VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

Department of Computer Engineering

Project Report on

FarmImpact: Impact of Climate Change on Agriculture in India

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

Academic Year 2024-2025

Dr. Gresha Bhatia

Submitted by

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

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S
INSTITUTE OF TECHNOLOGY

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 has satisfactorily presented the project on "***FarmImpact: Impact of Climate Change on Agriculture in India*** " as a part of the coursework of PROJECT-I for Semester-VII under the guidance of ***Dr. Gresha Bhatia*** in the year 2024-2025.

Date

Internal Examiner

External Examiner

Dr. Gresha Bhatia

Head of the Department
Dr. Mrs. Nupur Giri

Principal
Dr. J. M. Nair

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 Assistant Professor **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 Head of the Computer 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 a 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.

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to:-

Course Outcome	Description of the Course Outcome
CO 1	Do literature survey/industrial visit and identify the problem of the selected project topic.
CO2	Apply basic engineering fundamental in the domain of practical applications FORproblem identification, formulation and solution
CO 3	Attempt & Design a problem solution in a right approach to complex problems
CO 4	Cultivate the habit of working in a team
CO 5	Correlate the theoretical and experimental/simulations results and draw the proper inferences
CO 6	Demonstrate the knowledge, skills and attitudes of a professional engineer & Prepare report as per the standard guidelines.

ABSTRACT

Climate change poses a multifaceted and growing threat to agriculture in India, a sector that not only provides livelihoods for a significant portion of the population but also plays a critical role in the nation's economic stability and food security. As climate patterns shift, the intricate relationship between environmental factors and agricultural productivity becomes increasingly complex. This study aims to analyze how various climatic and environmental variables—such as temperature fluctuations, rainfall patterns, irrigation practices, atmospheric CO_2 concentrations, the frequency of extreme weather events (e.g., droughts, floods, cyclones), pesticide and fertilizer application rates, and solar irradiance components (DNI, DHI, GHI)—influence agricultural yield and production.

The research leverages statistical and machine learning techniques, including ARIMA (AutoRegressive Integrated Moving Average) for time-series forecasting, Ridge regression and Lasso regression for identifying key predictors while addressing multicollinearity, and Generalized Additive Models (GAM) for capturing non-linear relationships between environmental factors and agricultural output. By integrating both historical data and real-time environmental metrics, the study aims to provide a comprehensive analysis of the direct and indirect effects of these variables on crop yields across different regions of India.

Through data-driven insights and predictive modeling, this research highlights the most influential factors affecting agricultural production under changing climate conditions. The results will inform policymakers, farmers, and stakeholders, providing evidence-based recommendations to enhance climate resilience, optimize resource allocation, and improve sustainable farming practices. Ultimately, the findings will contribute to the development of adaptive strategies that address the adverse effects of climate change, safeguarding India's food security and promoting long-term agricultural sustainability.

INDEX

Chapter No.	Title	Page No.
1	Introduction 1.1 Introduction to the project 1.2 Motivation for the project 1.3 Drawbacks of the existing system 1.4 Problem Definition 1.5 Relevance of the Project 1.6 Methodology used	3
2.	Literature Survey 2.1 Survey of Existing Systems a. Abstract of the research paper b. Inference drawn from the paper 2.2 Existing System 2.3 Lacuna in Existing System 2.4 Comparison of existing system and proposed system 2.5 Focus Area	6
3.	Requirement Of Proposed System 3.1 Functional Requirements 3.2 Non-Functional Requirements 3.3 Constraints 3.4 Hardware & Software Requirements 3.5 Techniques utilized 3.6 Algorithms Utilized	15
4.	Proposed Design 4.1 Block diagram 4.2 Modular diagram 4.3 Design of the proposed system: Dataflow, Flowchart, State transition	20
5.	Proposed Results and Discussions 5.1 Results 5.2 Evaluation Parameters 5.3 Reports on sensitivity analysis	26

6.	Plan Of Action For the Next Semester 6.1 Work done 6.2 Plan of action	32
7.	Conclusion	33
8.	References	34
9.	Appendix 9.1 List Of Figures 9.2 List Of Tables 9.3 List Of Abbreviations 9.4 Paper Publications a. Draft of the paper / Paper Published b. Plagiarism report of the paper c. Xerox of project review sheet (1 per student)	36

Chapter 1: Introduction

1.1 Introduction to the Project

FarmImpact is a project that aims to assess and mitigate the impact of climate change on agriculture in India. Given that agriculture is the backbone of the Indian economy—employing over 50% of the population and contributing to 17-18% of the GDP—the effects of climate variability pose significant challenges. The project uses statistical and machine learning models such as ARIMA (AutoRegressive Integrated Moving Average), Ridge Regression, Lasso Regression, and Generalized Additive Models (GAM) to analyze and predict agricultural outcomes based on various climate parameters.

This project focuses on identifying key climate variables, such as temperature fluctuations, rainfall patterns, solar radiation, CO2 levels, and the frequency of extreme weather events (droughts, floods, heatwaves). The goal is to forecast the potential impact of these factors on agricultural production and yield, particularly in vulnerable rainfed areas, which constitute about 60% of India's cultivated land.

By leveraging machine learning, *FarmImpact* aims to deliver insights that can inform the development of adaptive strategies for farmers, policymakers, and the agricultural sector at large. These insights will support the formulation of climate-resilient agricultural practices, better resource allocation, and improved crop management systems, thereby enhancing food security and rural livelihoods.

1.2 Motivation for the Project

The motivation for *FarmImpact* stems from the growing global awareness of climate change and its disproportionate impact on developing countries, especially those with agriculture-centric economies like India. Indian agriculture is highly vulnerable to climate variability, particularly due to the reliance on monsoonal rainfall. Several key factors drive the need for this project:

1. **Rising Temperatures**: Climate projections suggest that by 2050, global temperatures could increase by 1.5°C to 2.0°C, which could drastically reduce crop yields and threaten food security. Certain crops, such as wheat and rice, are particularly sensitive to temperature changes during critical growth stages.
2. **Erratic Rainfall Patterns**: India's rainfed agricultural areas are highly dependent on seasonal rainfall, and any deviations, such as delayed or excessive monsoons, can lead to either droughts or floods. This variability poses a severe risk to crop yields and the livelihoods of farmers.
3. **Frequency of Extreme Weather Events**: The increasing frequency of extreme weather events, including droughts, heatwaves, and floods, has exacerbated the uncertainty in agricultural productivity. These events result in crop failures, reduced soil fertility, and economic instability for rural communities.
4. **Food Security Concerns**: As India's population continues to grow, projected to surpass 1.5 billion by 2030, ensuring a stable food supply becomes critical. Climate change poses a significant risk to achieving this, with potentially severe consequences for national and global food security.
5. **Technological Advancements**: With the advancement of data analytics, machine learning, and predictive modeling, there is an unprecedented opportunity to apply these technologies to agriculture. *FarmImpact* is motivated by the potential to use these tools to predict outcomes, optimize farming practices, and build resilience against climate variability.

1.3 Drawbacks of the Existing System

India's agricultural system, while vast, suffers from several drawbacks that make it ill-equipped to handle the pressures of climate change:

1. **Low Resilience to Climate Variability**: The existing farming infrastructure in India, particularly in rainfed areas, lacks resilience. Small and marginal farmers, who make up over 80% of the agricultural workforce, have limited access to advanced farming techniques, irrigation systems, or climate-resilient crop varieties.
2. **Outdated Agricultural Practices**: A significant portion of Indian agriculture relies on traditional methods, which are not optimized for changing environmental conditions. These methods fail to account for shifts in weather patterns, resulting in inefficient water use, soil degradation, and lower yields.
3. **Inefficient Use of Resources**: Current agricultural systems do not efficiently manage resources such as water, fertilizers, and pesticides. Excessive pesticide use, for example, has led to soil and water contamination, affecting not only crop quality but also public health. This has worsened the long-term sustainability of the sector.
4. **Limited Data-Driven Decision Making**: Indian farmers, particularly in rural areas, often lack access to real-time data and modern forecasting tools that could inform better decision-making. Without reliable predictions on weather conditions, farmers are unable to take preventive measures to protect their crops, leading to avoidable losses.
5. **Fragmented Land Holdings**: A structural issue in Indian agriculture is the fragmentation of land holdings. Small plots reduce economies of scale, making it difficult for individual farmers to adopt advanced technologies or invest in climate-resilient farming practices.
6. **Poor Irrigation Infrastructure**: Although some regions benefit from irrigation, 60% of India's cultivated land is still dependent on monsoonal rainfall. Inadequate irrigation systems further exacerbate the impacts of climate variability, especially in drought-prone areas.

1.4 Problem Definition

The central problem addressed by *FarmImpact* is the increasing uncertainty in agricultural productivity caused by climate change. The project aims to identify and quantify the impact of climate variability (e.g., temperature fluctuations, unpredictable rainfall, and extreme weather events) on agricultural yield and production in India. Specifically, the project addresses the following key issues:

- How do temperature, rainfall, CO_2 emissions, and extreme weather events affect crop productivity?
- What machine learning models are most effective in predicting agricultural yield based on climate data?
- How can the insights from these models be translated into practical, adaptive farming strategies to mitigate the effects of climate change?

The overall objective is to use data-driven approaches to provide actionable insights that can enhance climate resilience in Indian agriculture.

1.5 Relevance of the Project

FarmImpact is highly relevant to several national and global challenges:

1. **Food Security**: With a growing population and shrinking agricultural output due to climate change, ensuring food security has become a pressing concern. This project provides predictive insights that can help farmers and policymakers plan better and mitigate the risks associated with reduced crop yields.
2. **Climate Change Adaptation**: As climate change continues to alter weather patterns and agricultural conditions, the development of adaptive strategies becomes critical. The insights provided by this project are directly applicable to creating climate-resilient agricultural systems, improving both short-term coping mechanisms and long-term resilience.
3. **Sustainable Agriculture**: *FarmImpact* supports sustainable agricultural practices by identifying key stressors on crop productivity. By promoting data-driven decisions, the project aligns with global sustainability goals, such as reducing water consumption, minimizing pesticide use, and promoting environmentally friendly farming techniques.
4. **Economic Stability for Farmers**: Small and marginal farmers, who constitute the majority of the agricultural workforce, are particularly vulnerable to climate shocks. By providing predictive models and adaptive strategies, *FarmImpact* can help reduce the economic vulnerability of farmers, ensuring more stable incomes and better livelihood opportunities.

1.6 Methodology Used

- Agile methodology can be effectively implemented in the FarmImpact project to manage complexity and ensure adaptive progress across its multiple phases. The project can be divided into short, iterative sprints, where each sprint focuses on specific components such as data collection, preprocessing, model development, and result interpretation. Agile's iterative nature allows the team to handle the complexities of integrating multiple data sources (e.g., government datasets, climate platforms) while addressing challenges like missing data or discrepancies through real-time feedback and collaboration. This approach ensures that the team remains flexible, continuously refining processes based on insights gathered during each sprint, and reducing risks associated with delays or unexpected challenges.
- Furthermore, Agile fosters a collaborative environment that allows for continuous stakeholder engagement. For instance, after each sprint, team members can review the model performance (e.g., ARIMA forecasts or regression analyses) and gather feedback to fine-tune the project objectives. This iterative feedback loop ensures that critical aspects, such as the accuracy of climate impact models, are addressed early and improved incrementally. Agile's emphasis on adaptive planning and incremental delivery ensures that the project remains aligned with evolving goals, enabling the team to respond swiftly to changes, deliver value progressively, and optimize outcomes, such as predictive insights for policy recommendations and risk mitigation strategies for farmers.

Chapter 2: Literature Survey

The literature survey provides an in-depth review of existing research, books, articles, and expert interactions relevant to the project. It highlights the knowledge gaps that *FarmImpact* seeks to address while supporting the project's methodology and problem definition.

2.1 Research Papers

1. A Machine Learning Approach to Assess Implications of Climate Risk Factors on Agriculture: The Indian Case

- Abstract: This paper investigates the impact of climate risk factors, such as temperature fluctuations and rainfall variability, on Indian agriculture. It aims to forecast crop yield disruptions and provides strategies to manage agricultural risks.
- Methodology: Random Forest and XGBoost models were used with climate data from CMIP6 and historical agricultural productivity records. The models are evaluated based on accuracy, and cross-validation is applied for robustness.
- Lacuna: The model does not account for socioeconomic factors, such as farmer income or access to technology, which could influence adaptive capacity. Regional variability in climate impacts is also underexplored.

2. Analysis of Weather Parameters Using Machine Learning

- Abstract: This study aims to improve weather forecasting accuracy by analyzing rainfall, temperature, and humidity using machine learning. The improved predictions are intended to help farmers with agricultural planning.
- Methodology: SVM and LSTM models were applied to historical weather datasets from the Indian Meteorological Department, and performance was evaluated using MAE and RMSE.
- Lacuna: The interaction between different weather parameters is not explored, and the analysis is limited to individual effects. Moreover, regional climate variations across India are not sufficiently addressed.

3. Analyzing Trend and Forecasting of Rainfall Changes in India Using Non-Parametric and Machine Learning Approaches

- Abstract: This paper analyzes historical rainfall data to identify long-term trends and forecasts future changes using non-parametric and machine learning techniques.
- Methodology: Mann-Kendall trend analysis and Sen's slope estimator were employed alongside machine learning models like ANN and Decision Trees for rainfall forecasting.
- Lacuna: While rainfall trends are analyzed, the study does not integrate other climatic factors such as temperature or wind speed, which may affect the overall prediction. The practical application of forecasts in agriculture is also not discussed.

4. Analysis of Various Climate Change Parameters in India Using Machine Learning

- Abstract: The study explores the use of machine learning to analyze climate change parameters such as temperature and precipitation patterns in India and their impact on agriculture.
- Methodology: Gradient Boosting and k-NN algorithms were applied to climate data from satellite observations and ground-based sensors. The models were validated using k-fold cross-validation.
- Lacuna: The research does not consider the adaptive strategies farmers could employ to mitigate climate impacts. Additionally, the uncertainty in long-term climate models is not adequately addressed.

