iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: X Month of publication: October 2025

DOI: https://doi.org/10.22214/ijraset.2025.72170

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

Climate-Aware Predictive Forecasting for Maharashtra Crops: Integrating Statistical and Machine Learning Approaches

Dr. Gresha Bhatia[1], Vishakha Singh[2], Manasi Sharma[3], Anushka Shirode[4]
Department of Computer Engineering, VESIT, Mumbai

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

I. INTRODUCTION

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

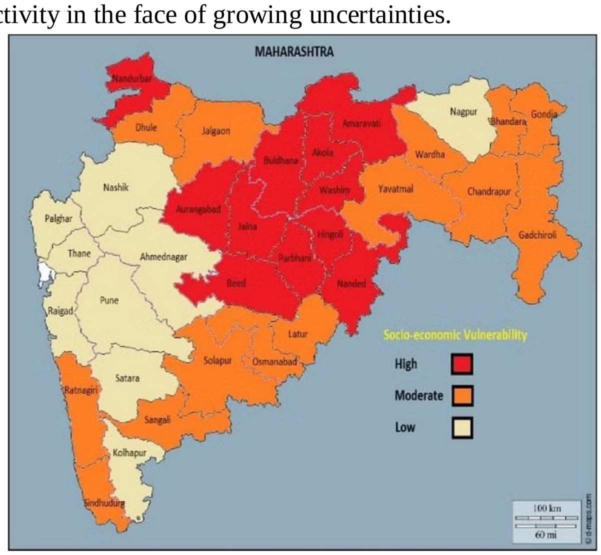


Fig.1 District-wise socioeconomic vulnerability in Maharashtra (Source: Deccan Herald)

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

Using data from 1966 to 2023, this study applies correlation and STL decomposition, and the Apriori algorithm to identify key trends and associations in agriculture. It also evaluates forecasting models like RFR, SVR, and GBR to determine their accuracy in predicting crop yield and production. By comparing their performance, the study offers insights into the most effective methods for agricultural forecasting under changing climatic conditions.

II. RELATED WORK

Paper [1], titled "Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques" employs a Multilayer Perceptron (MLP) to predict crop yield using rainfall data from 1901 to 2002, with the dataset split into 60%-40% for training and evaluation, assessed via MSE and NMSE.

The study's merits include its ability to capture nonlinear relationships in climate data and its emphasis on data preprocessing for effective feature extraction. However, it has notable demerits, such as the limited dataset, which ends in 2002, making predictions less relevant for current climate trends, and the lack of comparison with other advanced ML models like Random Forest or XGBoost, which could have provided a more comprehensive analysis.

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

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

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

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

Our study is important because it gives a clear, district-wise picture of how climate change is affecting farming in Maharashtra—something that hasn’t been done in this detail before. While most research looks at state-level trends, this study breaks it down by crop, district, and weather patterns, making it useful for both farmers and decision-makers. It also combines advanced tools from data science, like machine learning models (GBR, SVR, RFR), pattern mining (Apriori), and STL analysis to study seasonal trends and compare real data with forecasts. This kind of detailed checking is rare in farm research. Most studies look only at the past, but this one looks ahead to help prevent future problems, making it a smart and forward-thinking approach.

III.METHODOLOGY

To analyze the impact of NPK fertilizers and weather on yield and production, correlation is used. SHAP helps explain the importance of features and how each parameter influences yield.

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

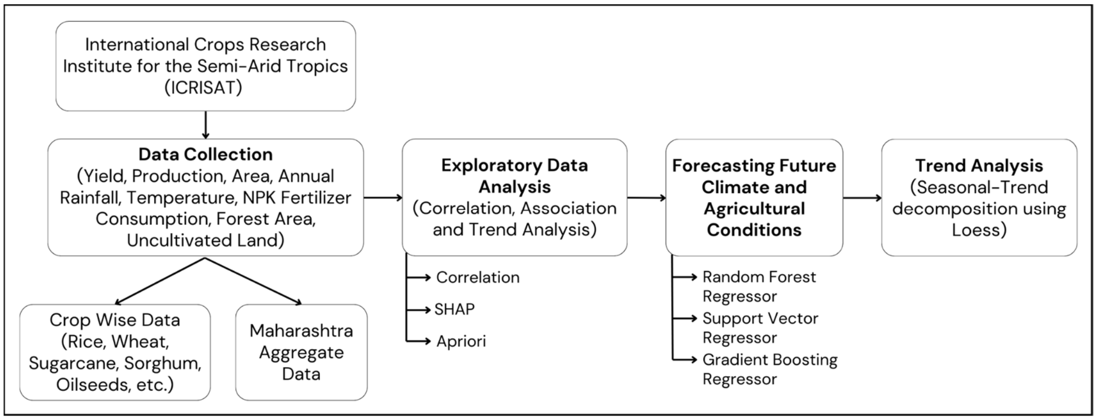


Fig.2 Block Diagram

Fig. 2 outlines the methodology used to analyze and forecast the impact of climatic and agricultural factors on crop yield and production. The study begins with data collection from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), focusing on Maharashtra's agricultural and environmental data, including yield, production, area, rainfall, temperature, NPK fertilizer consumption.

