Identifying Workforce Size Determinants in Rebana and Arumanis Using Least Absolute Shrinkage and Selection Operator (LASSO) Regression

Ika Nur Laily Fitriana[(1)], Fonda Leviany[(2)], Kurnia Sari Kasmiarno[(3)], Nuramaliyah[(4)]

[1,4] Faculty of Science and Technology, Statistics Study Program, Universitas Terbuka

[2] Faculty of Science and Technology, Data Science Study Program, Universitas Terbuka

[3] Faculty of Economic and Business, Sharia Economic Study Program, Universitas Terbuka

Jl. Cabe Raya, Pondok Cabe, Pamulang, Tangerang Selatan 15437
e-mail: ika.nur@ecampus.ut.ac.id[(1)], fonda.leviany@ecampus.ut.ac.id[(2)], kurnia.sari@ecampus.ut.ac.id[(3)], nuramaliyah@ecampus.ut.ac.id[(4)]

ABSTRAK

Pengembangan Kawasan Rebana dan Arumanis di Jawa Barat diarahkan untuk memperkuat perekonomian dan memperluas kesempatan kerja. Identifikasi faktor-faktor yang berpengaruh sangat penting untuk merumuskan kebijakan yang efektif dan terarah bagi penyerapan tenaga kerja di kedua kawasan tersebut. Penelitian ini bertujuan untuk menganalisis dan mengidentifikasi faktor-faktor utama yang memengaruhi jumlah tenaga kerja di Wilayah Rebana dan Arumanis menggunakan metode *Least Absolute Shrinkage and Selection Operator Regression* (LASSO). Untuk menentukan penalti optimal (L1) dalam LASSO, dilakukan *cross-validation* (k=5), dan menghasilkan penalti optimal (L1) sebesar 363,03. Hasil analisis menunjukkan bahwa urutan faktor dari yang paling penting dalam penyerapan tenaga kerja di kawasan Rebana dan Arumanis adalah: (1) Jumlah UMKM, (2) Indeks Pembangunan Manusia (IPM), (3) Produk Domestik Regional Bruto (PDRB), (4) Upah Minimum Regional, (5) Penanaman Modal Asing (PMA), (6) Panjang Jalan, dan (7) Penanaman Modal Dalam Negeri (PMDN). Kinerja model dinilai menggunakan R-kuadrat rata-rata 83,09%. Temuan ini menegaskan pentingnya penguatan ekosistem UMKM dan perencanaan pembangunan daerah yang berorientasi pada produktivitas untuk mendorong penyerapan tenaga kerja di Rebana dan Arumanis.

Kata kunci: Kawasan Arumanis; Faktor Utama; Penyerapan Tenaga Kerja, Regresi LASSO, Kawasan Rebana.

ABSTRACT

Development of West Java's Rebana and Arumanis Areas is intended to strengthen the economy and expand employment opportunities. Identifying influential factors is essential for formulating effective and well-targeted policies to enhance workforce absorption in these two areas. This study aims to analyze and identify the main factors affecting workforce size in the Rebana and Arumanis Areas using the Least Absolute Shrinkage and Selection Operator (LASSO) regression method. To determine the optimal (L1) penalty in LASSO, 5-fold cross-validation (k = 5) was applied, yielding an optimal penalty value of 363.03. The results indicate that the factors ranked from most to least important for workforce absorption in Rebana and Arumanis are: (1) the number of MSMEs, (2) the Human Development Index (HDI), (3) Gross Regional Domestic Product (GRDP), (4) the regional minimum wage, (5) realized Foreign Direct Investment (FDI), (6) road length, and (7) realized Domestic Direct Investment (DDI). Model performance was evaluated using an average (R^2) of 83.09%. These findings highlight the importance of strengthening the MSME ecosystem

and implementing productivity-oriented regional development planning to promote workforce absorption in Rebana and Arumanis.
Keywords: *Arumanis Area; Rebana Area; Variable Selection; Workforce size; LASSO Regression.*

INTRODUCTION

West Java Province faces unique challenges in addressing employment issues. According to data from the Central Bureau of Statistics (BPS) as of February 2024, West Java Province ranks third highest in terms of the Open Unemployment Rate (OUR) in Indonesia (Figure 1). BPS also noted that West Java's OUR has consistently exceeded the national OUR over the past five years. In 2024, the national OUR was 4.82%, while West Java's OUR was 6.91%.

To address workforce absorption challenges, the government is advancing the development of a special economic zone to establish a new economic hub in West Java. The Rebana and Arumanis Areas are key strategic initiatives aimed at diversifying the regional economy, attracting investment, and generating employment opportunities [1]. The President of Indonesia has directed the acceleration of an integrated and comprehensive development plan for these regions within the Regional Development Master Plan. This directive is outlined in Presidential Regulation (Peraturan Presiden) Numer 87 of 2021, which pertains to accelerating the development of the Rebana and Arumanis Areas. Additionally, these regions have been classified as National Strategic Projects (PSN) under Presidential Regulation Number 87 of 2021 and Regulation Number 21 of 2022 issued by the Coordinating Minister for Economic Affairs.

The Rebana region is a new metropolitan area in the northern and eastern parts of West Java, encompassing seven areas: Cirebon Regency, Subang Regency, Majalengka Regency, Indramayu Regency, Kuningan Regency, Sumedang Regency, and Cirebon City [1]. Spanning 43,913 hectares, the plan includes the construction of 13 new industrial cities over the next 10 to 30 years. Meanwhile, development in the Arumanis region, stretching from Sukabumi to Pangandaran, will prioritize transportation, tourism, disaster mitigation, irrigation, and marine and fisheries [2].

Within the Rebana region, Patimban Port in Subang Regency and West Java International Airport (BIJB) Kertajati in Majalengka Regency serve as connectivity and logistics hubs. In the coming years, the Rebana region is expected to become a key driver of West Java’s economic growth through the development of integrated, innovative, collaborative, competitive, and sustainable industrial zones. In contrast, the Arumanis region has 81 development initiatives, including 59 infrastructure projects, 5 agribusiness units, 8 marine and fisheries units, and 9 tourism units. This region aims to drive new economic growth in West Java by focusing on tourism, fisheries, marine, and agribusiness sectors [2].

