Analysis of Factors Affecting Minimum Salary of Workers in Indonesia Using Binary Logistic Regression

Surjohadi [a,1,*], Sahat Renaldi [b,2], Trimono [b,3], I Gede Susrama Mas Diyasa [b,4]

[a] Department of Industrial Engineering, Faculty of Engineering, University of Yos Sudarso Surabaya
[b] Department of Data Science, Faculty of Computer Science, University of Pembangunan Nasional Veteran Jawa Timur
[1] 20083010026@student.upnjatim.ac.id*; [2] bestsuryo@yahoo.com*; [3] trimono.sd@upnjatim.ac.id, igsusrama.if@upnjatim.ac.id
* corresponding author

ARTICLE INFO

Keywords
Salary,
hypothesis testing,
logistic
regression model

ABSTRACT

Salary is an important indicator used to measure the compensation and recognition individuals receive for their contributions to the workforce. Investigating the factors that influence salary levels is an intriguing research area. This study uses a logistic regression approach to analyze the relationship and influence of job field, job level, company location, and tenure on workers' salaries in Indonesia. The research findings reveal that the variables of job level and company location have a significant relationship with the minimum salary level received by workers. Based on the logistic regression modeling results, the variables that influence the minimum salary level are the company location (foreign) and average tenure.

1. Introduction

Salary refers to the value received by individuals for their contribution to work. It reflects the rewards given by the company to its workers for the effort, time and skills invested in carrying out job duties and responsibilities. Salary not only acts as a reward for a worker's performance and dedication, but also as a factor that can motivate and encourage worker productivity. In practice, there are many factors that influence the level of salary received by a worker. Understanding the factors that affect a worker's salary is one of the important things to be known by job seekers and prospective workers. Understanding the factors that affect workers' salary levels is crucial for individuals who are currently working and those who are looking for work. By understanding these factors, individuals can have realistic expectations regarding the level of salary they can earn, and they can develop appropriate strategies to increase their salary value in the future.

Previous research related to salary has been conducted by Nuril et al. (2021) which examines the influence of several variables on employee performance using the Partial Least Square (PLS) method. The results showed that salary is one of the variables that has a significant influence on employee performance [1]. One of the factors that influence salary is the length of time worked, as researched by Natania et al. (2021) who analysed the factors that influence labour wages using multiple linear regression methods. Based on this research, it was found that the length of service has a positive and significant effect on the wage level. This shows that the longer an employee works, the higher the wage rate received [2]. Salary plays an important role in influencing the interest of prospective workers to join a company. The level of salary offered can be a factor that influences a person's decision to apply and accept a job at a particular company. This has been proven by Cynthia & Rifqi (2017) in their research which analyses what factors influence students' interest in working in Islamic financial institutions. The test results in this study show that financial / salary rewards are

one of the factors that have a significant influence on student interest in working in Islamic financial institutions [3].

Based on the results of the review of previous studies, researchers are interested in continuing research with more specific objectives. The researcher will conduct logistic regression modelling to analyse the effect of job field, position level, company location, and tenure on workers' salaries in Indonesia. The data used in this study is taken from the results of a survey of worker salaries in Indonesia conducted by HRDbacot, this survey has presented many insights from thousands of data that have been collected [4]. In this study, researchers used the logistic regression method to analyse the effect of the independent variables, namely field of work, position level, company location, and tenure on the dependent variable, namely worker salary. In this context, with the existing data, the field of work variable will be categorised into 3 different groups, namely Management & Administration, Service & Operations, and Technology & Marketing. Furthermore, the position level variable covers the hierarchical position within the company organisation, which is grouped into 3 categories, namely Low for low position level, Middle for medium position level, and High for top position level. Then the company location variable is grouped into categories of company location in Java, outside Java and abroad. Finally, the dependent variable concerns the average length of service in years. The logistic regression method was chosen because it is suitable for analysing the relationship between the independent variables to be used and the dependent variable, namely the minimum worker salary, which is a binary variable. The minimum wage variable will be categorised into 2 category classes. The minimum salary category class used in this study refers to the average Regional Minimum Wage (UMR) of workers in 2023. The Regional Minimum Wage (UMR) is a standard applied by employers and industry players to set the level of wages that must be given to employees, employees, or labourers in their work environment. The use of UMR aims to protect the rights of labourers so that they receive a fair wage that is in line with the level of work they do. In 2023, the average Regional Minimum Wage (UMR) for workers in Indonesia is IDR 2,844,193. This value is the result of calculating the average minimum wage of all provinces in Indonesia. In this study, the value of the minimum salary variable used will be grouped into 2 categories, namely "Above the average UMR" and "Below the average UMR". The researcher aims to identify the relationship and influence of each observed variable on the probability of an individual being in one of the two categories using a logistic regression approach.