Exploratory Data Analysis (EDA) is performed to identify trends and relationships [6]. Correlation between yield and production with climatic features is derived. SHAP is a game-theoretic approach to explain the output of any machine learning model. It is used to find the top 10 most influential factors for each crop's yield and production. The Apriori algorithm is used to uncover associations between key attributes, helping to identify influential factors in agricultural productivity. For forecasting, predictive models like Random Forest (RFR) and Support Vector Regressor (SVR) capture complex relationships, while Gradient Boosting (GBR) enhances accuracy by optimizing weak models. The performance of these models is compared to determine the most effective approach for estimating future agricultural trends, and Seasonal-Trend Decomposition using Loess (STL) captures seasonal variations. This methodology offers a structured framework for understanding past patterns and making informed decisions about future agricultural planning.

IV. IMPLEMENTATION

A. Data Collection & Preprocessing

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

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

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

To handle missing values in the dataset, forward fill (ffill) imputation is applied to maintain temporal consistency across the time series.

B. Statistical Modeling & Pattern Analysis

To understand the relationships between agricultural factors, correlation analysis was applied to determine the strength and direction of associations between target attributes and Yield and Production. SHAP [8] was then used to assess how different attributes influence the two target variables, Yield and Production, providing a more granular understanding of feature importance and their impact on the predictions. To uncover hidden associations, the Apriori algorithm [9] was applied to identify frequent itemsets between Area, Irrigation, and Nutrient Consumption with Yield and Production.

Continuous data was binarized, and association rules were generated with minimum support (0.3) and confidence (0.5). These relationships were visualized through directed graphs, providing insights into key agricultural dependencies.

C. Regression Models

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

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

D. Model Evaluation & Forecasting

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

E. STL Decomposition for Trend Comparison

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

V. RESULTS

A. Correlation

1) Aggregate Dataset

The correlation analysis for the aggregate Yield data reveals a positive correlation with Nitrogen Consumption (tons) and a negative correlation with Minimum Temperature (°C). For Production, there is a positive correlation with Total Consumption (tons) and a negative correlation with Minimum Temperature (°C).

2) Crop Dataset

TABLE I. CORRELATION OF CROP YIELDS WITH CLIMATIC FACTORS AND FERTILIZER USE

Crop	Strongest Positive Correlation	Value	Strongest Negative Correlation	Value
Rice	Precipitation	0.498	Max Temperature	-0.407
Wheat	Nitrogen per HA of NCA	0.556	Nitrogen Share in NPK	-0.229
Sorghum	Nitrogen per HA of NCA	0.124	Precipitation	-0.556
Pearl Millet	Irrigated Area	0.257	Precipitation	-0.148
Maize	Nitrogen per HA of NCA	0.484	Max Temperature	-0.030
Chickpea	Phosphate per HA of GCA	0.493	Nitrogen Share in NPK	-0.303
Pigeonpea	Phosphate per HA of GCA	0.492	Potash Share in NPK	-0.161
Minor Pulses	Nitrogen per HA of NCA	0.258	Potash Share in NPK	-0.106
Groundnut	Phosphate per HA of GCA	0.337	Max Temperature	-0.465
Sesamum	Phosphate per HA of GCA	0.392	Max Temperature	-0.397
Oilseeds	Nitrogen per HA of NCA	0.454	Max Temperature	-0.060
Sugarcane	Phosphate per HA of GCA	0.394	Max Temperature	-0.391
Cotton	Phosphate per HA of GCA	0.336	Max Temperature	-0.046

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

Overall, Table I shows that nitrogen and phosphate usage generally support crop yields, while extreme temperatures and imbalanced fertilizer shares often suppress them.

B. SHAP (SHapley Additive Explanations)

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

1) Aggregate Dataset

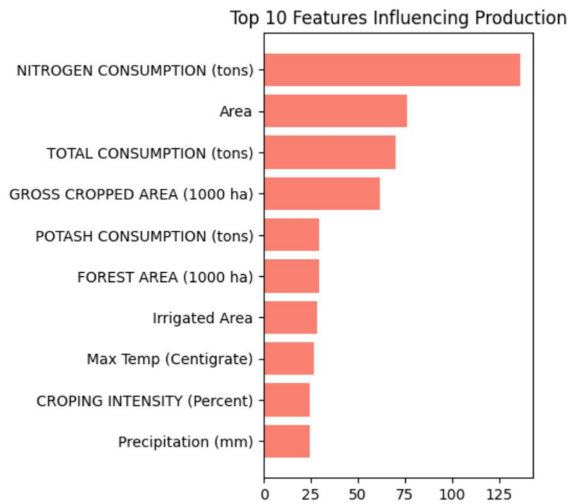


Fig.3 Top 10 features influencing Production

The most important factors for Production have been found as shown in Fig. 3. The x-axis represents the magnitude of feature impact, where larger values indicate that these features play a stronger role in influencing the model's production predictions. It shows that features such as Nitrogen Consumption (tons) and Area have the most significant impact, as indicated by their longer bars.

