Indonesia Downstreaming Investment and Labor Absorption Regional-Sectoral Analysis Using Gradient Boosting and SHAP

Nauval Alfarizi, Muhammad Iqbal

Abstract

This study examines investment realization and domestic labor absorption in Indonesia using BKPM data for Q2–Q3 2026. The two datasets initially comprised 63,663 records; auditing and preprocessing retained 36,801 active records. Exploratory and spatial analyses investigated investment distribution, dominant sectors, downstreaming intensity, and regional labor absorption. DummyRegressor, Random Forest, and XGBoost regression models were evaluated using MAE, RMSE, and R^2. XGBoost with the raw TKI target achieved the highest R^2 (0.333) and lowest RMSE (193.391), whereas log1p-transformed XGBoost achieved the lowest MAE (37.972). SHAP interpretation identified investment value as the most influential feature, followed by business sector, country of origin, and province. The findings reveal regional and sectoral heterogeneity and a nonlinear relationship between investment scale and predicted labor absorption. This interpretable predictive approach provides insights into how investment characteristics relate to domestic employment patterns across Indonesian regions and sectors. The analysis does not establish causal effects or impacts.

***Keywords**:* Investment, Downstreaming, Labor Absorption, XGBoost, SHAP.

Nauval Alfarizi[1]
[1]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: nauvalalfarizi026@gmail.com [1]

Muhammad Iqbal[2]
[2]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Indonesia
e-mail: muhammadiqbal@dosen.pancabudi.ac.id [2]
2nd International Conference on Islamic Community Studies (ICICS)
Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era
https://proceeding.pancabudi.ac.id/index.php/ICIE/index

Introduction

Indonesia has increasingly positioned downstream industrialization as a strategic instrument for accelerating economic transformation and strengthening the domestic industrial base. Through the processing of natural resources into higher-value products, the downstreaming strategy seeks to reduce dependence on raw commodity exports, enhance domestic value creation, and improve Indonesia's competitiveness in global markets [1]. The implementation of this strategy requires substantial capital investment to support industrial infrastructure, processing facilities, technological development, and the expansion of manufacturing activities. Consequently, investment plays an important role in facilitating industrialization and strengthening domestic production capacity [2].

Beyond its contribution to economic growth and industrial development, the success of downstream industrialization should also be examined through its capacity to generate employment opportunities and improve labor absorption. Previous empirical research has demonstrated that investment can contribute positively to employment generation, although the relationship may vary depending on regional and economic characteristics [3]

The relationship between investment and labor absorption is not always proportional, as the employment outcomes of investment depend on the production characteristics and labor requirements of individual economic sectors. Although capital investment can expand production capacity and stimulate industrial development, investments concentrated in capital-intensive industries may generate substantial economic output without creating an equivalent number of employment opportunities[4].

In contrast, labor-intensive industries generally require a larger workforce to sustain their production activities, potentially resulting in higher employment absorption relative to the amount of capital invested. Evidence from Indonesia also indicates considerable differences in employment contributions across economic sectors. For instance, the mining sector accounted for approximately 7.4% of national GDP in 2023 but absorbed only around 1.2% of the national workforce, illustrating the relatively limited labor absorption capacity of capital-intensive activities [5]. In the context of Indonesia's investment landscape, regional and sectoral dimensions provide important perspectives for examining the distribution of investment and its relationship with domestic employment absorption. Previous research has demonstrated that investment patterns differ across geographical regions and economic sectors, indicating the importance of disaggregated analysis rather than relying exclusively on national-level aggregates [6].

Accordingly, this study focuses on the realization of investment in Indonesia during the second and third quarters of 2026 (Q2–Q3 2026), using official investment realization records published by the Ministry of Investment and Downstreaming/BKPM [7]. The analysis incorporates two regional classifications, namely Java and Outside Java, as well as two industrial classifications, downstreaming and non-downstreaming sectors, based on the classification provided in the dataset. The primary economic indicators consist of realized investment value and the number of Indonesian workers absorbed (Tenaga Kerja Indonesia/TKI) [8]. By integrating these regional, sectoral, and economic dimensions, this study aims to examine variations in investment distribution and domestic labor absorption across different categories of economic activities in Indonesia [9].

Investigated the distributional implications of economic zone development in Indonesia and identified considerable regional differences in its effects on local income inequality. More recently, machine learning approaches have been increasingly adopted to capture complex and nonlinear economic relationships Wu (2023) [10]. integrated XGBoost and SHAP to evaluate provincial economic resilience in China and identify the contribution of important economic indicators. Tang and Wei (2023) [11] . also combined machine learning to predict enterprise digital transformation and interpret the influence of financial and organizational characteristics. Furthermore, Pyo and Choi (2025) applied, alongside conventional econometric approaches to investigate nonlinear relationships between regional innovation and economic growth.

Collectively, these studies demonstrate the relevance of combining regional and sectoral economic analysis with interpretable machine learning techniques to understand complex economic patterns [12] .

Despite the growing body of research on investment, employment, and regional economic development, several research gaps remain. Previous studies have examined sectoral economic growth and labor absorption [5], as well as the relationship between foreign direct investment and regional economic growth across Indonesian provinces and sectors [6]. However, these studies have not fully integrated investment realization, downstreaming classification, regional distribution, and domestic labor absorption within a single analytical framework. Furthermore, conventional econometric approaches, including panel regression and autoregressive models, remain valuable for examining statistical relationships, but may require additional specifications to capture complex nonlinear patterns and heterogeneous interactions among economic variables [13]. Recent developments in explainable machine learning have demonstrated the potential of combining XGBoost and SHAP to identify nonlinear relationships and interpret the contributions of individual features in regional economic and industrial development. Nevertheless, the application of these approaches to explain domestic labor absorption based on official investment realization records, particularly by integrating downstreaming status and Java–Outside Java classifications [14].