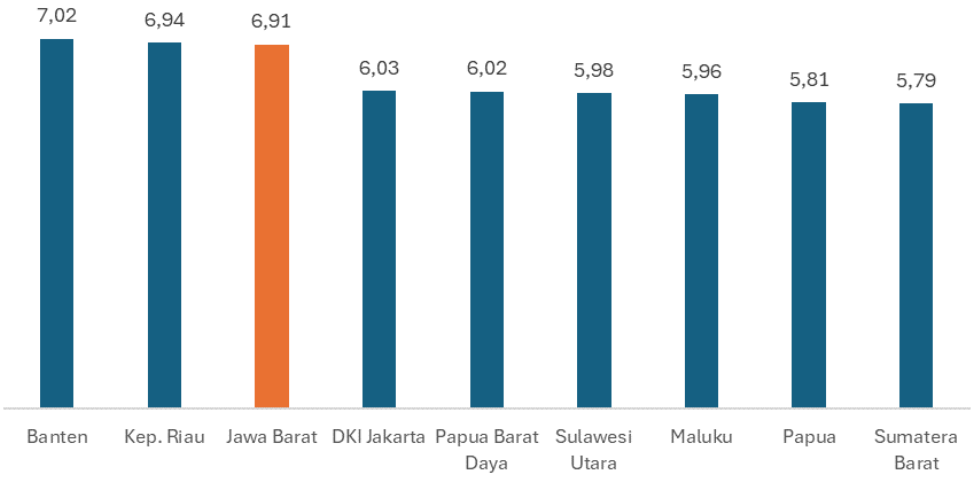

Figure 1. The Nine Regions with Highest Open Unemployment Rates (OUR) in Indonesia

The Rebana and Arumanis Areas are anticipated to play a pivotal role in absorbing a substantial number of workers. According to research by the West Java Provincial Government, workforce absorption in these areas is projected to reach 4-5 million people [2]. To realize this goal, strategic actions are essential to maximize workforce absorption in these regions. Identifying the key factors influencing workforce absorption is vital to create effective and targeted policies. Therefore, determining the hierarchy of the most influential factors in workforce absorption is necessary. One effective approach from machine learning for establishing such a ranking is the Least Absolute Shrinkage and Selection Operator Regression (LASSO regression).

LASSO regression offers advantages in efficient variable selection and managing overfitting, a common issue in regression models with numerous predictors [3]. Unlike traditional linear regression, LASSO adds an absolute penalty to the total sum of coefficient values, shrinking some coefficients to zero [4]. This feature enables LASSO to automatically select a relevant subset of variables from a large pool, excluding insignificant ones and thus simplifying the model [5]. Additionally, LASSO ranks variables by importance, providing clear insight into the variables most impacting the model [6].

In this study's context, applying LASSO regression to identify primary factors for workforce absorption in the Rebana and Arumanis Areas is highly relevant. This method facilitates focusing on the most impactful variables, ranked by each variable's influence on workforce numbers. As a result, the policies formulated will be more efficient and effective in achieving the desired objectives.

Previous research on identifying the factors influencing workforce absorption in West Java has been widely conducted, including studies by [7], [8] and [9]. These studies applied various social and economic variables affecting workforce absorption using traditional regression approaches. Ferdiansyah [10] specifically examined workforce absorption factors in the Rebana Triangle Development Area using panel data regression. However, there has been no research specifically identifying the primary factors for workforce absorption in both the Rebana and Arumanis Areas using a Machine Learning approach, particularly LASSO Regression.

In general, this study aims to analyze and identify the primary factors influencing workforce absorption in the Rebana and Arumanis Areas using a Machine Learning approach with the Least Absolute Shrinkage and Selection Operator Regression (LASSO Regression) method. This method establishes a prioritized ranking of influential variables in workforce absorption. With insights on primary factors that can enhance workforce absorption in the Rebana and Arumanis Areas is identified, this study will provide policy recommendations based on its findings.

METHOD

Dataset Description and Variables

This study utilizes secondary data collected from multiple sources. It examines eight research variables, with workforce size as the dependent variable and the other factors treated as independent variables. The unit of analysis in this study is the regency or city within the Rebana and Arumanis Areas of West Java Province. The Rebana region includes 7 regencies/cities: Cirebon City, Subang Regency, Indramayu Regency, Cirebon Regency, Majalengka Regency, Sumedang Regency, and Kuningan Regency. While, the Arumanis region includes 6 regencies: Sukabumi, Cianjur, Garut, Tasikmalaya, Ciamis, and Pangandaran. The operational definitions for each variable are provided in Table 1.

Table 1. Research Variables

Symbol	Name of Variable	Operational Definition	Unit	Data Source
Y	Workforce Size	Number of people aged 15 years and over who worked in the last week (2023)	People	West Java in Figures (BPS West Java)
X_1	Regional Minimum Wage	Minimum wage rates applicable in the Regency/City (2024)	Rupiah	West Java in Figures (BPS West Java)
X_2	Human Development Index (HDI)	Index that measures human development based on three main dimensions: health, education, and standard of living (2023)	-	West Java in Figures (BPS West Java)
X_3	Road Length	Total length of roads including asphalt, unpaved, and other roads (2023)	Km	West Java in Figures (BPS West Java)
X_4	Gross Regional Domestic Product (GRDP) at Constant Prices	Value added of goods and services calculated using prices prevailing in a specific year as the base year (2023)	Billion Rupiah	West Java in Figures (BPS West Java)
X_5	Amount of Realized Foreign Direct	The amount of investment from abroad that has been realized to operate a business in Indonesia in 2023	Rupiah	Investment Realization Report of West Java Province 2023,

Symbol	Name of Variable	Operational Definition	Unit	Data Source
	Investment (FDI)			West Java Provincial Investment Service
X_6	Amount of Realized Domestic Direct Investment (DDI)	The total amount of domestic investment that has been realized to operate a business domestically in 2023	Rupiah	Investment Realization Report of West Java Province 2023, West Java Provincial Investment Service
X_7	Projected Number of MSMEs	Projected number of Micro, Small, and Medium Enterprises in the area (2023)	Units	Cooperatives and Small Business Service

Data Preprocessing

Because LASSO is sensitive to the scale of predictors, all predictors were standardized (z-score) using statistics computed from the training set to avoid information leakage. For each predictor j, the standardized value is:

$$x_{ij}^{*} = \frac{x_{ij}-\mu_j}{s_j} \quad (1)$$

where μ_j and s_j are the mean and standard deviation of predictor j calculated from the training data. The same μ_j and s_j were applied to standardize the test data. The dataset was randomly split into training and testing sets (80%:20%) using a fixed random seed to ensure reproducibility.