2) Crop Dataset

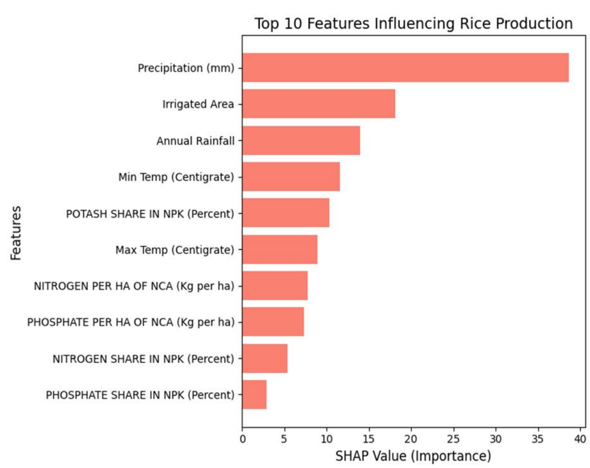


Fig.4 Top 10 features influencing Rice Production

Fig. 4 depicts that Precipitation (mm) and Irrigated Area have the most significant impact on Rice Production. Similarly, this has been done for the other crops as listed in the implementation part.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

C. Apriori Algorithm

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

1) Aggregate Dataset

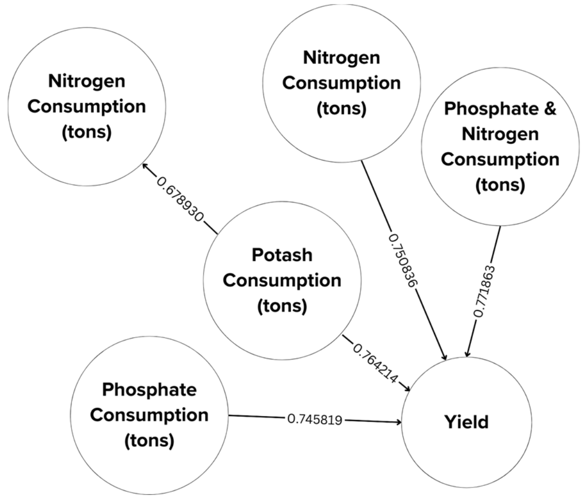


Fig.5 Apriori Rules for Yield

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

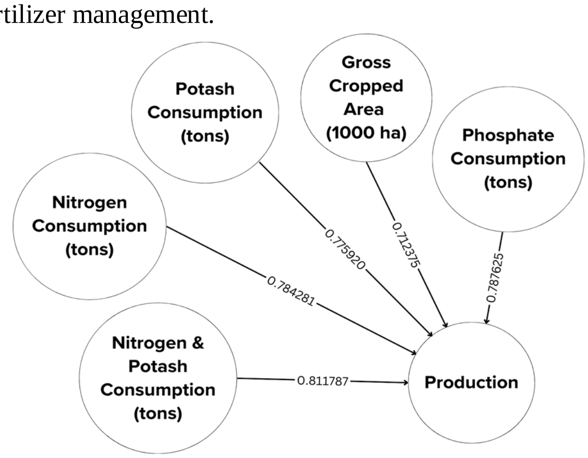


Fig.6 Apriori Rules for Production

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

2) Crop Dataset

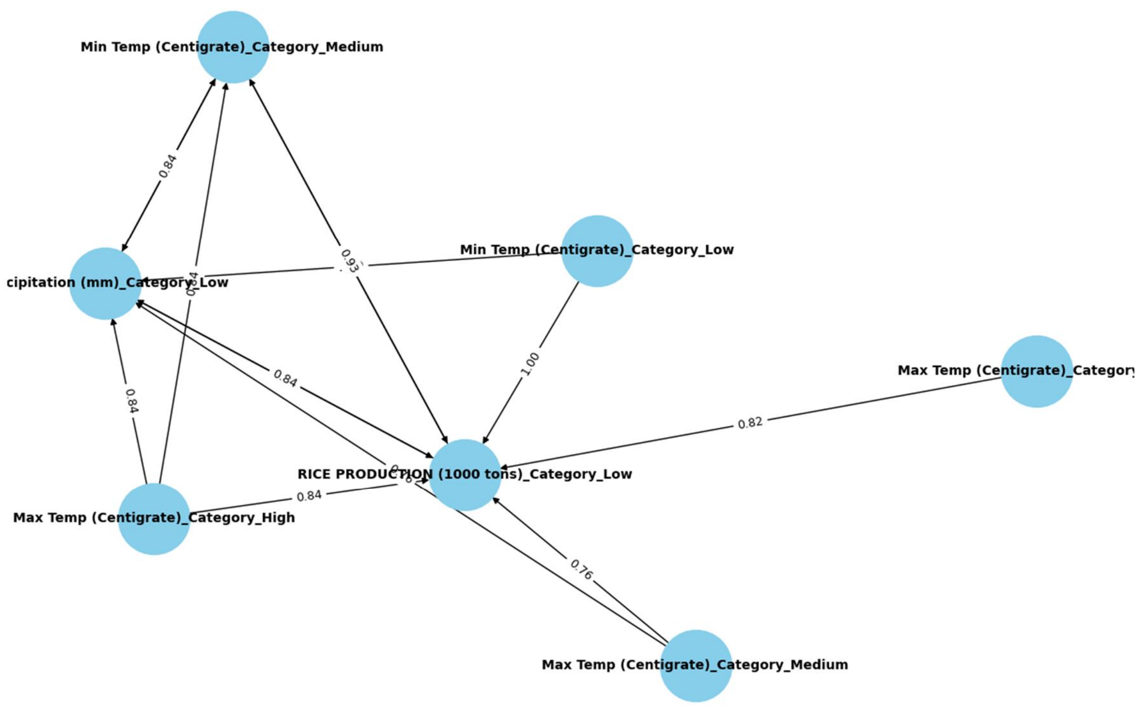


Fig.7 Apriori Rules for Rice Production

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

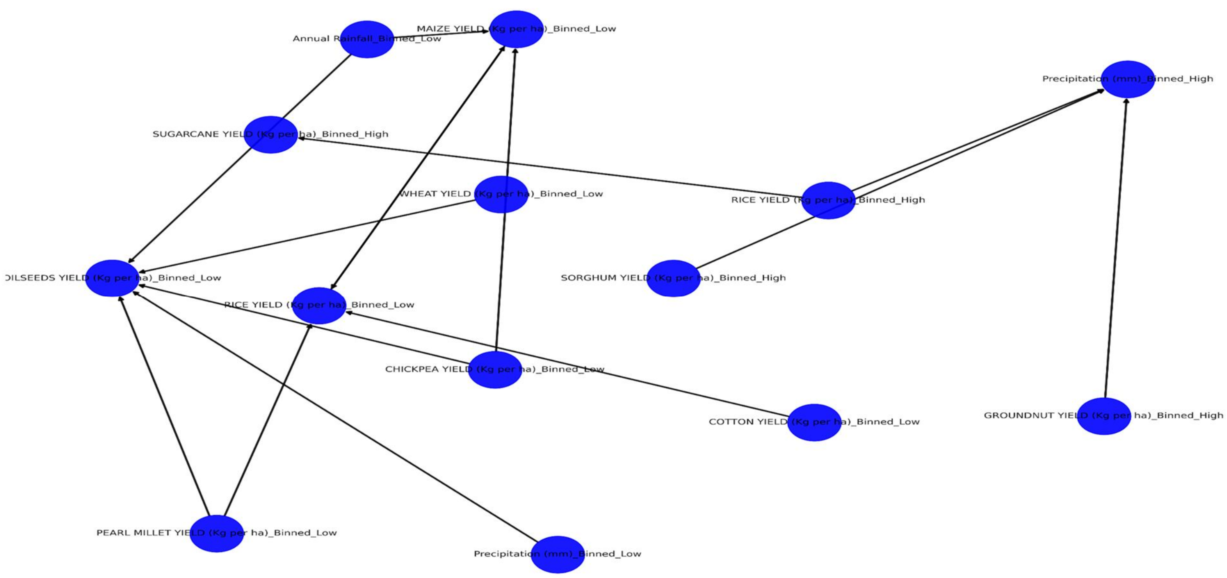


Fig.8 Apriori Rules for Ahmednagar District

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

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

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

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

D. Time Series Forecasting Using Regression: Support Vector, Random Forest, and Gradient Boosting

TABLE II. Comparison Of Forecasting Algorithms

	Aggregate data						Crop data					
	RFR		SVR		GBR		RFR		SVR		GBR	
	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
Ahmednagar	.939	36.42	.990	99.32	.988	5.14	.972	1.35	.992	1.39	.978	9.42
Akola	.943	41.21	.989	61.08	.982	3.72	.970	0.69	.970	5.30	.983	8.12
Amarawati	.927	44.33	.993	57.68	.964	7.06	.953	4.71	.986	9.20	.979	0.28
Beed	.909	36.96	.990	96.18	.957	2.00	.969	9.05	.970	4.31	.967	4.04
Bhandara	.900	05.38	.991	27.73	.949	8.72	.954	7.54	.944	7.77	.951	5.77
Buldhana	.943	38.75	.991	35.51	.958	0.31	.966	9.60	.969	5.44	.972	5.18
Chandrapur	.934	92.70	.993	61.73	.967	8.83	.965	8.23	.967	4.28	.972	4.92
Dhule	.931	91.47	.994	60.82	.958	10.7	.983	0.73	.987	6.10	.988	0.22