5. Impact of Climate Change on Agricultural Production: Issues, Challenges, and Opportunities in Asia

- Abstract: This paper reviews the effects of climate change on agricultural production in Asia, with a focus on key challenges like water scarcity and temperature shifts, and explores adaptation opportunities.
- Methodology: A mixed-method approach was used, including literature review and interviews with agricultural stakeholders in Asia. Climate models and agricultural data from FAO were analyzed to assess vulnerabilities.
- Lacuna: The reliance on secondary data weakens the paper's contribution to empirical research. Additionally, there is limited discussion on how different countries within Asia are responding to these challenges through policy or technology.

6. Chemical Fertilizers and Pesticides in Indian Agriculture

- Abstract: This paper examines the use of chemical fertilizers and pesticides in Indian agriculture and their environmental and health impacts. It also discusses the potential of organic farming as an alternative.
- Methodology: A survey-based study was conducted with farmers in various Indian states to gather data on fertilizer and pesticide usage. Statistical tools were used to analyze trends.
- Lacuna: The paper does not address how transitioning to organic farming could be economically feasible for small-scale farmers. Moreover, there is limited discussion on alternative policy solutions to reduce chemical dependency.

7. An Analysis of Climate Change Based on Machine Learning and an Endoreversible Model

- Abstract: This study combines machine learning with an endoreversible thermodynamic model to predict the impact of climate change on global weather patterns, particularly focusing on agriculture.
- Methodology: Hybrid models combining machine learning with thermodynamic principles were developed and trained on global climate datasets. Model accuracy was compared against standard climate prediction tools.
- Lacuna: While the model is innovative, its practical applications in agriculture are not fully explored. The hybrid approach has not been validated for regional predictions, making it less applicable in localized contexts.

8. Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review

- Abstract: This review highlights the key impacts of climate change on agriculture and explores various mitigation strategies, such as agroforestry and sustainable farming practices.
- Methodology: The paper conducts a meta-analysis of existing research on climate change impacts and mitigation strategies, drawing on global case studies and expert opinions.
- Lacuna: The review is largely qualitative, lacking quantitative data to support its conclusions. Additionally, the potential economic implications of the proposed mitigation strategies for farmers are not sufficiently addressed.

9. Influence of Climate Change on Agricultural Sustainability in India: A State-Wise Panel Data Analysis

- Abstract: This study conducts a state-wise panel data analysis of the effect of climate change on agricultural sustainability in India, focusing on variations in temperature and rainfall.
- Methodology: A fixed-effects regression model was used on a panel dataset of Indian states over a 20-year period. The analysis incorporates climate variables such as temperature and precipitation, alongside agricultural output data.

- Lacuna: The study does not consider the role of technological innovations in enhancing agricultural sustainability. Additionally, the analysis could benefit from a more granular exploration of district-level impacts.

10. Climate Change and Agriculture in India: Report by NMSKCC

- Abstract: This report evaluates the current and projected impacts of climate change on Indian agriculture, supported by data from the National Mission on Strategic Knowledge for Climate Change (NMSKCC).
- Methodology: The report uses climate models and agricultural yield data to assess the impact of temperature rise and monsoon variability on key crops. Policy recommendations are provided to enhance adaptive capacity.
- Lacuna: The report offers a broad overview but lacks detailed implementation strategies for the proposed policies. Moreover, there is minimal discussion on how smallholder farmers can be supported in adapting to these changes.

11. Global Warming and Agriculture: Impact Estimates by Country

- Abstract: This book provides country-specific estimates of the impact of global warming on agricultural productivity. The analysis highlights variations in vulnerability across different regions of the world.
- Methodology: The author uses economic modeling techniques and climate projections from IPCC reports to estimate the impact of global warming on agricultural output at the country level.
- Lacuna: The book focuses primarily on large-scale trends and does not address micro-level interventions that could mitigate the impacts of global warming at the farm level. Moreover, more recent climate data is not incorporated.

12. Indian Agriculture Under Climate Change: The Competing Effect of Temperature and Rainfall Anomalies

- Abstract: This paper explores the competing effects of temperature rise and rainfall anomalies on Indian agricultural productivity, particularly for staple crops like rice and wheat.
- Methodology: The authors employ econometric models to analyze the relationship between temperature anomalies, rainfall, and crop yields. Data from government agricultural reports and weather stations were used.
- Lacuna: The study does not consider potential adaptation strategies such as crop diversification or water conservation techniques. Additionally, the focus is limited to a few staple crops, leaving out other important agricultural sectors.

13. Planet-Friendly Agriculture: Farming for People and the Planet

- Abstract: This paper discusses sustainable agricultural practices that are beneficial both to human societies and the environment, with a focus on reducing greenhouse gas emissions from farming activities.
- Methodology: The authors analyze various sustainable farming techniques, such as no-till farming and organic farming, and assess their environmental and economic benefits. Data from field experiments in different regions is used.
- Lacuna: The paper does not address the scalability of these sustainable practices for large-scale agriculture. Additionally, the economic feasibility of transitioning to such methods for smallholder farmers is not explored.

14. Impact of Climate Change on Agriculture and Indian Economy: A Quantitative Research Perspective from 1980 to 2016

- Abstract: This study quantitatively assesses the impact of climate change on Indian agriculture and its broader effects on the national economy over a 36-year period.
- Methodology: Using time-series data and regression analysis, the authors evaluate the relationship between key climate variables (temperature, rainfall) and agricultural GDP. Data from government economic reports was used.
- Lacuna: The paper does not explore the impact of non-climatic factors, such as policy interventions or technological advances, on the agricultural economy. Furthermore, adaptation measures are not discussed in detail.

15. Climate Change, Agricultural Production, and Poverty in India

- Abstract: This chapter examines the links between climate change, agricultural productivity, and poverty in India, arguing that the adverse effects of climate change disproportionately affect the poor.
- Methodology: A mixed-methods approach was employed, combining econometric analysis of agricultural output data with qualitative case studies on poverty in rural India. The study draws on data from national economic and agricultural surveys.
- Lacuna: The analysis does not consider the role of government interventions, such as subsidies or agricultural insurance, in mitigating the impacts of climate change on the poor. Additionally, the model's predictive power is not thoroughly tested.

16. Climate Change and Rice Production in India: A Statistical Analysis

- Abstract: This paper statistically analyzes the impact of climate change on rice production in India, focusing on changing monsoon patterns and rising temperatures.
- Methodology: The authors used multiple linear regression to analyze the impact of climate variables on rice yield, based on data from the Indian Ministry of Agriculture. Climate data was sourced from meteorological stations.
- Lacuna: The paper is limited to rice and does not consider other crops that might be more resilient or susceptible to climate change. Moreover, adaptation strategies, such as altering planting dates or using drought-resistant rice varieties, are not explored.

17. Building Climate Resilience in Indian Agriculture: A Policy Perspective

- Abstract: This policy paper discusses strategies to enhance climate resilience in Indian agriculture, focusing on technological innovations, water conservation, and climate-smart agricultural practices.
- Methodology: The paper synthesizes data from government policy reports, international case studies, and expert interviews to propose a comprehensive policy framework for building agricultural resilience.
- Lacuna: The paper provides high-level policy recommendations but lacks specific, actionable steps for implementation at the grassroots level. Additionally, the role of private sector involvement in promoting climate-smart agriculture is underexplored.

18. Agricultural Adaptation to Climate Change: Lessons from Global Case Studies

- Abstract: This book reviews case studies from around the world where agricultural practices have successfully adapted to the challenges posed by climate change. The focus is on technological innovations and sustainable practices.
- Methodology: Each chapter presents a case study that combines field data, interviews with farmers, and government policy assessments. The cases span various regions, including Africa, Asia, and Latin America.

- Lacuna: The book does not provide a unified framework for scaling these adaptation strategies across different regions. Moreover, there is a lack of quantitative analysis on the long-term economic impacts of these adaptations.

19. A Statistical Review of Climate Change Impact on Crop Yield in India

- Abstract: This paper provides a statistical review of how climate change has affected crop yields in India over the last three decades, with a focus on temperature rise and changing rainfall patterns.
- Methodology: The study uses descriptive statistics and time-series analysis to examine the impact of temperature and rainfall changes on crop yields. Data was collected from Indian agricultural surveys and climate records.
- Lacuna: The study does not offer any predictive modeling or future projections of how ongoing climate changes will continue to affect crop yields. Additionally, the role of agricultural technologies in mitigating these impacts is not discussed.

20. Sustainable Water Management Practices for Agriculture in the Face of Climate Change

- Abstract: This paper explores sustainable water management practices that can help farmers cope with water scarcity exacerbated by climate change, with a focus on rainwater harvesting, drip irrigation, and efficient water use techniques.
- Methodology: The study uses data from field trials and surveys to assess the effectiveness of various water management practices. A cost-benefit analysis is also provided to evaluate the economic viability of these practices.
- Lacuna: The paper does not address the potential barriers to widespread adoption of these practices, such as the cost of implementation for smallholder farmers. Moreover, it lacks a discussion on how government policies could support water management at a larger scale.

2.2 Existing System

The current agricultural forecasting and climate analysis systems integrate various machine learning and statistical models to predict crop yields and evaluate climate risks. These models primarily rely on climate variables like temperature, rainfall, and CO2 levels, as well as agricultural data such as yield, production, and irrigation levels. The commonly employed models and techniques include ARIMA for time series forecasting, Random Forest and SVM for classification, and Neural Networks for capturing complex nonlinear relationships. Several new machine learning algorithms have also been introduced to enhance forecasting capabilities. Some of the key models and techniques in use are:

1. **ARIMA (AutoRegressive Integrated Moving Average)**: A time series forecasting model widely applied to predict climate variables such as rainfall and temperature. However, ARIMA struggles with nonlinear relationships and is best suited for linear trends in climate data.
2. **Random Forest**: A popular classification and regression algorithm, Random Forest is effective for identifying high-risk regions prone to crop failure. Its ensemble nature ensures high accuracy and robustness, but it may struggle with interpretability in large datasets.
3. **Support Vector Machine (SVM)**: Often used for classification tasks like climate risk assessment, SVM is effective in handling large feature spaces. However, it requires careful tuning of parameters and struggles with imbalanced datasets.
4. **Neural Networks (NN)**: Deep learning models, particularly Long Short-Term Memory (LSTM) networks, are used to capture long-term dependencies in climate data. They are useful for modeling nonlinear and temporal relationships, though they demand significant computational resources.

5. **Gradient Boosting Machines (GBM)**: Algorithms like XGBoost are becoming increasingly popular for their ability to handle complex interactions between climate variables and agricultural outputs. These models often outperform simpler techniques but are prone to overfitting.
6. **K-Nearest Neighbors (k-NN)**: A non-parametric method applied to spatial climate data for classification and regression tasks. Its performance heavily depends on the choice of 'k' and the distance metric used.
7. **Long Short-Term Memory (LSTM)**: A variant of Recurrent Neural Networks (RNNs) that excels in time series predictions, LSTMs can capture long-term dependencies in data, making them ideal for predicting seasonal agricultural outcomes based on historical climate data.
8. **Generalized Additive Models (GAM)**: Used to model nonlinear relationships between climate variables and crop yields. GAMs are interpretable and effective in isolating the effects of specific climate factors, but they may not capture interactions between variables as well as more complex models.
9. **Mann-Kendall Trend Analysis**: A non-parametric test used to detect trends in climate variables, such as rainfall, over time. It is often combined with machine learning models to improve the understanding of climate variability.
10. **GIS (Geographic Information Systems)**: GIS tools are occasionally integrated for spatial mapping of climate data, though they are not widely adopted in national agricultural forecasting systems. GIS aids in understanding regional climate patterns and their effects on agriculture.

Despite the advances in these techniques, existing systems face several challenges:

- **ARIMA** is limited by its inability to model nonlinear relationships in complex climate data.
- **Random Forest** and **SVM** models, while accurate, may struggle to generalize to new regions without extensive parameter tuning.
- **Neural Networks** like **LSTM** are resource-intensive, requiring large datasets and significant computational power, which limits their use in real-time systems.

In addition, spatial analysis through **GIS** is underutilized in most national models, and extreme weather events like floods and droughts are not consistently accounted for. These limitations highlight the need for more advanced, real-time, and region-specific forecasting systems that can better inform agricultural practices across India's diverse landscape.

2.3 Lacuna in Existing System

While existing systems provide a strong foundation for agricultural forecasting, several critical gaps remain, hindering their effectiveness in addressing the full scope of climate-related agricultural challenges.

1. No Real-Time Data Integration: Many existing models depend on historical datasets, which don't incorporate real-time updates. This reduces the ability to make timely, accurate predictions, especially when immediate action is needed for events like unexpected weather changes or pest outbreaks.
2. Lack of Multi-Scale Analysis: Current systems often produce generalized national-level forecasts, overlooking regional climate variations. State-level and district-level differences in weather patterns and agricultural practices are crucial for tailoring forecasts to specific needs.
3. Underutilization of Advanced Models: While basic machine learning models like ARIMA or Random Forest are frequently employed, few systems leverage advanced techniques such as deep learning or hybrid models that can capture more complex and dynamic relationships between climate factors and agriculture.
4. Limited Extreme Weather Forecasting: Most models focus on regular climate variables like rainfall and temperature, but do not account adequately for extreme events like floods, droughts, or heatwaves, which are becoming more frequent and disruptive.