LASSO Regression Model

Least Absolute Shrinkage and Selection Operator (LASSO) regression is a machine learning technique used for selecting variables and applying regularization in linear regression models [3]. Linear regression is a statistical approach for modelling and analyzing the relationship between a dependent variable and one or more independent variables [11]. LASSO regression enhances the prediction accuracy and model interpretability by simultaneously performing variable selection and regularization [12].

LASSO regression is also effective in addressing multicollinearity [13]. Multicollinearity occurs when predictors are linearly dependent (or nearly so), i.e., one predictor is (approximately) a linear combination of others. Multicollinearity can lead to inefficient estimates of regression coefficients [14]. LASSO can mitigate multicollinearity by shrinking some coefficients to zero, yielding a sparse model.

Let y_i denote workforce size in region i, and let $x_i = \left(x_{i1}, \ldots, x_{ip}\right)^T$ be a vector of p regional predictors. The LASSO estimator is obtained by solving:

$$(\hat{\beta}_0, \hat{\beta}) = \underset{\beta_0,\beta}{\operatorname{argmin}} \left(\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \beta_0 - x_i^T\beta\right)^2 + \lambda \sum_{j=1}^{p}\left|\beta_j\right|\right) \quad (2)$$

Where:

n: number of data

y_i: observed value of the dependent variable for the i-th data

X_i: vector of observed values for the independent variable

β: vector of regression coefficients

λ: penalty parameter that controls the level of regularization [15].

The optimal λ was selected using k-fold cross-validation (k=5) on the training set. The training data were divided into five folds, and then the model was trained on four folds and validated on the remaining fold, repeated five times so each fold served as validation once. We chose k = 5 because it provides a practical balance between estimation stability and computational efficiency [16]. After selecting optimal λ, the final LASSO model was refit on the full training set and evaluated on the held-out test set. Predictive performance was assessed using RMSE and coefficient of determination (R^2).

RESULT AND DISCUSSION

Data Exploration

Before conducting further analysis, data exploration was carried out to obtain a general overview of the data. The descriptive statistics results are presented in Table 2 and the boxplot is shown in Figure 2. According to the 2023 data from West Java BPS, the average number of workers in the Rebana and Arumanis Areas was 813,357, with Cirebon City having the smallest workforce at 167,037, and Sukabumi having the largest at 1,344,044 workers. Sukabumi's larger workforce can be attributed to its larger area, high population, diverse economy, and well-developed infrastructure, which attracts investment and migration. In contrast, Cirebon City's smaller workforce is due to its limited area, smaller population, less diversified economy, and fewer job opportunities. Sukabumi Regency is the largest area in West Java, covering 11.72% of the total provincial land, whereas Cirebon City is the smallest, covering only 0.11% of the total area [2].

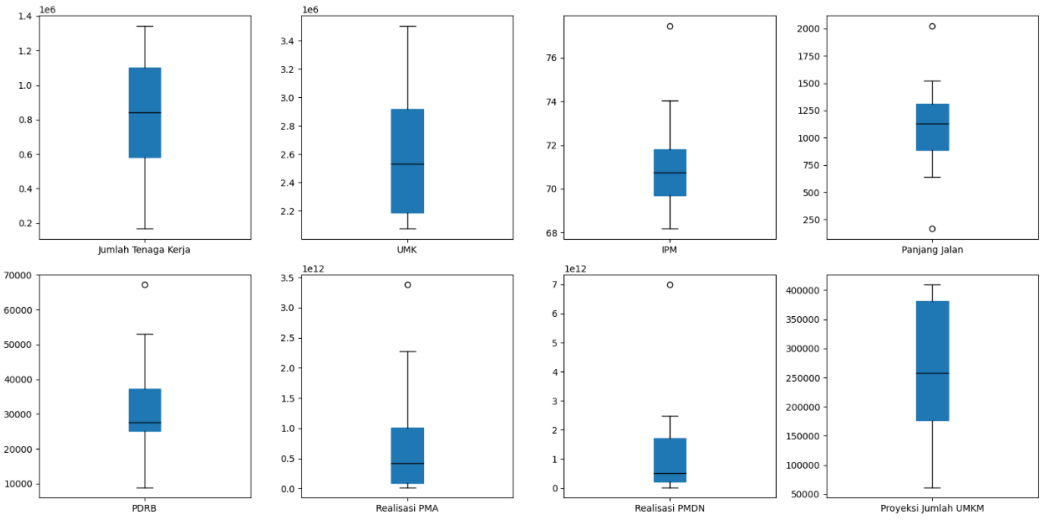

Figure 2. The Boxplot of Each Variables

Table 2. Descriptive Statistics

Variables	Mean	Standard Deviation	Minimum	Maximum
Y	813,356.69	380,943.13	167,037.00	1,344,044.00
X_1	2,615,586.08	509,609.98	2,074,666.00	3,504,308.00
X_2	71.29	2.44	68.18	77.45
X_3	1,116.17	444.25	165.72	2,026.03
X_4	32,470.30	15,459.00	8,869.86	67,237.69
X_5	845,855,126,084.69	1,017,463,255,864.00	14,588,344,491.00	3,381,206,040,592.00
X_6	1,278,785,892,307.69	1,877,642,609,655.81	15,970,700,000.00	6,995,185,900,000.00
X_7	256,289.85	116,550.39	61,234.00	409,507.00

Source: Author's Calculation

Table 2 shows that Kuningan district has the lowest minimum wage (X_1) in the Rebana and Arumanis Areas at Rp 2,074,666, while Sumedang district has the highest at Rp 3,504,308. The average minimum wage (X_1) in Rebana and Arumanis Areas is Rp 2,615,586.08. The Human Development Index (X_2) averages 71.29, ranging from 68.18 to 77.45. The HDI for West Java in 2023 was recorded at 74.24. Cirebon City exceeds the West Java 2023 provincial HDI (74.24) and appears as an outlier in Figure 2. Road length (X_3) averages 1,116.17 km, ranging from 165.72 km (Cirebon City) to 2,026.03 km (Sukabumi), reflecting differences in geographic area.

