the trend of the forecasted values with the actual data values obtained in the dataset. By decomposing variables such as Area, Rainfall, and Irrigation into trend, seasonal, and residual components, we validated the accuracy of the model's predictions and identified temporal patterns impacting agricultural outcomes.

V. Results

The correlation for aggregate wise data reveals for Yield, there is a positive correlation with Nitrogen Consumption (tons) and a negative correlation with Minimum Temperature (Celsius). For Production, it is positively correlated with Total Consumption (tons) and negatively correlated with Minimum Temperature (Celsius).

The correlation for crop-wise data reveals key relationships between climatic factors, fertilizer use, and crop yields. For rice, yield is most positively correlated with precipitation (0.498) and negatively with max temperature (-0.407). Wheat shows the strongest positive correlation with nitrogen per HA of NCA (0.556) but a negative correlation with nitrogen share in NPK (-0.229). Sorghum benefits from nitrogen per HA of NCA (0.124) but is negatively impacted by precipitation (-0.556). Pearl millet is positively linked to irrigated area (0.257) but negatively to precipitation (-0.148). Maize responds positively to nitrogen per HA of NCA (0.484) but negatively to max temperature (-0.030). Chickpea and pigeonpea both show strong positive correlations with phosphate per HA of GCA (0.493 and 0.492, respectively) but negative correlations with nitrogen share in NPK (-0.303) and potash share in NPK (-0.161). Minor pulses benefit from nitrogen per HA of NCA (0.258) but show a negative link with potash share in NPK (-0.106). Groundnut yields are positively tied to phosphate per HA of GCA (0.337) and negatively to max temperature (-0.465). Sesamum shows a similar pattern, positively correlated with phosphate per HA of GCA (0.392) and negatively with max temperature (-0.397). Oilseeds respond positively to nitrogen per HA of NCA (0.454) but slightly negatively to max temperature (-0.060). Sugarcane yields rise with phosphate per HA of GCA (0.394) but drop with max temperature (-0.391). Lastly, cotton shows a positive correlation with phosphate per HA of GCA (0.336) and a mild negative link to max temperature (-0.046). Overall, nitrogen and phosphate usage generally support crop yields, while extreme temperatures and imbalanced fertilizer shares often suppress them.

SHAP is a model-agnostic technique based on game theory that explains the contribution of each feature to a model's prediction. It assigns Shapley values to input features, indicating their impact (positive or negative) on the output.

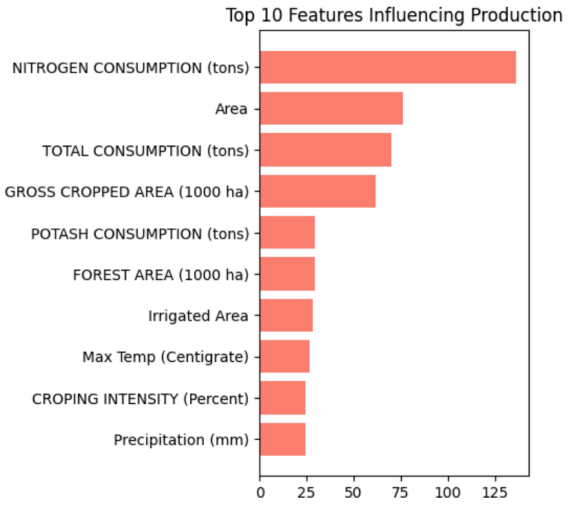


Fig.3 Top 10 features influencing Production

Fig. 3 shows the mean absolute SHAP values for each feature with respect to production predictions. Features such as Nitrogen Consumption (tons) and Area have the most significant impact, as indicated by their longer bars. The x-axis represents the magnitude of feature impact, where larger values indicate that these features play a stronger role in influencing the model's production predictions.

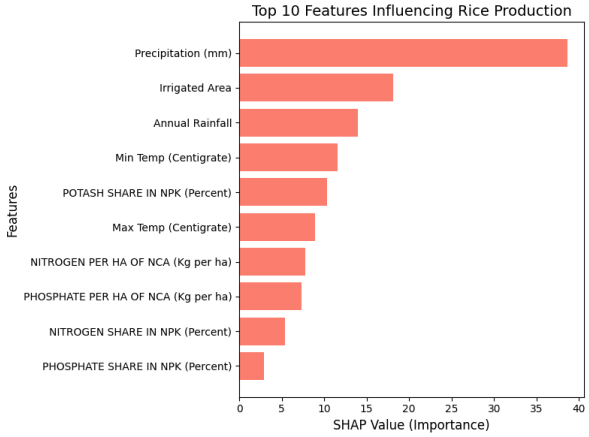


Fig.4 Top 10 features influencing Rice Production

Fig. 4 shows that Precipitation (mm) and Irrigated Area have the most significant impact on Rice Production.

The key climatic factors impacting agriculture in Maharashtra include rising maximum and minimum temperatures, leading to increased heat stress on crops, with extreme heat events further reducing grain-filling periods and increasing pest vulnerability. Rainfall fluctuations, including sharp declines and erratic precipitation patterns, contribute to inconsistent soil moisture retention and disrupt sowing schedules. The occurrence of alternating drought and flood cycles further exacerbates crop stress, particularly in water-intensive crops. These climatic shifts collectively affect irrigation availability, crop resilience, and long-term agricultural productivity, necessitating adaptive strategies to mitigate their impact.

The Apriori algorithm is used in data mining to identify frequent itemsets and generate association rules from transactional data. It works by iteratively finding subsets of items that frequently occur together, based on a minimum support threshold. The Association Rules Graph visualizes these rules, where nodes represent items (like temperature categories) and edges represent associations between them, with edge weights indicating the confidence level of the rule. In this graph, temperature categories (e.g., Min Temp_Medium, Max Temp_High) are connected based on their co-occurrence patterns in rice production data, helping to identify relationships between different temperature conditions and their impact on production.

