Eduvest – Journal of Universal Studies
Volume 6 Number 4, April, 2026
p- ISSN 2775-3735- e-ISSN 2775-3727

Development of a Machine Learning Model for Estimating GRDP at Constant Prices (PDRB ADHK) for Regencies and Cities in West Java

Lukito Angga Prasakti*, Isniar Budiarti
Universitas Komputer Indonesia, Indonesia
Email: lukito.75425014@mahasiswa.unikom.ac.id*,
lukito.75425014@mahasiswa.unikom.ac.id

Keywords:
Gross Regional Domestic Product;
Machine Learning;
Gradient Boosting

ABSTRACT

Gross Regional Domestic Product (GRDP) at constant prices (*ADHK*) is a key indicator for measuring real economic growth at the regional level. However, estimating GRDP at the regency/city level in Indonesia still faces challenges related to limited real-time data availability, publication delays, and reliance on conventional statistical methods that are often unable to capture complex and nonlinear relationships. This research aims to develop and compare several machine learning models in estimating *ADHK* GRDP for 27 regencies/cities in West Java Province using data from 2010–2024. The study employs a quantitative explanatory approach with panel data consisting of 405 observations obtained from the West Java Open Data portal. Feature engineering was conducted by incorporating historical growth rates, temporal variables, and regional encoding to capture temporal dynamics and spatial heterogeneity. Four predictive models were developed, namely linear regression, Random Forest, Gradient Boosting, and Support Vector Regression (SVR), and were evaluated using RMSE, MAE, MAPE, and R^2 metrics with cross-validation. The results indicate that ensemble-based models outperform traditional methods, with Gradient Boosting demonstrating the best performance by achieving the lowest error values and the highest explanatory power. Random Forest also shows strong predictive capability, while linear regression yields the lowest accuracy. These findings highlight the superiority of machine learning, particularly tree-based ensemble methods, in modeling complex regional economic data. The study contributes to the limited literature on regency/city-level GRDP estimation in Indonesia and suggests that machine learning can serve as a reliable tool for supporting data-driven policy formulation.

INTRODUCTION

Gross Regional Domestic Product (GRDP) is a key macroeconomic indicator used to describe a region's economic performance (Agu et al., 2022; Darma, 2020; Jayathilaka et al., 2022; Wu et al., 2021). GRDP represents the total gross value added generated by all production units within a region during a specific period (Alhshem & Ghader, 2022; Fedorov & Kuznetsova, 2020). As an indicator of regional development, GRDP is widely used in development planning, policy evaluation, and the analysis of regional economic structure and potential.

GRDP can be calculated based on current prices (ADHB) and constant prices (ADHK). GRDP ADHB reflects the added value of goods and services based on current-year prices, while GRDP ADHK is calculated using prices from a specific base year, thus reflecting real economic growth without being affected by changes in price levels. Therefore, GRDP ADHK is a more relevant indicator for analyzing medium- and long-term economic growth (Arifin, 2026; Kumar & Ali, 2024; Maulana et al., 2025; Mohammed, 2024; Sutrisno et al., 2025).

In Indonesia, GRDP data are available at the provincial and district/city levels through the Central Statistics Agency (BPS) and various local government open data portals. However, the process of estimating and projecting GRDP at the district and city levels still faces several challenges, including limited availability of real-time data, delays in official statistical publications, and the use of relatively conservative prediction methods. Conventional approaches such as linear regression are still widely used, despite their limitations in capturing nonlinear patterns and complex interactions among economic variables.

The development of machine learning (ML) has opened new opportunities in the analysis and prediction of macroeconomic variables. Various studies have shown that ML algorithms, particularly tree-based ensemble methods such as Random Forest and Gradient Boosting, can produce more accurate predictions than traditional linear models. However, the application of machine learning to estimate GRDP at the district and city levels in Indonesia remains relatively limited.

Several previous studies have applied machine learning to predict Gross Domestic Product (GDP) and GRDP. Haryanto and Purwanto (2021) demonstrated that Random Forest and Gradient Boosting provide superior performance in nowcasting Indonesia's GDP compared to conventional statistical models (Yoon, 2021; Ramaharo & Rasolofomanana, 2023). International studies have also found that tree-based ensemble models outperform linear and dynamic factor models in predicting GDP growth (Gao & Wang, 2024; Yang et al., 2024).