Indramayu Regency exhibits the highest GRDP at constant prices (X_4), contributing Rp 67,237.69 billion in 2023. In contrast, Pangandaran recorded the lowest GRDP at constant prices, amounting to Rp 8,869.86 billion. Indramayu Regency's comparatively high GRDP is supported by a diversified economic base, with major contributions from oil and gas as well as agriculture (including intensive rice farming) and fisheries. In contrast, Pangandaran's lower GRDP is consistent with a smaller and less diversified economy that is more concentrated in tourism and agriculture, where tourism generally generates lower value added than Indramayu's oil and gas sector.

Investment indicators show high dispersion: average realized FDI (X_5) is Rp 845,855,126,084.69 and average realized DDI (X_6) is Rp 1,278,785,892,307.69, with Sukabumi identified as a high outlier (Figure 2). Finally, projected MSMEs (X_7) total 3,331,768 across the two areas; at the district/city level they range from 61,234 (Cirebon City) to 409,507 (Sukabumi), with a mean of 256,289.85 units. Figure 3 visualizes the spatial distribution of the workforce in the Rebana and Arumanis Areas. This visualization clearly illustrates the workforce distribution across both regions, which together account for 45% of the total workforce in West Java Province, 10,573,637 out of 23,503,598 workers.

The map highlights that the Arumanis Region, which includes Cianjur, Garut, and Sukabumi Regencies, has a higher number of workforces compared to other districts and cities. This phenomenon can be explained by the high population in the Arumanis region which directly affects the number of available workers. In addition, the developing agricultural, plantation, and tourism sectors in this region also play a significant role in absorbing local labor. Conversely, the Rebana

Region has a lower number of workers, likely due to its smaller population relative to the Arumanis Region.

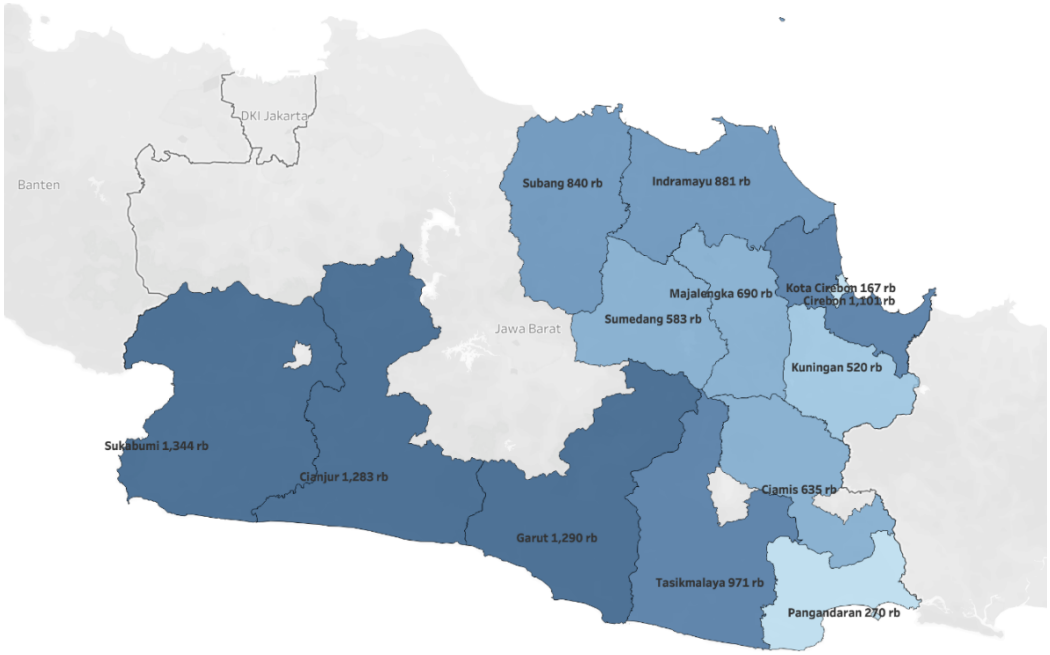

Figure 3. Number of Workers in Rebana and Arumanis Areas

Lasso Regression

Before we conduct regression, a correlation analysis was performed between the variables, as illustrated in Figure 4. The Correlation Matrix reveals that the dependent variable, Size of Workforce, exhibits a strong positive correlation with several independent variables, including Road Length (X_3), GRDP (X_4), Amount of Realized FDI (X_5), Amount of Realized DDI (X_6), and Projected Number of MSMEs (X_7). On the other hand, the HDI variable (X_2) exhibits a strong negative correlation with workforce size. The high correlation suggests a high relationship between these independent variables and workforce size. Specifically, an increase in the projected number of MSMEs, realized FDI, GRDP, and road length is associated with a growth in the workforce. The growth of MSMEs tends to increase the demand for workers, while a high level of FDI usually indicates the initiation of new projects that require a large workforce. Additionally, a high GRDP typically reflects a more developed economy, which is associated with greater workforce absorption. The Road Length variable shows a strong correlation with the workforce size, suggesting that better road infrastructure improves economic access and distribution, potentially creating more job opportunities.