Jalgaon	.935	12.98	.985	23.15	.966	5.69	.966	7.03	.986	0.38	.973	3.22

Kolhapur	.926	26.53	.987	45.14	.957	3.06	.978	4.66	.959	7.95	.950	3.76
Nagpur	.928	00.54	.989	54.46	.972	5.90	.972	2.52	.972	5.64	.965	9.46
Nanded	.909	21.52	.988	37.51	.963	7.82	.959	3.06	.968	3.94	.970	9.08
Nashik	.922	32.58	.990	14.08	.972	5.55	.974	14.8	.958	1.95	.962	1.23
Osmanabad	.938	25.43	.990	01.85	.975	5.82	.953	9.08	.973	9.07	.976	4.02
Parbhani	.934	53.12	.992	59.75	.954	4.18	.981	3.28	.986	5.33	.987	0.50
Pune	.904	57.07	.985	05.92	.945	9.73	.971	5.64	.982	3.16	.984	0.43
Sangli	.922	007.0	.987	46.06	.962	7.81	.973	9.26	.983	8.92	.987	4.32
Satara	.932	31.01	.992	56.83	0.96	9.47	.958	4.14	.969	7.96	.952	0.16
Solapur	.919	42.82	.985	65.94	.954	2.76	.975	3.43	.975	0.26	.985	0.27
Yeotmal	.922	58.13	.988	32.27	.974	7.66	.961	7.04	.977	7.99	.966	4.96

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

For Aggregate data, while SVR (Support Vector Regressor) occasionally achieves marginally higher R^2 values (e.g., 0.990 compared to GBR's 0.988 in Ahmednagar), GBR demonstrates significantly lower RMSE values (e.g., 15.14 vs. SVR's 799.32), highlighting its superior practical accuracy and reliability. Similarly, for Crop data, GBR consistently outperforms both RFR (Random Forest Regressor) and SVR, achieving the highest R^2 values (e.g., 0.978 in Ahmednagar) and the lowest RMSE values (e.g., 9.42) across most regions.
GBR's consistent performance underscores its robustness and generalization across diverse datasets, making it the preferred choice for accurate and reliable agricultural yield and production forecasting.

TABLE III. FORECASTED VALUES FOR YIELD, PRODUCTION, AND AREA

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

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

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

TABLE IV. Fertilizer And Climate Factor Risk Analysis For Districts