5. Farmer-Centric Usability Gaps: The technical nature of many forecasting systems makes it difficult for farmers to interpret the data or translate it into actionable strategies. There's often no user-friendly interface that provides practical advice, limiting the systems' real-world applicability.
6. Overemphasis on Yield Prediction: While crop yield prediction is a central focus, other critical aspects like soil health, pest resistance, and sustainability are underexplored. This one-dimensional focus neglects the broader factors that influence long-term agricultural success.
7. Data Quality and Availability Issues: Accurate, localized data is often either unavailable or inconsistent, particularly for remote regions. This leads to gaps in the forecasting model's performance at a granular level, especially in areas with limited climate stations.
8. Limited Policy Integration: Many models operate in academic or research settings without translating into actionable policies. This disconnect reduces the impact these tools can have on shaping real-world agricultural practices and government decision-making.
9. High Computational and Resource Demands: Advanced machine learning models, especially deep learning or neural networks, require significant computational resources. This poses a challenge for small-scale farmers or organizations in rural areas that do not have access to high-end technology.

2.4 Comparison of existing system and proposed system

Existing Systems:

Current agricultural forecasting systems focus primarily on either broad national-level data or limited region-specific data for a few crop types, like foodgrains or oilseeds. These models typically use machine learning techniques like ARIMA for time-series forecasting and Random Forest or SVM for classification. However, they often lack the granularity needed to address India's diverse agricultural landscape.

- Climate Variability at State Level: Most models rely on national averages, overlooking the significant climatic differences between states, which can affect regional agriculture.
- Limited Spatial Analysis: Few models integrate GIS-based spatial analysis, making it difficult to capture regional differences in climate and agriculture.
- Narrow Crop Focus: Existing systems mostly focus on a limited set of crops, excluding many others like pulses or horticultural crops.
- Underuse of Advanced Models: There's limited adoption of more sophisticated models like deep learning, which could better capture complex relationships between climate factors and agricultural outcomes.

As a result, these systems often miss out on providing more localized predictions, accounting for extreme weather events, and delivering actionable insights for diverse regions across India.

Proposed System:

The proposed system aims to address these limitations by incorporating state-wise agricultural data and a broader range of climate variables, offering more accurate and region-specific predictions.

- State-Level Data Integration: By using state-level data on production, yield, rainfall, and temperature, the system provides more tailored forecasts that reflect India's diverse climate and agriculture across regions.
- Expanded Climate Variables: In addition to basic parameters like rainfall and temperature, the system will include CO_2 emissions, extreme weather events, and solar radiation data (GHI, DNI, DHI) to provide a more comprehensive understanding of climate impacts.
- Advanced Forecasting Models: Models such as ARIMA, Ridge, Lasso, and GAM will be employed for more accurate predictions by handling complex interactions between climate variables and agricultural outputs.
- Focus on Crop Diversity: The system will cover a wider range of crops beyond just foodgrains and oilseeds, including pulses, cash crops, and horticulture, making the predictions useful for diverse agricultural sectors.

- Extreme Weather Forecasting: By incorporating extreme climate events, the system will be better equipped to provide early warnings for disasters like floods or droughts, helping farmers and policymakers take proactive steps.

Table I: Comparison of existing system and proposed system

Aspect	**Existing Systems**	**Proposed System**
Real-Time Data	Largely reliant on historical data, limiting responsiveness to real-time changes.	Incorporates real-time climate and agricultural data for more dynamic and timely forecasts.
Scale of Analysis	Primarily focused on national-level trends, with limited regional and state-level insights.	Supports multi-scale analysis, including region-specific, state, and district-level forecasting.
Advanced Models	Underutilization of deep learning and hybrid models, despite their potential for more accurate predictions.	Utilizes advanced machine learning techniques, including deep learning and hybrid models, to capture complex relationships.
Extreme Weather Forecasting	Inadequate in addressing extreme weather events like floods and droughts, which are increasingly frequent.	Emphasizes prediction and risk assessment for extreme weather events, offering early warnings for disaster management and mitigation strategies.
Farmer-Centric Usability	Lacks user-friendly interfaces, making it difficult for farmers to interpret and act on the predictions.	Proposes a farmer-friendly interface with actionable insights, enabling farmers to make informed decisions.
Focus on Yield Prediction	Primarily centered around crop yield prediction, neglecting other important factors like soil health and pest management.	Takes a holistic approach, incorporating other agricultural factors such as soil quality, pest resistance, and water use efficiency.
Data Availability	Suffers from limited availability and inconsistent quality of localized data, particularly in remote regions.	Employs data aggregation techniques and partnerships with local organizations to ensure high-quality, localized data collection and availability.
Policy Integration	Limited integration into government policy frameworks, reducing real-world impact.	Designed for closer collaboration with policymakers to influence agricultural policy and provide decision support.

Computational Requirements	Requires significant computational power, posing challenges for small-scale farmers in rural areas.	Optimized to run on lower computational resources, with potential integration into cloud-based or mobile platforms accessible to small-scale users.

This approach offers more localized, accurate, and actionable insights, better suited for addressing regional agricultural challenges across India.

2.5 Focus Area

The primary focus of this project is to develop an integrated system that can predict agricultural productivity based on a wide range of climatic factors at both the national and state levels. By incorporating state-wise data and using a combination of machine learning and spatial analysis models, the project aims to deliver detailed insights into how regional climate variability affects yield and production. Special attention will be paid to improving prediction accuracy and addressing current challenges in forecasting agricultural trends under changing climate conditions.

Chapter 3: Requirements for the Proposed System

The requirements for *FarmImpact* comprise functional, non-functional, and technical specifications to ensure the system's success in addressing climate impacts on agriculture.

3.1 Functional Requirements

- **Predictive Models**: The system should use ARIMA, Ridge, and Lasso regression models to predict the impact of climate variables on agricultural yield.
- **Data Collection and Analysis**: The system must gather climate and agricultural data, process it for missing values, and apply machine learning models to generate predictions.
- **Visualization**: The system should generate graphs and charts illustrating the trends in agricultural productivity based on climate forecasts.

3.2 Non-Functional Requirements

- **Accuracy**: The system should provide high accuracy in its predictions, with error margins within acceptable ranges for agricultural forecasting.
- **Scalability**: The system must handle large datasets from diverse sources, such as meteorological data and agricultural records.
- **Usability**: The interface should be user-friendly, providing farmers, policymakers, and researchers with easy access to predictive insights.

3.3 Constraints

- **Data Availability**: The accuracy of predictions is dependent on the availability and quality of climate and agricultural data.

3.4 Hardware & Software Requirements

- **Hardware**:
  - Processor (CPU): Intel Core i5
  - RAM : 16 GB
- **Software**:
  1. **Python**
     a. **Pandas**: For handling and preprocessing time series data efficiently.
     b. **NumPy**: For numerical operations and matrix computations in regression models.
     c. **Statsmodels**: Specifically for fitting ARIMA models and time series analysis.
     d. **Scikit-learn**: For implementing Ridge and Lasso regression models to predict production and yield.
     e. **Matplotlib/Seaborn**: For visualizing forecasted trends and model performance.
  2. **Jupyter Notebooks**: For interactive development, allowing step-by-step analysis and visualization.

3.5 Techniques Utilized

- **Data Preparation:**
  - Load the dataset from a CSV file using pandas.
  - Several time series features are extracted: CO_2, DNI, DHI, GHI, and agricultural features like Rain, Production(MillionTonnes), etc.
  - Missing values are handled using dropna() and replaced where necessary.
- **ARIMA Models for Forecasting:**
  - ARIMA models are used for time series forecasting on various features like CO_2, DNI, DHI, GHI, and others, with ($p=2$, $d=1$, $q=2$) as the parameter set.
  - Forecasts are added back to the dataset for specific years (e.g., back-predicting CO_2 levels from 1950 to 1960, predicting values for GHI/DNI/DHI until 2023, etc.).
- **Prediction for Features in 2023:**
  - ARIMA models are also used to predict other features such as Fertilizer Use, Pesticide Use, MeanTemperature, etc., up to 2023. The forecasted values are inserted into the original dataset.
- **Regression Models:**
  - Ridge and Lasso regression models are applied to predict agricultural production (Production(MillionTonnes)) and yield (Yield(KgPerHectare)).
  - The dataset is split into training and testing sets.
  - Both models' performance is evaluated using Mean Squared Error (MSE) and R^2 score.
  - Visualizations for Ridge and Lasso predictions versus actual values are generated.
- **Scaling and Further Regression:**
  - Features are scaled using StandardScaler to normalize the data for the regression models.
  - Ridge and Lasso regression are then applied to scaled data for predicting Production(MillionTonnes) and Yield(KgPerHectare).

3.6 Algorithms Utilized

- **ARIMA (AutoRegressive Integrated Moving Average):**
  ARIMA is a time-series forecasting model that captures the temporal dependencies in data. It uses three components: autoregressive (AR) terms, differencing (I for "integrated") to make the series stationary, and moving averages (MA) to capture past errors. ARIMA is effective in modeling historical trends and forecasting future values based on the patterns in past data. In the context of your agricultural project, ARIMA is used to forecast variables like CO_2, rainfall, and yield by analyzing the past 70+ years of data (1950–2023) and projecting future values. By doing this, you can anticipate trends in production or yield influenced by long-term environmental changes.

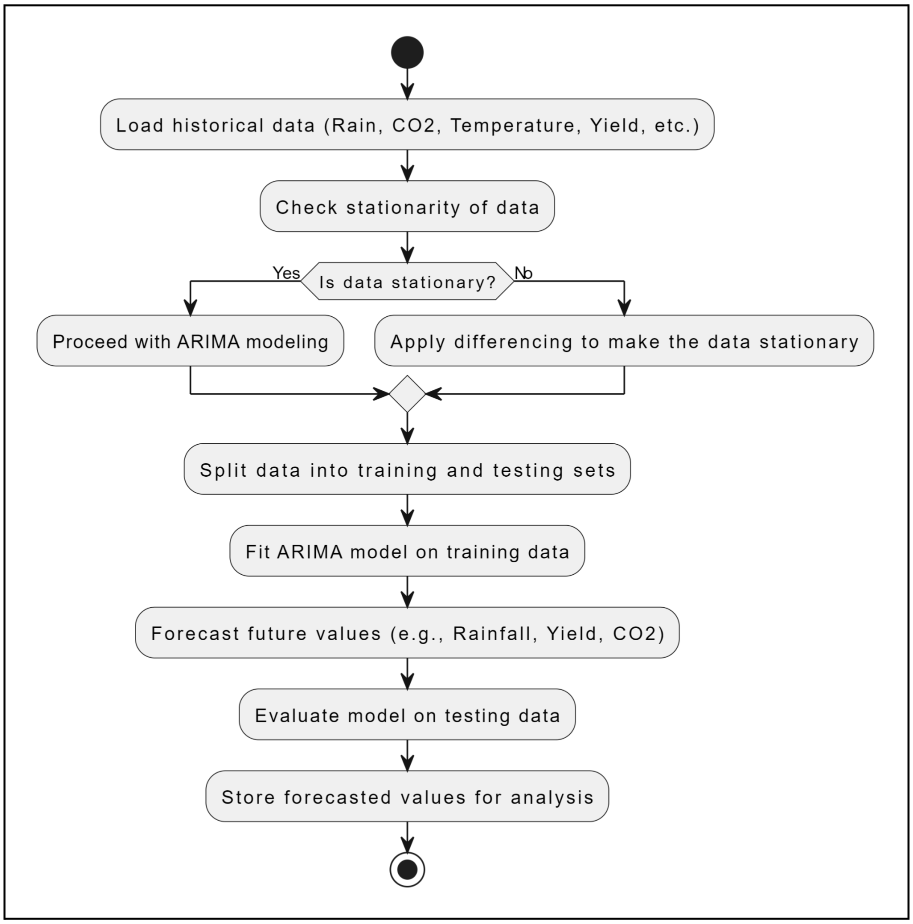

Fig 1. ARIMA Methodology Flowchart

- **GAM (Generalized Additive Model):**
  GAM is a flexible regression model that allows for non-linear relationships between the dependent and independent variables. It fits smooth functions to the data, making it particularly useful in capturing complex relationships between features. In agriculture, where variables like rainfall or temperature can have non-linear effects on crop yield, GAMs provide a more nuanced understanding of how these factors interact with yield. In your project, GAM is likely applied to investigate how features like rainfall, mean temperature, and CO2 impact production and yield over time. It helps identify thresholds or non-linear effects that simpler linear models might miss, allowing you to capture more accurate trends.

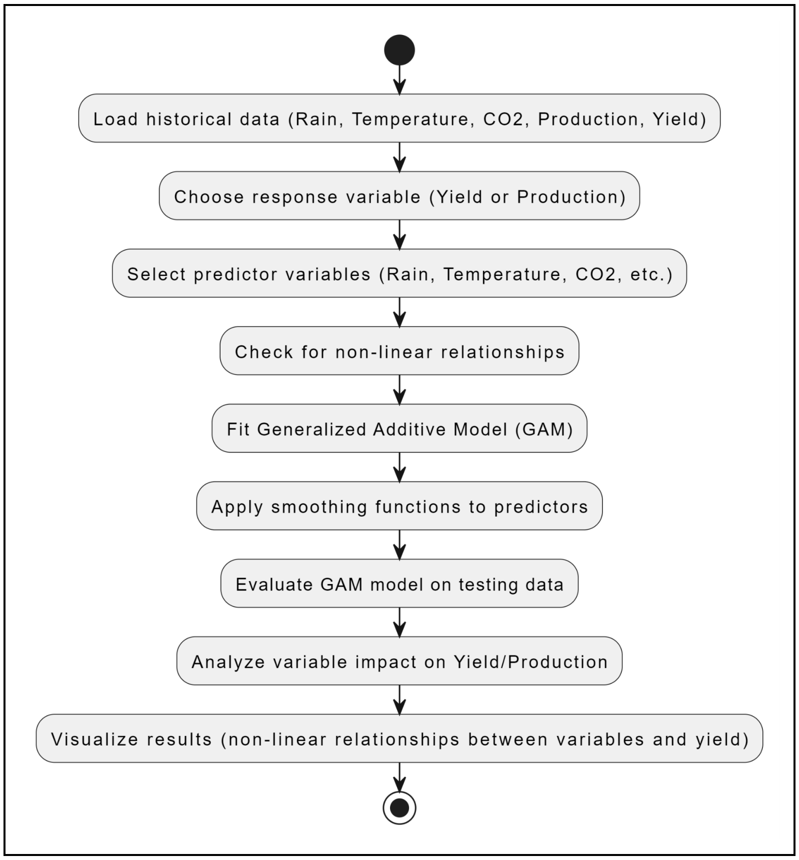

Fig 2. GAM Methodology Flowchart

- **Ridge and Lasso Regression:**
  Ridge and Lasso are linear regression models with regularization techniques that prevent overfitting by penalizing the size of coefficients. Ridge shrinks all coefficients uniformly, while Lasso can shrink some coefficients to zero, effectively selecting important features. In the agricultural context, these models are useful for predicting production and yield based on multiple features like rainfall, temperature, and fertilizer use. By applying Ridge or Lasso, you're able to create robust models that account for multicollinearity between predictors, and potentially identify which factors are most influential in determining crop yield. This is especially important in your project where numerous environmental variables are being considered for prediction.