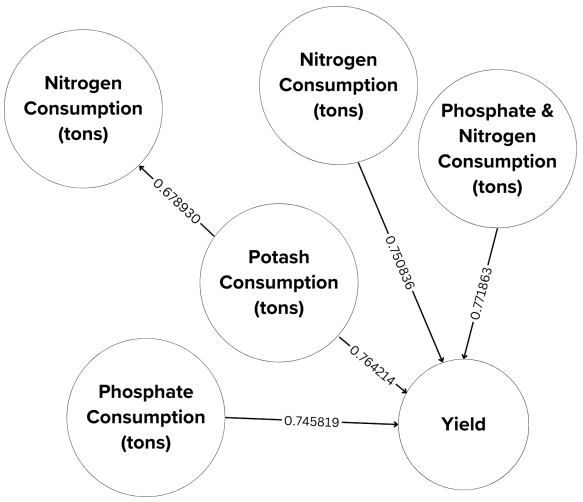


Fig.5 Apriori Rules for Yield

Fig 5 shows that nitrogen consumption boosts yield (0.75) but reduces potash use (-0.68), highlighting the need for balanced fertilization. Phosphate consumption positively impacts yield (0.75) with no direct link to other nutrients, indicating an independent effect. Higher potash use is linked to lower yields (-0.76) and is negatively associated with nitrogen use (-0.68). The combined effect of nitrogen and phosphate consumption has the strongest impact on yield (0.77), emphasizing their synergistic importance, while potash may hinder yield, illustrating the complexity of fertilizer management.

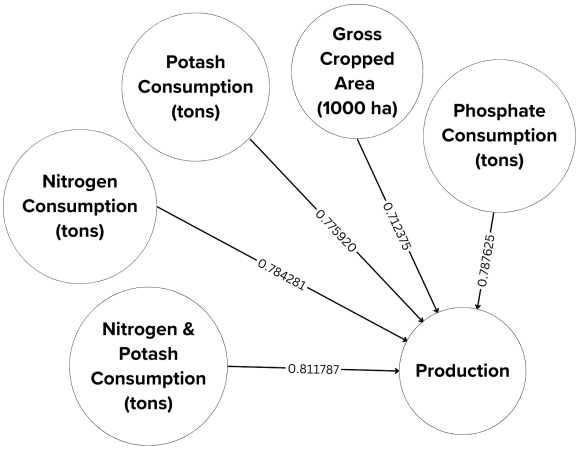


Fig.6 Apriori Rules for Production

Fig 6 shows that nitrogen consumption (0.78) significantly boosts production, while phosphate (-0.78) and potash (-0.77) have negative impacts, likely due to over-application or imbalances. Gross cropped area (0.71) positively affects production, as more cultivated land typically increases output. The combination of nitrogen and potash (0.81) shows the strongest positive impact on production, highlighting their combined importance. Overall, production is influenced by fertilizer use and cultivated area, with nitrogen and the nitrogen-potash combination having positive effects, while phosphate and potash alone show negative correlations.

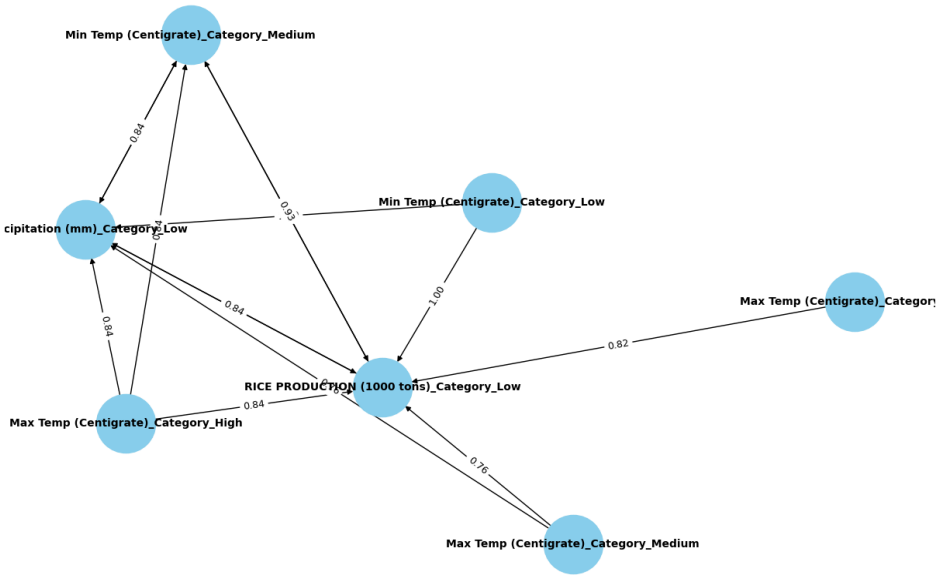


Fig.7 Apriori Rules for Rice Production

Fig. 7 shows that low minimum temperatures (1.00) have a very strong negative impact on rice production, with medium minimum temperatures (0.93) also hindering yields. High maximum temperatures (0.84) are more detrimental than medium ones (0.76), and overall, maximum temperatures strongly affect production (0.82). Low rainfall (0.84) is closely linked to low rice production and unfavorable temperature conditions. In summary, rice production is highly sensitive to low temperatures and requires sufficient rainfall.

The Apriori rules analysis for the Ahmednagar district in Fig. 8 reveals a strong interdependence between crop yields and rainfall, highlighting precipitation as a key factor influencing agricultural outcomes. High rice yields are frequently associated with high sorghum yields (confidence 89%, lift 1.42), indicating shared favorable conditions such as adequate rainfall and suitable soil quality. Conversely, low wheat yields consistently predict low cotton yields (confidence 83%, lift 1.64) and are linked to both low rice yields and poor precipitation (confidence 100%, lift 1.82), suggesting that wheat performance serves as an indicator of widespread agricultural stress, often due to drought. Groundnut yields are especially sensitive to rainfall, with low yields always corresponding to low precipitation levels (confidence 100%, lift 1.65). Overall, the district exhibits high vulnerability to climatic factors, as multiple major crops experience simultaneous yield declines during poor rainfall years, emphasizing the need for strategies focused on water management and climate resilience.