The scientific contribution of this study is established through the integration of recent investment realization data, downstreaming classification, regional distribution, and domestic labor absorption within an interpretable machine learning framework. Recent research on Indonesia's resource-based industrialization has highlighted the complexity of translating large-scale industrial investment into proportional employment growth, particularly in regions experiencing rapid downstream industrial development [15].

Literature Review

Investment is widely recognized as an important component of economic development because it supports capital formation, expands production capacity, and facilitates industrial transformation. In resource-based economies such as Indonesia, downstream industrialization represents an important strategy for increasing domestic value added through the processing of natural resources. However, the expansion of industrial investment does not necessarily generate proportional employment growth, particularly when investment is concentrated in capital-intensive processing activities . examined Indonesia's nickel downstreaming strategy using input-output analysis and found that the labor multiplier declined from approximately 6,514 to 3,366 jobs per IDR 1 trillion of final demand, indicating a reduction in employment intensity despite the expansion of industrial value creation. Similarly, Muhyiddin et al (2026). [13] investigated labor market dynamics in Morowali and reported relatively low employment elasticity during the region's rapid nickel-based industrialization. These findings suggest that investment performance should not be evaluated solely through its financial value or contribution to industrial output, but also through its capacity to generate domestic employment. Therefore, examining the relationship between investment realization and labor absorption is essential for understanding the employment implications of Indonesia's downstream industrialization strategy [16].

the relationship between foreign direct investment and economic growth in Indonesia varies across provinces and economic sectors, highlighting the importance of examining investment beyond national aggregate statistics. More recently, Franse and Maichal investigated employment determinants across Indonesian regional clusters using provincial panel data from 2011 to 2024. Their findings revealed substantial regional heterogeneity, indicating that the relationships between foreign direct investment, economic growth, and employment differ across geographical groups [17].

Regional and sectoral differences are important considerations in understanding the distribution of investment and its broader economic implications. analyzed foreign direct

investment and industrial agglomeration across the Western and Eastern regions of Indonesia, identifying regional differences in the relationship between industrial concentration and income inequality. These findings indicate that investment outcomes are closely associated with geographical characteristics and local industrial structures. Consequently, regional and sectoral classifications provide an important analytical foundation for examining investment realization and domestic labor absorption in Indonesia.[18].

Extreme Gradient Boosting (XGBoost) is an ensemble machine learning algorithm based on the gradient boosting framework, which constructs a predictive model by sequentially combining multiple decision trees. Unlike conventional statistical approaches that primarily represent linear relationships, XGBoost can capture complex nonlinear patterns and interactions among explanatory variables. The algorithm improves predictive performance through gradient-based optimization, regularization, and efficient tree construction, making it suitable for structured datasets containing heterogeneous numerical and categorical features (Wu, Zhan 2023) [19].

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$$

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$

Figure 1. XGBoost Formula

The first equation explains that the model's final prediction is obtained by aggregating the contributions of all decision trees constructed sequentially. The second equation represents the objective function, which aims to minimize prediction errors while controlling model complexity through regularization. The third equation illustrates that the regularization term consists of a penalty on the number of leaf nodes and the magnitude of their weights to reduce the risk of overfitting. Finally, the fourth equation describes the calculation of optimal leaf weights based on the accumulated first-order gradients, second-order gradients, and the regularization parameter[20] .

Machine learning models are utilized to automatically classify public complaint reports in Indonesia. Various machine learning algorithms and approaches have been developed to improve the efficiency and accuracy of the classification process. Previous studies have employed Recurrent Neural Network (RNN) approaches [21].

Through this mechanism, XGBoost can develop an optimized predictive model through an iterative learning process while maintaining model complexity and improving its generalization capability on unseen data. utilized XGBoost to investigate factors associated with industrial transformation, highlighting its capability to identify relationships among multidimensional economic indicators. In this study, XGBoost is implemented as a regression algorithm to estimate domestic labor absorption (TKI) based on investment realization, regional characteristics, and sectoral classifications. The model is expected to capture nonlinear relationships between investment characteristics and labor absorption, providing a predictive basis for further interpretation using SHAP [22].

In this study, SHAP is integrated with the XGBoost regression model to interpret how investment realization, downstreaming classification, regional distribution, and sectoral characteristics contribute to predicted domestic labor absorption (TKI). Through global and individual feature attribution, SHAP provides a more transparent understanding of the model's predictions and helps identify the variables employment patterns in Indonesia [23]. The approach combines Lasso regression for feature selection, optimized ensemble learning for predictive performance, and SHAP analysis to quantify feature importance and identify interaction effects. This framework enables effective modeling of high-dimensional data, nonlinear relationships, and model interpretability beyond the capabilities of conventional methods [24]. Shapley value analysis indicates that machine learning models trained using all available features do not produce sparse representations. Among the three best-performing models—Random Forest, Support Vector Regression (SVR), and Neural Network—the top 10 features account for only 41%, 32%, and 34% of the variance in their respective predictions [25].

Research Methodology

This study adopts a quantitative research approach using a machine learning framework to analyze domestic labor absorption (TKI) in relation to investment realization in Indonesia. The research is designed as a predictive and explanatory analysis, employing Extreme Gradient Boosting (XGBoost) to model the relationship between investment characteristics and labor absorption. Furthermore, SHapley Additive exPlanations (SHAP) is applied to interpret the contribution of individual features to the model predictions. This approach enables the identification of regional and sectoral patterns associated with domestic labor absorption, particularly in the context of Indonesia's downstreaming policy.