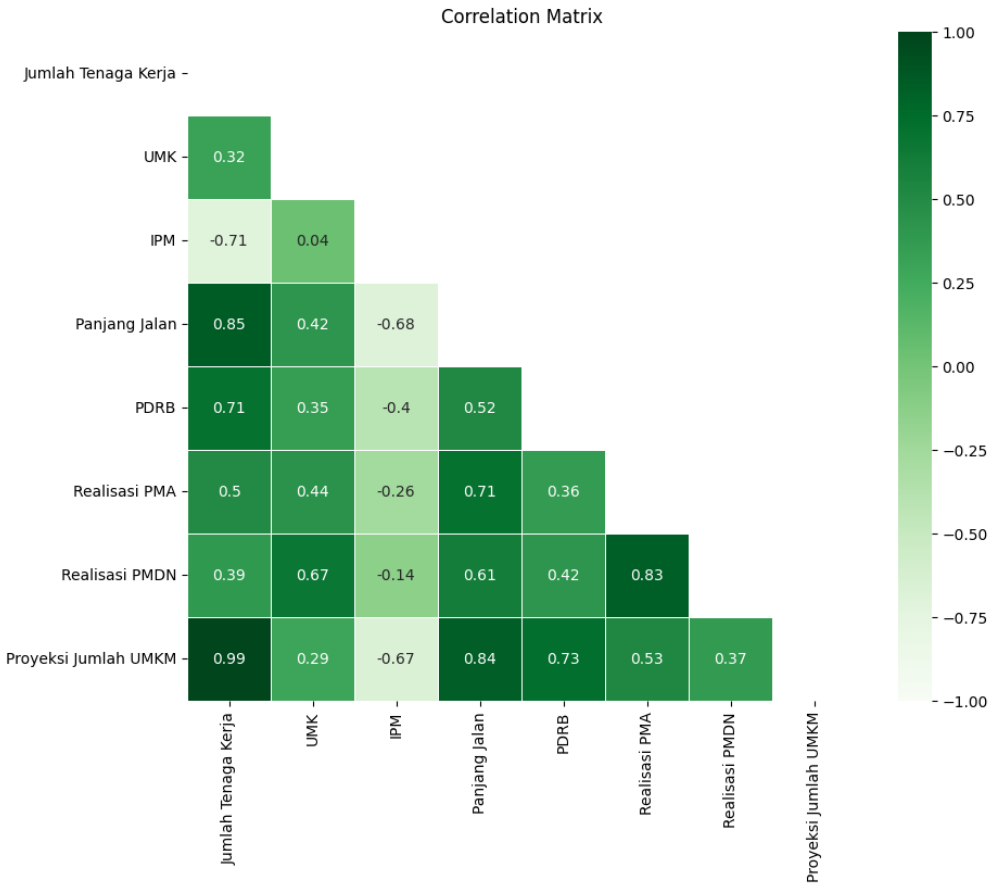

Figure 4. Correlation Matrix of Variables

In this study, the high correlation between independent variables and the workforce size also provides a basis for further in-depth analysis using regression methods. However, there is high multicollinearity between the independent variables which can be seen in the high correlation between the independent variables. Therefore, the LASSO Regression method is appropriate to use in this study, to overcome the high multicollinearity in the data. By applying the LASSO Regression method, it becomes possible to identify which variables significantly influence the dependent variable and determine the extent of their impact. This will help in formulating more targeted policies to increase workforce absorption in the Rebana and Arumanis Areas.

A LASSO regression analysis was subsequently performed using standardized predictors. LASSO operates by applying an L1 penalty on the absolute values of regression coefficients, which can shrink some coefficients to exactly zero. The optimal penalty parameter (λ) was selected via 5-fold cross-validation, and the resulting LASSO coefficients are presented in Table 3.

Table 3. LASSO Regression Coefficients

Symbol	Variable	Regression Coefficient
X_1	Regional Minimum Wage	34.521,99
X_2	Human Development Index (HDI)	-86.969,64

Symbol	Variable	Regression Coefficient
X_3	Road Length	-29.982,84
X_4	Gross Regional Domestic Product (GRDP) at Constant Prices (GRDP at Constant Prices)	51.807,78
X_5	Amount of Realized Foreign Direct Investment (FDI)	-34.005,20
X_6	Amount of Realized Domestic Direct Investment (DDI)	-23.300,77
X_7	Projected Number of MSMEs	280.123,24

Source: Author's Calculation

In LASSO regression, variables can be ranked based on the magnitude of their regression coefficients. Coefficients that are exactly zero indicate that a predictor is excluded under the selected λ. Unlike ordinary least squares, LASSO does not directly provide classical hypothesis tests (p-values); therefore, non-zero coefficients should be interpreted as selected predictors that contribute to model prediction, not as definitive evidence of statistical significance.

In this study's LASSO results, all seven predictors retain non-zero coefficients under the cross-validated λ, indicating that the selected regularization strength did not eliminate any predictor. The relative magnitudes of the coefficients (Table 3 and Figure 5) are used to discuss which variables have stronger contributions to explaining variation in workforce size.

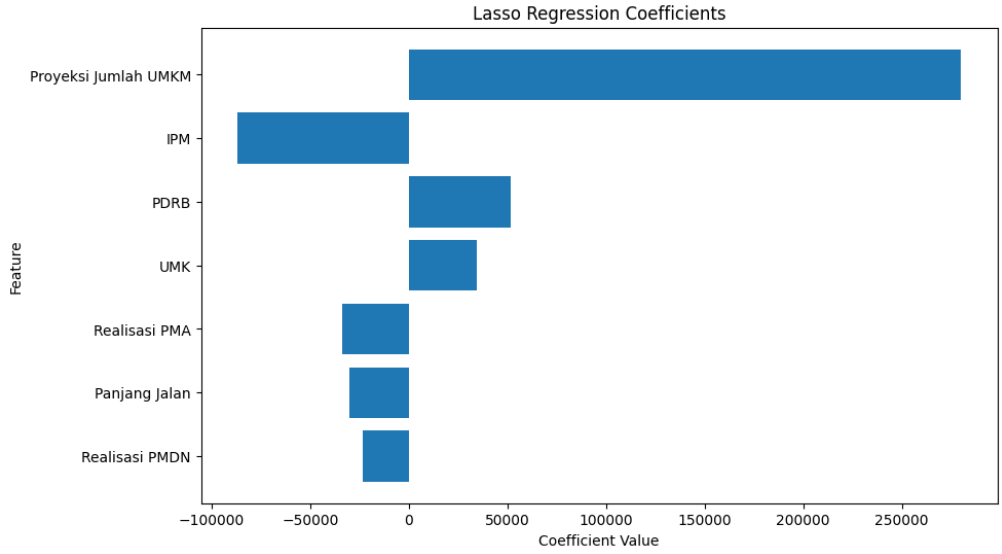

Figure 5. Plot of Lasso Regression Coefficients.