2. Method

2.1. Research Flow

The flow taken to achieve the objectives of this research refers to Figure 1, which starts from the process of data preparation, modelling, hypothesis testing, model selection and interpretation.

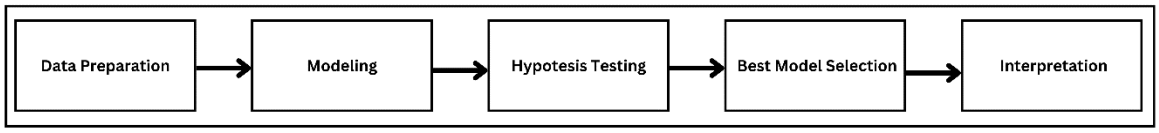

Fig. 1. Series of research flows

2.2. Data Preparation

The data used in this study was obtained through the results of a salary survey conducted by HRDbacot, a community of HR professionals in Indonesia. The survey involved the participation of respondents consisting of workers in the formal sector at various levels of positions and fields of work. The community sampled data through the distribution of questionnaires on various social media platforms. The questionnaires distributed through social media included questions that elicited information from workers. The data collected through this survey was then statistically analysed to get a more complete picture of the salary levels of workers in Indonesia. HRDbacot's survey results report includes a summary of all the data collected, providing information on various aspects of workers' salaries. The report also includes information on the respondent's field of work, position, level, company location, average length of employment, median salary, average salary, minimum salary, and maximum salary. In the context of this research, the data from the HRDbacot survey

report is used as an information base to conduct statistical modelling and analysis to understand the relationship between certain variables and workers' minimum salary levels.

Job Field	Job Title	Job level	Company Location	Avg. YOW	Median Salary	Average Salary	Minimum Salary	Maximum Salary
Account Management	Account Executive	Staff/Officer	DKI Jakarta	3	6,500,000	7,080,467	4,300,000	20,000,000
	Account Executive	Team Leader/Supervisor/Coordinato	DKI Jakarta	4	7,050,000	8,250,000	6,700,000	11,000,000
	Account Manager	Assistant Manager/Manager	DKI Jakarta	7	14,000,000	14,180,000	10,400,000	18,000,000
	Account Officer	Staff/Officer	DKI Jakarta	3	7,700,000	7,768,333	5,100,000	10,300,000
	Account Officer	Staff/Officer	DKI Jakarta	9	7,750,000	7,850,000	5,800,000	10,500,000
	Relationship Manager	Staff/Officer	DKI Jakarta	3	10,787,500	9,893,750	6,000,000	12,000,000

Fig. 2. HRDbacot survey report data snippet

Figure 2 above is a data snapshot of the salary survey report that has been conducted by HRDbacot. The variables that will be used in this study take information from the report, namely the variables of job field, position level, location of change, Avg. YOW (average length of service), and minimum salary. However, there are simplifications made in this study related to the information in Fig. 2. For the categories in the variables of company location, position level and field of work will be narrowed down to only a few categories with the aim of simplifying the analysis and allowing focus on clearer patterns later. Details of the adjustments made are contained in Table 1, Table 2 and Table 3 below:

Table 1. Simplification of company location categories

Company Location	Category
"DKI Jakarta", "Banten", "East Java", "West Java", "DI Yogyakarta", "Central Java"	Within Java Island
"South Sulawesi", "Riau", "East Kalimantan", "North Sumatra", "Bali", "NTB"	Outside Java Island
"Abroad - Australia", "Abroad - Europe", "Abroad - Asia"	Abroad

Table 2. Simplification of Job Level Categories.