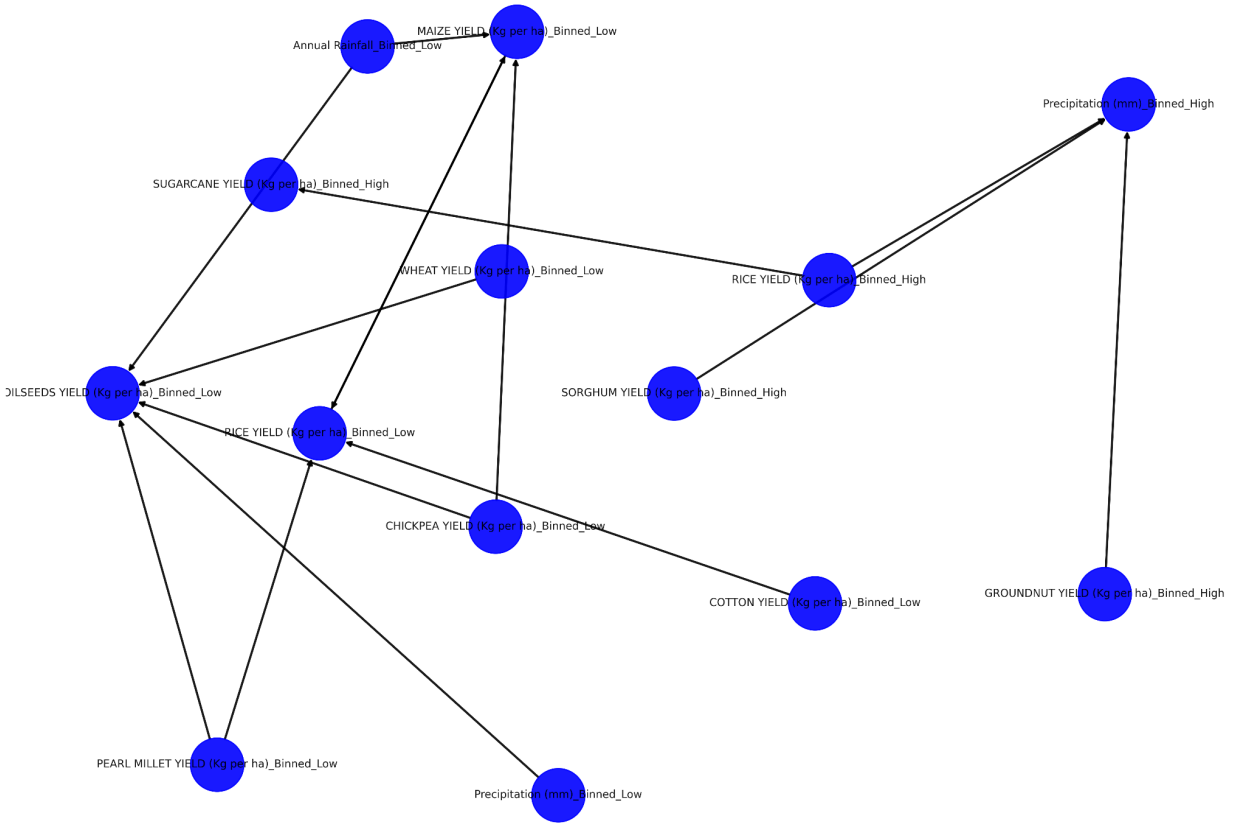


Fig.8 Apriori Rules for Ahmednagar District

Apriori analysis on 20 districts of Maharashtra reveals critical agricultural patterns driven by climate change. Rising maximum and minimum temperatures across all districts indicate intensifying heat stress, significantly impacting wheat and pulses, which are highly sensitive to temperature fluctuations. The increasing frequency of extreme heat events could lead to reduced grain filling periods, lower yields, and greater vulnerability to pests and diseases. Rainfall fluctuations are a major concern, with districts like Jalgaon and Satara experiencing sharp declines, while others face erratic precipitation patterns, leading to alternating drought and flood cycles. Such inconsistencies disrupt sowing schedules, delay crop maturity, and result in uneven soil moisture retention, affecting long-term productivity. A steady decline in irrigated area across several districts is shifting dependency toward rainfall, posing risks to water-intensive crops like rice and sugarcane. Reduced irrigation access in districts such as Ahmednagar, Aurangabad, and Amravati is already contributing to declining rice and wheat production, further exacerbated by rising temperatures. Fertilizer usage trends show declining nitrogen and phosphate application in multiple regions, which could lead to soil nutrient depletion, reduced crop resilience, and a long-term decline in agricultural output. The depletion of essential soil nutrients without adequate replenishment could reduce productivity, particularly in regions that already suffer from lower organic matter content. Maize production, however, shows fluctuations, with some districts like Akola recording increased yields, possibly due to its better adaptation to variable rainfall patterns. Meanwhile, sugarcane and cotton remain stable, with sugarcane exhibiting resilience to shifting climate conditions, and cotton benefiting from its drought-resistant properties, allowing it to sustain yields even in water-scarce districts.

These findings highlight a critical need for improved irrigation infrastructure, adaptive crop management strategies, and balanced fertilizer application to sustain long-term agricultural productivity. Without proactive interventions, the increasing unpredictability of climate factors could lead to heightened risks for staple crops, potentially threatening food security and farmer livelihoods in the region.

Table I. Comparison of forecasting algorithms.