District	Rainfall Risk	Precipitation Risk	Max Temp Risk	Min Temp Risk	Irrigation Risk	Nitrogen Risk	Phosphate Risk	Potash Risk	Overall Climate Risk
Ahmednagar	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	High
Akola	Increasing Variability	Stable	Rising	Rising	Stable	Declining	Declining	Stable	Medium
Nashik	Declining	Declining	Rising	Rising	Stable	Declining	Stable	Stable	Medium
Pune	Increasing Variability	Stable	Rising	Rising	Declining	Stable	Declining	Stable	Medium
Jalgaon	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Stable	High
Kolhapur	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	Low
Satara	Declining	Declining	Rising	Rising	Stable	Stable	Stable	Stable	Medium
Solapur	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	High
Nagpur	Increasing Variability	Stable	Rising	Rising	Stable	Declining	Stable	Stable	Low
Amrawati	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	High
Nanded	Declining	Declining	Rising	Rising	Declining	Declining	Declining	Stable	High
Osmanabad	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	High
Buldhana	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	Medium
Chandrapur	Declining	Declining	Rising	Rising	Declining	Stable	Stable	Declining	High
Beed	Declining	Declining	Rising	Rising	Declining	Declining	Stable	Declining	High
Yeotmal	Increasing Variability	Stable	Rising	Rising	Stable	Stable	Stable	Stable	Medium

Table IV shows that most districts face declining rainfall, increasing drought risk, while Akola, Pune, and Kolhapur face flood risks due to rainfall variability. Rising temperatures across all districts intensify heat stress on crops. High-risk districts like Ahmednagar and Solapur show declining irrigation, worsening climate vulnerability. Stable irrigation in Nagpur, Kolhapur, and others helps reduce this risk. Fertilizer use, especially nitrogen and phosphate, is falling in many areas, impacting soil fertility. Overall, high-risk districts include Ahmednagar and Aurangabad, while Kolhapur and Nagpur are low-risk.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