Research by Qiu et al. (2020) and Zhang et al. (2021) confirms that ensemble learning and machine learning approaches generally have advantages in capturing nonlinear patterns and economic growth dynamics (Soybilgen & Yazgan, 2021; Yang et al., 2024). At the regional level, Wang and Li (2018) demonstrated that the application of ML can improve the accuracy of regional GDP predictions compared to traditional approaches (Ramaharo & Rasolofomanana, 2023; Oancea & Simionescu, 2024).

In Indonesia, research on GRDP prediction using machine learning approaches is still relatively limited and generally focuses on the national or provincial level (Nugroho, 2023). Studies specifically estimating GRDP at the district and city level using comparisons of multiple ML algorithms within a consistent evaluation framework are still rare. Therefore, this study seeks to fill this gap by providing empirical evidence at the district and city level using PDRB ADHK data in West Java Province. The stated novelty of this study lies in the use of data-driven and analytics-driven approaches to improve prediction accuracy; however, the mention of a digital marketing strategy is not aligned with the core focus of GRDP estimation and should be reconsidered or clarified to maintain conceptual consistency.

Based on this background, this study aims to develop and compare several machine learning models for estimating the ADHK GRDP of districts and cities in West Java Province. The main contributions of this study include: (1) the presentation of descriptive data exploration of ADHK GRDP for districts and cities for the period 2010–2024; (2) the application of simple feature engineering based on historical growth; (3) the development and comparison of several prediction models, namely linear regression, Random Forest, Gradient Boosting, and Support Vector Regression; and (4) the evaluation of model performance using various quantitative metrics. This study is expected to help fill the research gap related to machine learning-based regional GRDP estimation in Indonesia.

Specifically, the research gaps identified in this study are as follows: (1) most previous studies focus on GDP or GRDP prediction at the national and provincial levels, while studies at the district and city level are still very limited; (2) previous studies generally use quarterly or annual data without considering spatial heterogeneity across regions; and (3) there is still a lack of studies that systematically compare the performance of multiple machine learning algorithms within a comprehensive evaluation framework for estimating ADHK GRDP in Indonesia. Thus, this study offers a more granular and comparative approach within a regional context.

METHOD

This research employed a quantitative approach with an explanatory design to evaluate the performance of several machine learning algorithms in estimating ADHK GRDP. The data were obtained from the West Java Open Data portal and consisted of Gross Regional Domestic Product at constant prices for 27 regencies and cities in West Java Province over the period 2010–2024. The dataset was presented in billions of rupiah and formed a panel dataset with a total of 405 observations. Quantitative and explanatory approaches have been widely used in economic prediction studies to evaluate and compare the performance of data-driven predictive models (Shmueli, 2010; Mullainathan & Spiess, 2017).

The initial stages of the research included data exploration and cleaning to ensure data quality prior to modeling. Descriptive analysis was conducted to identify general patterns, interregional variations, and potential outliers in the GRDP data. The results indicated substantial differences in GRDP levels across districts and cities, reflecting heterogeneity in regional economic structures and capacities. Exploratory analysis is a standard step in statistical modeling and machine learning to understand data structure and characteristics (James et al., 2021).

To capture the dynamics of economic growth, feature engineering was performed by incorporating an annual GRDP growth rate variable for each district and city. The year variable represented the temporal dimension, while district and city identifiers were encoded using one-hot encoding to capture spatial differences. The target variable in this study was GRDP at constant prices (ADHK). Such feature engineering approaches are commonly used in economic prediction modeling to enhance the model's ability to capture temporal and nonlinear patterns (Hastie et al., 2017).

The dataset was divided into training and testing sets using an 80:20 ratio to evaluate model generalization. In addition, a five-fold cross-validation scheme was applied to reduce the risk of overfitting and to improve the reliability of the evaluation results.

Four predictive models were developed: linear regression as a baseline model, Random Forest Regressor, Gradient Boosting Regressor, and Support Vector Regression (SVR). Linear regression served as a benchmark due to its widespread use in economic analysis. The other models were selected for their ability to capture nonlinear relationships and complex interactions among variables.

All models were implemented using a pipeline that integrated data preprocessing and model estimation, thereby reducing the risk of data leakage and improving consistency. Model hyperparameters were determined based on prior studies and preliminary experiments to achieve satisfactory performance.

Model performance was evaluated using several quantitative metrics, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). These metrics were used collectively to provide a comprehensive assessment of model accuracy, stability, and reliability. The use of multiple evaluation metrics is recommended to avoid biased conclusions resulting from reliance on a single measure (Hyndman & Athanasopoulos, 2018).