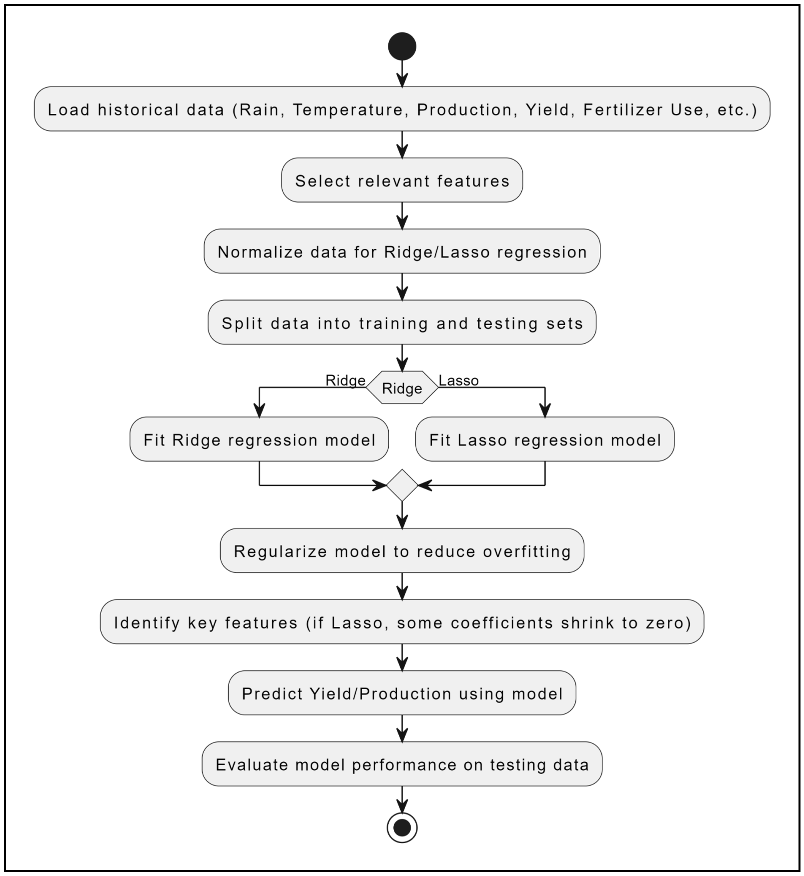

Fig 3. Ridge-Lasso Regression Methodology Flowchart

Chapter 4: Proposed Design

4.1 Block diagram

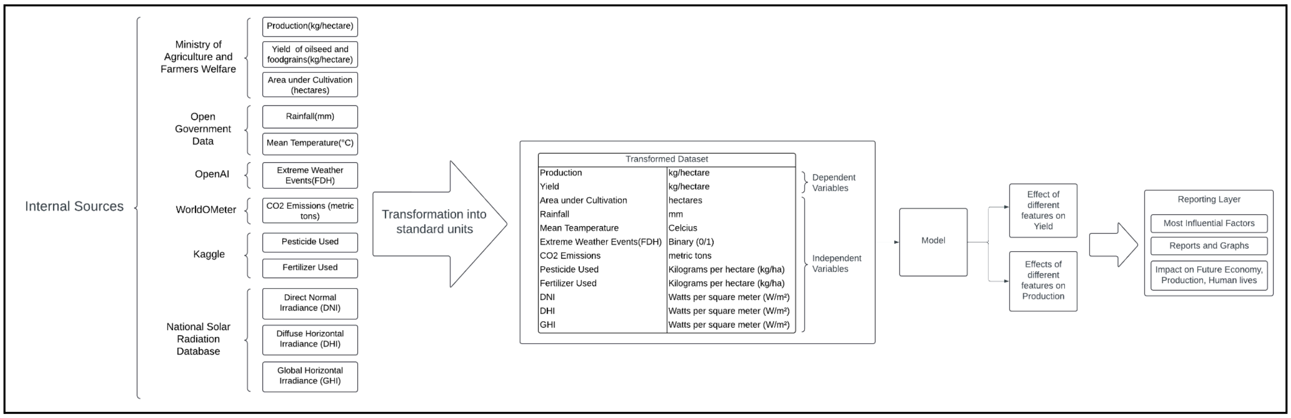

Fig 4. Block Diagram

Fig. 4 shows the methodology used in the project to analyze the impact of various climatic factors on agricultural yield and production.

The dataset used in this project aggregates data from various reliable sources, including Open Government Data from the Ministry of Agriculture & Farmers' Welfare, WorldOMeter, Our World in Data, Kaggle, FAOStat, OpenAI, and the National Solar Radiation Database (NSRDB) by NREL. The dataset spans several decades, capturing yearly data on key agricultural and environmental variables. The features include:

- **Rainfall** (from Open Government Data) and **Mean Temperature** records, which represent essential climatic conditions affecting agriculture.
- **Area (Million Hectares)**, **Production (Million Tonnes)**, and **Yield (Kg per Hectare)** that represent key agricultural productivity indicators.
- **Area Under Irrigation** and **Irrigated Area**, both crucial for understanding water resource usage in agriculture.
- **CO2 emissions** (from Our World in Data), providing insight into environmental changes and their impact on crop production.
- **AQI (Air Quality Index)** and **Extreme Weather Events (FDH)**, capturing air quality and weather anomalies.
- **Pesticide Use** and **Fertilizer Use**, which influence agricultural output and sustainability.
- Solar radiation metrics such as **DNI (Direct Normal Irradiance)**, **DHI (Diffuse Horizontal Irradiance)**, and **GHI (Global Horizontal Irradiance)** from NSRDB, which are crucial for analyzing solar energy potential and its impact on farming practices.

The provided code performs a comprehensive analysis of time series data using ARIMA models for forecasting and regression analysis (Ridge and Lasso). The data is loaded into a pandas DataFrame, with specific columns like CO2, DNI, DHI, GHI extracted for time series modeling. The ARIMA model, with parameters (p=2, d=1, q=2), is applied to these series to forecast historical values for CO2 (1950-1960) and future values for the other indicators (1950-2000). This is repeated for columns like Fertilizer Use, Pesticide Use, and Mean Temperature. The predicted values are added back to the dataset.

Following time series forecasting, the code proceeds with regression analysis for features affecting agricultural production and yield. After splitting the dataset into training and testing sets, Ridge and Lasso regression models are fitted. The models help predict the production (Million Tonnes) and yield (Kg per Hectare), using features like Rain, Mean Temperature, and CO2 levels. The final section extracts the feature importance from both regression techniques, highlighting which variables are more influential on the target variables. Coefficients from both Ridge and Lasso regressions are presented, showcasing the strength of the relationships between input features and agricultural outcomes.

1. **Data Loading and Preprocessing**: Load the dataset, handle missing values using dropna(), and prepare time series data.
2. **ARIMA Modeling**: Implement ARIMA models to forecast past and future data for CO2, DNI, DHI, and GHI, as well as other environmental variables like temperature and rainfall.
3. **Data Integration**: Update the original dataset with the predicted values for missing or future years.
4. **Regression Analysis**: Apply Ridge and Lasso regression on agricultural data, evaluating the impact of environmental factors (features) on agricultural production and yield.
5. **Feature Importance**: Analyze the coefficients from the regression models to understand the influence of each feature on production and yield.
6. **Visualization**: The final stage visualizes both actual and predicted values using matplotlib, providing insights into the time series forecast and regression model predictions.

4.2 Modular Diagram

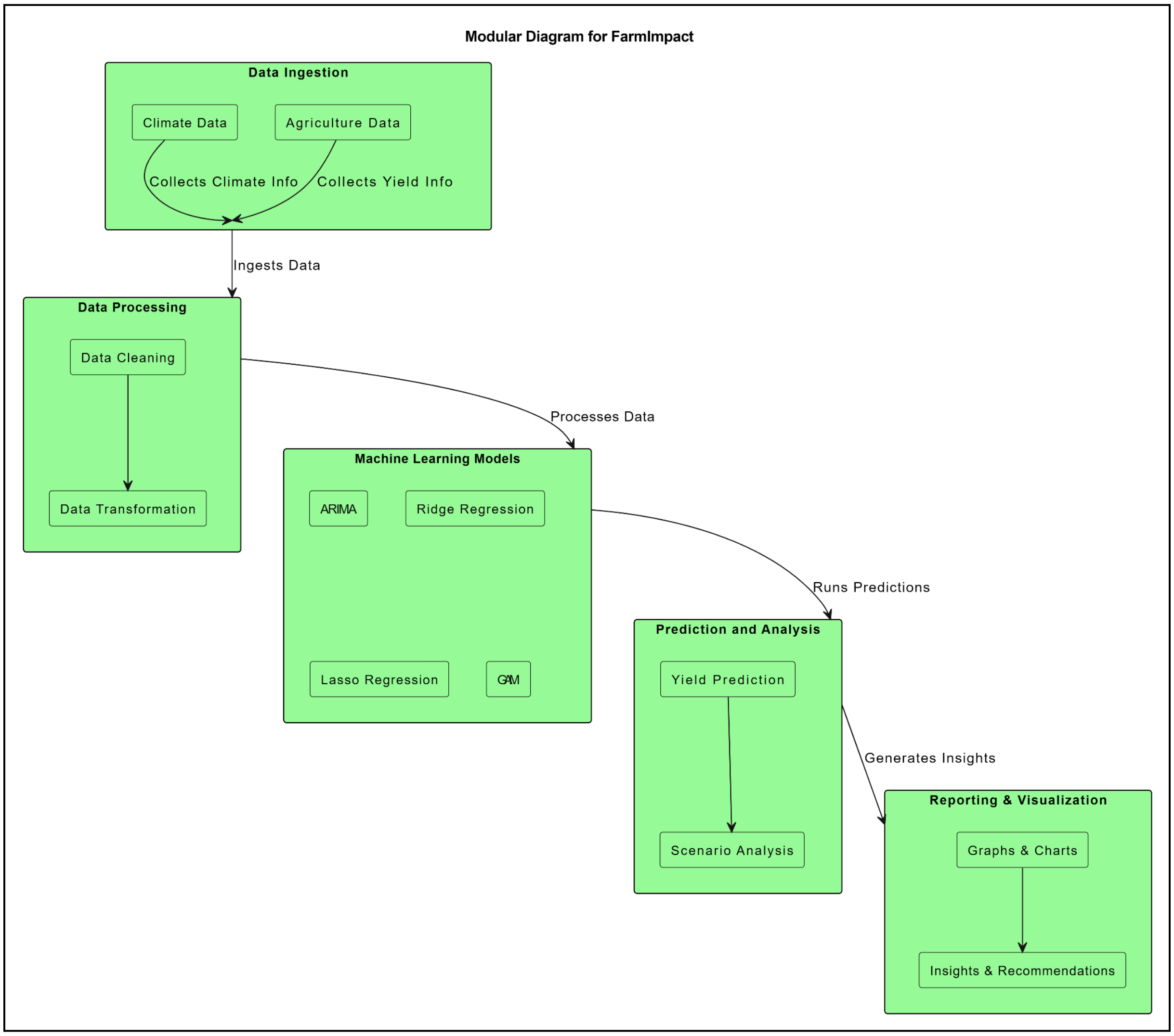

Fig 5. Modular Diagram

Fig. 5 outlines the 'FarmImpact' system:

1. **Data Ingestion**: Collects climate and agriculture data related to yields and environmental conditions.
2. **Data Processing**: Cleans and transforms the data to ensure it's ready for analysis.
3. **Machine Learning Models**: Uses models like ARIMA, Ridge, and Lasso Regression to process data and make predictions.
4. **Prediction and Analysis**: Focuses on yield prediction and scenario analysis based on model outputs.
5. **Reporting & Visualization**: Generates graphs, charts, insights, and recommendations to help users make informed decisions.