Job Level	Category
"Intern / Magang", "Staff Officer"	Low
"Assistant Manager / Manager", "Team Leader / Supervisor / Coordinator"	Middle
"President Director/CEO"	High

Table 3. Simplification of Job Field Categories

Job Field	Category
Account Management, Administrative Services, Accounting, Advisory, Advisory/Consultant, Auditing, Compliance, Corporate Affairs/Communications/Public Relations, Credit and Collections, Customer Service/Customer Support, General Affair, Human Resources, Internal Audit, Legal, Admin, Document Controller, Drafter, Consultant, Management Trainee, Manager, Staff	Management & Administration
Creative, Data Science and Business Intelligence, Design, Animation & Illustration, Digital Marketing, Information Technology, Marketing, Product Development, Sales, social media & Content Creation	Technology & Marketing

Business Development, Engineering non-IT, Environmental Health and Safety, Finance & Tax, Frontliners, Logistic, Maintenance/Technician, Manufacturing Operations, Medical (Doctor, Nurse), Medical Others, Operations, PPIC, Purchasing/Procurement/Buyer, Quality Assurance, Quality Control, Research and Development, Survey, Teaching/Training, Technical Service, Transportation Services, Underwriting, Journalist, Operator, Owner, Personal Assistant, Quantity Surveyor, Reporter, Supervisor, Teller, Translator .	Operational Services

Before modeling, descriptive statistical analysis and examination of the relationship between independent variables that are categorical and the dependent variable, minimum salary, will be conducted. This is done using the chi-square test or independence test. The chi-square test can be used to test hypotheses when the population consists of two or more classes where the data is categorical [5]. The hypothesis formulation is as follows:

- H0: There is no significant relationship between the two variables
- H1: There is a significant relationship between the two variables

Similar to parametric tests, non-parametric tests, including the χ^2 (chi-square) test, assume that the data are obtained through random selection [6]. Additionally, there are several prerequisites that must be met before performing a chi-square test for independence. Here is the translated text:

1. The data used must be in the form of frequencies or counts, not percentages or other transformations of the data.
2. The observed variables must have mutually exclusive levels or categories.
3. Each subject or case should contribute data to only one cell in the contingency table.
4. The study groups being compared must be independent of each other. If there is a linkage between groups, a different test should be used.
5. For a 2 x 2 contingency table, no cell should have an expected count less than 5. For tables larger than 2 x 2, such as 2 x 3, the number of cells with an expected count less than 5 should not exceed 20%. If this condition is not met, an alternative that can be used is the Fisher's exact test [7].

Equation (1) below represents the chi-square test statistic for independence:

$$\chi^2 = \sum \frac{(\mathrm{Oi} - \mathrm{Ei})^2}{\mathrm{Ei}} \quad (1)$$

Description:
Oi = Observed value for the i-th observation
Ei = Expected value for the i-th observation
The null hypothesis H_0 is rejected if $\chi^2 > \chi^2_{(\alpha,\, v)}$, where $\chi^2_{(\alpha,\, v)}$ is the value obtained from the chi-square table at the significance level (α) and degrees of freedom (v) .. Based on the p-value criterion, H_0 is rejected when the p-value < 0.05.

2.3. Modeling

This study uses a logistic regression approach to model the relationship between independent variables (job field, job level, company location, and length of service) and the dependent variable (minimum worker salary). Regression methods are integral to any data analysis process that involves explaining the relationship between a response variable and one or more explanatory variables. One commonly used regression method for qualitative or categorical data is logistic regression. There are three types of logistic regression: binary logistic regression, ordinal logistic regression, and multinomial logistic regression. The difference among these three types of logistic regression lies in the nature of the dependent variable. Binary logistic regression is used when the dependent variable has only two categories. Ordinal logistic regression is used when the dependent variable consists of three or more categories with a specific order or ranking. Multinomial logistic regression is used

when the dependent variable consists of three or more categories without any specific order or ranking. Applying logistic regression is important in the analysis of qualitative or categorical data because it allows us to understand the relationship between independent variables and the likelihood of a certain event or category occurring. The values predicted by the logistic regression model are probabilities, which must be within the range of 0 and 1. To describe this relationship, logistic regression uses a logistic curve, also known as the sigmoid function. At very low levels of the independent variable, the probability approaches 0, meaning the likelihood of a positive event is very small. However, the probability never reaches 0 exactly due to the asymptotic nature of the logistic curve. Conversely, as the independent variable increases, the predicted probability also increases and approaches 1. However, the probability also never reaches 1 exactly because the logistic curve is asymptotic at the value of 1 [8]. Logistic regression uses the logistic function, which produces output values within the range (0; 1) and forms an S-shaped curve [9]. Observed values of the random response variable in binary logistic regression have 2 categories, namely 0 and 1, thus following the Bernoulli distribution with a probability distribution function as shown in equation (2) below:

$$P(Y = y) = \pi^{y}(1-\pi)^{1-y}, \text{To } y = 0,\ 1.$$

In binary logistic regression, a commonly used link function is the logit function. The logit function transforms the log-odds (the logarithmic ratio between the probability of an event and the probability of non-occurrence) into a linear value that can be related to the independent variables. Mathematically, the logit function is expressed as follows:

$$Logit(p) = ln\left(\frac{p}{1-p}\right)$$

Logit(p) in equation (3) is the log-odds value of the probability of an event (p), where p is the probability of the event and (1 - p) is the probability of non-occurrence. To model the relationship between the independent variables and the dependent variable, the binary logistic regression equation is used. This equation can be written as follows:

$$Logit(p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

In equation (4), β_0, β_1, β_2, ..., β_r are the estimated regression coefficients, and x_1, x_2, ..., x_r are the values of the independent variables. Logistic regression estimates these coefficients using the Maximum Likelihood Estimation (MLE) method, which provides point estimates for regression coefficients and other metrics [10]. The MLE method in logistic regression is used to find the most likely (maximum likelihood) estimates of the regression coefficients and other parameters based on the observed data. In addition to estimating regression coefficients, MLE can also produce other metrics such as standard errors, confidence intervals, and test statistics used to evaluate the statistical significance of the regression coefficients. These estimates allow for hypothesis testing of each independent variable's statistical significance and help understand the extent to which these variables contribute to the probability of a minimum salary being above or below the average UMR.

2.4. Hypothesis Testing

To determine whether there is an effect of the independent variables on the dependent variable in logistic regression analysis, tests are performed on the logistic regression model parameters, both simultaneously and individually. The simultaneous testing of parameters is conducted using the likelihood ratio test statistic (G) with the following testing procedure:

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_n = 0$

(This means that, collectively, the independent variables have no effect on the dependent variable.)

$H_1 :$ At least one $\beta_j \neq 0$, To $j = 1,\ 2,\ \ldots\ ,\ n.$

(This means that, collectively, the independent variables do affect the dependent variable.)

The test statistic for the likelihood-ratio test used in this examination is given by Equation (5) [11].

The test statistic G follows a chi-square distribution. The null hypothesis (H_0) in this test will be rejected $G > \chi^2_{(\alpha,\, v)}$. The critical value of chi-square can be obtained from the chi-square table with the significance level (α) and degrees of freedom (v), which is equal to the number of independent variables in the analysis.

Individual parameter testing is conducted to determine whether each independent variable has a significant effect on the dependent variable. The hypotheses for this testing are as follows:

$H_0 : \beta_j = 0$

(The to – j independent variable has no effect on the dependent variable.)

$H_1 : \beta_j \neq 0$, To $j = 1, 2, \cdots, p.$

(The to - j independent variable does affect the dependent variable.)

For individual testing, the Wald test statistic (W) is used, as given by Equation (6) below:

$$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2$$

The Wald test statistic follows a chi-square distribution. The null hypothesis H_0 is rejected if $W > \chi^2_{(\alpha;\, 1)}$, where the value of $\chi^2_{(\alpha;\, 1)}$ can be obtained from the chi-square table.

Next, to measure how well the model explains the dependent variable, a goodness-of-fit test is performed. This is done using the Hosmer-Lemeshow test statistic ($\hat{C}$)as given by Equation (7), with the testing procedure as follows:

H_0:The model fits well (There is no difference between the observed results and the predicted results.

H_1:The model does not fit well (There is a difference between the observed results and the predicted results.)

$$\hat{C} = \sum_{k=1}^{g} \frac{\left(o_k - n_k' \bar{\pi}_k\right)^2}{n_k' \bar{\pi}_k \left(1 - \bar{\pi}_k\right)}$$

where:

g = Number of groups

n_k' = Number of subjects in the group to -k

$o_k = \sum_{j=1}^{c_k} y_j$, Number of independent variable values in c_k combination of independent variables.

$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}(x_j)}{n_k'}$, Average estimated probability with m_j is the number of subjects.

dengan c_k kombinasi variabel independen.

H_0 ditolak bila $\hat{C} > \chi^2_{(\alpha,\, v)}$, dengan $\chi^2_{(\alpha,\, v)}$ nilai yang dapat diperoleh dari tabel chi-square yang mempunyai tingkat signifikansi (α) dan derajat bebas (v) adalah $g-2$.