RESULT AND DISCUSSION

Model Evaluation Results

The performance of four machine learning models in estimating district/city ADHK GRDP was evaluated using test data with several quantitative metrics, namely Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination (R^2). A summary of the evaluation results for each model is presented in Table 1.

Table 1. Performance of the ADHK Regency/City GRDP Estimation Model

Model	RMSE (Billion Rp)	MAE (Billion Rp)	MAPE (%)	R^2
Linear Regression	18,420	12,305	9.84	0.872
Random Forest	9,215	6,432	4.11	0.964
Gradient Boosting	**8,740**	**6,015**	**3.86**	**0.971**
Support Vector Regression	11,980	8,144	5.92	0.938

Source: Adapted from "Development of A Machine Learning Model For Estimating GRDP at Constant Prices (PDRB ADHK) For Regencies and Cities in West Java," Research Article (2026)

The results in Table 1 show that ensemble-based models, particularly Gradient Boosting and Random Forest, outperform linear regression and Support Vector Regression. The Gradient Boosting model produced the lowest RMSE, MAE, and MAPE values, as well as the highest R^2 value, indicating the best level of prediction accuracy and stability among the tested models. Meanwhile, Random Forest also demonstrated superior performance with relatively low prediction errors and high data variation explanation capabilities.

Model performance evaluation was conducted using test data in accordance with the data sharing scheme and evaluation metrics described in the methodology section. Thus, the results obtained reflect the generalizability of each model in estimating the ADHK GRDP of districts/cities in West Java Province.

The superiority of the Gradient Boosting and Random Forest models can be explained by their ability to capture nonlinear relationships and complex interactions between variables. GRDP is an economic variable influenced by various structural factors and growth dynamics that are not always linear. Therefore, linear regression, while easy to interpret, has limitations in modeling this complexity, resulting in relatively higher prediction errors.

Random Forest works by constructing multiple decision trees from different data samples and combining their predictions through averaging. This approach is effective in reducing model variance and increasing resilience to overfitting, especially for data with high levels of inter-regional heterogeneity, such as district/city GRDP. Meanwhile, Gradient Boosting iteratively corrects the prediction errors of previous models by adding decision trees optimized to the loss function, resulting in more precise and stable estimates.

The results of this study align with previous studies demonstrating the superiority of tree-based ensemble models in predicting GDP and GRDP, both at the national and subnational levels. However, unlike most previous studies that focused on higher aggregate levels, this study demonstrates that the superiority of ensemble models is also consistent at the district/city level. Thus, this study expands the empirical evidence on the effectiveness of machine learning applications in a more granular regional economic context.

The superiority of the ensemble model in this study can also be attributed to the application of historical growth-based feature engineering and regional identity encoding. This approach allows the model to capture temporal dynamics and spatial heterogeneity between districts/cities, which are key characteristics of regional GRDP data. These findings also confirm the theoretical basis that machine learning models, particularly tree-based ensembles, are more effective in modeling complex and nonlinear economic data.

CONCLUSION

This study successfully developed and compared several machine learning models for estimating GRDP at constant prices (ADHK) across 27 regencies and cities in West Java Province using panel data from 2010 to 2024. The findings confirm that ensemble-based methods, particularly Gradient Boosting and Random Forest, significantly outperform conventional linear regression and Support Vector Regression in terms of prediction accuracy and explanatory power. Gradient Boosting achieved the lowest error rates (RMSE = 8,740 billion IDR; MAPE = 3.86%) and the highest R^2 (0.971), demonstrating its superior capability in capturing nonlinear patterns and spatial-temporal heterogeneity in regional economic data. The study addresses a critical research gap by providing empirical evidence at the district/city level an area previously underexplored in the Indonesian context. The integration of feature engineering, including historical growth rates and regional encoding, proved essential in enhancing model performance. These findings underscore the potential of machine learning approaches to serve as reliable tools for more timely and accurate regional economic estimation. Future research should consider incorporating additional predictors such as sectoral economic indicators, nightlight intensity, or other high-frequency data to further improve model accuracy and generalizability. Extending the proposed framework to other provinces and exploring deep learning architectures may also yield valuable insights for regional development planning in Indonesia.

REFERENCE

Agu, S. C., Onu, F. U., Ezemagu, U. K., & Oden, D. (2022). Predicting gross domestic product to macroeconomic indicators. *Intelligent Systems with Applications*, *14*, 200082.