3.1 Data and Instruments

The research uses secondary data obtained from the investment realization reports published by the Ministry of Investment and Downstream Industry (BKPM) for the second and third quarters of 2026. The dataset contains investment realization records with several attributes, including investment value, domestic labor absorption (TKI), investment status, business sector, province, regency or city, geographical classification, and downstreaming classification. The geographical classification distinguishes between Java and regions outside Java, while the downstreaming classification differentiates downstreaming and non-downstreaming activities based on the official BKPM classification.

Data processing and model development are conducted using Python. The primary instruments include Pandas and NumPy for data processing, Scikit-learn for preprocessing and model evaluation, XGBoost for regression modeling, and SHAP for model interpretation. The investment value and relevant categorical variables are used as model input features, while the number of domestic workers (TKI) is defined as the target variable.

3.2 Problem-Solving Analysis

The problem-solving analysis is conducted through a systematic workflow consisting of data collection, preprocessing, feature engineering, model development, evaluation, and interpretation. First, the Q2 and Q3 2026 BKPM datasets are integrated into a single analytical dataset. Data preprocessing is then performed to identify missing values, duplicate records, inconsistent formats, and invalid observations. Records containing zero or negative values are examined according to their data validity and analytical relevance before the modeling stage.

Next, feature engineering is performed to prepare the independent variables, including investment value, downstreaming classification, geographical classification, business sector, and reporting period. Categorical variables are transformed into numerical representations to ensure compatibility with the XGBoost algorithm. The dataset is subsequently divided into training and testing subsets, using an 80:20 ratio, respectively. The XGBoost regression model

is trained using the training data, followed by hyperparameter optimization to obtain the best-performing configuration.

Model performance is evaluated using the coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics are used to assess the model's predictive performance on the testing dataset. Finally, SHAP analysis is conducted to explain the model predictions through global and local interpretations. Global interpretation identifies the features that contribute most significantly to the model predictions, while local interpretation explains how individual feature values contribute to a particular prediction. The resulting analysis is used to examine regional and sectoral patterns of investment and domestic labor absorption in Indonesia.

The method includes a detailed description of the methods, instruments, and research analysis techniques used in solving problems. If it is *the result of a* literature review, then the order after the introduction is *a problem-solving analysis*. Problem Solving Analysis includes an objective description of problem solving. The distance between the subheading and the preceding text is one space.

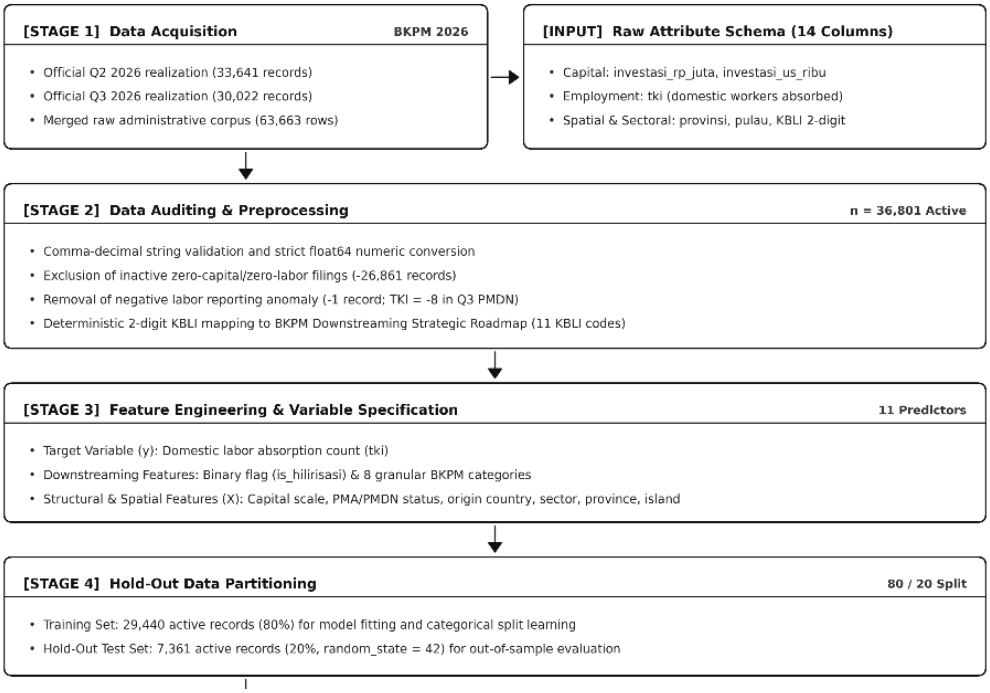


Figure 2. Research Framework Gradient-SHAP Methodology (1)