Source: Author's calculations

The projected number of MSMEs has the largest and positive coefficient (280,123.24), indicating that an increase in the number of MSMEs can have a substansial impact on increasing job opportunities in the Rebana and Arumanis Areas. MSMEs are often described as the economic backbone because of their role in creating local jobs. MSMEs are usually more labor-intensive than

larger companies, so they can absorb more workers. In addition, MSMEs also strengthen the local economy by diversifying activities and raising purchasing power.

The HDI variable has a large negative regression coefficient (-86,969.64), indicating that, under the conditions of this model, an increase in HDI is associated with a decrease in job opportunities. This seems inconsistent with the expectations of classical economic theory, which assumes that an increase in HDI supports economic growth and job creation. According to modern economic and socio-economic theories, this can still make sense because, based on Endogenous Growth Theory, it states that improvements in education and health infrastructure affect long-term productivity and innovation, potentially driving automation that reduces the need for manual labour.

The GRDP variable has a positive regression coefficient. GRDP is an indicator that describes the total value of goods and services produced in a region. An increase in GRDP generally indicates positive economic growth and, directly, creates more job opportunities. A growing economic sector will need more labor to support its production. An increase in GRDP also strengthens people's purchasing power, which in turn increases demand for goods and services and creates new jobs. The results of this study are also in line with Hafiz's research (2021) [8] which states that GRDP has a significant impact on workforce absorption.

The regression coefficient of the Regional Minimum Wage variable is positive. This indicates that the higher the minimum wages, the more it will increase workforce absorption. Setting a fair minimum wage aligned with local economic conditions can enhance worker welfare and motivation. Additionally, competitive minimum wages help attract workers from other areas.

In addition, three selected predictors display negative coefficients in the LASSO model: Road Length, Foreign Direct Investment (FDI), and Domestic Direct Investment (DDI). In penalized regressions, coefficient signs should be interpreted as conditional associations after shrinkage and variable selection; when predictors are correlated, LASSO may attribute shared variation to one variable and the remaining coefficients (including their signs) can differ from simple bivariate relationships [15][17][18].

The Road Length coefficient is negative (-29,982.84). A negative sign can be reasonable because total road length measures the extent of the road network, not necessarily the effective accessibility or quality of transport connections. The transport-development literature emphasizes that economic and labor outcomes are more closely tied to access or proximity to all-season roads (e.g., the share of people living within about 2 km of a good-quality road) and to road quality than to kilometers of roads alone [19]. In the Rebana and Arumanis context, districts with longer total roads may also be geographically larger and more rural, with dispersed settlements and lower employment density; more kilometers of road are required to connect villages, but this does not automatically translate into a larger local workforce. Hence, conditional on MSMEs, GRDP and HDI, the negative coefficient can be interpreted as capturing a ‘spatial dispersion or rurality’ effect rather than implying that roads reduce employment.

The FDI coefficient is also negative (-20,922.85). Empirically, the employment effect of FDI is heterogeneous: FDI can create jobs in foreign affiliates and supplier firms, but the net effect may be limited, or even negative, when investments concentrate in capital and technology intensive activities, shift production toward automation, or intensify competition that displaces domestic firms [20]. Evidence summarized by the World Bank notes that some FDI episodes are used for

capital-intensive, technology-driven production with limited labor demand, and that effects can differ by sector and skill level. Therefore, a negative coefficient in the LASSO model plausibly suggests that FDI in the study area is concentrated in relatively capital-intensive industries (e.g., heavy manufacturing or resource-based activities), where additional investment does not proportionally expand local workforce size, or where job creation is offset by displacement effects in smaller local firms.

Similarly, the DDI coefficient is negative (-25,825.16). Domestic investment may not always be labor-absorbing, especially when it is directed to capital-intensive sectors and machinery purchases that raise output productivity but reduce labor demand. For example, evidence from Indonesia shows that investment can be negatively associated with employment when investment realization is dominated by capital-intensive sectors such as mining and when firms prioritize machines to improve productivity [21]. Related evidence also suggests that DDI can be oriented toward sectors that do not strongly absorb labor, limiting its contribution to expanding employment opportunities [22].

Model fit was evaluated using Root Mean Square Error (RMSE) and the coefficient of determination (R^2) on the training and testing sets. The RMSE was 7,106.46 on the training data and 180,823.73 on the testing data, while the R^2 value was 99.96% on the training data and 83.09% on the testing data (Table 4).

Table 4. Performance of the LASSO model

Dataset	RMSE	R^2
Training	7,106.46	99.96%
Testing	180,823.73	83.09%

Discussion

This study identifies the principal determinants of workforce absorption in the Rebana and Arumanis Areas. The positive association of MSMEs with workforce size is consistent with the broader literature showing that SMEs are major employment contributors and central to job creation, particularly in developing contexts where small firms account for a large share of businesses and a substantial portion of employment. This supports the interpretation that areas with more MSME activity tend to require and absorb more workers, reinforcing MSMEs' role as a key channel linking local economic dynamism to employment outcomes. MSMEs function as critical entry points for employment. In Indonesia, MSMEs account for a very large share of output and jobs; official and international sources consistently report contributions of ~60% of GDP and ~97% of employment [23][24]. Expanding access to finance (e.g., KUR microcredit and interest subsidies), paired with managerial support and technology adoption, is expected to raise operational efficiency and market reach, thereby increasing labor demand. Recent assessments of KUR indicate that while credit access improves firm- and household-level outcomes, the macro-level effect on MSME value added is often modest unless bundled with non-financial upgrading [25][26][27][28].