E. STL (Seasonal and Trend decomposition using Loess)

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

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

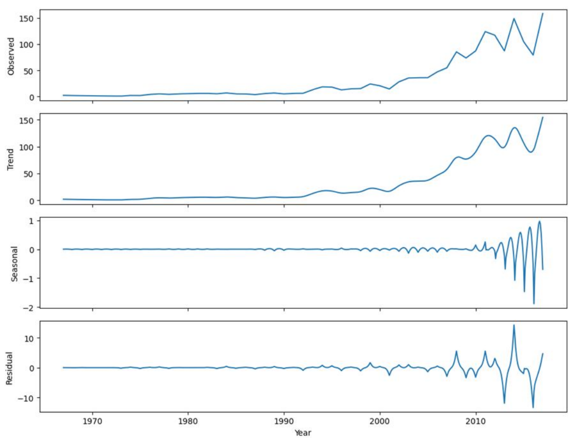


Fig.9 STL for Maize Production for original data

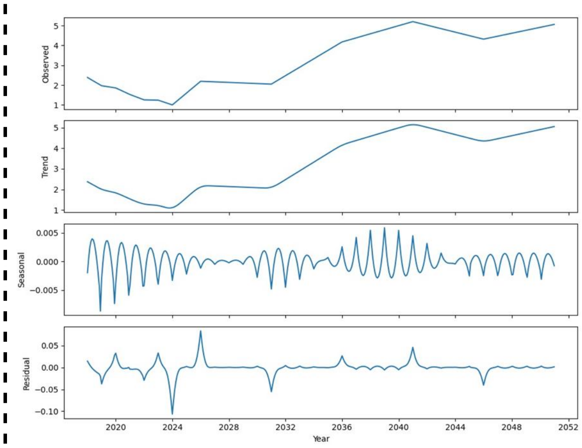


Fig.10 STL for Maize Production for future forecast

Fig. 9 shows the STL decomposition of current maize production, showcasing a significant upward trend in recent years with noticeable seasonality and fluctuations. The decomposition separates the observed production into its trend, seasonal, and residual components, providing insights into the underlying patterns and potential future production. Fig. 10 validates the result, and the forecast follows the upward trend with minimal residual.

VI. CONCLUSION

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

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

Regression models, including RFR, SVR, and GBR, were employed to predict future values, with GBR emerging as the most effective across both datasets, and GBR's predictions align with the original data's trend, ensuring the model's accuracy and reliability. For the Aggregate dataset, the R^2 scores were 0.927 for RFR, 0.989 for SVR, and 0.963 for GBR, while the RMSE values were 378 for RFR, 573.33 for SVR, and 17.16 for GBR. For the Crop dataset, the R^2 scores were 0.967 for RFR, 0.976 for SVR, and 0.974 for GBR, with RMSE values of 21.23 for RFR, 72.23 for SVR, and 17.5 for GBR. GBR is the best model overall because it consistently achieves the lowest RMSE while maintaining high R^2 scores across both datasets.

By applying STL (Seasonal and Trend decomposition using Loess) to GBR's predictions, the data is decomposed into trend, seasonal, and residual components. The alignment with the trend shows accuracy and reliability, ensuring that GBR's forecasts are consistent with historical patterns and capable of providing meaningful insights for future agricultural planning. The study presents a comprehensive evaluation of agricultural trends across 20 districts in Maharashtra, revealing mounting climate-related risks and production stress. The analysis identifies declining crop yields, climate variability, soil fertility concerns, and irrigation challenges that have begun to reshape the agricultural landscape of the region.

These insights offer critical guidance for stakeholders seeking to ensure long-term agricultural sustainability. One of the most prominent trends observed is the consistent decline in rice and wheat production across several districts, including Ahmednagar, Aurangabad, Amravati, Jalgaon, and Solapur. This reduction appears to stem from rising temperatures, unpredictable rainfall, and a steady decrease in irrigated land. These crops, especially wheat, are highly sensitive to heat stress and require stable water availability, making them particularly vulnerable to the current climate conditions. In contrast, maize shows a more complex pattern. While some districts report a decline, others show increased production. This inconsistency may be due to maize's greater resilience to varying rainfall patterns, suggesting that it may be a more adaptable cereal crop in the context of climate change. Sugarcane and cotton, in comparison, demonstrate relatively greater resilience. Cotton's inherent drought resistance makes it less dependent on consistent irrigation, allowing it to thrive even in water-scarce regions. Sugarcane, though water-intensive, has maintained a more stable output, though it remains at risk if irrigation continues to decline.

