ISSN 1660-6795
www.nano-ntp.com

Farming for Tomorrow: Sustainability amidst Climate Change in India

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

Abstract:

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

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

Introduction:

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

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

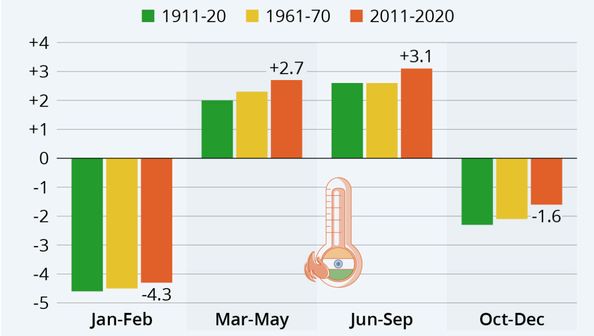

Fig.1 Average divergence from mean temperature

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

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

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

Related Work:

The authors in [1] explored the MARS model and the nonlinear relationships between climate factors and agricultural productivity in India, offering flexibility but facing computational challenges. Machine learning models predicted temperature and rainfall in Marathwada, excelling in seasonal data but struggling with irregular patterns [2]. In [3] ANN-MLP models examined rainfall trends in India post-1960, effectively handling nonlinear data but limited by reliance on historical data. Regression models assessed CO2 emissions and population growth impacts on climate but oversimplified the complexity of broader climate systems were elaborated by the authors in [4]. The authors of [5] observed that the Simulation models in Asia highlighted climate impacts on agriculture, though real-world variability was not fully captured.The study critiqued the Green Revolution's chemical reliance, advocating organic farming, though it faces adoption challenges due to higher labor and lower yields[6]. Integrating thermodynamics with machine learning improved climate predictions but was hindered by high computational demands, The authors in [7] Climate-smart agriculture strategies offered sustainability for developing nations but required significant policy and financial support mentioned in [8]. Authors of [9] indicated that rising temperatures threatened agricultural sustainability in Indian states, with solutions facing implementation challenges in marginalized regions. Water management and climate-resilient crops were essential for rainfed regions, but progress depends on long-term investments [10]. Further the authors in [10] focussed on Global studies that showed rising temperatures reducing yields, especially in developing countries, but lacked regional adaptability insights. District-level rice yield forecasts under climate scenarios offered valuable data but faced uncertainties due to potential future advances was mentioned in [12]. Sustainable farming practices were essential for food security but constrained by socioeconomic barriers like access to technology and support was specified in [13]. Authors of [14] clearly mentioned that climate variables adversely impacted India's economic growth, though adaptive measures were not fully accounted for. Paper [15][16] elaborated that climate change worsened rural poverty and productivity issues, but the study lacked insights into urban-rural interactions. Global food system impacts from climate change require adaptive strategies, though practical application across diverse contexts remains challenging, mentioned the authors of [16][18] Water management and infrastructure improvements were critical for India's agriculture, but financial and logistical barriers persist[19]. The authors of [20][21] identified that ICT and supply chain management could reduce post-harvest losses but require investment and accessibility for smallholder farmers. Further the authors of [22][23] mentioned that modern agricultural practices are essential for India's growth, but financial and technological barriers limit widespread adoption. Authors of [24][25] expressed that investments in agriculture through targeted policies and infrastructure boost productivity but depend on political commitment and equitable resource distribution.

Methodology:

Fig.2 shows the methodology used in the project to analyze the impact of various climatic factors on agricultural yield and production. It begins with data collection from diverse sources, including government websites, open databases, and environmental data platforms, such as the Ministry of Agriculture, WorldOMeter, and the National Solar Radiation Database. The collected data, including factors like production, CO_2 emissions, rainfall, and solar irradiance (DNI, DHI, GHI), is transformed into standard units to ensure uniformity across the dataset.The transformed data is divided into dependent and independent variables for further analysis. Machine learning models like ARIMA, Ridge, Lasso, and GAM are applied to evaluate the relationships between climatic variables and agricultural outcomes.