2.5. Best Model Selection

At this stage, variables that are not related or do not have a significant effect on the minimum salary will be discarded and not used in subsequent modeling. The model to be formed will combine the backward and forward elimination algorithms. The backward elimination algorithm is used to remove variables that have the smallest impact on the model. Initially, all remaining variables are included in the model. Then, the variable with the smallest coefficient (the least impact) will be removed one by one until no variable with a small significant impact remains in the model. Backward elimination involves regressing all predictor variables with the response variable and eliminating predictor variables based on the smallest F (partial) value [12]. The forward selection method is the opposite of the backward method [13]. This method starts with no variables in the model and adds variables one by one until no other variable can contribute significantly to the model's results. Afterward, evaluation is conducted by examining the AIC (Akaike Information Criterion) value. The AIC value displayed indicates the relative quality of a model in explaining the data. The best regression model is the one with the smallest AIC value [14].

2.6. Interpretation

After evaluating and selecting the model based on the lowest AIC value, the next step is interpretation. In logistic regression modeling, the goal of parameter interpretation is to understand the meaning of the estimated parameter values for predictor variables. One limitation of logistic regression is that its results cannot be interpreted directly as probabilities [15].

To interpret the logistic regression parameters for categorical variables, odds ratios are used. The odds ratio is the comparison between the odds of a category and the odds of a reference category. To obtain the odds ratio from logistic regression coefficients, we take the exponent of the coefficient. For example, if we have a logistic regression coefficient of β, then the odds ratio is exp(β). The interpretation of this odds ratio is that if the independent variable increases by one unit, the odds of experiencing the desired outcome will increase by exp(β) times compared to the other category within the same variable. For interpreting logistic regression parameters of continuous variables, an approach involving the assumption that continuous independent variables have a linear relationship with the logit (log odds) of the dependent variable is used. This means that each unit change in the independent variable will result in a proportional change in the log odds of the dependent variable.

3. Results and Discussion

3.1. Descriptive Statistical Analysis and Independence Testing

This section presents the descriptive statistical analysis and independence testing. The purpose of this analysis is to provide a deeper understanding of the characteristics of the data and the relationships between the variables studied. Table 4 displays the independent and dependent variables used in this research. The independent variables include the field of work, job level, company location, and average length of employment, while the dependent variable is the minimum salary. The minimum salary variable has been categorized into two groups, 0 and 1, with the interpretation that 0 indicates a salary below the average Regional Minimum Wage (UMR) / less than IDR 2,844,193, while 1 indicates a salary above the average UMR.

The next step in this section involves conducting descriptive statistical analysis and independence testing. The purpose of descriptive statistical analysis is to describe the characteristics of the data for each variable, such as mean, median, and other basic values. Additionally, independence testing is conducted to examine whether there is a significant relationship between the observed variables. This is aimed at gaining a deeper understanding of the data used in this research.

Table 5 presents a descriptive statistical summary for the numeric variable "Average Length of Employment." The "Average Length of Employment" variable represents the average working tenure of respondents in years. As shown in Table 5, the minimum value for this variable is 0, indicating that some respondents have less than one year of work experience or no prior work experience. Based on this descriptive statistical summary, it can be concluded that the majority of respondents have relatively short average working tenures, with 75% of observations (Q3) averaging 4 years of employment, and an overall range between 0 and 10 years.

Table 4. Research Variables

Variable Type	Variable Name	Data Type & Scale
Independent	Field of Work	Categorical Nominal
	Job Level : Low (1) Middle (2) High (3)	Categorical Ordinal
	Company Location	Categorical Nominal
	Average Length of Employment	Numerical
Dependent	inimum Salary Above average UMR (1) Below average UMR (2)	Categorical

Table 5. Summary of Numeric Variables

Variabel	Min	Q1	Mean	Median	Q3	Max
Average Length of Employment	0	2	2.87	3	4	10

Table 6. Frequency Distribution of Categorical Variables

Characteristic	Frequency (n)	Percentage (%)
Field of Work:		
Management & Administration	133	35.66
Technology & Marketing	134	35.92
Operational Services	106	28.42
Total	373	100
Job Level:		
Low	320	88.47
Middle	42	11.26
High	1	0.27
Total	373	100

Company Location:		
Within Java Island	335	89.81
Outside Java Island	9	2.41
Overseas	29	7.78
Jumlah	373	100
Minimum Salary:		
Above Average UMR	298	79.89
Below Average UMR	75	20.11
Total	373	100

Table 6 shows that the data used in this study consists of 373 records. The highest record counts for each variable are in the Technology & Marketing field of work with 134 records (35.92%), Low job level (88.47%), company location within Java Island (89.81%), and minimum salary category above the average UMR (79.89%).