Climate trends across the districts show a troubling rise in both maximum and minimum temperatures. This temperature increase exacerbates heat stress and directly affects yields, especially for temperature-sensitive crops like wheat. In parallel, rainfall patterns have become increasingly erratic. Districts such as Jalgaon and Satara face sharp rainfall declines, while others like Akola, Pune, and Kolhapur are experiencing irregular cycles of drought and flooding. This climate unpredictability disrupts planting and harvesting schedules and threatens to destabilize annual crop output. The situation is further aggravated by the ongoing reduction in irrigation coverage, especially in districts such as Ahmednagar, Solapur, Aurangabad, and Amravati. As more land becomes reliant on rainfall, water-intensive crops face heightened production risks. Simultaneously, the use of fertilizers—particularly nitrogen, phosphate, and potash—has been decreasing in several districts. This decline may be attributed to shifts in farming practices, rising costs, or underlying soil health problems, which could contribute to long-term reductions in yield due to nutrient deficiencies. Based on the combined effect of these variables, districts have been classified into different risk categories. High-risk districts include Ahmednagar, Aurangabad, Jalgaon, Solapur, Amravati, Nanded, Osmanabad, Chandrapur, Beed, and Gadchiroli. These areas suffer from a convergence of declining rice and wheat yields, increasing temperatures, reduced irrigation, and rainfall inconsistency. Medium-risk districts such as Akola, Nashik, Pune, Satara, Wardha, Latur, Buldhana, and Yavatmal experience moderate but growing impacts from climate variability, though some of them maintain relatively better irrigation or soil health. Kolhapur and Nagpur emerge as low-risk districts, with more stable climate trends, irrigation availability, and crop production. This study reinforces that climate change is no longer a distant threat—it is already affecting agricultural systems in tangible ways. Without timely interventions, the situation will worsen, particularly for smallholder farmers who lack access to adaptive technologies or data-driven support systems. The analysis highlights the urgent need for better water management, widespread adoption of climate-resilient crop varieties, and investment in precision agriculture. Monitoring soil health and managing fertilizer use effectively are equally critical to preserving long-term productivity. The findings of this study are valuable for a broad range of stakeholders. Farmers and agricultural cooperatives can use these insights to adjust their cropping strategies, adopt efficient irrigation practices, and anticipate yield risks based on district-specific climate patterns. In many high-risk regions, farmers may benefit from transitioning to less water-intensive crops or introducing drought-tolerant varieties. For policymakers and government agencies, the results offer a roadmap for designing proactive agricultural policies. These could include targeted subsidies, improved irrigation infrastructure, climate-smart extension services, and early warning systems for extreme weather. Most importantly, this research underscores the need to shift from reactive to preventive agricultural planning to secure food and economic security for rural communities.

REFERENCES

[1] Rainfall Prediction for Enhancing Crop-Yield based on Machine Learning Techniques, S. Malathy, C. N. Vanitha, Kotteswari, S. V, S. P and M. E, 2022. [Online]. Available: https://doi.org/10.1109/ICAAIC53929.2022.9792793

[2] A Creative Use of Machine Learning for Crop Prediction and Analysis, R. C. Mahore and N. G. Gadge, 2022. [Online]. Available: https://doi.org/10.1109/ICWITE57052.2022.10176256

[3] Influence of Causal Inference For Crop Prediction, K. Srivastava, A. Sawarkar, S. Nawale and A. R. Agrawal, 2024. [Online]. Available: https://doi.org/10.1109/ICCCNT61001.2024.10725416