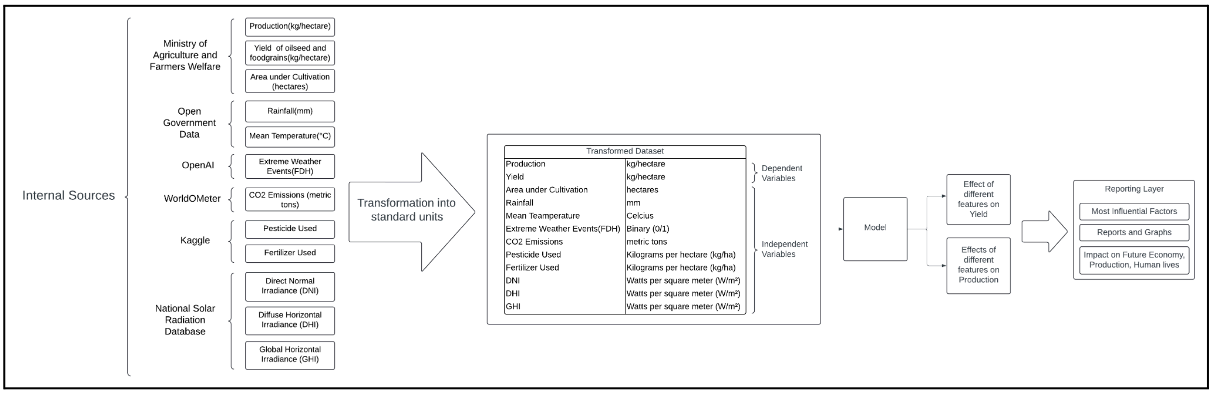

Fig.2 Block Diagram representation

Implementation:

A. Identification of Influential Climatic Factors

1) Data Collection:

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

- Dependent Variables: Production (kg/ha), Yield (kg/ha), Area under Cultivation (hectares).
- Independent Variables: Rainfall (mm), Mean Temperature (°C), Extreme Weather Events (binary), CO_2 Emissions (metric tons), Pesticide Use (kg/ha), Fertilizer Use (kg/ha), DNI, DHI, GHI (W/m²).

2) Data Transformation:

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

3) Exploratory Data Analysis (EDA):

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

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

4) Model Selection for Impact Analysis:

The analysis of influential climatic factors employs several statistical models.

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

B. Forecasting Climatic Factors and Yield

1) Time Series Forecasting:

A time series forecasting approach is adopted to predict key climatic parameters that significantly influence agricultural outcomes. The ARIMA model is employed to forecast essential variables such as rainfall, temperature, CO_2 emissions, and solar radiation metrics (DNI, DHI, GHI) for the years 2023 to 2030. The ARIMA models are trained on historical data, and validation techniques, including cross-validation, are used to ensure the accuracy and reliability of the forecasts. The predicted values of these independent variables will be utilized in the next step to assess their impact on agricultural yield [25][29].

2) Yield Prediction Based on Forecasts:

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

C. Inference and Risk Mitigation

1) Trend Analysis:

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

2) Risk Identification and Mitigation:

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

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

3) Reporting and Insights:

The final reporting layer will provide insights on:

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

Results:

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

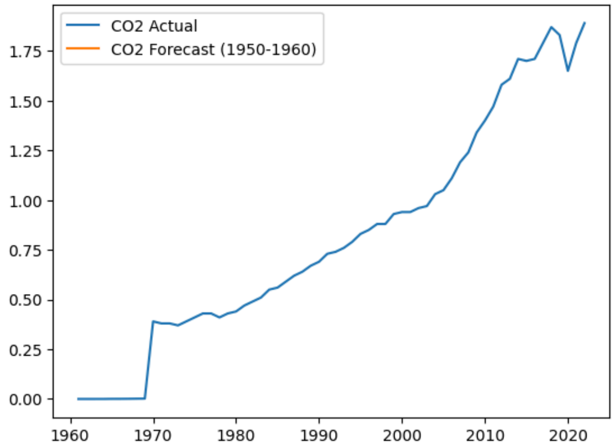

Fig.3 ARIMA for CO2 Prediction of Missing Values

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

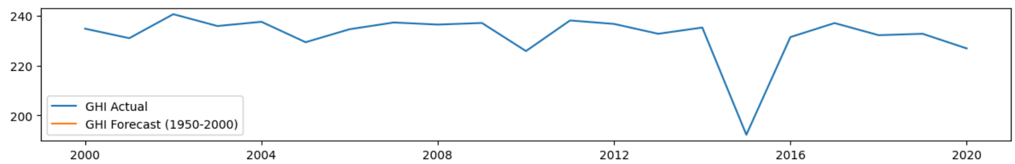

Fig.4 ARIMA's predicted Global Horizontal Irradiance (GHI) values

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

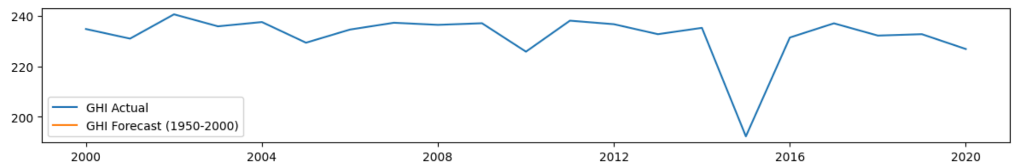

Fig.5 ARIMA for DNI Prediction of Missing Values

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

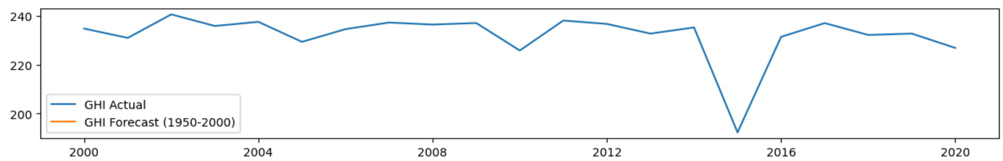

Fig.6 ARIMA for DHI Prediction of Missing Values

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

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

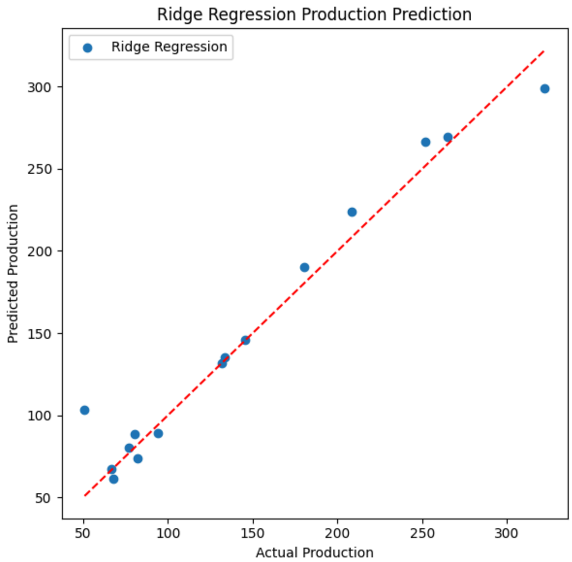

Fig.7 Ridge Regression

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

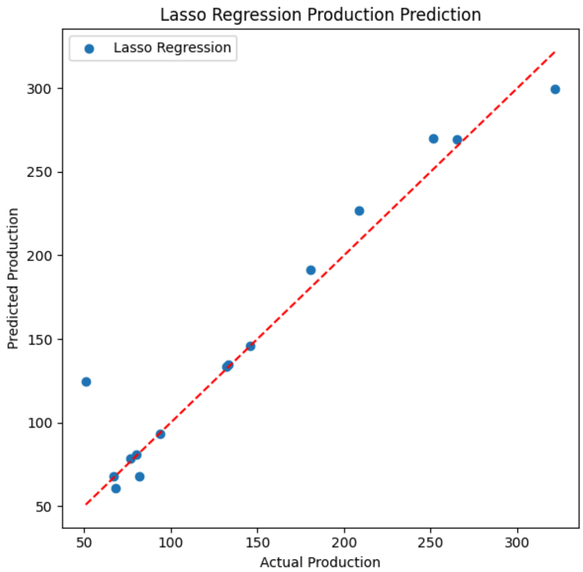

Fig.8 Lasso Regression