Table 7. Contingency Between Field of Work and Minimum Salary

Field of Work	Management & Administration	Technology & Marketing	Operational Services	Jumlah
0	28	22	25	75
1	105	112	81	298
Total	133	134	106	373

Table 7 depicts the contingency between the field of work and minimum salary variables. In the Management & Administration field, there are 28 observations with salaries below the average UMR and 105 observations with salaries above the average UMR. In the Technology & Marketing field, there are 22 observations with salaries below the average UMR and 112 observations with salaries above the average UMR. Meanwhile, in the Operational Services field, there are 25 observations with salaries below the average UMR and 81 observations with salaries above the average UMR. In summary, it can be concluded that the Technology and Marketing field has a higher frequency of observations with salaries above the average UMR compared to other fields of work.

Table 8. Contingency Between Field of Work and Minimum Salary

Test Type	χ^2	p-value
Chi-square	2.0074	0.3665

The results of the independence test between the field of work and minimum salary variables can be seen in Table 8. The p-value of 0.3665 is greater than the significance level used (0.05). Therefore, there is not enough evidence to reject the null hypothesis (H0). The conclusion is that there is no significant relationship between the field of work and minimum salary variables.

Table 9. Contingency Between Job Level and Minimum Salary

Job Level	Low	Middle	High	Total
0	73	1	1	75
1	257	41	0	298

Total	330	42	1	373

Table 9 shows the contingency between the job level and minimum salary variables. There are 330 observations (88.47%) with a low job level, 42 observations (11.26%) with a middle job level, and 1 observation (0.26%) with a high job level. For independence testing, the Fisher exact test will be conducted because more than 20% of cells have expected frequencies < 5.

Table 10. Independence Test Between Job Level and Minimum Salary Variables

Test Type	p-value
Fisher exact	0.0002695

The independence test between job level and minimum salary variables using the Fisher exact test method, as shown in Table 10, results in a p-value of 0.0002695. Since the p-value is less than 0.05, the null hypothesis (H0) is rejected, indicating a significant relationship between job level and minimum salary variables.

Table 11. Contingency Between Company Location and Minimum Salary

Company Location	Within Java Island	Outside Java Island	Overseas	Total
0	59	3	13	75
1	276	6	16	298
Total	335	9	29	373

Table 11 shows the contingency between company location (Within Java Island, Outside Java Island, Overseas) and minimum salary. There are 335 observations with company locations within Java Island, 9 observations with locations outside Java Island, and the remaining 29 observations are located overseas.

Table 12. dependence Test Between Company Location and Minimum Salary Variables

Test Type	p-value
Fisher exact	0.001583

Based on the independence test between company location and minimum salary variables using the Fisher exact test method in Table 12, the p-value is 0.001583. Since the p-value is less than 0.05, this indicates a significant relationship between company location and minimum salary variables.

3.2. Modeling and Model Testing

After conducting descriptive analysis and independence testing, logistic regression modeling and model testing are then performed to observe how each independent variable influences the dependent variable. Based on the parameter estimation results in Table 13, the logistic regression model is obtained as shown in Equation 8 below:

$$Logit(p) = 0.08074 + 0.14083x_{1\,(1)} + 0.49691x_{1\,(2)} + 0.30499x_2 - 1.30226x_{3\,(1)} - 0.55424x_{3\,(2)} + 0.35506x_4$$

Table 13. below shows the output of the initial modeling using all available variables.

Table 13. Parameter Estimation Results of the Base Model

Parameter	estimasi	Standar Error	Wald	p-value
Intercept	0.08074	0.69172	0.01362	0.90707

Field of Work - Management & Administration	0.14083	0.33551	0.17619	0.67466
Field of Work - Technology & Marketing	0.49691	0.33946	2.14284	0.14324
Job Level	0.30499	0.65201	0.21881	0.63995
Company Location - Overseas	-1.30226	0.41632	9.7843	0.00176 **
Company Location - Outside Java Island	-0.55424	0.76231	0.5286	0.4672
Average Length of Employment	0.35506	0.10887	10.3503	0.00111 **

Based on the model in equation (8), simultaneous parameter testing will be conducted. The test is performed using the likelihood ratio test (G) statistic shown in Table 14.

Table 14. Likelihood Ratio Test (G)

Test Type	G	p-value
Likelihood Test	30.84	0.00002

The likelihood ratio test results in Table 14 show a p-value (0.00002) < 0.05. Therefore, at the 5% significance level, there is sufficient evidence to reject the null hypothesis. The conclusion is that at least one independent variable in the model has a significant effect on the minimum salary.