4.3 Design of the proposed system

- **Data Flow Diagram**

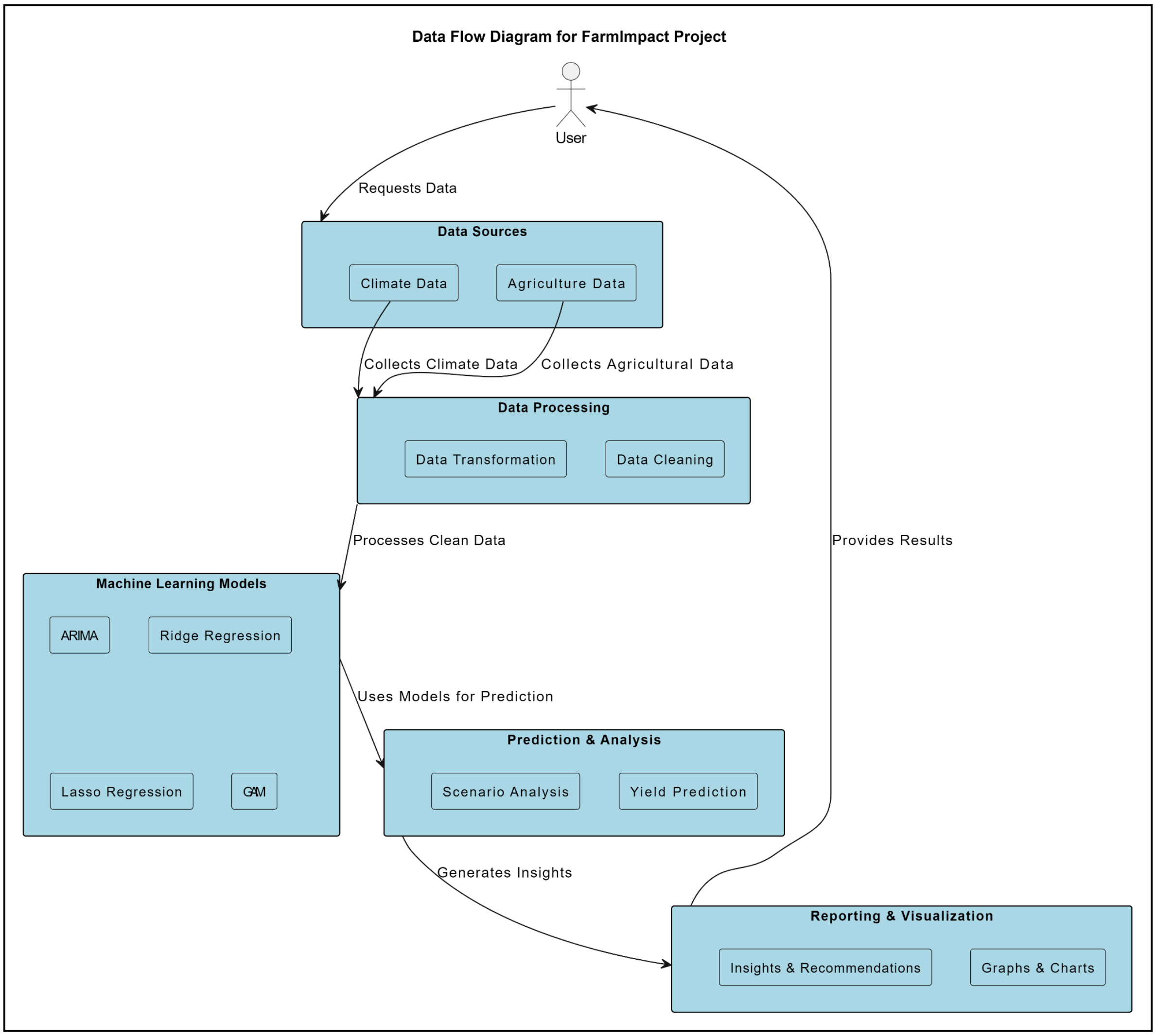

Fig 6. Data Flow Diagram

Fig. 6 shows the data flow for the 'FarmImpact' project,

1. **User Interaction**: The process begins with the user requesting data, which triggers the collection of **Climate Data** and **Agriculture Data**.
2. **Data Sources**: These data sources gather relevant information, including weather conditions and agricultural metrics like crop yields.
3. **Data Processing**: The raw data undergoes **Data Cleaning** to remove inconsistencies and **Data Transformation** to prepare it for model inputs.
4. **Machine Learning Models**: Processed data is then used by various models like **ARIMA**, **Ridge Regression**, **Lasso Regression**, and **Generalized Additive Models (GAM)** to make predictions.
5. **Prediction & Analysis**: These models focus on **Yield Prediction** and **Scenario Analysis** to foresee crop outputs and assess different agricultural strategies.

6. **Reporting & Visualization**: The predictions are transformed into meaningful insights, visualized through **Graphs & Charts**, and delivered as **Insights & Recommendations** to the user for decision-making.

- **Flowchart Diagram**

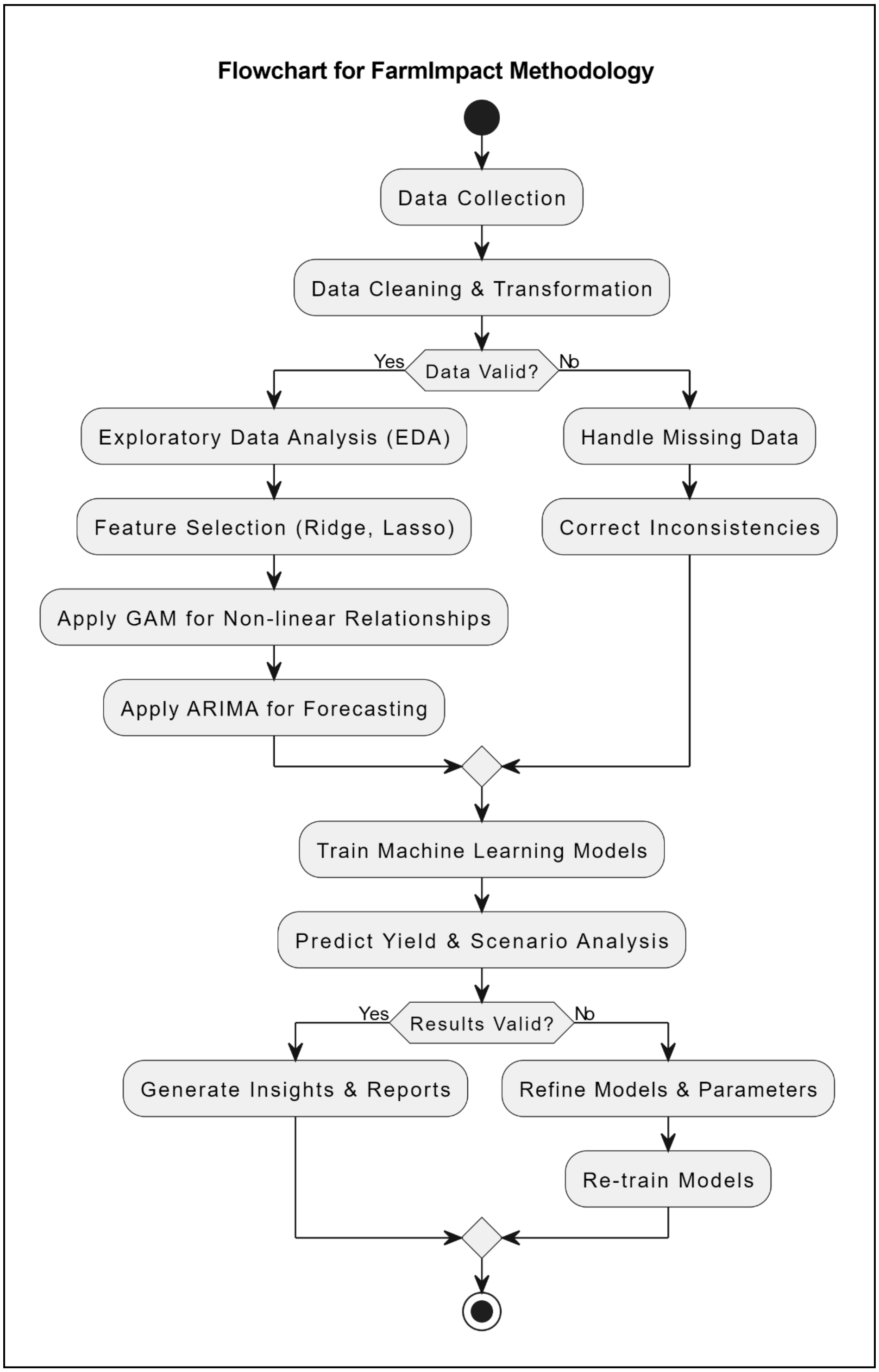

Fig 7. Flowchart Diagram

The flowchart in Fig. 7 outlines the 'FarmImpact' methodology. This flowchart outlines the methodology for a system called "FarmImpact." The process begins with data collection, followed by data cleaning and transformation. If the data is valid, it proceeds to exploratory data analysis (EDA). If not, missing data is handled, and inconsistencies are corrected. Feature selection techniques such as Ridge and Lasso are applied, followed by the application of Generalized Additive Models (GAM) for non-linear relationships. ARIMA is then used for forecasting. Machine learning models are trained, and yield predictions and scenario analyses are performed. If the results are valid, insights and reports are generated. If not, the models and parameters are refined, and the models are retrained, looping back to revalidate the results.

- **State Transition Diagram**

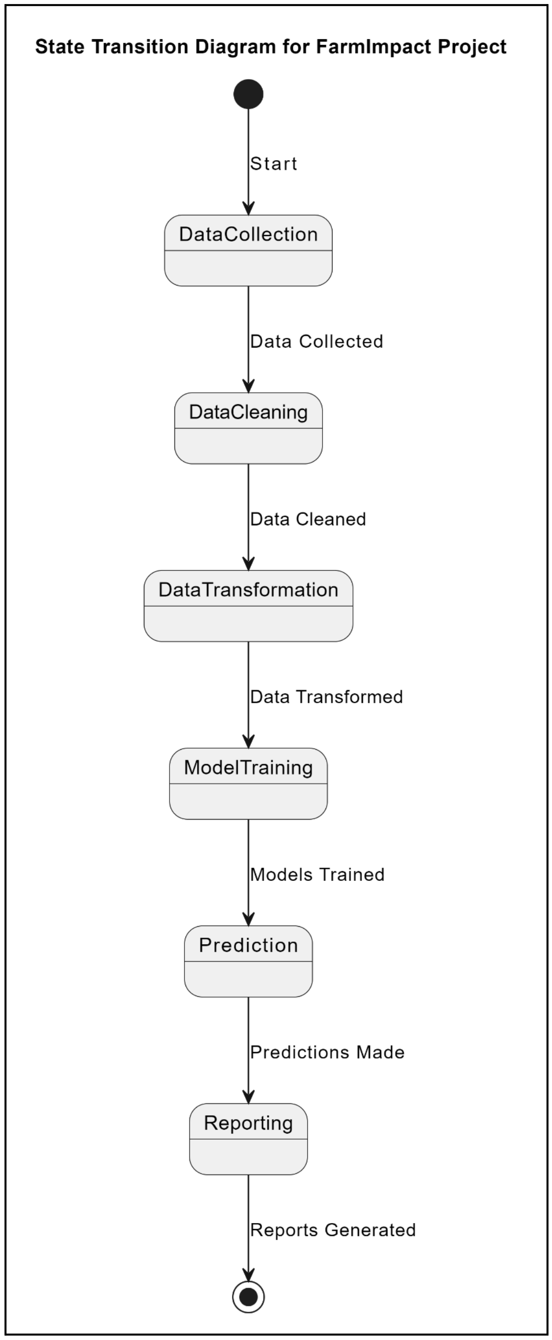

Fig 8. State Transition Diagram

Fig. 8 shows the process flow for the 'FarmImpact' project, beginning with data collection. After the data is cleaned and transformed, models are trained to make predictions. Finally, reports are generated based on the predictions, marking the completion of the process.

Chapter 5: Proposed Results

The results include various graphs and models showcasing key findings. ARIMA models were used to predict missing values for CO2, GHI, DNI, and DHI. Ridge and Lasso regression plots compare predicted and actual values for production, while feature importance graphs highlight that the area under irrigation is the most influential factor in both production and yield. Spline plots using GAM illustrate how factors like rainfall, temperature, and CO2 affect production and yield, with moderate levels being beneficial and extremes reducing output. A bar plot compares MSE and R-squared values, showing Ridge and Lasso outperforming GAM.

5.1 Results

Dataset:

Sources	Open Government Data	Open Government Data	Ministry of Agriculture & Farmers' Welfare					WorldOMeter our world in data	OpenAI	Kaggle FaoStat		NSRDB: National Solar Radiation Database (NREL)		
Year	Rain	MeanTemperature	Area(MillionHectors)	Production(MillionTonnes)	Yield(KgPerHectare)	AreaUnderIrrigation	IrrigatedArea	CO2	ExtremeWeatherEvents(FDH)	Pesticide use	Fertilizer Use	DNI	DHI	GHI
1950	1172.5	23.71	97.32	50.82	522	18.1	17.6	0	0	37349.96	11466439	0	0	0
1951	1061.3	24.22	96.96	51.99	536	18.41	17.9	8.03E-06	1	24264.39	8266627	89.27014	208.2716	234.7729
1952	1108.2	24.34	102.09	59.2	580	18.09	18.5	2.66E-05	1	16848.95	6622885	92.18967	198.3147	232.7654
1953	1222.9	24.57	109.07	69.82	640	18.13	19.8	2.49E-05	0	13631.15	5146806	93.83682	204.3453	235.7494
1954	1180.3	24.13	107.86	68.03	631	18.42	19.9	5.07E-05	1	14012.84	5504350	93.39251	199.773	234.8703
1955	1298.3	23.97	110.56	66.85	605	18.53	20.5	0.00064	0	12665.74	4796264	93.97383	200.9734	236.7548
1956	1386.2	23.96	111.14	69.86	629	18.15	20.2	0.000657	0	12462.94	5139923	93.57216	196.7028	234.2759
1957	1178.2	23.97	109.48	64.31	587	19.28	21.1	0.001039	0	8935.074	4609647	94.44763	197.5109	236.0721
1958	1331	24.62	114.76	77.14	672	18.69	21.4	0.001518	0	10209.11	5829889	94.82441	195.1078	234.1336
1959	1382.1	24.3	115.82	76.67	662	18.77	21.7	0.001988	0	14425.64	8374082	95.77048	193.1135	235.9598
1960	1149	24.29	115.58	82.02	710	19.09	22.1	0.42911	0	10380.21	8654497	96.40638	191.9811	234.9756

Fig 9. Dataset

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).