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

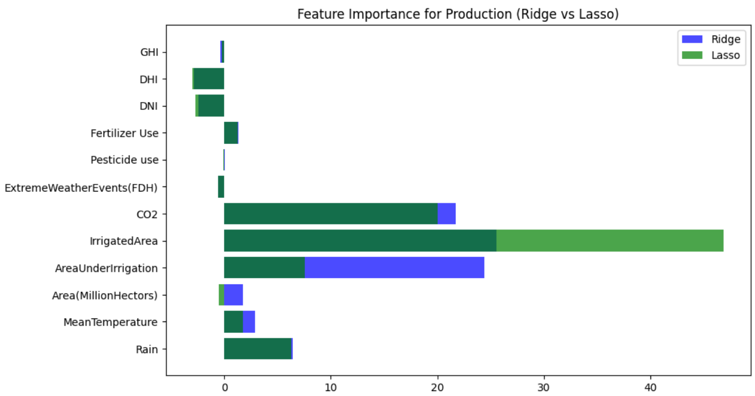

Fig.9 Feature Importance for Production

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

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

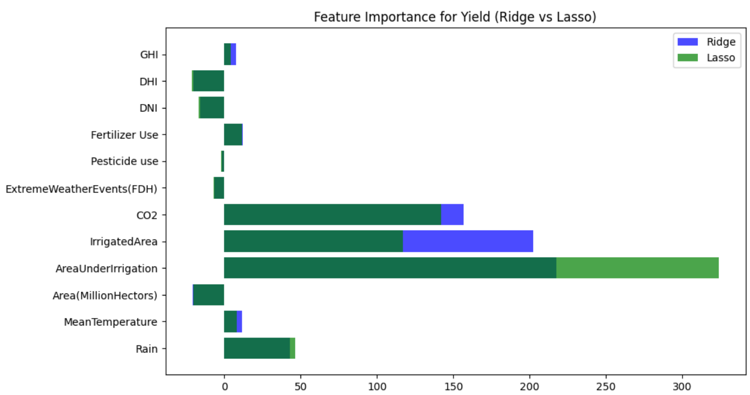

Fig.10 Feature Importance for Yield

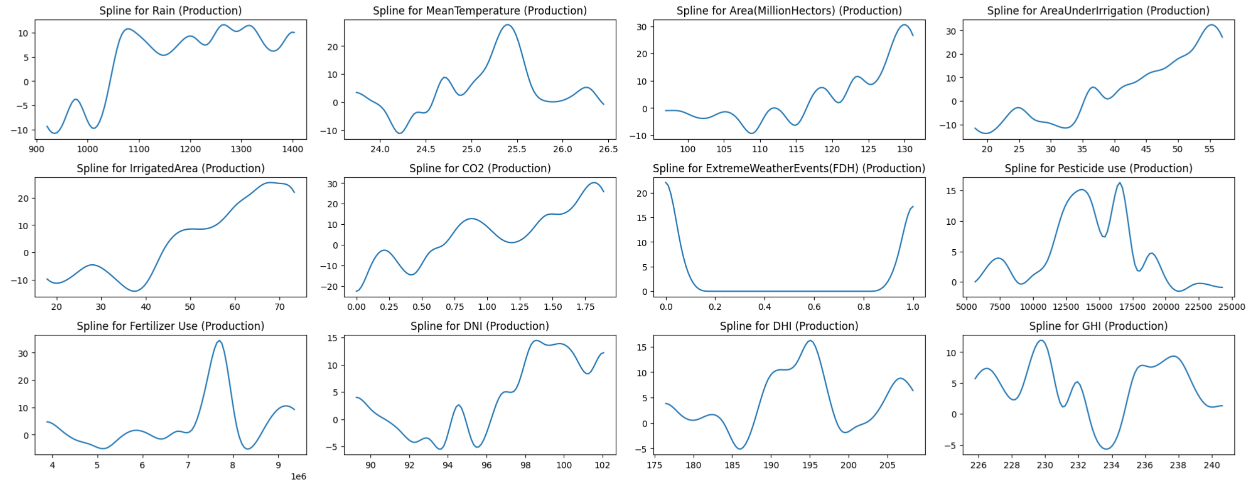

Fig.11 Impact of Individual Features on Production

In Fig.11 the spline plots [30] indicate that agricultural production is influenced by several factors. Moderate rainfall boosts production, but too much or too little rain reduces it, emphasizing the need for balance. Optimal temperatures, around 25.5°C, support higher production, while extremes negatively impact output. An increase in cultivated area steadily raises production, and irrigation also helps, but beyond a certain limit, the benefits diminish. Rising CO_2 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.