	Aggregate data						Crop data					
	RFR		SVR		GBR		RFR		SVR		GBR	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Ahmednagar	0.939	586.42	0.990	799.32	0.988	15.14	0.972	11.35	0.992	11.39	0.978	9.42
Akola	0.943	441.21	0.989	361.08	0.982	3.72	0.970	20.69	0.970	55.30	0.983	18.12
Amarawati	0.927	344.33	0.993	157.68	0.964	67.06	0.953	14.71	0.986	19.20	0.979	10.28
Beed	0.909	386.96	0.990	396.18	0.957	22.00	0.969	29.05	0.970	34.31	0.967	24.04
Bhandara	0.900	405.38	0.991	227.73	0.949	8.72	0.954	27.54	0.944	47.77	0.951	25.77
Buldhana	0.943	388.75	0.991	335.51	0.958	10.31	0.966	39.60	0.969	45.44	0.972	35.18
Chandrapur	0.934	292.70	0.993	251.73	0.967	28.83	0.965	18.23	0.967	24.28	0.972	14.92
Dhule	0.931	291.47	0.994	360.82	0.958	10.7	0.983	10.73	0.987	16.10	0.988	10.22
Jalgaon	0.935	212.98	0.985	1523.15	0.966	25.69	0.966	17.03	0.986	20.38	0.973	13.22
Kolhapur	0.926	326.53	0.987	445.14	0.957	3.06	0.978	14.66	0.959	7.95	0.950	8.76
Nagpur	0.928	400.54	0.989	264.46	0.972	5.90	0.972	12.52	0.972	15.64	0.965	9.46
Nanded	0.909	121.52	0.988	737.51	0.963	17.82	0.959	63.06	0.968	73.94	0.970	59.08
Nasik	0.922	482.58	0.990	614.08	0.972	5.55	0.974	14.8	0.958	21.95	0.962	11.23
Osmanabad	0.938	325.43	0.990	401.85	0.975	5.82	0.953	29.08	0.973	29.07	0.976	24.02
Parbhani	0.934	153.12	0.992	269.75	0.954	24.18	0.981	13.28	0.986	15.33	0.987	10.50
Pune	0.904	567.07	0.985	1205.92	0.945	19.73	0.971	15.64	0.982	23.16	0.984	10.43
Sangli	0.922	1007.0	0.987	546.06	0.962	7.81	0.973	19.26	0.983	18.92	0.987	14.32
Satara	0.932	281.01	0.992	256.83	0.96	9.47	0.958	14.14	0.969	7.96	0.952	10.16
Solapur	0.919	342.82	0.985	1265.94	0.954	12.76	0.975	13.43	0.975	20.26	0.985	10.27
Yeotmal	0.922	358.13	0.988	432.27	0.974	7.66	0.961	17.04	0.977	17.99	0.966	14.96

Table I shows that GBR (Gradient Boosting Regressor) emerges as the best-performing model for both Aggregate and Crop data due to its optimal balance between high explanatory power (R^2) and low prediction error (RMSE).

For Aggregate data, while SVR (Support Vector Regressor) occasionally achieves marginally higher R^2 values (e.g., 0.990 compared to GBR's 0.988 in Ahmednagar), GBR demonstrates significantly lower RMSE values (e.g., 15.14 vs. SVR's 799.32), highlighting its superior practical accuracy and reliability. Similarly, for Crop data, GBR consistently outperforms both RFR (Random Forest Regressor) and SVR, achieving the highest R^2 values (e.g., 0.978 in Ahmednagar) and the lowest RMSE values (e.g., 9.42) across most regions.

GBR's consistent performance underscores its robustness and generalization across diverse datasets, making it the preferred choice for accurate and reliable agricultural yield and production forecasting.

Table II. Forecasted values for Yield, Production, and Area.

District	Major Crop	Year	Yield (Kg per ha)	Production (tons)	Area (1000 ha)
Bhandara	Rice (85.5% of total production)	2030	766.76	215.99	281.00
		2035	1585.15	448.99	283.00
		2040	1127.00	325.99	289.00
Nanded	Sorghum (42.7% of total production)	2030	665.00	173.05	260.00
		2035	408.00	116.95	287.00
		2040	845.99	254.01	300.00
Solapur	Sugarcane (48.4% of total production)	2030	8721.07	133.99	15.01
		2035	8447.17	185.04	21.98
		2040	8715.37	176.00	19.98
Nagpur	Oilseeds (29.2% of total production)	2030	315.00	24.00	75.00
		2035	330.00	21.99	65.99
		2040	231.01	15.00	67.00

Table II shows the forecasted values for Yield, Production, and Area for selected districts and their major crops in 2030, 2035, and 2040. The districts are chosen based on the highest percentage contribution of a specific crop to total production. The forecasted values are obtained using the Gradient Boosting Regressor (GBR) model.

The table indicates an overall increasing trend in Area for most districts, suggesting potential expansion of cultivation. However, Yield and Production show fluctuations, highlighting the influence of various environmental and agronomic factors. In Bhandara, Rice Yield and Production peaked in 2035 before slightly declining in 2040. Nanded's Sorghum shows a dip in 2035 but recovers by 2040. Solapur's Sugarcane maintains a high Yield with minor variations in Production, while Nagpur's Oilseed Production exhibits a gradual decline, possibly due to environmental constraints. These trends highlight the dynamic nature of agricultural patterns.

STL (Seasonal and Trend decomposition using Loess) is a time series decomposition method that separates data into three components: Seasonal, Trend, and Residual. It helps identify underlying patterns, such as long-term trends and recurring seasonal effects, in time series data.

When applied to GBR's forecasted data, STL decomposes the predictions into these components. If the original data shows an upward trend, GBR's predictions will also reflect this pattern, as GBR captures the underlying relationships in the data. By analyzing the Trend component from STL, we can confirm that GBR's predictions align with the original data's trend, ensuring the model's accuracy and reliability.