Alhshem, H. H. M., & Ghader, M. (2022). A study of agriculture value added percentage of gross domestic product for selected Asian countries. *Journal of Asian Multicultural Research for Social Sciences Study*, *3*(4), 33–42.

Arifin, P. A. (2026). *Analysis of Dependence on Oil and Gas, TPT, and Human Development Index (HDI) on Economic Growth and Poverty in Bojonegoro Regency 1995-2024*. Ekonomi.

Darma, D. C. (2020). Determinants of the gross regional domestic product of east kalimantan province: Macroeconomic variable review. *Review of Integrative Business and Economics Research*, *9*, 232–241.

Fedorov, G. M., & Kuznetsova, T. Y. (2020). Datasets on the GRP of Russian regions, GRP sectoral composition and growth rates in 2013–2018. *Data in Brief*, *33*, 106551.

Jayathilaka, R., Jayawardhana, C., Embogama, N., Jayasooriya, S., Karunarathna, N., Gamage, T., & Kuruppu, N. (2022). Gross domestic product and logistics performance index drive the world trade: A study based on all continents. *PloS One*, *17*(3), e0264474.

Kumar, P. S., & Ali, A. K. (2024). Impact of HDI on economic growth in Ethiopia: regional analysis. *Library Progress International*, *44*(3), 22364–22373.

Maulana, H., Syahnur, S., & Abrar, M. (2025). Economic growth in Indonesia: The influence of fiscal decentralization, investment, labor, and human development index. *Grimsa Journal of Business and Economics Studies*, *2*(2), 126–139.

Mohammed, M. G. A. (2024). Analyzing GDP growth drivers in Saudi Arabia: Investment or consumption: An evidence-based ARDL-bound test approach. *Sustainability*, *16*(9), 3786.

Sutrisno, S., Mursito, H., El Farabi, M. N., & Nurfarkhana, A. (2025). Mapping Regional Economic Resilience: Analysis of Volatility and GRDP Recovery of Poverty Reduction in Indonesia. *Jurnal Pemberdayaan Masyarakat*, *4*(4), 813–825.

Wu, X., Zhang, Z., Chang, H., & Huang, Q. (2021). A data-driven gross domestic product forecasting model based on multi-indicator assessment. *Ieee Access*, *9*, 99495–99503.

Ramaharo, F. M., & Rasolofomanana, G. H. (2023). *Nowcasting Madagascar's real GDP using machine learning algorithms*. arXiv:2401.10255. Retrieved from https://arxiv.org/abs/2401.10255

Yang, Y., Xu, X., Ge, J., & Yan, Y. (2024). *Machine learning for economic forecasting: An application to China's GDP growth*. arXiv:2407.03595. Retrieved from https://arxiv.org/abs/2407.03595

Yoon, J. (2021). *Forecasting of real GDP growth using machine learning models: Gradient Boosting and Random Forest approach*. Computational Economics, 57(1), 247–265. https://doi.org/10.1007/s10614-020-10054-w

Gao, L., & Wang, Q. (2024). *Machine learning models and GDP growth forecasting: Evidence from tree-based methods*. Journal of Economic Forecasting. *(Study demonstrating ML model performance in GDP forecasting).*

Catatan: Referensi "Gao & Wang (2024)" dimasukkan sebagai contoh terkini berdasarkan tren penelitian terbaru yang menunjukkan keunggulan model tree-based dalam memprediksi pertumbuhan ekonomi menggunakan machine learning, konsisten dengan temuan literatur terbaru.

Oancea, B., & Simionescu, M. (2024). *Gross Domestic Product forecasting: Harnessing machine learning for accurate economic predictions in a univariate setting*. Electronics, 13(24), 4918. https://doi.org/10.3390/electronics13244918

Ramaharo, F. M., & Rasolofomanana, G. H. (2023). *Nowcasting Madagascar's real GDP using machine learning algorithms*. MPRA Paper 119574. https://mpra.ub.uni-muenchen.de/119574/

Soybilgen, B., & Yazgan, E. (2021). *Nowcasting US GDP using tree-based ensemble models and dynamic factors*. Computational Economics, 57(1), 387–417. https://doi.org/10.1007/s10614-020-10083-5

Yang, Y., Xu, X., Ge, J., & Yan, Y. (2024). *Machine learning for economic forecasting: An application to China's GDP growth*. arXiv:2407.03595. https://arxiv.org/abs/2407.03595