In contrast, the negative coefficient on HDI suggests that districts with higher human development levels may not necessarily exhibit larger workforces in the industries captured by our outcome variable. Importantly, this sign should not be interpreted as "human development reduces jobs," but rather as a conditional association given other covariates in the model. One plausible

mechanism is skills mismatch: as education and overall human capital improve, workers may seek higher-skill, higher-quality jobs, while local firms may still predominantly offer lower- to mid-skill positions. When the skills supplied by workers do not align with the skills demanded by employers, labor absorption can weaken even in places with relatively high HDI. When the skills demanded by employers do not align with the skills possessed by workers, labor absorption can weaken despite higher HDI, producing a negative empirical association. Evidence from Indonesia's Occupational Employment & Vacancy Survey (OEVS) and complementary SAKERNAS-based studies shows substantial vertical and horizontal mismatch, with overeducation linked to lower earnings and longer unemployment spells [29][30]. This explanation aligns with evidence that mismatch between skill supply and demand can constrain employment outcomes when structural transformation and job upgrading lag behind improvements in human development.

The positive GRDP coefficient supports standard labor-demand intuition and prior research by World Bank [31] that larger regional economic activity often coincides with larger employment levels. For Arumanis, tourism and fisheries have strong backward–forward linkages and meaningful output/employment multipliers in Indonesian IO/CGE work [32]. For Rebana, eco-industrial parks (EIPs) can consolidate supply chains and catalyze "green" manufacturing clusters with measurable job spillovers when properly governed.

The coefficient of regional minimum wage is positive. In cross-sectional regional settings, wages often capture broader differences in local productivity and agglomeration rather than acting purely as a marginal cost variable. Empirical evidence on the urban wage premium shows that workers in denser and more developed urban areas tend to earn higher wages, partly reflecting sorting and productivity advantages in cities [33]. Related local labor-market research also highlights persistent spatial differences in wages and employment conditions driven by firm composition, productivity, and mobility across places [34][35]. Therefore, the positive sign of minimum wage in our LASSO model is best interpreted as a proxy for regional economic maturity or agglomeration and higher-productivity job structures, which can coincide with larger labor markets.

The negative coefficients on realized FDI and DDI contradict the simplified narrative that investment inflows automatically expand employment. In the Rebana–Arumanis context, these negative signs are therefore consistent with an investment-composition interpretation: investment may raise output and efficiency without proportionally expanding employment, or may shift activity toward sectors with lower labor intensity [20].

The negative coefficient on road length contrasts with a common expectation in the infrastructure development literature that transport investments tend to support productivity and, indirectly, labor demand [36]. However, spatial-economy evidence suggests transport networks can generate reallocation (core–periphery) effects, where improved connectivity can concentrate activity in core nodes and weaken outcomes in peripheral areas [37][38]. In the Rebana and Arumanis context, districts with longer total roads may be geographically larger and more rural with dispersed settlements and lower employment density; thus, conditional on MSMEs, GRDP, and HDI, the negative coefficient is most plausibly interpreted as capturing a spatial dispersion/rurality proxy rather than implying that roads reduce employment.

CONCLUSION

This study used LASSO regression with 5-fold cross-validation to examine regional determinants of workforce size in the Rebana and Arumanis areas. The findings suggest that the projected number of MSMEs is the strongest positive predictor of workforce size, alongside regional economic output (GRDP) and the regional minimum wage, while HDI and several investment/infrastructure indicators show negative associations under the selected specification. In terms of model performance, the testing-set R^2 of 83.09% indicates that the selected indicators explain a substantial share of regional variation in workforce size.

From a regional-planning perspective, the results highlight the importance of strengthening the MSME ecosystem and supporting productivity-oriented local industry development to expand employment capacity, while also ensuring that human capital improvements translate into job matching and labor absorption in local sectors. This study has limitations that the analysis is based on a cross-sectional dataset with a relatively small number of regions, and the regularized coefficients reflect statistical associations rather than causal effects. Future studies could extend this work using panel or spatial econometric models, alternative regularization (e.g., elastic net), and robustness checks against Ridge estimates to better understand mechanisms and improve generalization.

REFERENCE

[1] E. F. Setiawan and T. M. Chalil, "Strategi Pengembangan Kawasan Metropolitan REBANA Menggunakan Interpretative Structural Modelling (Studi Kasus: Kabupaten Cirebon, Kabupaten Majalengka dan Kota Cirebon)," Jurnal Multidisiplin West Science, vol. 2, no. 7, p. 568 ~ 579, 2023.

[2] D. Hartoyo and dkk., Pengembangan Kawasan Rebana dan Jawa Barat Bagian Selatan, Jakarta: Deputi Bidang Koordinasi Infrastruktur dan Transportasi, Kementerian Koordinator Bidang Kemaritiman dan Investasi, 2023.

[3] J. M. Pereira, M. Basto and A. F. d. Silva, "The logistic lasso and ridge regression in predicting corporate failure," Procedia Economics and Finance, vol. 39, pp. 634-641, 2016.

[4] Z. Mei and Z. Shi, "On LASSO for high dimensional predictive regression," Journal of Econometrics, vol. 242, no. 2, pp. 1-19, 2024.

[5] G. Hesamian, A. Johannssen and N. Chukhrova, "An explainable fused lasso regression model for handling high-dimensional fuzzy data," Journal of Computational and Applied Mathematics, vol. 441, 2024.

[6] M. S. H. Shaon, T. Karim, M. S. Shakil and M. Z. Hasan, "A comparative study of machine learning models with LASSO and SHAP feature selection for breast cancer prediction," Healthcare Analytics, vol. 6, pp. 1-13, 2024.

[7] M. H. Setiawan, Analisis Penyerapan Tenaga Kerja di Jawa Barat, Yogyakarta: Universitas Islam Indonesia, 2019.

[8] E. A. Hafiz and R. Haryatiningsih, "Pengaruh PDRB, UMK, IPM terhadap Penyerapan Tenaga Kerja Kabupaten/Kota Jawa Barat 2010-2020," Journal Riset Ilmu Ekonomi, vol. 1, no. 1, pp. 55-65, 2021.