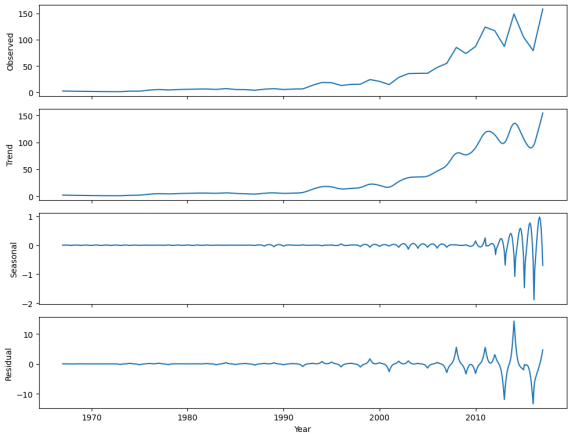


Fig.9 STL for Maize Production for original data

Fig. 9 shows the STL decomposition of current maize production, showcasing a significant upward trend in recent years with noticeable seasonality and fluctuations. The decomposition separates the observed production into its trend, seasonal, and

residual components, providing insights into the underlying patterns and potential future production.

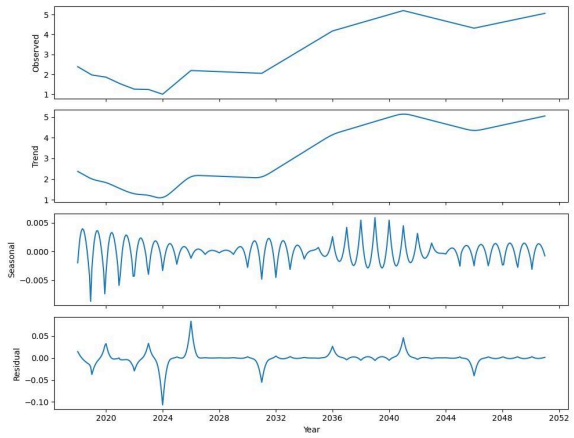


Fig.10 STL for Maize Production for future forecast

Fig. 10 follows the STL decomposition of maize production as observed in the original data by dividing it into observed, trend, seasonal, and residual components, revealing a general upward trend with noticeable seasonality and fluctuations.

VI. Conclusion And Future Scope

This study analyzed the impact of climatic and agricultural factors on crop yield and production using two datasets: an aggregate dataset and a crop-specific dataset. SHAP is applied to find the most influential features per attributes, highlighting importance. The Apriori algorithm identified key associations between attributes.

For Yield, there is a positive correlation with Nitrogen Consumption (tons) and a negative correlation with Minimum Temperature (Celsius). For Production, it is positively correlated with Total Consumption (tons) and negatively correlated with Minimum Temperature (Celsius).

Regression models, including RFR, SVR, and GBR, were employed to predict future values, with GBR emerging as the most effective across both datasets and GBR's predictions align with the original data's trend, ensuring the model's accuracy and reliability.

By applying STL (Seasonal and Trend decomposition using Loess) to GBR's predictions, data is decomposed into trend, seasonal, and residual components. The alignment with trend shows accuracy and reliability, ensuring that GBR's forecasts are consistent with historical patterns and capable of providing meaningful insights for future agricultural planning.

The data reveals significant crop-related risks across districts. Crop production is impacted, with districts experiencing declining rice, wheat, and maize yields. Fertilizer use trends indicate decreasing nitrogen and phosphate levels in several regions, potentially affecting soil fertility and long-term crop productivity.

Based on the risk observed, the districts have been classified into high-risk districts—Ahmednagar, Aurangabad, Jalgaon, Solapur, Amravati, Nanded, Osmanabad, Chandrapur, Beed, and Gadchiroli—face severe climate-induced agricultural challenges. Medium-risk districts—Akola, Nashik, Pune, Satara, Wardha, Latur, Buldhana, and Yavatmal—experience moderate declines in production and climate fluctuations. Low-risk districts—Kolhapur and Nagpur—demonstrate relative stability in crop output and climatic conditions. These insights are crucial for policymakers and farmers to develop targeted mitigation strategies, such as improved irrigation management, adaptive cropping patterns, and climate-resilient agricultural practices.

References

[1] S. Malathy, C. N. Vanitha, Kotteswari, S. V, S. P and M. E, "Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022,pp.437-442,doi:10.1109/ICAAIC53929.2022.9792793.

[2] R. C. Mahore and N. G. Gadge, "A Creative Use of Machine Learning for Crop Prediction and Analysis," 2022 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE), Bangalore, India, 2022, pp. 1-6, doi: 10.1109/ICWITE57052.2022.10176256.

[3] K. Srivastava, A. Sawarkar, S. Nawale and A. R. Agrawal, "Influence of Causal Inference For Crop Prediction," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10725416.