In addition to the likelihood ratio test, individual testing is also performed using the Wald test to determine which variables significantly affect the model. A variable is considered to have a significant effect if the p-value < 0.05. Based on Table 13, the variables that significantly affect the minimum salary are Company Location - Overseas ($X_{3\,(1)}$)and Average Length of Employment (X_4).

Considering the results of the independence test and Wald test, the field of work variable will be excluded from the model. The field of work variable is excluded from the model because its relationship and influence are not significant on the minimum salary. Subsequently, modeling is performed again using the variables Job Level, Company Location, Average Length of Employment, and Minimum Salary. Table 15 below displays the AIC values obtained from the modeling using backward and forward methods.

Table 15. AIC Model Comparison

Model	AIC
Backward	354.0655
Forward	355.8858

The backward model shows better quality compared to the forward model based on the given AIC value, which is 354.0655. Therefore, the backward model is the final model to be interpreted. However, before that, a model fit test will be conducted using the Hosmer-Lemeshow test ($\hat{C}$).

Table 16. Hosmer-Lemeshow Test ($\hat{C}$)

Test Type	$\hat{C}$	p-value
Hosmer-Lemeshow	3.9574	0.8609

At the 0.05 significance level, there is not enough statistical evidence to reject the null hypothesis because the p-value in Table 16 is 0.8609. Therefore, based on the test results, it can be concluded that the backward model built is a good fit for the observed data.

3.3. Interpretation

```
Call:
glm(formula = Minimum.Gaji ~ Lokasi.Perusahaan + Rata.rata.Lama.Bekerja,
    family = binomial(link = "logit"), data = datamodel)

Coefficients:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                       0.5862     0.2723   2.153 0.031346 *
Lokasi.PerusahaanLuar Negeri     -1.3148     0.4131  -3.183 0.001460 **
Lokasi.PerusahaanLuar Pulau Jawa -0.6483     0.7382  -0.878 0.379857
Rata.rata.Lama.Bekerja            0.3727     0.1007   3.699 0.000216 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 374.41  on 372  degrees of freedom
Residual deviance: 346.07  on 369  degrees of freedom
AIC: 354.07

Number of Fisher Scoring iterations: 5
```

Fig. 3. Results of the Parameter Estimation for the Backward Model

Referring to Figure 3, the backward logistic regression model formed is as follows:

$$Logit(p) = 0.5862 - 1.3148x_{3\,(1)} - 0.6483x_{3\,(2)} + 0.3727x_4$$

Table 17 below shows the odds of each variable, indicated by the exp(β) value of the log odds ratio.

Table 17. Odds Against the Target

Parameter	log odds	exp(β)
Intercept	0.5862	1.79
Lokasi.PerusahaanLuar Negeri $(x_{3\,(1)})$	-1.3147	0.26
Lokasi.PerusahaanLuar Pulau Jawa $(x_{3\,(2)})$	-0.6482	0.52
Rata.rata.Lama.Bekerja (x_4)	0.3726	1.45

Based on the exp(β) values in Table 17, it can be interpreted that the odds of a worker employed overseas earning at least a salary above the regional minimum wage (UMR) is 0.26 times that of a worker employed in Java Island. Meanwhile, the odds of a worker employed outside Java Island (but still within the country) earning at least a salary above the UMR is 0.52 times that of a worker employed in Java Island. Lastly, for each 1-point increase in the average length of employment, the odds of the worker earning at least a salary above the UMR increases by 1.45 times.

4. Conclusion

Understanding the factors that contribute to salary levels is crucial for both workers and job seekers. This study aims to identify the factors that impact the salary levels received by workers by using information such as field of work, job level, company location, and average length of employment as predictor variables. The independence test results indicate that job level and company location variables have a significant relationship with the minimum wage level discussed in this study. Additionally, logistic regression modeling concludes that workers employed overseas have a 0.26

times higher chance of earning a minimum salary above the regional minimum wage (UMR) compared to workers employed on Java Island. Furthermore, workers employed outside Java Island (but still within the country) have a 0.52 times higher chance of earning a minimum salary above the UMR compared to workers on Java Island. Lastly, each 1-point (year) increase in the average length of employment increases the likelihood of a worker earning a minimum salary above the UMR by 1.45 times.