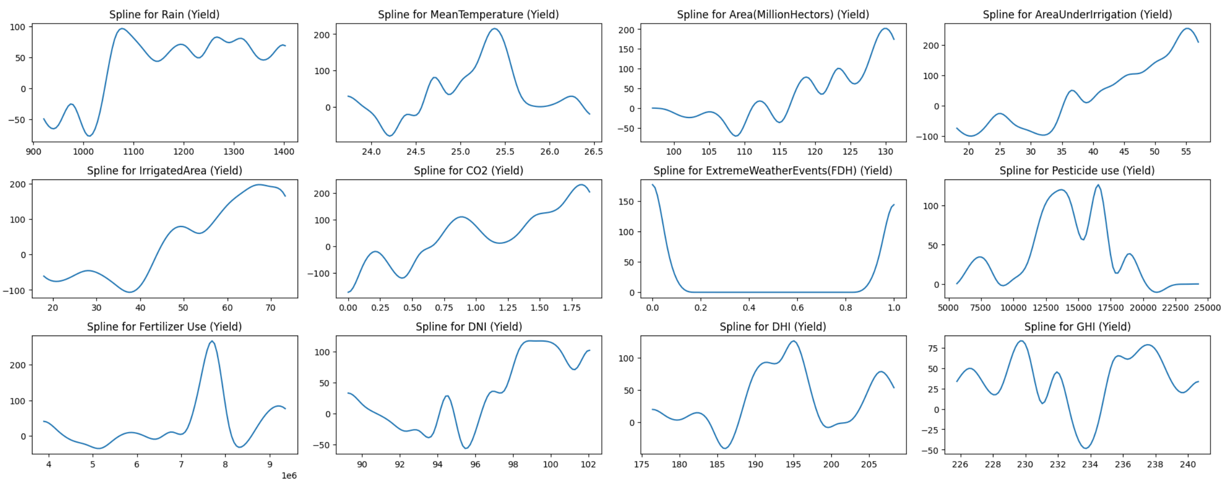

Fig.12 Impact of Individual Features on Yield

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 CO_2 emissions. The model considers the available historical data to identify trends, seasonality, and noise, and uses this information to forecast future values. By comparing the predicted values with the actual recorded data, ARIMA showed minimal error, validating its effectiveness in handling time-series data. The close alignment of predicted values with the actual values indicated that ARIMA is reliable in estimating missing data and projecting future trends, allowing for more accurate forecasting and analysis.

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

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

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

Table I. Comparison of evaluation of algorithms

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

Conclusion and Future Scope:

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

References:

[1] Jha, Paritosh, Sona Chinngaihlian, and Priyanka Upreti. "A Machine Learning Approach to Assess Implications of Climate Risk Factors on Agriculture: The Indian Case." Department of Economic and Policy Research, Reserve Bank of India, Mumbai, India., 2021.
[2] Ramdas D. Gore and Bharti W. Gawali, ―Analysis of Weather Parameters Using Machine Learning‖, July 2023, URL (https://www.atlantis-press.com/article/125989847.pdf)
[3] Bapi Raju et al., ―Analyzing Trend and Forecasting of Rainfall Changes in India Using Non-Parametric and Machine Learning Approaches,‖, June 2020, URL (https://www.nature.com/articles/s41598-020-67228-7).
[4] P. Roy et al., ―Analysis of Various Climate Change Parameters in India Using Machine Learning,‖, Jan 2022, URL (https://arxiv.org/pdf/2201.10123)
[5] Habib-ur-Rahman, Muhammad, Ashfaq Ahmad, Ahsan Raza, Muhammad Usama