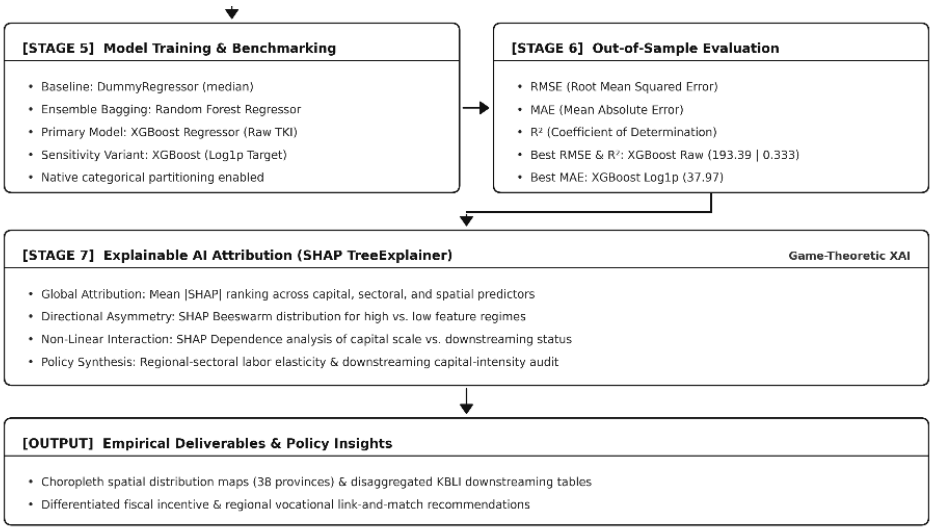


Figure 3. Research Framework Gradient-SHAP Methodology (2)

The research methodology consists of seven stages, beginning with the acquisition and integration of BKPM investment realization data for Q2 and Q3 2026. Data auditing and preprocessing are conducted to obtain 36,801 active records, followed by feature engineering to define 11 predictors and domestic labor absorption (TKI) as the target variable. The dataset is divided into training (80%) and testing (20%) subsets.

Model development involves a median Dummy Regressor as the baseline, Random Forest as a benchmarking model, and XGBoost as the primary regression model. Model performance is evaluated using RMSE, MAE, and R^2. Finally, SHAP TreeExplainer is applied to identify global feature importance, directional relationships, and nonlinear patterns influencing labor absorption across regional and sectoral categories.

Results

The initial dataset consists of two BKPM investment realization datasets, comprising 33,641 records from Q2 2026 and 30,022 records from Q3 2026. After integration, the combined dataset contains 63,663 records and 14 attributes. These attributes represent several dimensions of investment realization, including reporting period, investment status, regional classification, country of origin, business sector, province, regency or city, geographical grouping, investment value in Indonesian rupiah and US dollars, and domestic labor absorption (TKI). The dataset provides a comprehensive basis for examining investment distribution and its relationship with employment across different regions and economic sectors in Indonesia. Furthermore, the availability of geographical and sectoral attributes enables a comparative analysis of downstreaming and non-downstreaming investment patterns.

Table 1. Shape Q2 – Q3

No.	Name	Note
1	Shape Q2: (33641, 14)	(63663, 14)
2	Shape Q3: (30022, 14)	

The data auditing process identified one anomalous record containing a negative domestic labor absorption value of -8 workers, originating from a PMDN investment record in the Trade and Repair sector in East Nusa Tenggara, with zero recorded investment realization. This observation was flagged for validation because negative employment values are not consistent with the intended interpretation of labor absorption. Following the data validation and filtering process, 36,801 active records were retained from the initial dataset of 63,663 records.

Table 2. Dataset auditing results

No.	Indicator	Result
1	Initial combined records	63,663
2	Negative TKI records	1
3	Total negative TKI value	-8
4	Province of anomaly	Nusa Tenggara Timur
5	Investment status	PMDN
6	Business sector	Perdagangan dan Reparasi
7	Investment value of anomaly	Rp 0
8	Active records after filtering	36,801
9	Total investment realization	Rp 1,040,446,177.30 million
10	Total domestic labor absorption	1,487,935 workers

These records represent a total investment realization of IDR 1,040,446,177.30 million and an aggregate domestic labor absorption of 1,487,935 workers. The resulting dataset provides the analytical foundation for subsequent feature engineering and machine learning model development.

The exploratory data analysis reveals substantial differences in investment realization and domestic labor absorption across geographical and sectoral classifications. Although investment realization outside Java reached IDR 528.1 trillion, slightly exceeding Java's IDR 512.3 trillion, Java accounted for a considerably larger share of labor absorption, with approximately 1,080.1 thousand workers compared to 407.8 thousand outside Java.

A similar contrast is observed between downstreaming and non-downstreaming activities. Non-downstreaming sectors recorded IDR 670.7 trillion in investment and absorbed approximately 1,205.7 thousand workers, whereas downstreaming sectors recorded IDR 369.7 trillion and absorbed approximately 282.3 thousand workers. At the sectoral level, the Basic Metal and Metal Products Industry recorded the highest investment realization at IDR 199.8 trillion. Meanwhile, West Java, Jakarta, and Central Java were the three largest contributors to domestic labor absorption. Furthermore, the distribution of labor absorption per project is positively skewed, with a median of 11 workers and a mean of 79 workers, indicating substantial variation across investment projects. The scatter plot also illustrates heterogeneous relationships between capital investment and labor absorption. These descriptive findings highlight the importance of regional and sectoral characteristics in explaining labor absorption patterns and provide the empirical motivation for the subsequent XGBoost regression and SHAP interpretation.