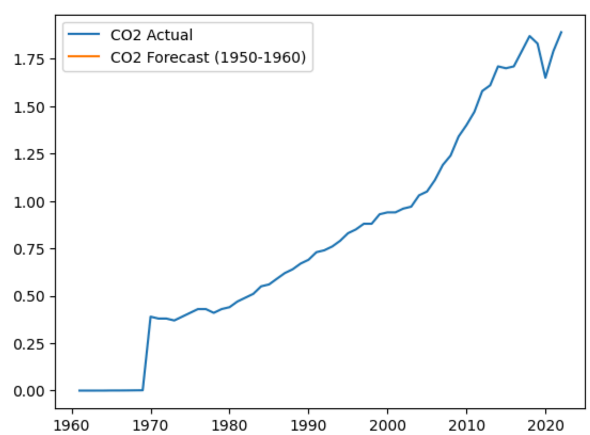

Fig 10. ARIMA for CO2 Prediction of Missing Values

The graph in Fig.10 illustrates the ARIMA model's predicted values for CO2, comparing the observed and predicted trends to estimate missing data accurately.

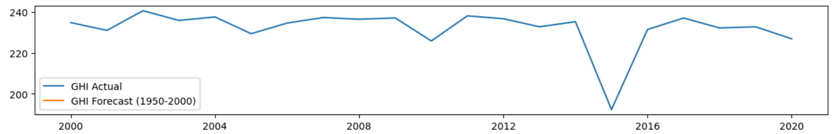

Fig 11. ARIMA for GHI Prediction of Missing Values

The graph in Fig.11 shows ARIMA's predicted Global Horizontal Irradiance (GHI) values, effectively filling in the missing values while maintaining the observed trend.

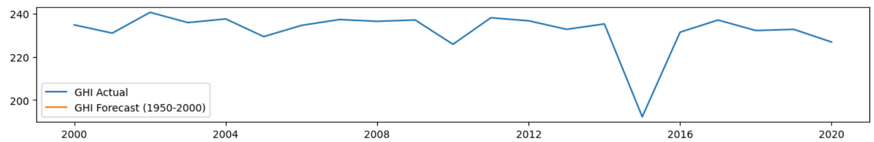

Fig 12. ARIMA for DNI Prediction of Missing Values

The graph in Fig.12 visualizes ARIMA's predictions for Direct Normal Irradiance (DNI), comparing forecasted values with the actual trend to fill in missing data points.

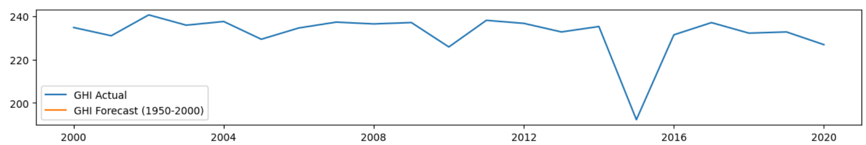

Fig 13. ARIMA for DHI Prediction of Missing Values

ARIMA predictions for Diffuse Horizontal Irradiance (DHI) are presented in Fig.13, where missing values are forecasted based on historical patterns.

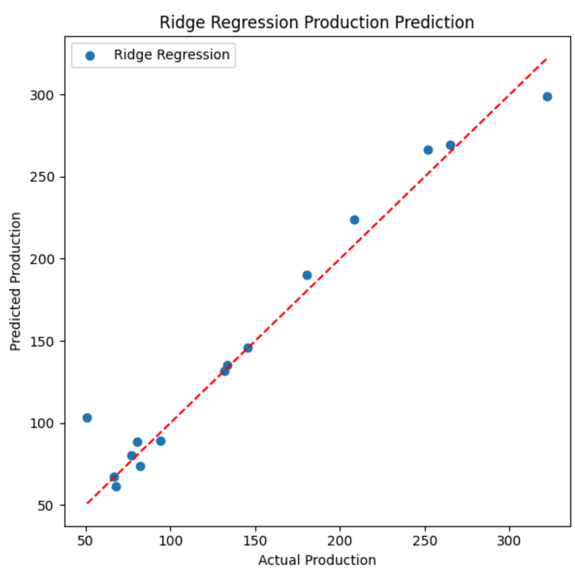

Fig 14. Ridge Regression

Ridge and Lasso regression models in Fig.14 and Fig.15 below assess the relationships between various climatic parameters and yield/production. The red dashed in Fig.14 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.

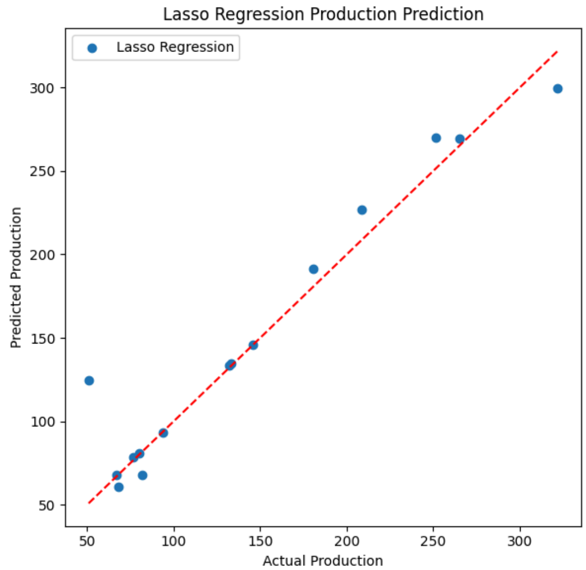

Fig 15. Lasso Regression

Similarly in Fig.15, 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.

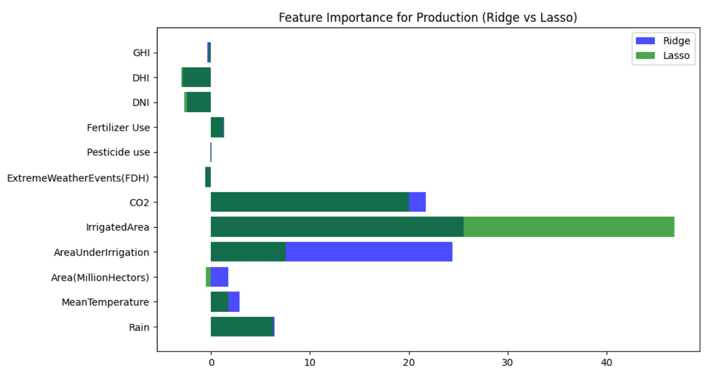

Fig 16. Feature Importance for Production

Each bar in Fig.16 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.

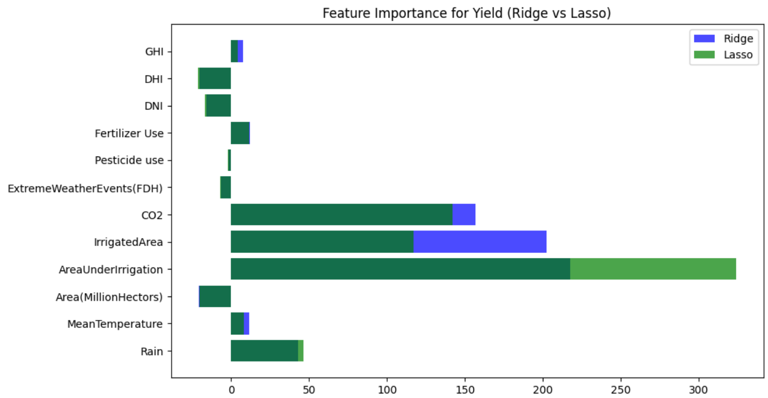

Fig 17. Feature Importance for Yield

Similar to the production analysis in Fig.16 in Fig.17, each bar corresponds to a feature from the dataset, with the horizontal axis showing the magnitude of each feature's impact on yield. Area Under Irrigation is by far the most important factor in determining agricultural yield. Ensuring sufficient irrigation infrastructure is key to boosting crop productivity. The inference for the remaining features is the same as that of production.Overall, this graph reinforces the importance of water availability and irrigation systems in agriculture, emphasizing their central role in ensuring higher crop yields.

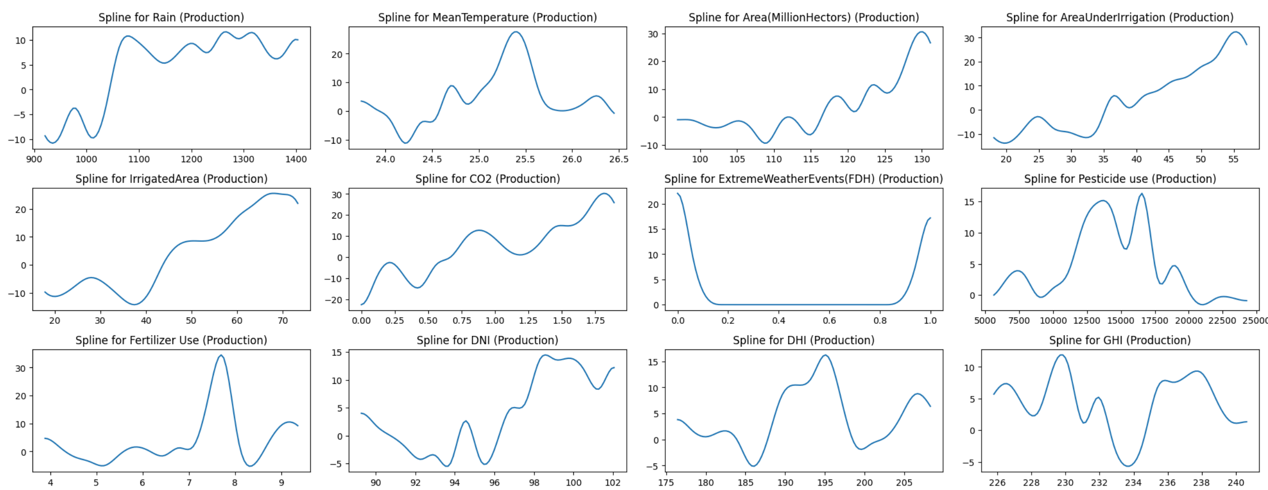

Fig 18. Impact of Individual Features on Production

In Fig.18 the spline plots(GAM) 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.

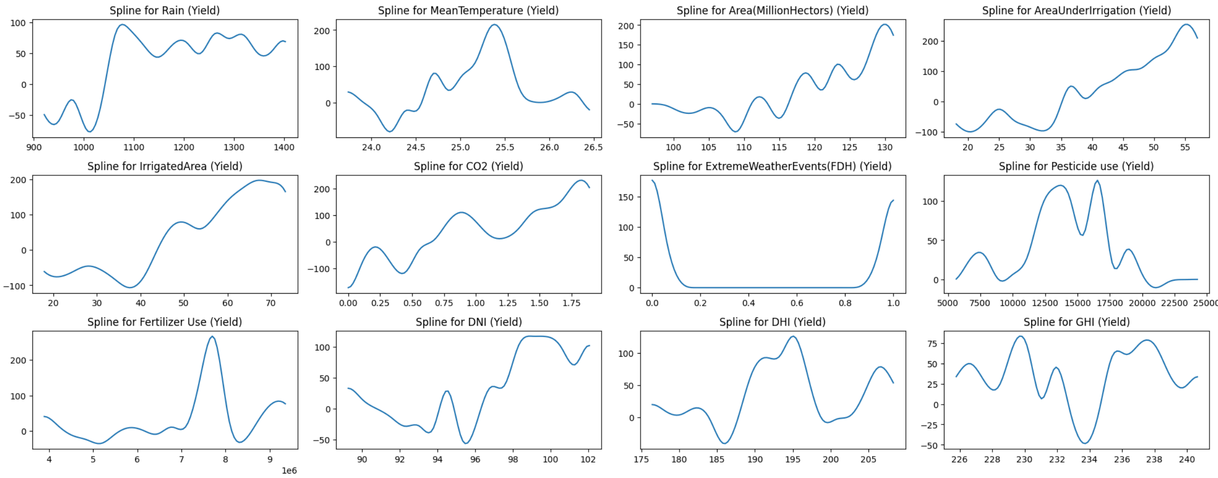

Fig 19. Impact of Individual Features on Yield

In Fig.19 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.

Table II: Comparison of evaluation of algorithms

Algorithm	Mean Square Error	R-Squared Error	Inference Drawn
Ridge Regression	2011.29	0.97	Irrigated Area
Lasso Regression	2112.12	0.96	Irrigated Area
Generative Additive Model	61694.85	-8.34	Irrigated Area

Table I shows the evaluation metrics (MSE and R-Squared) for 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.

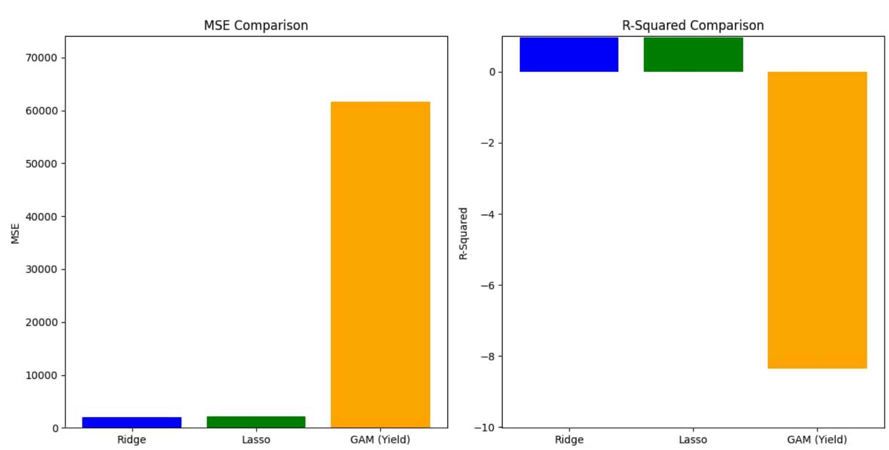

Fig 20. Comparison of MSE and R-Squared Values

Fig. 20 shows the bar plots comparing the evaluation metrics (MSE and R-Squared) for the Ridge, Lasso, and GAM algorithms. The left plot indicates that Ridge and Lasso have significantly lower MSE values, reflecting better predictive accuracy compared to GAM. The right plot highlights that both Ridge and Lasso achieve high R-Squared values, while GAM's low R-Squared suggests a poor fit for the data. This emphasizes that the most influential factor remains the Irrigated Area, with Ridge and Lasso demonstrating stronger predictive performance.