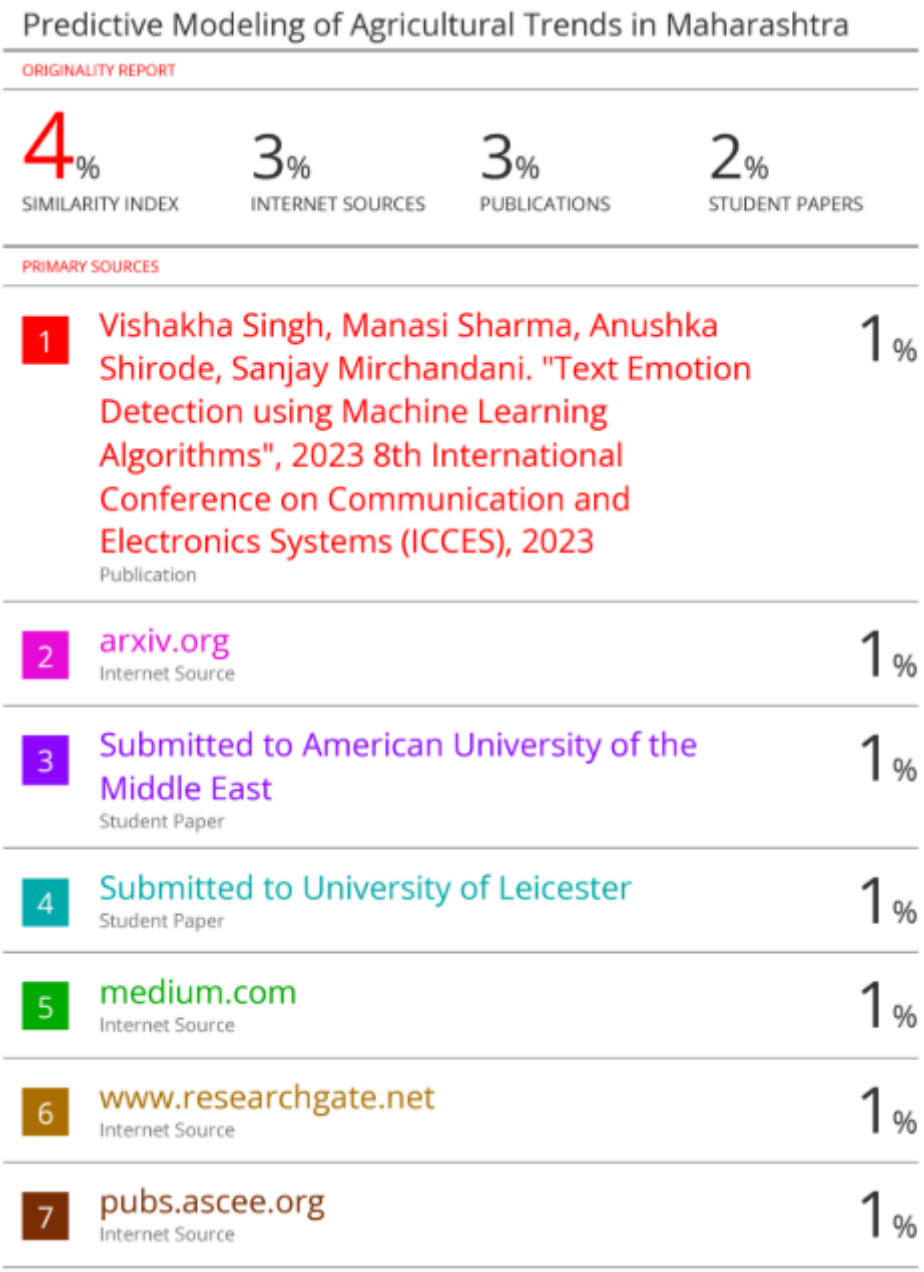
[4] S. A. Jiwani, J. Shinde, B. Bag, R. Nayak, R. K. Shial and U. Ghugar, "A Case Study on the Application of Machine Learning to the Process of Crop Forecasting," 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0, Raigarh, India, 2024, pp. 1-5, doi: 10.1109/OTCON60325.2024.10688298.

[5] P. Akhila, R. L. S. Anjana and M. Kavitha, "Climate Forecasting:Long short Term Memory Model using Global Temperature Data," 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2022, pp. 469-473, doi: 10.1109/ICCMC53470.2022.9753779.

[6] "What is Exploratory Data Analysis?," GeeksforGeeks, [Online]. Available: https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/.

[7] "ICRISAT - Science of Discovery to Science of Delivery," International Crops Research Institute for

the Semi-Arid Tropics, [Online]. Available: https://www.icrisat.org/.
[8] “SHAP,”[Online]. Available: https://www.geeksforgeeks.org/shap-a-comprehensive-guide-to-shapley-additive-explanations/
[9] "Apriori Algorithm," IBM, [Online]. Available: https://www.ibm.com/think/topics/apriori-algorithm.
[10] "Random Forest Regression," GeeksforGeeks, [Online]. Available: https://www.geeksforgeeks.org/random-forest-regression-in-python/.
[11] "RandomForestRegressor," [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.RandomForestRegressor.html.
[12] "Support Vector Regression (SVR)," GeeksforGeeks, [Online]. Available: https://www.geeksforgeeks.org/support-vector-regression-svr-using-linear-and-non-linear-kernels-in-scikit-learn/
[13] "SVR," scikit-learn, [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.svm.SVR.html.
[14] "Gradient Boosting in ML," GeeksforGeeks, [Online]. Available: https://www.geeksforgeeks.org/ml-gradient-boosting/.
[15] "Gradient Boosting Regressor," [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html#sklearn.ensemble.GradientBoostingRegressor.
[16] "Regression Metrics," GeeksforGeeks, [Online]. Available: https://www.geeksforgeeks.org/regression-metrics/.
[17] "STL," ArcGIS Insights, [Online]. Available: https://doc.arcgis.com/en/insights/latest/analyze/stl.htm