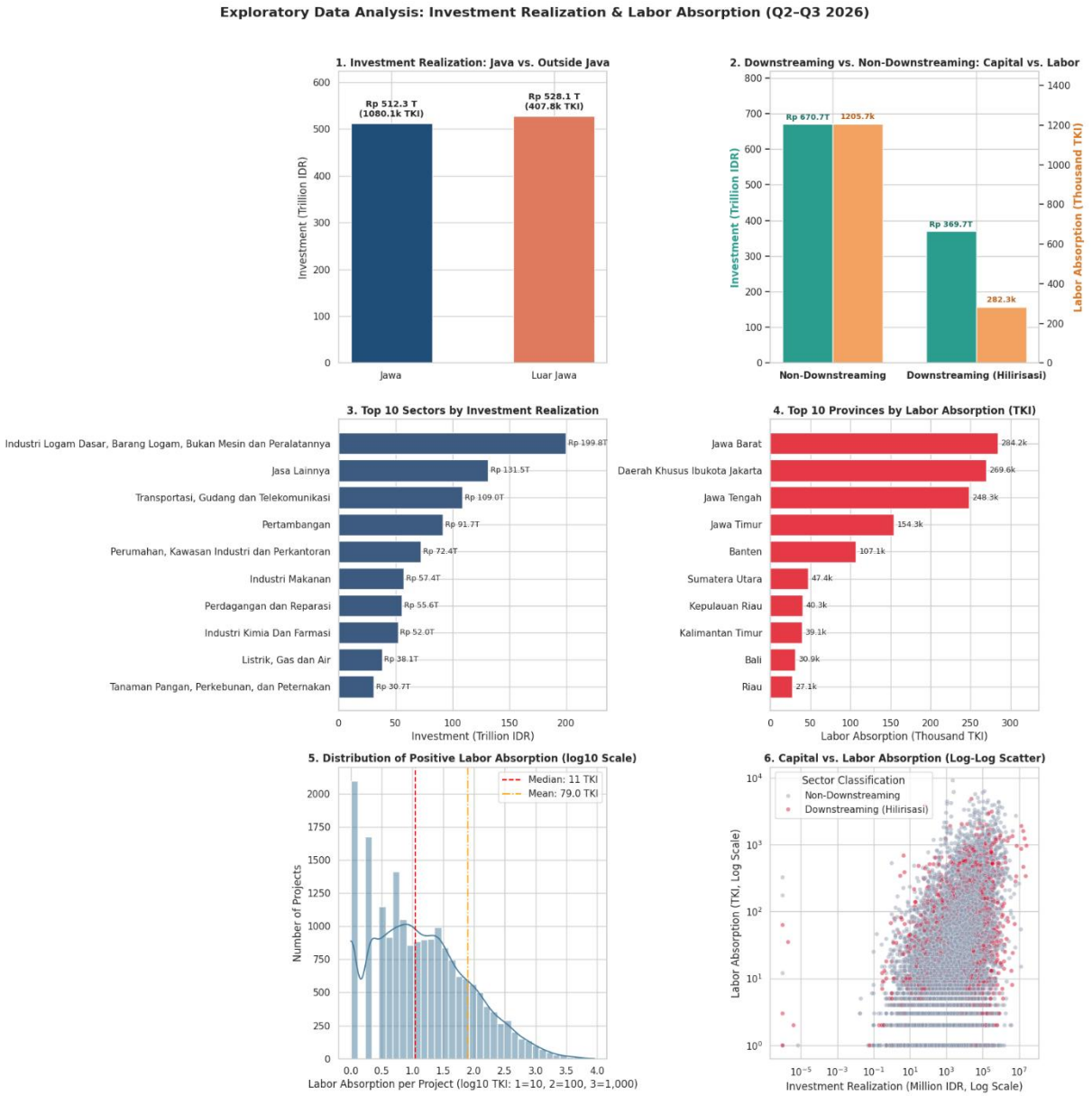


Figure 4. EDA Output Q2-Q3

The spatial analysis reveals considerable regional heterogeneity in investment realization, dominant economic sectors, and downstreaming intensity across Indonesian provinces. The distribution of investment indicates that provinces differ substantially in their investment scale and economic specialization. Among the highlighted provinces, Banten recorded the largest investment realization at IDR 58.5 trillion, with the chemical and pharmaceutical industry as its dominant sector and a downstreaming share of 37%. Meanwhile, Gorontalo exhibited the highest downstreaming intensity among the annotated provinces, reaching 42%, despite having a relatively smaller investment realization of IDR 4.4 trillion.

Jambi also demonstrated a notable downstreaming share of 29%, whereas Bali recorded only 2%, with hotels and restaurants as its dominant sector. These findings illustrate that investment magnitude and downstreaming intensity do not necessarily follow the same regional pattern. The spatial variation further emphasizes the importance of incorporating provincial and sectoral characteristics into the analysis of investment and domestic labor absorption.

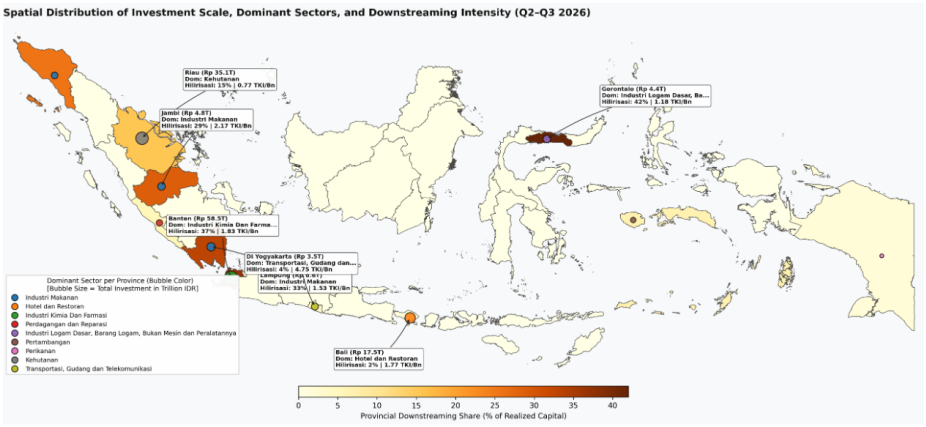


Figure 5. Spatial Distribution of Investment Scale

The regression performance evaluation demonstrates different predictive characteristics among the tested models. The median-based DummyRegressor baseline achieved an MAE of 42.90, an RMSE of 240.57, and an R^2 of -0.031. The XGBoost model trained using the raw TKI target obtained the highest coefficient of determination (R^2 = 0.333) and the lowest RMSE (193.39), indicating a better ability to explain target variability and reduce large prediction errors. However, its MAE of 48.12 was higher than that of the baseline model. In contrast, the log1p-transformed XGBoost model achieved the lowest MAE of 37.97, indicating more accurate predictions in terms of average absolute error, although its R^2 (0.188) and RMSE (213.45) were inferior to those of the raw-target model.

Table 3. Regression Performance Evaluation

Model	MAE	RMSE	R^2
DummyRegressor (Median)	42.8989	240.5654	-0.0313
XGBoost Regressor (Raw TKI Target)	48.1206	193.3910	0.3335
XGBoost Regressor (Log1p-Transformed Target)	37.9724	213.4543	0.1880

These results indicate a trade-off between minimizing average absolute prediction errors and explaining variations in labor absorption. Considering the higher R^2 and lower RMSE, the raw-target XGBoost model is selected as the primary model for subsequent SHAP interpretation, while the log1p model is retained as a sensitivity comparison.

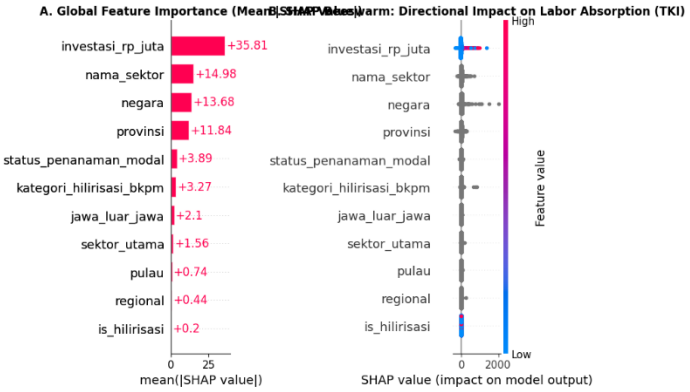


Figure 6. Spatial Distribution of Investment Scale

The SHAP analysis identifies investment value (investasi_rp_juta) as the most influential feature in predicting domestic labor absorption, with a mean absolute SHAP value of 35.81. It is followed by business sector (nama_sektor) at 14.98, country of origin (negara) at 13.68, and province (provinsi) at 11.84. These results indicate that investment scale and structural

characteristics of investment projects contribute substantially to the model's predictions. Meanwhile, investment status and downstreaming-related variables exhibit comparatively lower global importance, with mean absolute SHAP values of 3.89 for investment status and 3.27 for the BKPM downstreaming category.

Other geographical indicators, including Java versus outside Java, island, and regional classification, also demonstrate relatively smaller contributions. Overall, the findings suggest that labor absorption predictions are more strongly associated with investment magnitude and sectoral characteristics than with individual binary downstreaming or broad geographical indicators. However, SHAP importance represents the contribution of features to model predictions and should not be interpreted as evidence of causal relationships.

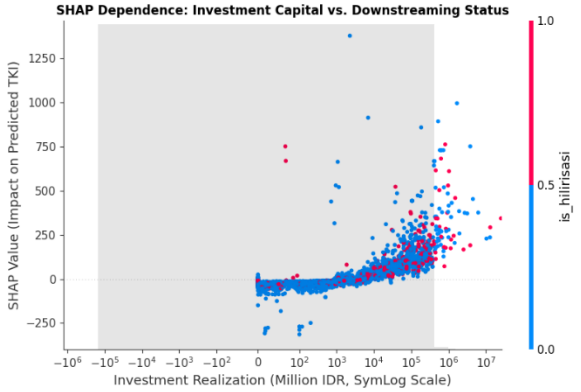


Figure 7. SHAP investment capital and Dowstreming Status

The SHAP dependence analysis illustrates a nonlinear relationship between investment realization and predicted domestic labor absorption. At relatively low investment levels, SHAP values are predominantly concentrated around zero or in the negative range, indicating limited or negative contributions to the model's predicted labor absorption. As investment increases, SHAP values generally shift toward positive contributions, with substantially greater variation observed among high-investment projects. This pattern suggests that investment magnitude plays an important role in the model's predictions, although its contribution varies across individual observations. Furthermore, the distribution of downstreaming and non-downstreaming observations exhibits considerable overlap across investment levels. Therefore, the visualization does not indicate a consistent separation in SHAP contributions based solely on downstreaming status. These findings highlight the nonlinear and heterogeneous relationship between investment scale and predicted labor absorption, suggesting that investment magnitude should be interpreted alongside other sectoral and regional characteristics.