[9] A. M. Rahmah and I. A. Juliannisa, "Pengaruh Tingkat Pendidikan, Upah Minimum, Dan PDRB Terhadap Penyerapan Tenaga Kerja Di Kabupaten/Kota Provinsi Jawa Barat," Ikraith-Ekonomika, vol. 5, no. 3, pp. 246-254, 2022.
[10] H. Ferdianysah, "Analisis Faktor Serapan Tenaga Kerja Di Kawasan Pengembangan Segitiga Rebana," CR Journal, vol. 7, no. 2, pp. 87-98, 2021.
[11] Y.-A. Wang, Q. Huang, Z. Yao and Y. Zhang, "On a class of linear regression methods," Journal of Complexity, 2024.
[12] Zhao, Y. and Huo, X., "A survey of numerical algorithms that can solve the Lasso problems," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 5, 2022.
[13] A. J. Mahalani and N. A. K. Rifai, "Least Absolute Shrinkage And Selection Operator (LASSO) untuk Mengatasi Multikolinearitas pada Model Regresi Linear Berganda," in Bandung Conference Series: Statistics, Bandung, 2022.
[14] I. Sartika, N. N. Debataraja and N. Imro'ah, "Analisis Regresi Dengan Metode Least Absolute Shrinkage And Selection Operator (LASSO) Dalam Mengatasi Multikolinearitas," BIMASTER : Buletin Ilmiah Matematika, Statistika dan Terapannya, vol. 9, no. 1, pp. 31-38, 2020.
[15] R. Tibshirani, "Regression Shrinkage and Selection via the LASSO," Journal of the Royal Statistical Society, vol. 58, no. 1, pp. 267-288, 1996.
[16] D. El Shahat, A. Tolba, M. Abouhawwash and M. Abdel Basset, "Machine learning and deep learning models based grid search cross validation for short-term solar irradiance forecasting," J Big Data, 2024.
[17] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of The Royal Statistical Society, vol. 67, no. 2, pp. 301-320, 2005.
[18] J. H. Friedman, T. Hastie and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," Journal of Statistical Software, vol. 33, no. 1, pp. 1-22, 2010.
[19] World Bank, Measuring Rural Access: Using New Technologies, Washington DC: World Bank Group, 2016.
[20] A. Saurav, Y. Liu and A. Sinha, Foreign Direct Investment and Employment Outcomes in Developing Countries, Washington DC: World Bank Group, 2020.
[21] Rustam, M. Pudjihardjo and S. Muljaningsih, "The Effect of Capital-Intensive Investment on Employment: A Case Study in East Borneo," International Journal of Business, Economics and Law, vol. 19, no. 5, pp. 322-327, 2019.
[22] B. I. Nasution, A. C. B. Tarigan and S. I. Siregar, "Investment and Unemployment Reduction: An Empirical Study of Indonesia using Panel Data Regression," in Proceedings of the 3rd International Conference on Digita Society (ICDS), 2020.
[23] C. S. o. t. R. o. Indonesia, "setkab.go.id," 1 January 2022. [Online]. Available: https://setkab.go.id/en/govt-to-maintain-msmes-role-as-economic-backbone/. [Accessed 3 2025].
[24] H. Palaon, R. P. Radikun, E. P. Pramudya, A. G. B. Brata and S. Triandaru, "Digitalizing MSMEs for Employment Creation and Sustainable Development," T20 Policy Brief, 2024.
[25] S. Alibhai, H. C. Johnson, C. T. Niang and F. Strobbe, "Can Public Credit Schemes Improve Access to Finance for Small Businesses? Evidence from Indonesia," World Bank, Jakarta, 2024.

[26] R. Y. Siregar, I. K. Rohman, R. R. Ronaldo, A. T. W. Massie and I. Naufal, "Economic Impact of Kredit Usaha Rakyat (KUR)," Indonesia Financial Group, Jakarta, 2024.
[27] R. A. Subagio and S. N. Aini, "Microfinance Unveiled: Analyzing The Effects Of Kredit Usaha Rakyat (KUR) On MSMES, Financial Discipline, And Business Expansion," Journal of Allied Research in Social and Development sciences, vol. 11, no. 2, pp. 27-37, 2023.
[28] D. Setyawati and D. Hartono, "Microcredit and household energy consumption: Exploring the impact of The Kredit Usaha Rakyat (KUR) program in Indonesia," Energy Nexus, vol. 18, pp. 1-9, 2025.
[29] World Bank, Bappenas and Australian Government, "Indonesia's Occupational Employment Outlook," World Bank, Jakarta, 2020.
[30] P. Wicaksono, I. Theresia and B. Al Aufa, "Education–occupation mismatch and its wage penalties: Evidence from Indonesia," Taylor anda Francis, vol. 10, no. 3, pp. 1-23, 2023.
[31] W. Bank, World Development Report, Washington, DC: World Bank, 2013.
[32] A. Hasudungan, D. G. S. Raeskyesa, E. N. Lukas and F. Ramadhanti, "Analysis Of The Tourism Sector In Indonesia Using The Input-Output And Error-Correction Model Approach," Jurnal Ekonomi Bisnis dan Kewirausahaan (JEBIK), vol. 10, pp. 73-90, 2021.
[33] P.-P. Combes, G. Duranton and L. Gobillon, "Spatial wage disparities: Sorting matters!," Journal of Urban Economics, vol. 63, no. 2, pp. 723-742, 2008.
[34] R. K. Putri and M. Athoillah, "Artificial intelligence and machine learning in digital transformation: Exploring the role of AI and ML in reshaping businesses and information systems," in Advances in Digital Transformation-Rise of Ultra-Smart Fully Automated Cyberspace, IntechOpen, 2024.
[35] M. Enrico, "Local Labor Markets," in Handbook of Labor Economics, Amsterdam, Elsevier, 2011, pp. 1237-1313.
[36] P. J. Gertler, M. Gonzalez-Navarro, T. Gračner and A. D. Rothenberg, "Road maintenance and local economic development: Evidence from Indonesia's highways," Journal of Urban Economics, vol. 143, p. 103687, 2024.
[37] B. Faber, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," The Review of Economic Studies, vol. 81, no. 3, pp. 1046-1070, 2014.
[38] S. J. Redding and M. A. Turner, "Transportation Costs and the Spatial Organization of Economic Activity," in Handbook of Regional and Urban Economics, Elsevier, 2015, pp. 1339-1398.