5.2 Evaluation Parameters

- Mean Square Error: It measures the average squared difference between actual and predicted values.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- y_i: Actual value
- $\hat{y}_i$: Predicted value
- n: Number of data points

- R-Squared Error: It represents the proportion of variance explained by the model.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

- y_i: Actual value
- $\hat{y}_i$: Predicted value
- $\bar{y}$: Mean of the actual values
- n: Number of data points

5.3 Report On Sensitivity Analysis

Sensitivity analysis is a critical aspect of understanding how variations in input parameters affect model outputs, particularly in the context of climate change and agriculture. This section summarizes findings from various studies that address the sensitivity of agricultural productivity to climate risk factors.

Methodologies: The studies employed various methodologies for sensitivity analysis, including:

- Regression-based approaches: Used to estimate the effect of independent climate variables on dependent agricultural outputs.
- Scenario analysis: Evaluated the impacts of different climate scenarios on agricultural productivity.
- Machine learning techniques: Enabled the quantification of sensitivity in a more dynamic context, allowing for a better understanding of complex interactions among variables.

Chapter 6: Plan of Action for Next Semester

6.1 Work Done Till Date

- **Data Collection**: Gathered climate and agricultural data.
- **Model Implementation**: Initial implementation of ARIMA, GAM and Ridge-Lasso regression models.

6.2 Plan of Action for Project II

- **Model Optimization**: Fine-tuning the machine learning models.

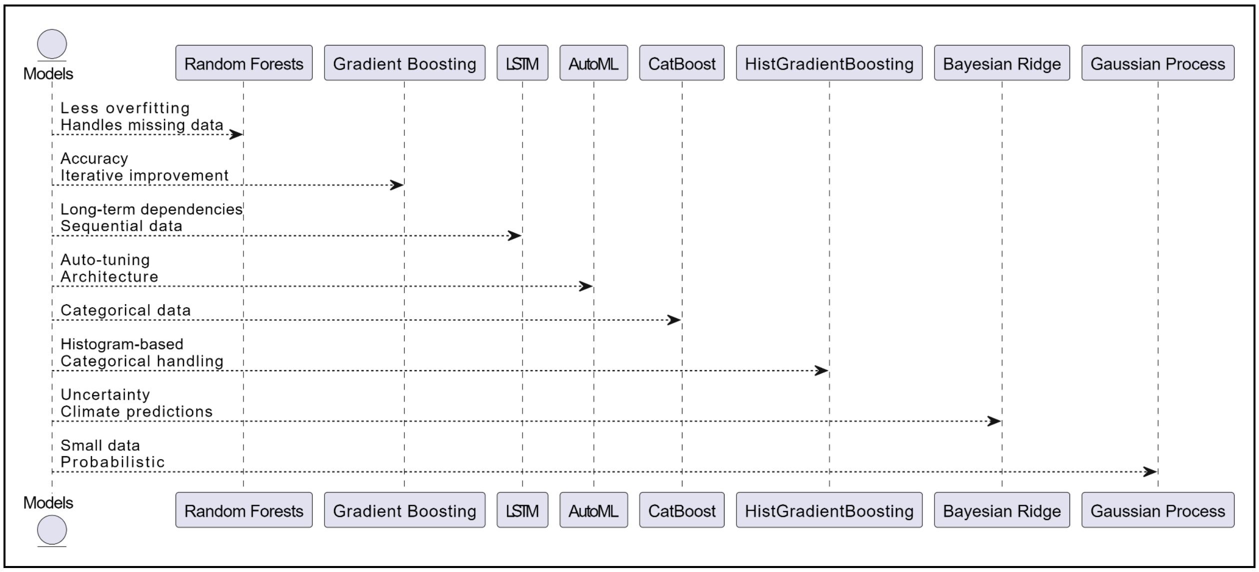

Fig 21. Algorithms

Chapter 7: Conclusion

FarmImpact represents a comprehensive approach to understanding and mitigating the impacts of climate change on agriculture in India. By leveraging machine learning techniques, the project aims to provide actionable insights to enhance agricultural productivity and resilience. Integrating climate data with predictive modeling allows for a deeper understanding of the complex interactions between climate variables and crop yields.

The findings from this research will be pivotal for policymakers and stakeholders in making informed decisions that prioritize sustainable agricultural practices. Through data-driven strategies, FarmImpact addresses the critical challenges posed by climate variability, ensuring food security in a rapidly changing environment.

The project emphasizes the need for innovative technologies and adaptive management strategies to build resilience in the agricultural sector. The methodologies employed, including ARIMA, Ridge, Lasso regression, and GAM, serve as powerful tools for forecasting and analysis.

By focusing on key climate factors, such as temperature and rainfall, the project aims to develop effective climate-resilient strategies. Ultimately, FarmImpact aspires to foster sustainable practices that improve productivity while safeguarding the livelihoods of farmers against the threats of climate change. Through continued research and collaboration, the project envisions a future where agriculture in India can thrive despite these challenges.

The analysis revealed that the "Irrigated Area" was the most significant driver of crop yield and production 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.

Chapter 8: 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. https://www.researchgate.net/publication/346655247.

[15] 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.

[16] 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.

[17] Cagliarini, Adam, and Anthony Rush. "Economic Development and Agriculture in India." Bulletin (June Quarter 2011): 15-22. Reserve Bank of Australia.

[18] National Institute of Agricultural Marketing. Agriculture and Economic Development in India. Final Report. Jaipur: National Institute of Agricultural Marketing, 2011.

[19] Palanivel, Prakash. A Study on Role of Agricultural Development in Indian Economy. Project Report, Ramakrishna Mission Vivekananda College, 2020.

[20] Mehta, Niti. "Agricultural Investments: Trends and Role in Enhancing Agricultural Output and Incomes." Indian Journal of Agricultural Economics 78, no. 4 (2023): 576-589.

[21] 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

[22] Kritika Banerjee. Handling missing values in EDA. Medium (https://medium.com/@kritiikaaa/handling-missing-values-in-eda-b12efc7da26d0)

[23] DataCamp. Lasso and Ridge Regression in Python Tutorial. https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression

[24] Wikipedia. Generalized additive model. https://en.wikipedia.org/wiki/Generalized_additive_model

[25] IBM. ARIMA model. (https://www.ibm.com/topics/arima-model)

[26] 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.

[27] 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

[28] Vaisala Energy Support, What is Global Horizontal Irradiance? https://www.3tier.com/en/support/solar-prospecting-tools/what-global-horizontal-irradiance-solar-prospecting/

[29] World Health Organization. Radiation: The ultraviolet (UV) index, https://www.who.int/news-room/questions-and-answers/item/radiation-the-ultraviolet-(uv)-index

[30] 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

Chapter 9: Appendix

List of Figures

Number	**Name**
Fig.1	ARIMA Methodology Flowchart
Fig.2	GAM Methodology Flowchart
Fig.3	Ridge-Lasso Regression Methodology Flowchart
Fig.4	Block Diagram
Fig.5	Modular Diagram
Fig.6	Data Flow Diagram
Fig.7	Flow Chart
Fig.8	State Transition Diagram
Fig.9	Dataset
Fig.10	ARIMA for CO2 Prediction of Missing Values
Fig.11	ARIMA for GHI Prediction of Missing Values
Fig.12	ARIMA for DNI Prediction of Missing Values
Fig.13	ARIMA for DHI Prediction of Missing Values
Fig.14	Ridge Regression
Fig.15	Lasso Regression
Fig.16	Feature Importance for Production
Fig.17	Feature Importance for Yield
Fig.18	Impact of Individual Features on Production
Fig.19	Impact of Individual Features on Yield
Fig.20	Comparison of MSE and R-Squared Values
Fig.21	Algorithms

List of Tables

Number	Name
Table I	Comparison of existing system and proposed system
Table II	Comparison of Evaluation of Algorithms

List of Abbreviations

Abbreviation	Full Form
ARIMA	AutoRegressive Integrated Moving Average
GAM	Generative Additive Models
GHI	Global Horizontal Irradiance
DNI	Direct Normal Irradiance
DHI	Diffuse Horizontal Irradiance

FarmImpact: Impact of Climate Change on Agriculture in India

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*—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 will provide insights into the complex interactions between these factors and their direct and indirect effects on crop yield and production. The findings will be 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, and production prediction.**

I. 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).

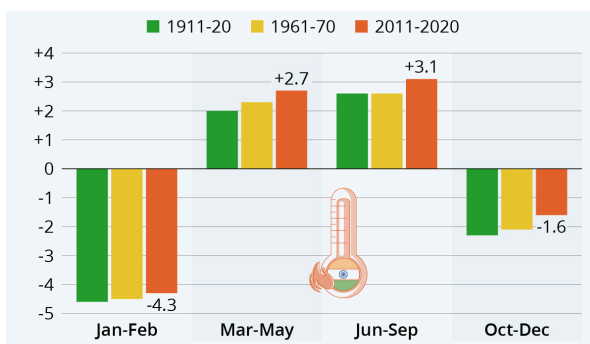

Fig.1 Average divergence from mean temperature (Sources: Indian Meteorological Department, Ministry of Earth Sciences)

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 tackle 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 project first aims to identify the most influential climate parameters affecting yield and production. Next, forecasting models like ARIMA are applied to predict future climate conditions and their corresponding effects on crop yield. The proposed work ultimately 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.

II. Related work

Paper [1]: The MARS model explored nonlinear relationships between climate factors and agricultural productivity in India, offering flexibility but facing computational challenges.
Paper [2]: Machine learning models predicted temperature and rainfall in Marathwada, excelling in seasonal data but struggling with irregular patterns.
Paper [3]: ANN-MLP models examined rainfall trends in India post-1960, effectively handling nonlinear data but limited by reliance on historical data.
Paper [4]: Regression models assessed CO2 emissions and population growth impacts on climate but oversimplified the complexity of broader climate systems.
Paper [5]: Simulation models in Asia highlighted climate impacts on agriculture, though real-world variability was not fully captured.
Paper [6]: The study critiqued the Green Revolution's chemical reliance, advocating organic farming, though it faces adoption challenges due to higher labor and lower yields.
Paper [7]: Integrating thermodynamics with machine learning improved climate predictions but was hindered by high computational demands.
Paper [8]: Climate-smart agriculture strategies offered sustainability for developing nations but require significant policy and financial support.

Paper [9]: Rising temperatures threatened agricultural sustainability in Indian states, with solutions facing implementation challenges in marginalized regions.
Paper [10]: Water management and climate-resilient crops were essential for rainfed regions, but progress depends on long-term investments.
Paper [11]: Global studies showed rising temperatures reducing yields, especially in developing countries, but lacked regional adaptability insights.
Paper [12]: District-level rice yield forecasts under climate scenarios offered valuable data but faced uncertainties due to potential future advances.
Paper [13]: Sustainable farming practices were essential for food security but constrained by socioeconomic barriers like access to technology and support.
Paper [14]: Climate variables adversely impacted India's economic growth, though adaptive measures were not fully accounted for.
Paper [15]: Climate change worsened rural poverty and productivity issues, but the study lacked insights into urban-rural interactions.
Paper [16]: Global food system impacts from climate change required adaptive strategies, though practical application across diverse contexts remains challenging.
Paper [17]: Water management and infrastructure improvements were critical for India's agriculture, but financial and logistical barriers persist.
Paper [18]: ICT and supply chain management could reduce post-harvest losses but require investment and accessibility for smallholder farmers.
Paper [19]: Modern agricultural practices are essential for India's growth, but financial and technological barriers limit widespread adoption.
Paper [20]: Investments in agriculture through targeted policies and infrastructure boost productivity but depend on political commitment and equitable resource distribution.

III. 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. 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.

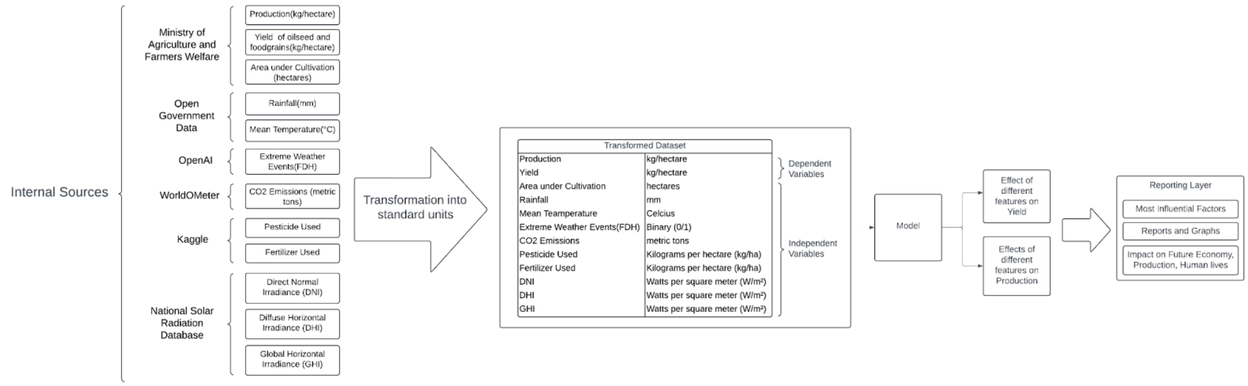

Fig.2 Block Diagram

IV. 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]

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]

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]
- 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]

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]

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. [27].