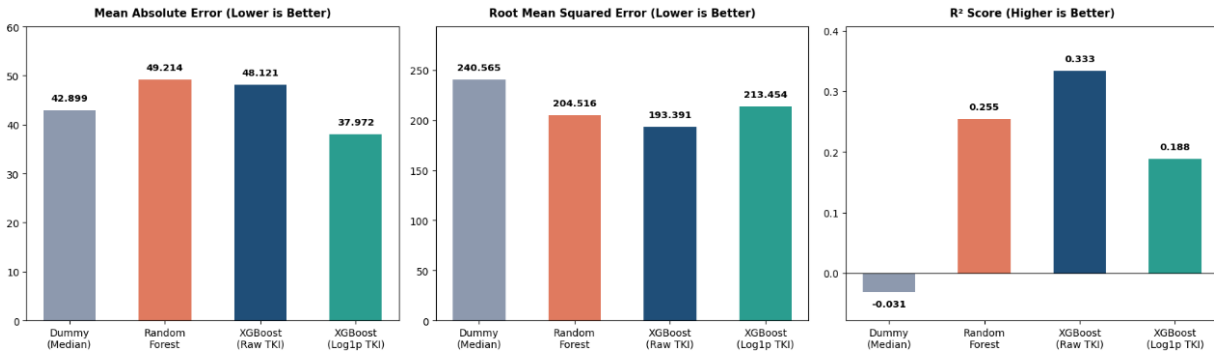


Figure 8. Comparative Performance Evaluation of Regression Models on the Test Dataset.

The comparative evaluation of four regression models reveals different levels of predictive performance in estimating domestic labor absorption. The DummyRegressor baseline achieved an MAE of 42.899, an RMSE of 240.565, and an R^2 of -0.031. Random Forest obtained an R^2 of 0.255 and an RMSE of 204.516, demonstrating improved performance compared with the baseline. Among the evaluated models, XGBoost using the raw TKI target achieved the highest R^2 of 0.333 and the lowest RMSE of 193.391, indicating its stronger ability to explain variations in labor absorption and reduce squared prediction errors. However, the XGBoost model with a log1p-transformed target achieved the lowest MAE of 37.972, indicating better average absolute prediction accuracy. These findings demonstrate a trade-off between absolute prediction error and the ability to explain target variability. Based on its superior R^2 and RMSE performance, the raw-target XGBoost model is selected as the primary model for subsequent SHAP-based interpretation, while the log1p-transformed model provides an alternative perspective on prediction accuracy.

Conclusion

This study examined the relationship between investment realization and domestic labor absorption in Indonesia using BKPM investment data from Q2–Q3 2026. The exploratory and spatial analyses revealed substantial regional and sectoral heterogeneity in investment distribution and labor absorption. Although investment realization outside Java was slightly higher than in Java, labor absorption remained considerably concentrated in Java. Furthermore, non-downstreaming activities recorded higher aggregate investment and labor absorption than downstreaming activities in the observed dataset. These findings indicate that investment magnitude, geographical distribution, and sectoral characteristics exhibit different patterns across Indonesian provinces.

The regression evaluation demonstrated that XGBoost with the raw TKI target achieved the highest R^2 of 0.333 and the lowest RMSE of 193.391 among the evaluated models. Meanwhile, the log1p-transformed XGBoost model obtained the lowest MAE of 37.972. This indicates a trade-off between reducing average absolute prediction errors and explaining variations in labor absorption. Based on its R^2 and RMSE performance, the raw-target XGBoost model was selected for the subsequent SHAP interpretation.

The SHAP results identified investment value as the most influential feature in predicting domestic labor absorption, followed by business sector, country of origin, and province. Downstreaming-related variables exhibited lower individual global importance compared with investment magnitude and sectoral characteristics. The dependence analysis further revealed a nonlinear pattern in which higher investment values generally contributed more positively to predicted labor absorption, although considerable variation remained across individual projects. The overlapping SHAP distributions of downstreaming and non-downstreaming observations also suggest that downstreaming status alone does not consistently distinguish the model's prediction contributions.

Overall, this study demonstrates the potential of combining XGBoost and SHAP to examine regional and sectoral patterns of investment and labor absorption using official BKPM data. The findings provide an interpretable empirical perspective for understanding differences in labor absorption across investment characteristics and geographical locations. However, the model's R^2 of 0.333 indicates that a substantial proportion of labor absorption variability remains unexplained. Future research should incorporate additional project-level characteristics, expand the observation period, and investigate interactions between investment scale, sectoral classification, regional conditions, and downstreaming status to improve predictive performance and provide deeper policy insights.

References

[1] N. L. Widyastuti, U. Widiatmoko, and H. Nugroho, "Downstream Industrialization and the Future of Indonesia's Resource- Based Economy: Resource Nationalism, Value Addition, and Structural Challenges," *J. Perenc. Pembang. Indones. J. Dev. Plan.*, vol. 10, no. 1, pp. 51–65, 2026, doi: 10.36574/jpp.v10i1.836.

[2] F. T. Wau, M. A. Kiton, M. Wau, and J. F. Fau, "ANALISIS STRATEGIS KEBIJAKAN HILIRISASI MINERAL: Implikasi Ekonomi dan Pengaruhnya terhadap Perekonomian Indonesia," *J. Publicuho*, vol. 7, no. 3, pp. 1215–1224, 2024, doi: 10.35817/publicuho.v7i3.481.

[3] M. Henny Ariska Anggrainin, "Analisis Pengaruh Investasi dan Inflasi terhadap Penyerapan Tenaga Kerja melalui Pertumbuhan Ekonomi .sebagai Variabel Intervening di Provinsi Bali Henny," vol. 7, no. 2, pp. 500–513, 2023.

[4] bank indonesia, "NATIONAL ECONOMIC PERFORMANCE AND OUTLOOK :," *Econ. Rep. Indones. 2024*, 2025.

[5] A. A. Awaludin, M. Cholil, and S. Rusgianto, "Impact of Sectoral Growth on Employment Absorption: Evidence From Key Indonesian Industries," *J. Ekon. Bisnis dan Kewirausahaan*, vol. 13, no. 2, pp. 277–298, 2024, doi: 10.26418/jebik.v13i2.79592.