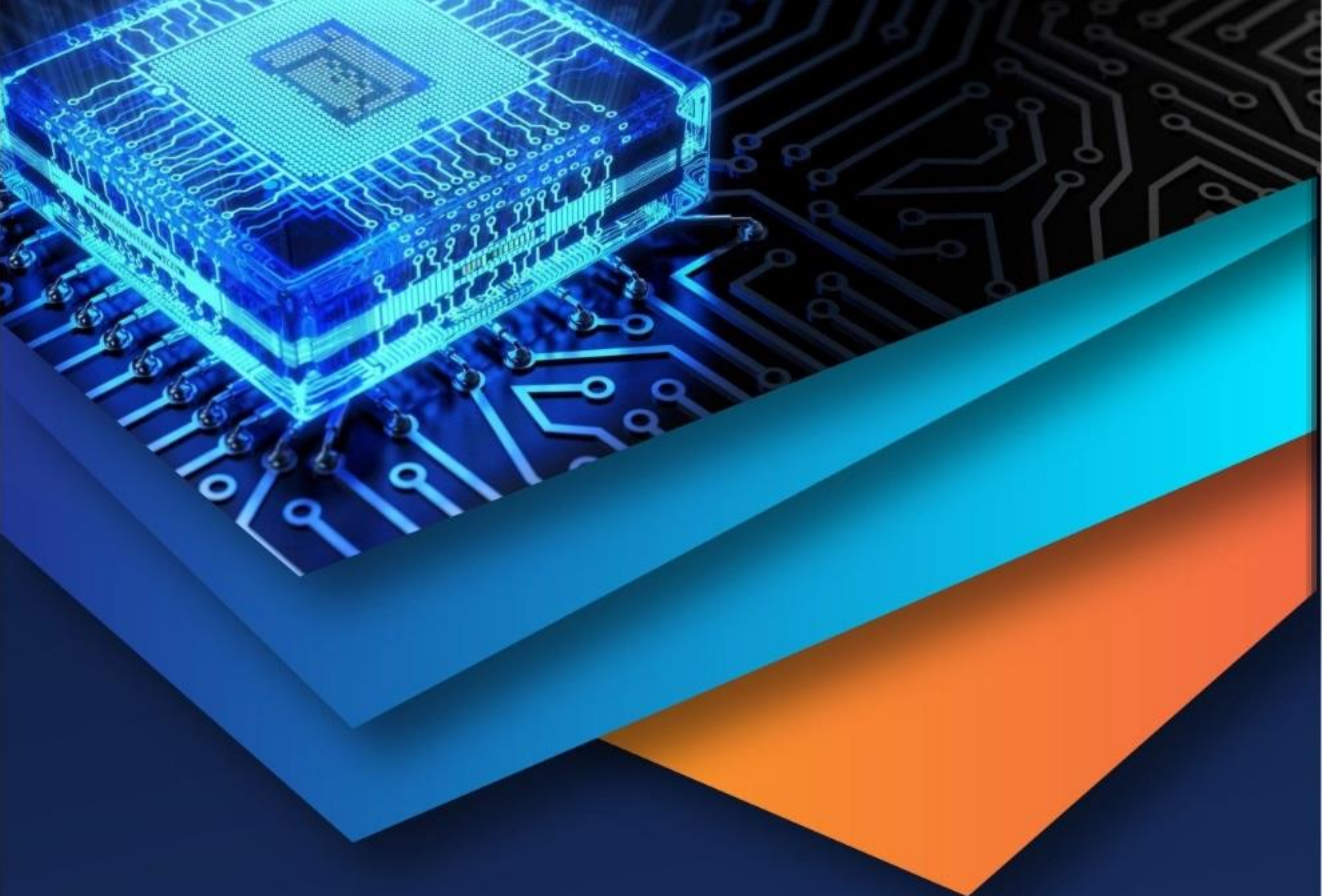
[4] A Case Study on the Application of Machine Learning to the Process of Crop Forecasting, S. A. Jiwani, J. Shinde, B. Bag, R. Nayak, R. K. Shial and U. Ghugar, 2024. [Online]. Available: https://doi.org/10.1109/OTCON60325.2024.10688298

[5] Climate Forecasting: Long Short Term Memory Model using Global Temperature Data, P. Akhila, R. L. S. Anjana and M. Kavitha, 2022. [Online]. Available: https://doi.org/10.1109/ICCMC53470.2022.9753779

[6] What is Exploratory Data Analysis?, GeeksforGeeks, 2024. [Online]. Available: https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/

[7] ICRISAT - Science of Discovery to Science of Delivery, International Crops Research Institute for the Semi-Arid Tropics, 2024. [Online]. Available: https://www.icrisat.org/

[8] SHAP: A Comprehensive Guide to Shapley Additive Explanations, GeeksforGeeks, 2024. [Online]. Available: https://www.geeksforgeeks.org/shap-a-comprehensive-guide-to-shapley-additive-explanations/
[9] Apriori Algorithm, IBM, 2024. [Online]. Available: https://www.ibm.com/think/topics/apriori-algorithm
[10] Random Forest Regression in Python, GeeksforGeeks, 2024. [Online]. Available: https://www.geeksforgeeks.org/random-forest-regression-in-python/
[11] RandomForestRegressor, scikit-learn, 2024. [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.RandomForestRegressor.html
[12] Support Vector Regression (SVR) Using Linear and Non-Linear Kernels, GeeksforGeeks, 2024. [Online]. Available: https://www.geeksforgeeks.org/support-vector-regression-svr-using-linear-and-non-linear-kernels-in-scikit-learn/
[13] SVR, scikit-learn, 2024. [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.svm.SVR.html
[14] Gradient Boosting in Machine Learning, GeeksforGeeks, 2024. [Online]. Available: https://www.geeksforgeeks.org/ml-gradient-boosting/
[15] GradientBoostingRegressor, scikit-learn, 2024. [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html#sklearn.ensemble.GradientBoostingRegressor
[16] Regression Metrics, GeeksforGeeks, 2024. [Online]. Available: https://www.geeksforgeeks.org/regression-metrics/
[17] STL, ArcGIS Insights, 2024. [Online]. Available: https://doc.arcgis.com/en/insights/latest/analyze/stl.htm



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY