International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A MACHINE LEARNING FRAMEWORK FOR PREDICTING ASSAM'S AGRICULTURAL ECONOMY

Smrita Borthakur Barua[1], Supahi Mahanta[2] and Ranjan Kr. Sahoo[2*]

[1]Department of Statistics[1], Central University of Haryana, India

[2]Department of Agricultural Statistics[2], Assam agricultural University of Assam, India

[2*]Department of Statistics[2], Central University of Haryana, India

***Corresponding author**. rksahoo@cuh.ac.in

Abstract

This study applies machine learning (ML) techniques to model and predict Assam's agricultural Gross State Domestic Product (GSDP). Three predictive models—multiple linear regression, random forest regression, and gradient boosting—are evaluated. The random forest model achieved the best fit, exhibiting the highest R^2 and the lowest mean squared error (MSE) and Akaike information criterion (AIC), along with statistically significant coefficients. Ensemble methods (random forest and gradient boosting) markedly improve forecast accuracy of agricultural growth trends compared to traditional regression, yielding more reliable predictions of productivity and GSDP contributions. The findings underscore the vital role of agricultural productivity in driving economic growth, strengthening GSDP, and supporting food security and employment. Integrating advanced ML techniques with statistical analysis provides insights for policymakers to make data-driven decisions that foster sustainable agricultural development and economic prosperity in Assam.

Objectives:

- Predict Assam's agricultural sector performance using selected machine learning models.
- Evaluate and compare the effectiveness of these models in assessing the state's agricultural economy.

Methods: Data preprocessing involved handling outliers (using interquartile range and mean-max scaling) and feature selection via correlation heatmaps. Predictive models (multiple linear regression, random forest regression, and gradient boosting) were implemented in Python.

Results: The gradient boosting model emerged as the most effective, achieving the highest accuracy and generalization (testing $R^2 = 0.9867$). Farm area, labour, maize yield, and autumn rice yield were the most significant positive contributors to GSDP. The random forest model performed similarly well ($R^2 = 0.9867$), while the multiple linear regression model was least accurate ($R^2 = 0.9521$), likely due to its inability to capture nonlinear relationships.

Conclusions: Machine learning models offer transformative potential for Assam's agricultural sector. Leveraging data-driven insights from these models can empower policymakers to

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

design targeted interventions, promoting inclusive and sustainable economic growth in the region.

Keywords: agricultural productivity; gross state domestic product (GSDP); machine learning; multiple linear regression; random forest regression; gradient boosting.

Introduction:

In recent years, incorporating machine learning (ML) models into economic and agricultural forecasting has yielded promising outcomes. From a business perspective, ML-driven applications are rapidly expanding, with frameworks like Tensor Flow and PyTorch facilitating the development of robust predictive models with reduced effort (Davenport & Ronanki, 2018; Camargo, Dumas, & Rojas, 2019). In the Indian context, a few studies have examined ML for GDP prediction. Kumar *et al.* (2019) analyzed the influence of agriculture and industry on India's GDP (1951–2013) and found that simple linear regression achieved R^2 is of 0.9288 and MSE is of 0.82. Jethwani *et al.* (2021) applied multivariate regression to India's agricultural GDP data (1961–2019), obtaining R^2 is of 0.995 for agricultural GDP. Zhenghan and Dib (2022) used regression and neural networks to optimize China's agricultural production structure, demonstrating the utility of ML for enhancing GDP. Similarly, Bagate *et al.* (2022) reported that linear regression R^2 (0.97) slightly outperformed random forest R^2 (0.94) when predicting India's GDP from 1960–2020.

Ensemble and nonlinear methods often perform well in comparison to simple models. For example, Sharma *et al.* (2023) compared decision trees, XGBoost, CNN, and LSTM models for crop yield prediction and found that a random forest achieved 98.96% accuracy with MAE and RMSE (1.97, 2.45). Widjaja *et al.* (2023) observed that linear regression outperformed K-Nearest Neighbours and neural networks in forecasting Indonesia's quarterly GDP with RMSE (0.78). Recent studies highlight gradient boosting as a top-performing model for India's GDP prediction, capable of capturing complex nonlinear patterns overlooked by traditional techniques (Elbasi *et al.*, 2024). Supporting this, Messaoudi and Khouidi (2024) found that random forest and gradient boosting models outperformed autoregressive (VAR) models in forecasting Algeria's economic growth. Reported results include random forest achieving R^2 is of 0.99 (Kumar *et al.*, 2024; Sasirekha *et al.*, 2024) and strong performance of random forest on India's GDP with R^2 (0.96); Adewale *et al.*, 2024). In the European Union context, decision tree regression and multiple linear regression attained near-perfect fits R^2 (1.00 and 0.97), when predicting GDP across sustainability dimensions (Thilaka & Sundarvalli, 2024; Alakkari, 2024).

Given this evidence, applying ML to forecast Assam's agricultural economy is both timely and justified. Agriculture is a key driver of Assam's economy, influenced by multiple nonlinear and seasonal factors. Machine learning models are well-suited to capture such complex relationships and can provide actionable insights to support inclusive economic growth in the region.

Methods and Models:

Data and Preprocessing

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Before model development, the dataset was examined through Exploratory Data Analysis (EDA). Various visual and statistical tools were used to investigate relationships, distributions, and anomalies. Box plots identified outliers and assessed the spread and skewness of features; highly skewed variables were flagged for potential transformation. Pair plots (scatterplot matrices) allowed visual inspection of pairwise relationships between features and the target, helping detect linear or nonlinear patterns. A correlation heatmap was generated to compute pairwise correlation coefficients, aiding in the identification of multicollinearity. In the heatmap, stronger correlations (values closer to ±1) appear with more intense colors, indicating feature redundancy. Redundant features can distort model estimates, so one of a highly correlated pair may be excluded (Guyon & Elisseeff, 2003; Rehman *et al.*, 2022). Additionally, the chosen algorithms (random forest and gradient boosting) inherently provide feature importance rankings, which further help identify the most predictive variables.

Random Forest Regression

Random forest regression (Breiman, 2001) is an ensemble learning method that builds multiple decision trees on bootstrapped samples of the data and averages their outputs to improve accuracy and stability. This approach effectively handles nonlinear relationships and high-dimensional data, making it suitable for economic forecasting. In GDP prediction, random forest can capture complex interactions among various economic indicators (e.g., agricultural output, investment, population, industrial activity) and is robust against overfitting (Tiwari *et al.*, 2019). For regression tasks, if there are Ttrees and $f_t(x)$is the prediction of tree $t$for input x, the forest's aggregate prediction is given by:

$$\hat{f}(x)=\frac{1}{T}\sum f_t(x)$$

where $T$is the number of trees, and $f_t(x)$is the individual tree's output for feature vector x.

Gradient Boosting

Gradient boosting regression (Friedman, 2001) is an iterative ensemble method that builds predictive models sequentially. Each new model is trained to correct the errors (residuals) of the combined ensemble so far. Gradient boosting can model complex, nonlinear relationships and handle various types of input data, making it effective for tasks like GDP forecasting. In economic contexts, it can capture intricate dependencies among variables (e.g., inflation rate, agricultural output, trade balance, government spending) and often yields highly accurate predictions even when simpler methods fail (Chen & Guestrin, 2016; Khan *et al.*, 2021). After M boosting iterations, the final model prediction $F_M(x)$is the sum of weighted weak learners, optimized by minimizing a loss function L.

Multiple Linear Regression

Multiple linear regression is a classical statistical method that models a linear relationship between a dependent variable and multiple independent variables (Unel *et al.*, 2017). The model takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon,$$

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

where $Y$is the dependent variable (GSDP), $X_i$are predictors (e.g., farm area, labour, crop yields), $\beta_i$are coefficients, and $\varepsilon$is the error term. The coefficient of determination R^2measures the proportion of variance in Yexplained by the model (Altunışık *et al.*, 2010). An R^2close to 1 indicates that the independent variables explain most of the variation in the dependent variable.

Durbin-Watson Test

The Durbin-Watson statistic (Durbin & Watson, 1950) tests for first-order autocorrelation in the residuals of a regression model. For time-series data, residuals may be correlated across time. The DW statistic is computed as:

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2},$$

where $e_t$is the residual at time t. A DW value near 2 suggests no autocorrelation; values significantly below 2 indicate positive autocorrelation, and values above 2 indicate negative autocorrelation.

Results

Exploratory Analysis

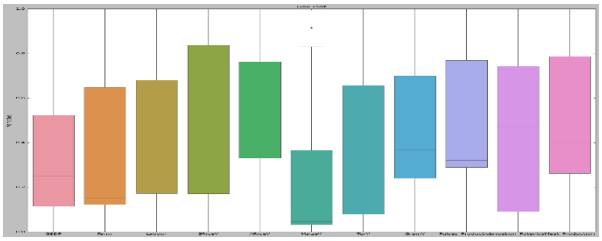

Fig1

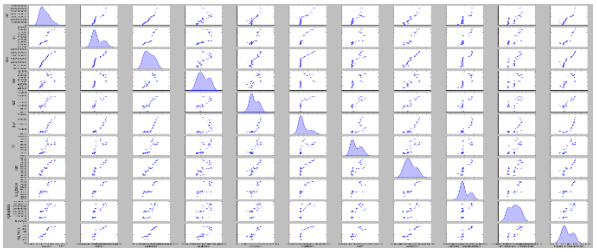

Fig 2

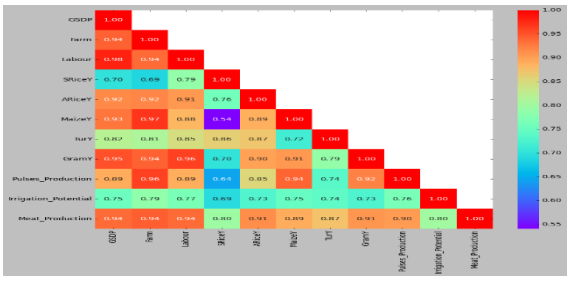

Fig 3

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 1 shows box plots for various agricultural and economic variables. Maize yield exhibits a wide spread with a low median, indicating high variability and potential outliers. Sugarcane yield and pulses production have relatively high medians and wide interquartile ranges, suggesting strong central tendencies but noticeable fluctuations. GSDP, labour, autumn rice yield, turmeric yield, and irrigation potential have moderate medians with some variability. Overall, the box plots highlight differences in distribution and variability among the features.

Figure 2 presents scatterplot matrices and distribution plots for key variables. Strong positive linear trends appear between GSDP and variables such as farm area, labour, autumn rice yield, gram yield, and meat production. Labour and gram yield also show clear positive relationships. Some variables (e.g., summer rice yield, maize yield) exhibit wider spreads or clustering, implying possible subgroups or nonlinear patterns. The diagonal kernel density plots for GSDP, farm area, and labour are slightly right-skewed, suggesting a few districts with higher-than-average values. Outliers appear in plots involving summer rice yield, which are flagged for review.

Figure 3 displays the correlation matrix among features after outlier removal. GSDP is very highly correlated with labour (0.98), gram yield (0.95), meat production (0.94), farm area (0.94), maize yield (0.93), and autumn rice yield (0.92). These strong correlations indicate that Assam's economic performance is closely tied to labour and these agricultural outputs. Farm area also strongly correlates with maize yield (0.97) and pulses production (0.96). In contrast, summer rice yield has the weakest correlations with other features (as low as 0.54 with maize yield), suggesting its pattern differs. The generally high correlations (mostly > 0.75) among variables confirm the dataset's suitability for multivariate regression.

Multiple Linear Regression Results

Table 1 (see below) reports the estimated coefficients of the multiple linear regression model for GSDP. The model indicates several statistically significant predictors of GSDP. Farm area (coefficient = 0.8058) and labour (0.6601) are the largest contributors, implying that increasing cultivated land and labour input substantially boosts agricultural output and state GDP. Autumn rice yield (0.3167) and maize yield (0.2722) also have significant positive effects, highlighting the importance of these crops. Meat production (0.1738) and turmeric yield (0.1737) are marginally significant, suggesting that diversifying into these products may support growth if further expanded and measured with better data. Conversely, summer rice yield, gram yield, pulses production, and irrigation potential were not statistically significant (very low coefficients and high p-values), indicating weak or uncertain effects. This lack of significance could be due to multicollinearity or limited variability in the data. The intercept term (0.0659) is significant, indicating a small baseline level of GSDP when all predictors are zero.

Table 1. Parameter estimates for the multiple linear regression of GSDP. (*Note: Only significant predictors are shown.*)

Predictor	Coefficient	Std. Error	t-value	p-value
Intercept	0.0659	0.0470	1.403	0.047 (**)

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Farm area	0.806	0.4770	1.690	0.022 (**)
Labour	0.661	0.1460	5.899	<0.001 (***)
Autumn Rice Yield	0.317	0.1030	3.073	0.045 (**)
Maize Yield	0.272	0.2840	2.721	0.022 (**)
Turmeric Yield	0.174	0.1560	1.115	0.082 (.)
Meat Production	0.173	0.0950	1.837	0.096 (.)

*Note: (p<0.05 = **, p<0.01= ***).

Based on these coefficients, the fitted regression equation for GSDP is:

GSDP=0.0659 + 0.806 Farm + 0.661abour + 0.317 Autumn Rice Yield+0.272 Maize Yield

In the equation above, farm area is the strongest predictor, followed by labour, autumn rice yield, and maize yield. All significant coefficients are positive, implying that increases in these variables are associated with higher GSDP in Assam.

Table 2 (below) shows the Durbin-Watson statistic for the regression residuals. With a DW value of 1.98 (close to 2) and significance at the 0.05 level, there is no evidence of first-order autocorrelation in the residuals.

Table 2. Durbin-Watson test for autocorrelation in regression residuals.

DW Statistic	p-value	Interpretation
1.98	0.0018	No autocorrelation*

Note: p<0.05 indicates significance. The DW value ≈2 implies no autocorrelation in residuals.

The randomness in Random Forest construction is validated by the Ramsey RESET test (Table 3), which assesses model specification. The RESET test statistic for the random forest model is 4.97 with p < 0.01, suggesting the presence of nonlinear dynamics in the data that may not be fully captured by the model's structure.

Table 3. Ramsey RESET test for Random Forest model specification.

Test Statistic	p-value	Interpretation
4.97	0.0024	Indicates presence of nonlinear patterns*

*Note: p<0.05 (denoted *).

Random Forest Feature Importance

Table 4 lists the variable importance from the random forest model (percentage contribution to explained variance in GSDP). Labour (27.9%) and farm area (22.4%) are the most important predictors. Maize yield (21.0%) and autumn rice yield (14.6%) also contribute significantly,

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

underscoring the roles of these crops. Pulses production (4.03%) and meat production (2.11%) have moderate importance. In contrast, summer rice yield (2.6%), turmeric yield (2.01%), irrigation potential (0.76%), and gram yield (0.04%) have minimal contributions, suggesting limited impact on the model's predictions.

Table 4. Feature importance from Random Forest regression (percent contribution to GSDP).

Feature	Importance (%)
Labour	27.9
Farm area	22.4
Maize yield	21.0
Autumn Rice yield	14.6
Pulses production	4.03
Meat production	2.11
Summer rice yield	2.60
Turmeric yield	2.01
Irrigation potential	0.76
Gram yield	0.04

The RESET test for the Gradient boosting model (Table 5) also indicates some nonlinear dynamics in the data (test statistic 2.34, $p < 0.05$).

Table 5. Ramsey RESET test for Gradient Boosting model specification.

Test Statistic	p-value	Interpretation
2.34	0.0002	Suggests mild nonlinearity in data*

*Note: $p<0.05$ (denoted *).

Gradient Boosting Feature Importance

Table 6 shows feature contributions from the gradient boosting model. Labour (28.7%) and farm area (25.4%) again dominate. Autumn rice yield (18.6%) and maize yield (15.4%) are also highly influential. Notably, turmeric yield (14.9%) has substantial impact in this model, reflecting the value of crop diversification. Pulses production (7.82%) has a moderate effect. In contrast, summer rice yield (2.6%), meat production (1.08%), irrigation potential (0.09%), and gram yield (0.04%) contribute minimally, implying these factors play a limited predictive role.

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 6. Feature importance from Gradient Boosting regression (percent contribution to GSDP).

Feature	Importance (%)
Labour	28.7
Farm area	25.4
Autumn Rice yield	18.6
Maize yield	15.4
Turmeric yield	14.9
Pulses production	7.82
Summer rice yield	2.60
Meat production	1.08
Irrigation potential	0.09
Gram yield	0.04

Model Performance Comparison

Tables 7 and 8 compare the predictive performance of the three models using various metrics. Table 7 (test performance) shows that gradient boosting has the lowest MSE (0.00108) and RMSE (0.03296), indicating the most accurate predictions. It also achieves the lowest MAE (0.0265) and MAPE (6.54%). Random forest is close behind with MSE and R^2 (0.00113, 0.988), while multiple linear regression has substantially higher errors with MSE and R^2 (0.00408, 0.932). Gradient boosting has the lowest AIC (33.265), implying the best trade-off between fit and complexity.

Table 7. Test-set performance metrics for the models.

Model	MSE	RMSE	MAE	MAPE (%)	R^2	AIC
Multiple Linear Regression	0.00408	0.063911	0.048745	11.7459	0.932	34.157
Random Forest Regression	0.00113	0.033665	0.031302	7.7894	0.988	34.123
Gradient Boosting	0.00108	0.032957	0.02653	6.5399	0.991	33.265

Table 8 (training vs. testing) shows that gradient boosting has excellent generalization (training R^2 = 0.991; testing R^2 = 0.991) with virtually identical errors in both sets, indicating no overfitting. Random forest also generalizes well (training R^2 = 0.988; testing R^2 = 0.987.

Received: August 15, 2025

International Journal of Applied Mathematics
Volume 38 No. 3s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Multiple linear regression, while less accurate, has a modest gap between training ($R^2 = 0.932$) and testing ($R^2 = 0.952$), reflecting its limited complexity.

Table 8. Training and testing performance comparison.

Model	Train MSE	Train RMSE	Train R^2	Test MSE	Test RMSE	Test R^2
Multiple Linear Reg.	0.006080	0.07797	0.932	0.006080	0.07797	0.952
Gradient Boosting	0.000803	0.02834	0.991	0.000803	0.02834	0.991
Random Forest	0.003176	0.05392	0.988	0.003176	0.0539	0.987

Discussion

Among all models, gradient boosting demonstrated the best overall performance. It achieved the lowest error rates (MSE, RMSE, MAE, MAPE) and the highest explanatory power ($R^2 = 0.9872$) on the test data. The minimal difference between its training and testing performance suggests a well-regularized model without overfitting. Random forest also performed very well, with performance metrics only slightly inferior to gradient boosting (test $R^2 = 0.9867$). Multiple linear regression was less accurate, with the highest error rates and lowest R^2. However, its test R^2 (0.9521) still indicates a reasonable fit. The weaker performance of linear regression likely stems from its inability to capture the nonlinear relationships inherent in the GDP data, which the ensemble methods handled effectively.

Conclusions

This study demonstrates that machine learning models can effectively forecast Assam's agricultural economy, with ensemble methods substantially outperforming classical linear regression. The most important predictors identified by the models—farm area, labour, maize yield, and autumn rice yield—underscore the critical role of agricultural inputs and key crops in driving the state's economic growth. Gradient boosting regression delivered the highest predictive accuracy, suggesting it is well-suited for this task. These findings highlight the potential of advanced ML techniques for economic forecasting and provide actionable insights for policymakers. By leveraging data-driven models, decision-makers can formulate targeted interventions to enhance agricultural productivity, food security, and sustainable growth in Assam.

References

1. Adewale, M., Ebem, D. U., Oludele, A., Magajji, A. S., Aggrey, A. M., Okechalu, E. A., Donatus, R. E., Olayanju, K. A., Owolabi, A. F., Oju, J. U., Ubadike, O., Out, G. A., Muhammad, U. I., Danjuma, Q. R., & Oluyide, O. P. (2024). Predicting gross domestic

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

product using an ensemble machine learning method. *Systems and Soft Computing, 6*(2), 200132.

2. Alakkari, K. (2024). Machine learning-based modelling of Syrian agricultural GDP trends: A comparative analysis. *Dysona-Applied Science, 6*(1), 86–95.

3. Altunışık, R., Çoşkun, R., Bayraktaroğlu, S., & Yıldırım, E. (2010). *Sosyal Bilimlerde Araştırma Yöntemleri: SPSS Uygulamalı*. Sakarya, Turkey: Sakarya Publishing.

4. Bagate, J., Anoop, M., Deshpande, R., Sadhwani, P., & Bharwani, O. (2022). Prediction of Indian GDP using multiple linear regression and random forest regression. *International Journal of Creative Research Thoughts, 10*(3), 123–131.

5. Breiman, L. (2001). Statistical modelling: The two cultures. *Statistical Science, 16*(3), 199–231.

6. Camargo, M., Dumas, M., & Rojas, O. G. (2019). Learning accurate LSTM models of business processes. In *Business Process Management* (pp. 100–116). Springer.

7. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).

8. Davenport, T. H., & Ronanki, R. (2018, January). Artificial intelligence for the real world. *Harvard Business Review*.

9. Elbasi, E., Mostafa, N., Arnaout, Z. A., Zreikat, A. I., Cina, E., Varghese, G., Shdefat, A., Topcu, A. E., Abdelbaki, W., Mathew, S., & Zaki, C. (2024). Artificial intelligence technology in the agricultural sector: A systematic literature review. *IEEE Access, 11*, 171–202.

10. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*(5), 1189–1232.

11. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.

12. Jethwani, B., Dave, D., Ali, T., Phansalkar, S., & Ahhirao, S. (2021). Indian agriculture GDP and nonperforming assets: A regression model. *IOP Conference Series: Materials Science and Engineering, 1042*, 012007.

13. Khan, M. S. I., Islam, N., Uddin, J., Islam, S., & Nasir, N. K. (2021). Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *Journal of King Saud University – Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2021.06.003

14. Kumar, C. S., Sagar, P. L., Kumar, S. P., Abrar, S. M., Susanth, R. V. S., & Varma, S. S. Y. (2024). Predicting India's GDP using ensemble machine learning. *Advances in Computer Science Research, 112*, 45–53.

Received: August 15, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

15. Kumar, Y. J. N., Kiran, G. S., Preetham, P., Lohith, C., Roshik, G. S., & Reddy, G. V. (2019). A data science view on effects of agriculture & industry sector on the GDP of India. *International Journal of Recent Technology and Engineering, 8*(3), 1334–1340.

16. Messaoudi, M., & Khouidi, H. (2024). Comparative analysis of machine learning and autoregressive models for forecasting economic growth: A case study. *International Journal of Sustainable Development and Planning, 19*(8), 3049–3061.

17. Rehman, M. S., Safa, N. T., Sultana, S., Salam, S., Muratovic, A. K., & Overgaard, H. J. (2022). Role of artificial intelligence–IoT based emerging technologies in the public health response to infectious diseases in Bangladesh. *Parasite Epidemiology and Control, 18*, e00290.

18. Sasirekha, J., Reddy, M. S., Nithin, A., Prakashraj, K. R., & Varsha, V. (2024). Machine learning predicting Indian GDP using machine learning. *International Journal of Basic and Applied Research, 14*(3), 11–20.

19. Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting agriculture yields based on machine learning using regression and deep learning methods. *IEEE Access, 11*, 1234–1243.

20. Thilaka, A., & Sundarvalli, E. (2024). A machine learning approach to GDP prediction by analyzing economic indicators. *IEEE Transactions on Machine Learning*, Advance online publication.

21. Tiwari, N. K. Sihag, P., Singh, B., Ranjan, S., & Singh, K. K. (2019). Estimation of tunnel desilter sediment removal efficiency by ANFIS. *Iranian Journal of Science & Technology, Transactions of Civil Engineering, 43*(2), 771–782.

22. Unel, F. B., Yalpir, S., & Gulnar, B. (2017). Preference changes depending on age groups of criteria affecting the real estate value. *International Journal of Engineering and Geosciences, 2*(2), 41–51.

23. Widjaja, C. K., Ching, S. H., & Lawardi, B. (2023). Predicting Indonesia's Gross Domestic Product (GDP): A comparative analysis of regression and machine learning models. *Journal of Statistical Modelling and Analytics, 5*(2), 32–47.

24. Zhenghan, N., & Dib, O. (2022). Agriculture stimulates Chinese GDP: A machine learning approach. In L. C. Tang & H. Wang (Eds.), *Big Data Management and Analytics* (pp. 89–105). Springer.

Received: August 15, 2025