[6] A. M. Fazaalloh, "FDI and economic growth in Indonesia: a provincial and sectoral analysis," *J. Econ. Struct.*, vol. 13, no. 1, 2024, doi: 10.1186/s40008-023-00323-w.

[7] C. Hornok and D. G. S. Raeskyesa, *Economic zones and local income inequality: Evidence from Indonesia*, vol. 22, no. 1. Springer US, 2024. doi: 10.1007/s10888-023-09581-x.

[8] M. M. Wihardja and W. Cunningham, "Pathways to Middle-Class Jobs in Indonesia," *Pathways to Middle-Class Jobs Indones.*, no. June, 2021, doi: 10.1596/35848.

[9] Y. Eriani, J. Roy, and D. Lestari, "Sectoral Investment, Economic Growth, and Labor Absorption in Resource-Based Regions: Empirical Evidence from Kutai Kartanegara, Indonesia," *Equity J. Ekon.*, vol. 14, no. 1, pp. 141–154, 2026, doi: 10.33019/equity.v14i1.738.

[10] Z. Wu, "Evaluation of Provincial Economic Resilience in China Based on the TOPSIS-XGBoost-SHAP Model," *J. Math.*, vol. 2023, 2023, doi: 10.1155/2023/6652800.

[11] D. Tang and J. Wei, "Prediction and Characteristic Analysis of Enterprise Digital Transformation Integrating XGBoost and SHAP," *J. Adv. Comput. Intell. Intell. Informatics*, vol. 27, no. 5, pp. 780–789, 2023, doi: 10.20965/jaciii.2023.p0780.

[12] S. Pyo and S. O. Choi, "Regional innovation and economic growth: Empirical insights from FGLS, FE-DKSE, and XGBoost-SHAP approach," *J. Open Innov. Technol. Mark. Complex.*, vol. 11, no. 2, p. 100524, 2025, doi: 10.1016/j.joitmc.2025.100524.

[13] E. F. Cahyono, W. A. S. Wan Ngah, S. H. Law, N. S. Mazlan, M. N. N. Ahmad, and A. Mohhammed, "Evaluating Asymmetric Impacts of Islamic Bank Financing on Employment In Indonesia," *Etikonomi*, vol. 24, no. 1, pp. 31–52, 2025, doi: 10.15408/etk.v24i1.40139.

[14] Franse and Maichal, "Regional Heterogeneity in Employment Determinants: Evidence from Indonesia," *Signifikan J. Ilmu Ekon.*, vol. 15, no. 1, pp. 139–156, 2026, doi: 10.15408/sjie.v15i1.49655.

[15] Muhyiddin and M. A. , Achmanto Mendatu, Muhammad Dio Rhiza Amrizal, "Jobless Growth or Structural Transformation? Labor Market Dynamics in Morowali, Indonesia's Nickel-Based Industrialization Hub Muhyiddin1," vol. X, no. 1, pp. 29–49, 2026.

[16] A. H. Asyono, P. L. Samputra, and H. Djatmiko, "Domestic Structural Transformation in a Critical Mineral Economy: A Multisectoral Assessment of Indonesia's Nickel Downstreaming Strategy," *Economies*, vol. 14, no. 4, 2026, doi: 10.3390/economies14040133.

[17] M. Maichal, P. G. Hartono, A. Firman, and I. M. E. K. Yudha, "The Influence of Gross Regional Domestic Product Per Capita and Foreign Direct Investment on Income

Inequality: An Empirical Study of 34 Provinces in Indonesia," *J. Econ. Res. Soc. Sci.*, vol. 8, no. 2, pp. 197–206, 2024, doi: 10.18196/jerss.v8i2.23256.

[18] R. Ananda and E. Hendrawaty, "E-journal Field of Economics , Business , and Entrepreneurship (EFEBE)," *E-journal F. Econ. Business, Entrep.*, vol. 3, no. 3, p. 579, 2025, doi: 10.23960/efebe.v4i2.352.

[19] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-Augu, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.

[20] D. Boldini, F. Grisoni, D. Kuhn, L. Friedrich, and S. A. Sieber, "Practical guidelines for the use of gradient boosting for molecular property prediction," *J. Cheminform.*, vol. 15, no. 1, pp. 1–13, 2023, doi: 10.1186/s13321-023-00743-7.

[21] V. Ayumi, D. Ramayanti, H. Noprisson, Y. Jumaryadi, and U. Salamah, "Model Extreme Gradient Boosting Berbasis Term Frequency (TFXGBoost) Untuk Pengolahan Laporan Pengaduan Masyarakat," vol. 6, no. 1, pp. 65–70, 2023.

[22] Y. Lai, G. Wan, and X. Qin, "Decoding China's new-type industrialization: Insights from an XGBoost-SHAP analysis," *J. Clean. Prod.*, vol. 478, no. May, p. 143927, 2024, doi: 10.1016/j.jclepro.2024.143927.

[23] M. Li, H. Sun, Y. Huang, and H. Chen, "Shapley value: from cooperative game to explainable artificial intelligence," *Auton. Intell. Syst.*, vol. 4, no. 1, 2024, doi: 10.1007/s43684-023-00060-8.

[24] Y. Wang, W. Wei, Z. Liu, J. Liu, Y. Lv, and X. Li, "Interpretable Machine Learning Framework for Corporate Financialization Prediction: A SHAP-Based Analysis of High-Dimensional Data," *Mathematics*, vol. 13, no. 15, pp. 1–27, 2025, doi: 10.3390/math13152526.

[25] M. Buckmann and A. Joseph, "An Interpretable Machine Learning Workflow with An Application to Economic Forecasting," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4130517.