Predictive Modeling of Agricultural Trends in Maharashtra

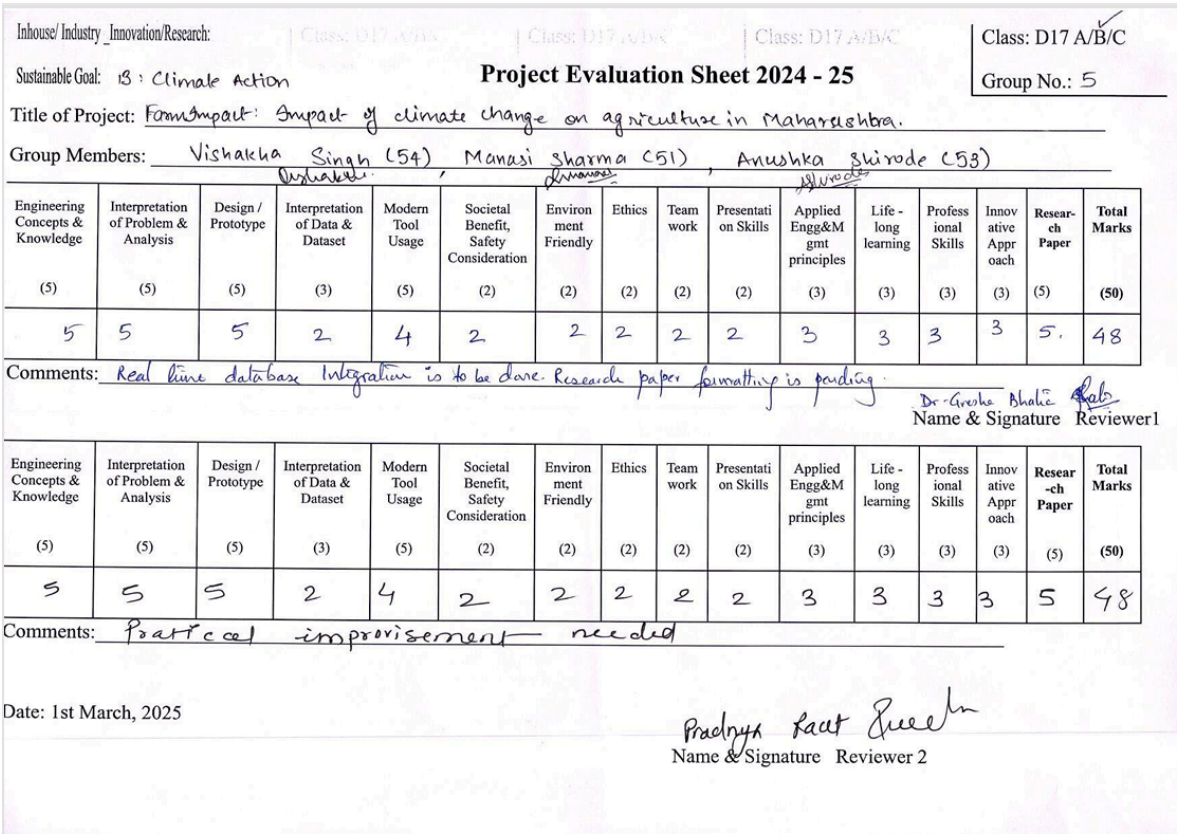
ORIGINALITY REPORT

4%	3%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Vishakha Singh, Manasi Sharma, Anushka Shirode, Sanjay Mirchandani. "Text Emotion Detection using Machine Learning Algorithms", 2023 8th International Conference on Communication and Electronics Systems (ICCES), 2023 Publication	1%
2	arxiv.org Internet Source	1%
3	Submitted to American University of the Middle East Student Paper	1%
4	Submitted to University of Leicester Student Paper	1%
5	medium.com Internet Source	1%
6	www.researchgate.net Internet Source	1%
7	pubs.ascee.org Internet Source	1%

Review 1 Sheet



Inhouse/ Industry _Innovation/Research:

Class: D17 A/B/C

Sustainable Goal: 13 : Climate Action

Project Evaluation Sheet 2024 - 25

Group No.: 5

Title of Project: FarmImpact: Impact of climate change on agriculture in Maharashtra.

Group Members: Vishakha Singh (54), Manasi Sharma (51), Anushka Shirode (53)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
5	5	5	2	4	2	2	2	2	2	3	3	3	3	5.	48

Comments: Real time database integration is to be done. Research paper formatting is pending.

Dr. Gresha Bhatia

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
5	5	5	2	4	2	2	2	2	2	3	3	3	3	5	48

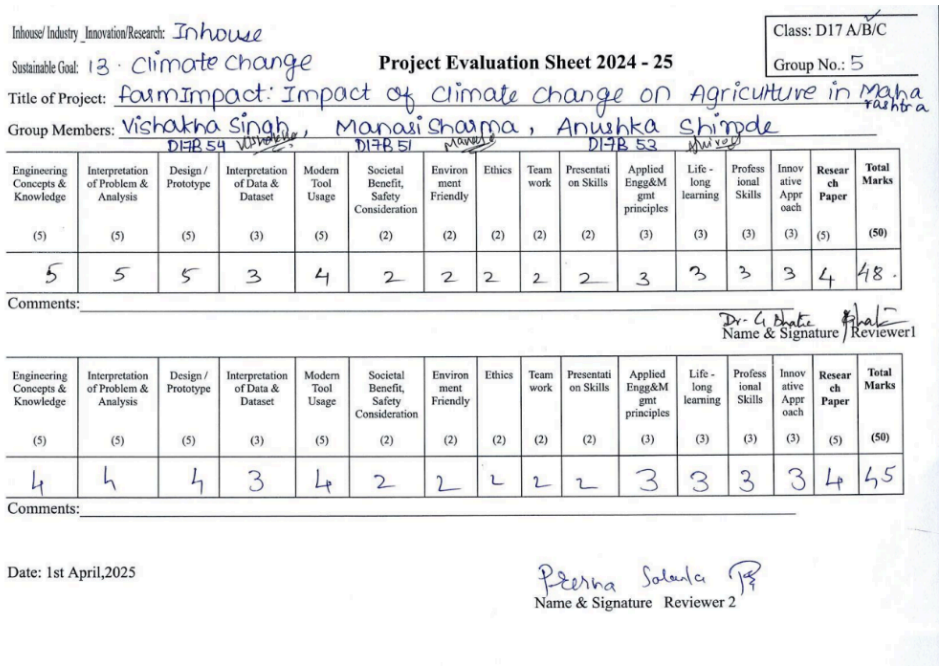
Comments: Practical improvisement needed

Date: 1st March, 2025

Pradnya Raut

Name & Signature Reviewer 2

Review 2 Sheet



Inhouse/ Industry _Innovation/Research: Inhouse

Class: D17 A/B/C

Sustainable Goal: 13 · Climate Change

Project Evaluation Sheet 2024 - 25

Group No.: 5

Title of Project: farmImpact: Impact of Climate Change on Agriculture in Maharashtra

Group Members: Vishakha Singh, Manasi Sharma, Anushka Shirode

D17B 54 D17B 51 D17B 53

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
5	5	5	3	4	2	2	2	2	2	3	3	3	3	4	48.