References

[1] N. Iksyaniyah, L. Hakim, dan A. Listyadi, "Effects of Work Experience, Education Level, and Wages on Employee Performance with Religiosity as Moderating Variables," Tech. Soc. Sci. J., vol. 16, hal. 355–365, 2021, [Daring]. Tersedia pada: https://techniumscience.com/index.php/socialsciences/article/view/2562

[2] N. I. Susanto et al., "An Analysis on the Factors Affecting the Worker Wages Level in the Agribusiness Industry in Tumpaan District, South Minahasa Regency (Case Study PT. Tropica Coco Prima)," Jurnal EMBA, vol. 9, no. 3, hal. 1152–1161, 2021. doi : 10.35794/emba.v9i3.35325

[3] C. Candraning dan R. Muhammad, "Faktor-faktor yang mempengaruhi minat mahasiswa bekerja di lembaga keuangan syariah," Jurnal Ekonomi & Keuangan Islam, vol. 3, no. 2, hal. 90–98, 2017, doi: 10.20885/jeki.vol3.iss2.art5

[4] M. S. Rizky dan A. Ekananta, "SALARY SURVEI Ala-ala HRDBacot," 2022. https://supersonality.com/wp-content/uploads/2021/12/Salary-Survey-Report-hrdbacot-2022-Final-Bungkus.pdf

[5] N. F. Rochmawati, W. H. Riyanto, dan I. Nuraini, "Hubungan Tingkat Pendidikan, Usia, Dan Pengalaman Keja Terhadap Pendapatan Pekerja Wanita Pada Industri Kerajinan Dompet Ida Collection Di Desa Pulo Kecamatan Tempeh Kabupaten Lumajang," Jurnal Ilmu Ekonomi, vol. 2, no. 3, hal. 399–408, 2018, doi: 10.22219/jie.v2i3.7098

[6] M. L. McHugh, "The Chi-square test of independence," *Biochemia Medica*, vol. 23, no. 2, hal. 143–149, 2012, doi: 10.11613/BM.2013.018.

[7] H. Kim, "Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test," Restorative Dentistry & Endodontics, vol. 42, no. 2, hal. 152-155, 2017, doi: 10.5395/rde.2017.42.2.152.

[8] E. Y. Boateng dan D. A. Abaye, "A Review of the Logistic Regression Model with Emphasis on Medical Research," Journal of Data Analysis and Information Processing, vol. 07, no. 04, hal. 190–207, 2019, doi: 10.4236/jdaip.2019.74012.

[9] A. Strzelecka, A. KurdyS-Kujawska, dan D. Zawadzka, "Application of logistic regression models to assess household financial decisions regarding debt," Procedia Comput. Sci., vol. 176, hal. 3418–3427, 2020, doi: 10.1016/j.procs.2020.09.055.

[10] K. C. Asanya, M. Kharrat, A. U. Udom, dan E. Torsen, "Robust Bayesian approach to logistic regression modeling in small sample size utilizing a weakly informative student's t prior distribution," Commun. Stat. - Theory Methods, vol. 52, no. 2, hal. 283–293, 2023, doi: 10.1080/03610926.2021.1912767.

[11] Aprilyani, V., "Analisis Regresi Logistik dan Aplikasinya pada Penyakit Anemia untuk Ibu Hamil di RSKD Ibu dan Anak Siti Fatimah Makassar," Skripsi, Universitas Negeri Makassar, Makassar, 2017.

[12] W. Pramesti, W. Utami, dan F. Fitriani, "Regresi Logistik Backward Elimination pada Risiko Penyebaran Covid-19 di Jawa Timur," VARIANSI J. Stat. Its …, vol. 3, no. 3, hal. 162–170, 2021, doi: 10.35580/variansiunm26132.

[13] M. Z. I. Chowdhury dan T. C. Turin, “Variable selection strategies and its importance in clinical prediction modelling,” Fam. Med. Community Heal., vol. 8, no. 1, hal. 1–7, 2020, doi: 10.1136/fmch-2019-000262.

[14] S. Annas, A. Aswi, M. Abdy, dan B. Poerwanto, “Binary Logistic Regression Model of Stroke Patients: A Case Study of Stroke Centre Hospital in Makassar,” Indonesian Journal of Statistics and Its Applications, vol. 6, no. 1, hal. 161–169, 2022, doi: 10.29244/ijsa.v6i1p161-169.

[15] E. C. Norton, B. E. Dowd, dan M. L. Maciejewski, "Odds Ratios - Current Best Practice and Use," JAMA, vol. 320, no. 1, hal. 84-85, Jul. 3, 2018, doi: 10.1001/jama.2018.6971.