V. 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.

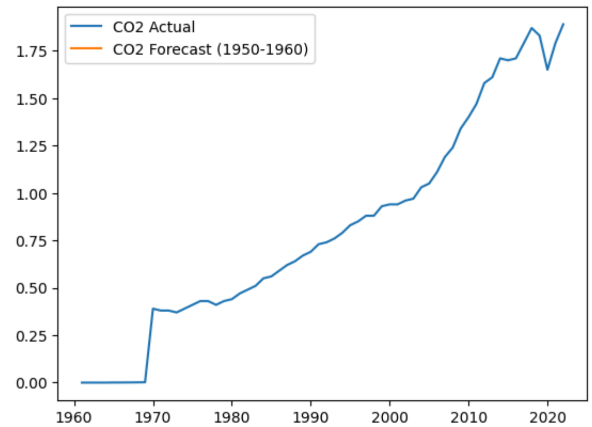

Fig.3 ARIMA for CO2 Prediction of Missing Values

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.

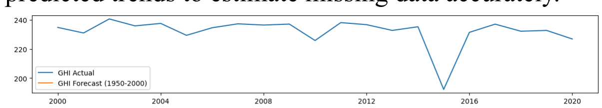

Fig.4 ARIMA for GHI Prediction of Missing Values

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]

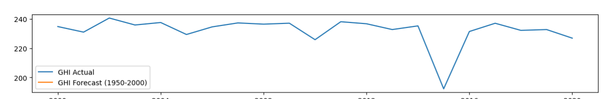

Fig.5 ARIMA for DNI Prediction of Missing Values

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]

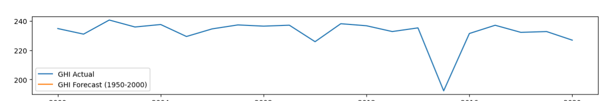

Fig.6 ARIMA for DHI Prediction of Missing Values

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.

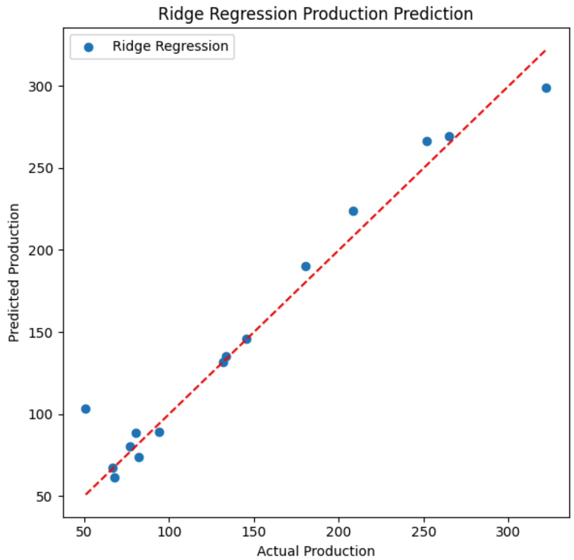

Fig.7 Ridge Regression

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.

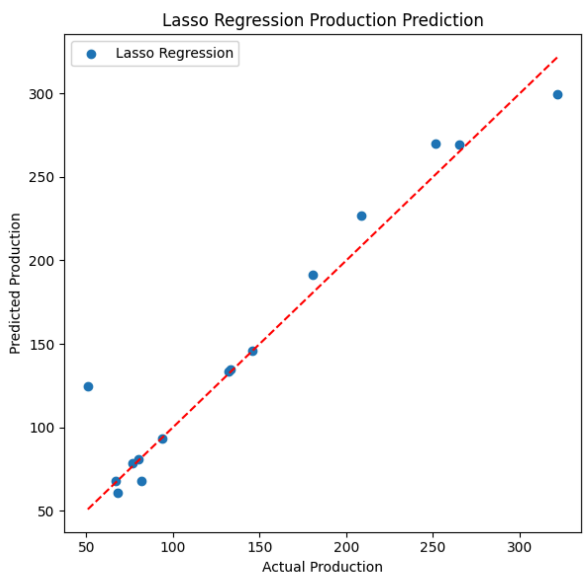

Fig.8 Lasso Regression

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.

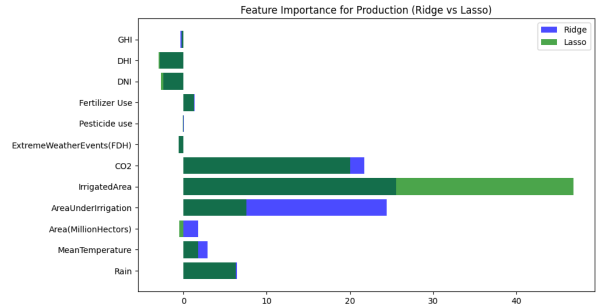

Fig.9 Feature Importance for Production

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.

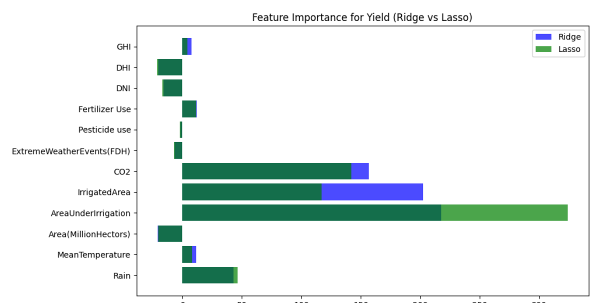

Fig.10 Feature Importance for Yield

Similar to the production analysis in Fig.9 in Fig.10, each bar corresponds to a feature from the dataset, with the horizontal axis showing the magnitude of each feature's impact on yield. Area Under Irrigation is by far the most important factor in determining agricultural yield. Ensuring sufficient irrigation infrastructure is key to boosting crop productivity. The inference for the remaining features is the same as that of production.

Overall, this graph reinforces the importance of water availability and irrigation systems in agriculture, emphasizing their central role in ensuring higher crop yields.

The Generalized Additive Model (GAM) offers flexibility by allowing non-linear relationships between the dependent and independent variables.

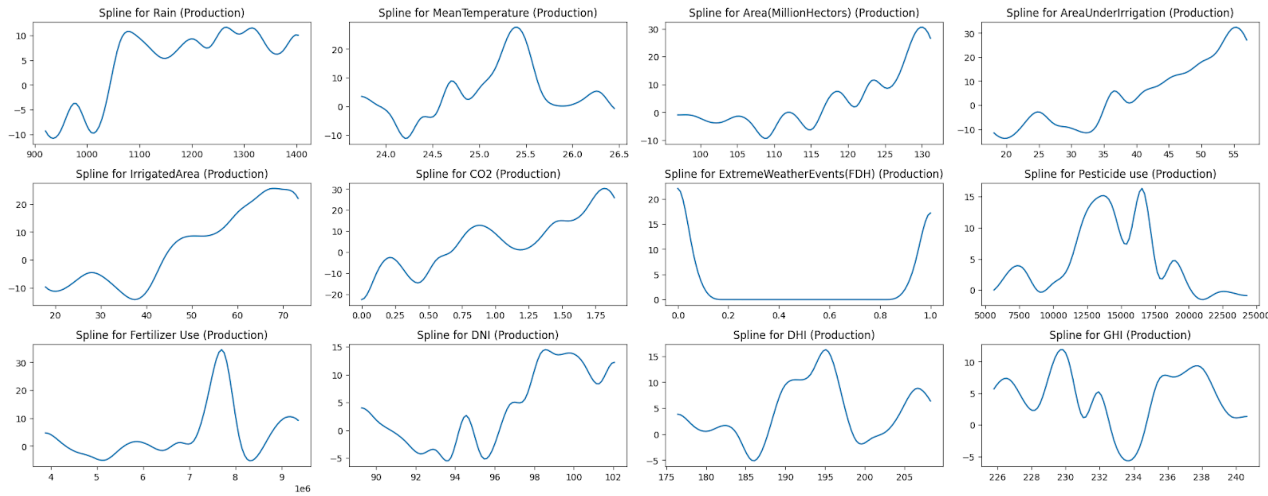

Fig.11 Impact of Individual Features on Production

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.

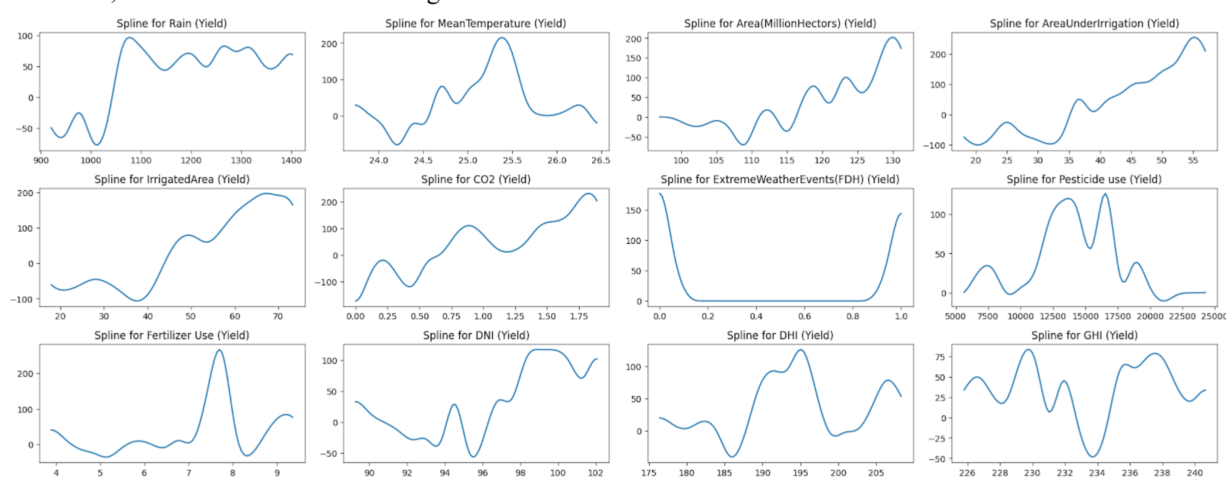

Fig.12 Impact of Individual Features on Yield

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.

Table I. Comparison of evaluation of algorithms

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 shows the evaluation metrics (MSE and R-Squared) for 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.

VI. Conclusion And Future Scope

The analysis revealed that the "Irrigated Area" was the most significant driver of crop yield and production 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. https://www.researchgate.net/publication/346655247.
[15] 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.
[16] 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.
[17] Cagliarini, Adam, and Anthony Rush. "Economic Development and Agriculture in India." Bulletin (June Quarter 2011): 15-22. Reserve Bank of Australia.
[18] National Institute of Agricultural Marketing. Agriculture and Economic Development in India. Final Report. Jaipur: National Institute of Agricultural Marketing, 2011.
[19] Palanivel, Prakash. A Study on Role of Agricultural Development in Indian Economy. Project Report, Ramakrishna Mission Vivekananda College, 2020.
[20] Mehta, Niti. "Agricultural Investments: Trends and Role in Enhancing Agricultural Output and Incomes." Indian Journal of Agricultural Economics 78, no. 4 (2023): 576-589.
[21] 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
[22] Kritika Banerjee. Handling missing values in EDA. Medium (https://medium.com/@kritiikaaa/handling-missing-values-in-eda-b12efc7da26d0)
[23] DataCamp. Lasso and Ridge Regression in Python Tutorial. https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression
[24] Wikipedia. Generalized additive model. https://en.wikipedia.org/wiki/Generalized_additive_model
[25] IBM. ARIMA model. (https://www.ibm.com/topics/arima-model)
[26] 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.
[27] 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
[28] Vaisala Energy Support, What is Global Horizontal Irradiance? https://www.3tier.com/en/support/solar-prospecting-tools/what-global-horizontal-irradiance-solar-prospecting/
[29] World Health Organization. Radiation: The ultraviolet (UV) index, https://www.who.int/news-room/questions-and-answers/item/radiation-the-ultraviolet-(uv)-index
[30] 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

FarmImpact

ORIGINALITY REPORT

5%	3%	3%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Food Security in a Developing World", Springer Science and Business Media LLC, 2024 Publication	1%
2	www.smujo.id Internet Source	1%
3	"The Climate-Health-Sustainability Nexus", Springer Science and Business Media LLC, 2024 Publication	1%
4	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%
5	link.springer.com Internet Source	1%
6	www.researchgate.net Internet Source	1%

REVIEW EVALUATION SHEET

SDG 13. Climate Action

3

Industry/Inhouse: **Project Evaluation Sheet 2024-25** **Class: D17 B**

Title of Project(Group no): 5. Farm Impact: Impact of climate change on agriculture in India

Group Members: Vishakha Singh (D17B,54), Manasi Sharma (D17B51), Anushka Shinde 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 (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	5	4	4	3	4	2	2	2	2	3	3	3	4	4	45

Comments: ML model needs to be finalized along with Notification models.

R. 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 (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	3	4	2	4	2	2	2	2	3	3	3	4	4	42

Comments:

Date: 23rd August, 2024

Name & Signature Reviewer2

SDG 13: Climate Action

5

Industry/Inhouse: **Project Evaluation Sheet 2024-25** **Class: D17 B**

Title of Project(Group no): 5 Farm Impact: Impact of climate change on Agriculture in India

Group Members: Vishakha Singh D17B (54), Manasi Sharma D17B (51), Anushka Shinde 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 (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	3	4	4	2	4	2	2	2	2	3	3	3	4	4	42

Comments: Dataset needs more details. ML algos needs to be finalized.

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 (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	3	3	4	2	4	2	2	2	2	3	3	3	3	4	40

Comments: Real world data is desired.

Date: 26th September, 2024

Name & Signature Reviewer2