Comments:

Dr. G Bhatia

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (2)	Applied Engg&Mgmt principles (3)	Life - long learning (3)	Professional Skills (3)	Innovative Approach (3)	Research Paper (5)	Total Marks (50)
4	4	4	3	4	2	2	2	2	2	3	3	3	3	4	45

Comments:

Date: 1st April,2025

Reena Solanki

Name & Signature Reviewer 2

2. Competition Certificates

Vivekanand Education Society's Institute Of Technology, Mumbai

An Autonomous Institute affiliated to University of Mumbai, approved by AICTE

In association with VESIT - IQAC & VESIT - IIC

Department Of Computer Engineering

CERTIFICATE OF EXCELLENCE

This to certify that Vishakha Singh
of D17B has won 2nd prize at Department Technology Day, **Pradarshini '25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI
HOD, COMPUTER DEPARTMENT

DR. MRS. GRESHA BHATIA
DHOD, COMPUTER DEPARTMENT

PROJECT COORDINATOR
COMPUTER DEPARTMENT

Vivekanand Education Society's Institute Of Technology, Mumbai

An Autonomous Institute affiliated to University of Mumbai, approved by AICTE

In association with VESIT - IQAC & VESIT - IIC

Department Of Computer Engineering

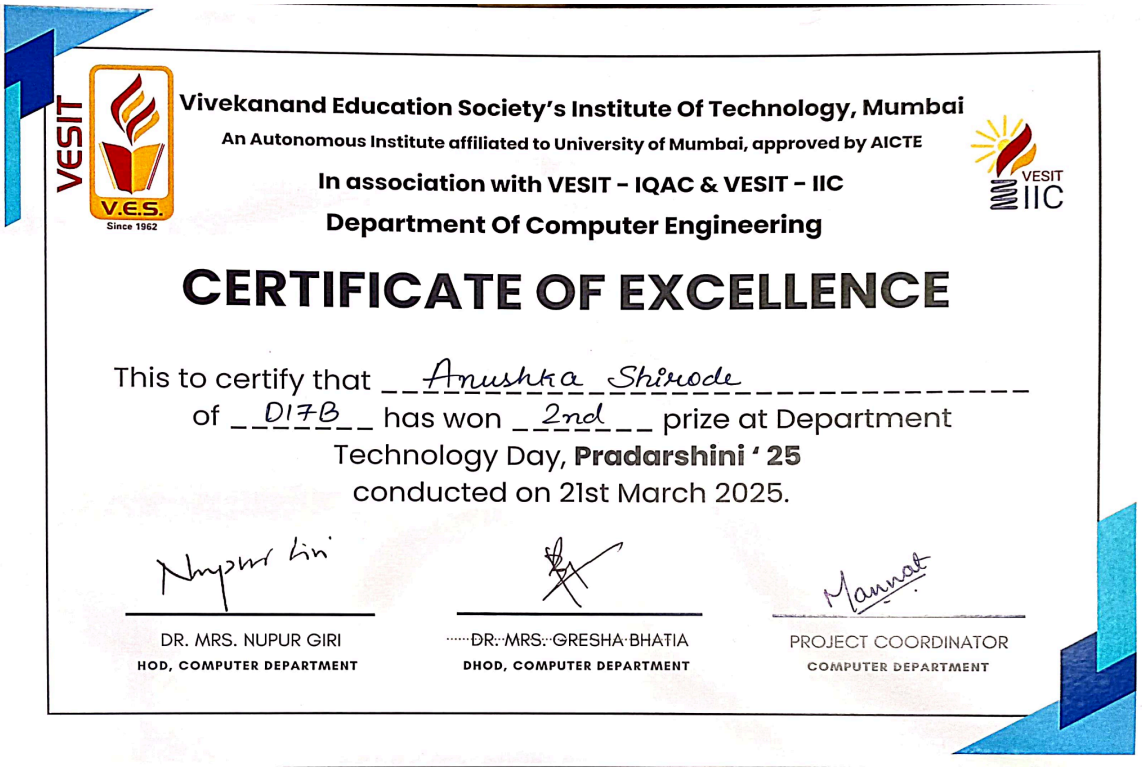
CERTIFICATE OF EXCELLENCE

This to certify that Manasi Sharma
of D17B has won 2nd prize at Department Technology Day, **Pradarshini '25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI
HOD, COMPUTER DEPARTMENT

DR. MRS. GRESHA BHATIA
DHOD, COMPUTER DEPARTMENT

PROJECT COORDINATOR
COMPUTER DEPARTMENT



Vivekanand Education Society's Institute Of Technology, Mumbai

An Autonomous Institute affiliated to University of Mumbai, approved by AICTE

In association with VESIT - IQAC & VESIT - IIC

Department Of Computer Engineering

CERTIFICATE OF EXCELLENCE

This to certify that Anushka Shirode of D17B has won 2nd prize at Department Technology Day, **Pradarshini '25** conducted on 21st March 2025.

DR. MRS. NUPUR GIRI
HOD, COMPUTER DEPARTMENT

DR. MRS. GRESHA BHATIA
DHOD, COMPUTER DEPARTMENT

PROJECT COORDINATOR
COMPUTER DEPARTMENT