Hasnain, Hesham F. Alharby, Yahya M. Alzahrani, Atif A. Bamagoos, Khalid Rehman Hakeem, Saeed Ahmad, Wajid Nasim, Shafaqat Ali, Fatma Mansour, and Ayman EL Sabagh. "Impact of Climate Change on Agricultural Production: Issues, Challenges, and Opportunities in Asia." 2022.
[6] Choudhary, D. K. (2021). Chemical Fertilizers and Pesticides in Indian Agriculture. International Journal of Research and Analysis in Science and Engineering, 1(6). Retrieved from https://www.iarj.in/index.php/ijrase/index
[7] Vázquez-Ramírez, S., Torres-Ruiz, M., Quintero, R., Chui, K.T., & Guzmán Sánchez-Mejorada, C. (2023). An Analysis of Climate Change Based on Machine Learning and an Endoreversible Model. Mathematics, 11(3060).
[8] Malhi, G.S., Kaur, M., & Kaushik, P. (2021). Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. Sustainability, 13(3), 1318.
[9] Singh, A.K., Kumar, S., & Jyoti, B. (2022). Influence of Climate Change on Agricultural Sustainability in India: A State-Wise Panel Data Analysis. Asian Journal of Agriculture, 6(1), 15-27.
[10] "Climate Change and Agriculture in India" report supported by the National Mission on Strategic Knowledge for Climate Change (NMSKCC)
[11] Cline, William R. Global Warming and Agriculture: Impact Estimates by Country. Washington: Center for Global Development and Peterson Institute for International Economics, 2007.
[12] Gallé, Johannes, and Anja Katzenberger. "Indian Agriculture Under Climate Change: The Competing Effect of Temperature and Rainfall Anomalies." Economics of Disasters and Climate Change (2024). https://doi.org/10.1007/s41885-024-00154-4.
[13] Dubey, Pradeep Kumar, Ajeet Singh, Rajan Chaurasia, Krishna Kumar Pandey, Amit Kumar Bundela, Rama Kant Dubey, and Purushothaman Chirakkuzhyil Abhilash. "Planet Friendly Agriculture: Farming for People and the Planet." Current Research in Environmental Sustainability 3 (2021): 100041. https://doi.org/10.1016/j.crsust.2021.100041.
[14] Husain, Uvesh, and Sarfaraz Javed. "Impact of Climate Change on Agriculture and Indian Economy: A Quantitative Research Perspective from 1980 to 2016." Industrial Engineering & Management 8, no. 2 (2019): 281.
[15] https://www.researchgate.net/publication/346655247.
[16] Kar, Saibal, and Nimai Das. "Climate Change, Agricultural Production, and Poverty in India." In Poverty Reduction Policies and Practices in Developing Asia, edited by A. Heshmati, 55–76. Singapore: Springer, 2015. https://doi.org/10.1007/978-981-287-420-7_4.
[17] Kumara, Lalit, Ngawang Chhogyel, Tharani Gopalakrishnan, Md Kamrul Hasan, Sadeeka Layomi Jayasinghe, Champika Shyamalie Kariyawasam, Benjamin Kipkemboi Kogo, and Sujith Ratnayake. "Climate Change and Future of Agri-Food Production." In Future Foods, edited by P.C. Keenan, 49–64. Elsevier, 2022. https://doi.org/10.1016/B978-0- 323-91001-9.00009-8.
[18] Cagliarini, Adam, and Anthony Rush. "Economic Development and Agriculture in India." Bulletin (June Quarter 2011): 15-22. Reserve Bank of Australia.
[19] National Institute of Agricultural Marketing. Agriculture and Economic Development in India. Final Report. Jaipur: National Institute of Agricultural Marketing, 2011.
[20] Palanivel, Prakash. A Study on Role of Agricultural Development in Indian Economy. Project Report, Ramakrishna Mission Vivekananda College, 2020.
[21] Mehta, Niti. "Agricultural Investments: Trends and Role in Enhancing Agricultural Output and Incomes." Indian Journal of Agricultural Economics 78, no. 4 (2023): 576-589.
[22] Aldoseri, Abdulaziz, Khalifa N. Al-Khalifa, and Abdel Magid Hamouda. 2023. "Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges" Applied Sciences 13, no. 12: 7082. https://doi.org/10.3390/app13127082
[23] Kritika Banerjee. Handling missing values in EDA. Medium (https://medium.com/@kritiikaaa/handling-missing-values-in-eda-b12efc7da26d0)
[24] DataCamp. Lasso and Ridge Regression in Python Tutorial.

https://www.datacamp.com/tutorial/tutorial-lasso-ridge-regression
[25] Wikipedia. Generalized additive model. https://en.wikipedia.org/wiki/Generalized_additive_model
[26] IBM. ARIMA model. (https://www.ibm.com/topics/arima-model)
[27] Raza A, Razzaq A, Mehmood SS, Zou X, Zhang X, Lv Y, Xu J. Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. Plants (Basel). 2019 Jan 30;8(2):34. doi: 10.3390/plants8020034. PMID: 30704089; PMCID: PMC6409995.
[28] Food and Agriculture Organization of the United Nations. (2024). Implications of Economic Policy for Food Security: A Training Manual. https://www.fao.org/4/X3936E/X3936E00.htm
[29] Vaisala Energy Support, What is Global Horizontal Irradiance? https://www.3tier.com/en/support/solar-prospecting-tools/what-global-horizontal-irradiance- solar-prospecting/
[30] World Health Organization. Radiation: The ultraviolet (UV) index, https://www.who.int/news-room/questions-and-answers/item/radiation-the-ultraviolet-(uv) - index
[31] R Development Core Team. plot.gam: Default GAM plotting in mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [rdrr.io]. Retrieved from https://rdrr.io/cran/mgcv/man/plot.gam.html