Multiple Linear Regression Analysis on Factors that Influence Employees Work Motivation

Gracia Theofani1)*, Eko Sediyono2)

1) Universitas Kristen Satya Wacana, Indonesia
2) 672018278@student.uksw.edu, 2)eko@uksw.edu

Submitted : June 9, 2022 | Accepted : June 16, 2022 | Published : July 2, 2022

Abstract: This study aims to analyze the factors that influence employee motivation. The analysis was carried out using four variables, namely Facilities, Work Environment, Salary and Jobdesk. In collecting data, this study used primary and secondary data. Primary data were collected by distributing questionnaires to employees using the Voluntary Sampling Technique, questionnaires were distributed to employees with a willingness to participate in the research. From the questionnaire obtained 71 data as primary data. Secondary data used results of previous studies and data from company which was given to assist the research. Furthermore, data processing is done by using multiple linear regression algorithm. From this research, it is concluded that the variables that have the most influence on employee motivation are Facilities and Jobdesk, where these variables have a positive effect on employee motivation. Meanwhile, the work environment and salary variables have a p-value < 0.05 so that they are eliminated by the backward elimination method. The results of the data analysis show that an increase in facilities will increase work motivation by 20.659% and jobdesk will increase work motivation by 27.901%. Through this research, it is hoped that the company can improve the facilities and distribution of jobdesk properly so that employee work motivation also increases.

Keywords: employee motivation, facility, jobdesk, RStudio, salary, work environment

INTRODUCTION

In a company that is established, of course it has Human Resources (HR) who are often called employees. In the Kamus Besar Bahasa Indonesia (KBBI), an employee is someone who works and gets a salary in an institution or company (KBBI, 2012). A company cannot stand alone without employees who help in it. Employees have an important role in the running of a company. Competition in today's business world is increasingly competitive. To deal with this, companies need employees who are experienced in terms of knowledge and skill. In improving their experience, employees need motivation to work. According to Ernest J. McCormick in a quote from A.A Prabu Mangkunegara (2005:94), work motivation is an influential condition in generating, directing and maintaining behavior related to the work environment (Agustininrum, 2010). With work motivation, it will give a sense of enthusiasm, desire, and sincerity to work as an employee. The higher the sense of enthusiasm and desire to work will make work more optimal, so that it can increase work productivity (LinoHR, n.d.).

Frederick Herzberg, the psychologist who founded the Hygiene Motivation theory, examines the question of what people want from their jobs. The results of the employee respondents stated that the responses of employees who were happy with the job and not happy with the job were different. Employees who are not happy with their work tend to respond from extrinsic factors such as salary, company policies, working conditions, and supervision (Coretanzone, 2017).

Giving work motivation can be done by two methods, namely direct motivation and indirect motivation. Direct motivation means that the leadership is actively involved in providing motivation to employees. For example, the leader gives praise and appreciation to employees. Indirect motivation is motivation carried out by the company by providing facilities that can increase employee morale. It is very good if superiors can
understand what motivates an employee in their work. Lack of direct or indirect motivation from the company can reduce the work motivation of an employee (Agustyawati, Djohar, & Su’udi, 2019). This can have an impact on the employee's desire to stay with the company or leave looking for another company.

One of the companies which is engaged in the insurance sector, experienced an increase in the number of employees leaving and only a small number of employees remained. The following is data on the total number of employees in the past few years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Early year employees</th>
<th>Employees leave</th>
<th>Year-end employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 2019</td>
<td>1012</td>
<td>305</td>
<td>707</td>
</tr>
<tr>
<td>2019</td>
<td>2271</td>
<td>996</td>
<td>1275</td>
</tr>
<tr>
<td>2020</td>
<td>1901</td>
<td>1058</td>
<td>843</td>
</tr>
<tr>
<td>2021</td>
<td>1150</td>
<td>502</td>
<td>648</td>
</tr>
</tbody>
</table>

It can be seen from Table 1, the total number of employees leaving before 2019 to 2020 has increased. Meanwhile, in 2021, the total number of employees who left began to decline. Employees leave due to many factors, some of which often occur, namely resignation and dismissal of employees.

Based on this background, the purpose of this study for companies is to be able find out what factors can affect an employee's work motivation and to find out how the influence of these variables on employees motivation. So that later the company that has the same case, can reduce the number of employees who leave by maximizing the work motivation of an employee.

**LITERATURE REVIEW**

The study, entitled Analysis of Factors Affecting Employee Work Motivation in the Regional Secretariat of Kaur Regency, was conducted by Sisardi in 2018. This study aims to determine the effect of job satisfaction and work environment on employee work motivation. The results of this study indicate that the work environment has an effect on work motivation and job satisfaction has an effect on work motivation (Sisardi, 2016).

Furthermore, research conducted by Alden Nelson (2021) regarding the Factors Influencing Employee Motivation in the Electronic Industry in Batam City. This research has 3 variables which were tested by quantitative descriptive method using Non Probability Sampling technique. The results of the hypothesis indicate that the three variables are proven to be factors that affect employee motivation (Sisvard, 2021).

Another study was conducted by Sisvana Damayanti with the title Factors Connected With Employee Employment Motivation In General Hospital Regional North Paser Regional East Kalimantan, Year 2014. This study aims to determine the motivation of employees to work at the institution. Of the 7 variables used, the achievement variable is a motivating factor for employees to work at the institution with a percentage of 87.3% (Management, Manajemen, & Islam, 2018).

The fourth study conducted by Bobby Roy Zano and Thomas Santos (2019) entitled Analysis of the Effect of Product Quality, Price and Advertising on Purchase Decisions for Yamaha Motorcycles at Pt Surya Timur Sakti, East Java, Surabaya. The study used multiple linear regression to determine the effect of product quality, price and advertising on motorcycle purchasing decisions. The result research in a real goal, that product quality, price and advertising had an effect. In addition, another very influential factor is advertising (Zano & Santos, 2019).

Another study conducted by Siska Ernida Wati, Djakaria Sebayang, and Rachmad Sitepu (2013) regarding the Comparison of Fuzzy Methods with Multiple Linear Regression in Forecasting Total Production. The purpose of this study is to compare the results of forecasting using the two methods. After processing the data, the results of the study show that the average relative error value of the multiple linear regression method is smaller than that of the fuzzy method. From this research, it can be concluded that forecasting using linear regression is better than fuzzy (Wati, Sebayang, & Sitepu, 2013).

The latest research from Fitria Norma Aula Zahro etc, entitled Analysis of Factors Affecting The Level Unemployment in North Sulawesi During Pandemic COVID-19 Year 2020. This study aims to determine the factors that influence the unemployment rate in that region using multiple linear regression. From the four variables used, it is obtained that the average length of schooling and population are the main factors affecting the unemployment rate with R² 73.61% (Ekonomi et al., 2021).

This study uses 5 factor variables in determining the effect of work motivation, including: Work Motivation, Facilities, Work Environment, Salary, Jobdesk. This research was processed using the multiple linear regression method because it looked at previous studies that used this method to produce good outputs. Data processing is done using RStudio.

*name of corresponding author
METHOD

The sources used in this research are primary data and secondary data. Primary data was obtained through the distribution of online questionnaires to employees. The questionnaire was made into two parts, the respondent’s identity section and the question section which was divided into five sub-chapters. The sub-chapter contains the variables used in the research, including Work Motivation, Facilities, Work Environment, Salary and also Jobdesk. Each sub-chapter has three Likert scale questions with 5 response options, namely Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), Strongly Disagree (SD). There are 71 correspondents from different divisions.

Secondary data obtained from the company. The data provided is raw data with information on the date of entry and date of employee exit. From the raw data provided by the company, it is processed to obtain data in the form of the total number of employees who enter and leave from <2019 to 2021. The results of the data processing serve as supporting data in writing the background of this research.

The algorithm used is the Multiple Linear Regression method, with a total of five variables, with one variable as the dependent variable, namely Work Motivation (Y) and four variables as independent variables: Facilities (X₁), Work Environment (X₂), Salary (X₃), and Jobdesk (X₄).

Data analysis was performed using multiple linear regression analysis and the tool used was Rstudio software. The following is a flow chart that explains the stages in conducting research using the Multiple Linear Regression method.

![Research Flowchart](image-url)

**Multiple Linear Regression**

Multiple Linear Regression has a similar concept to Simple Linear Regression. What distinguishes it is the number of independent variables. Simple Linear Regression has only one independent variable, while Multiple Linear Regression can have more than one independent variable. Simple Linear Regression and Multiple Linear Regression can be used to predict the future and determine the effect between variables (Abdilah, 2019).
Description:
\[ Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n \]  
(1)

\begin{align*}
Y &= \text{dependent variable} \\
X &= \text{independent variable} \\
a &= \text{constant (intercept)} \\
b &= \text{regression coefficient (slope)}
\end{align*}

Forming a Regression Model

In forming the regression model, the best model is selected. There are three kinds of methods for selecting the best model, including backward, forward, and stepwise. In this study, to choose the best model, the backward method was used. The backward method is performed backwards. Each variable is regressed first, then gradually reduces one by one the variables that are not significant or have a larger p-value (Samosir, Siagian, & Bangun, 2014).

Parameter Significance Test

Overall Test (F)

Overall test (F) is needed to determine the significance of the effect between all independent variables on the dependent variable (Sisardi, 2016). The model will be significant if the p-value < 0.05.

Partial Test (T)

Partial Test (T) is needed to determine the significance of the effect of each independent variable separately on the dependent variable (Sisardi, 2016). The model will be significant if the p-value < 0.05.

Coefficient of Determination Test (\(R^2\))

The Coefficient of Determination (\(R^2\)) test was conducted to determine the strength of the model in describing the variables. The higher the \(R^2\) value, the better the prediction model made in the study (Ghozali, 2018).

Classic Assumption Test

Normality Test

Normality test is used to determine whether there is a normal distribution in the model and its variables. The model is said to be good if the data distribution is normal with p-value > 0.05 (Ghozali, 2016). Tests in this study using the Shapiro-Wilk test.

Autocorrelation Test

Autocorrelation test was conducted to determine whether there was a correlation in each independent and dependent variable separately or in combination. The best model is a model that has no autocorrelation with p-value > 0.05 (Anwar Hidayat, 2017). Tests in this study using the Durbin Watson test.

Heteroscedasticity Test

Heteroscedasticity test was carried out to determine whether there was an inequality of residual variance. The regression model is said to be good if there is no heteroscedasticity or different variance and p-value > 0.05 (A & Irfan, 2014). Tests in this study using the Breusch-Pagan test.

Multicollinearity Test

Multicollinearity test is needed to find out whether there is a correlation between independent variables. The best model is a model that does not have multicollinearity, it can be seen from the value of Variance Inflation Factor (VIF) 10 (Zamora, 2017).

RESULT

Multiple Linear Regression Analysis

From the existing questionnaires, it is processed into data which will later be analyzed using RStudio Software. Before starting to analyze, first import the necessary data files into the RStudio Software.

Program Code 1 Multiple Linear Regression Model

```
1 library(lmtest)
2 modell <- lm(MotivasiKerja~.,data = DataKuisioner)
3 summary(modell)
```

*name of corresponding author
The first step is to import the libraries needed in the analysis process. The library needed for multiple linear regression analysis is lmtest. First create the first model by initializing model1. Then model1 is processed using Multiple Linear Regression as in Program Code 1. line-2. So that the output of Multiple Linear Regression in the first model is obtained as below.

| Table 2 Multiple Linear Regression Output |
|-----------------|--------|-------|-------|
| Estimate        | Std.Error | t-value | p-value |
| Intercept       | 7.31481  | 1.58423 | 4.617  | 1.85e-05 |
| Facilities (X_1)| 0.26603  | 0.10978 | 2.423  | 0.01813  |
| Work            | -0.15915 | 0.10132 | -1.571 | 0.12104  |
| Environment (X_2)| 0.06519  | 0.10992 | 0.593  | 0.55517  |
| Salary (X_3)    | 0.30206  | 0.09614 | 3.142  | 0.00251  |

From the results of the analysis, the estimated model is:

\[
\hat{Y} = 7.31481 + 0.26603X_1 - 0.15915X_2 + 0.06519X_3 + 0.30206X_4
\] (2)

Based on Table 2, it is known that the p-value is still not in accordance with the provisions. In determining the best model, it is necessary to follow the provisions of the p-value < 0.05. For this reason, it is necessary to do the elimination using the Backward Elimination method. Because the variable Salary (X_3) has the largest p-value than the other variables, this variable will be eliminated and reanalyzed.

Program Code 2 Multiple Linear Regression Model2

```r
model2 <- lm(MotivasiKerja ~ Fasilitas+LingkunganKerja+Jobdesk, data = DataKuisioner)
summary(model2)
```

The process carried out in the second model is no different from the first model. What makes the difference is that the library is only imported once, and the variables used are the variables that are left after elimination. In addition, the second model is initialized with model2 and the output of multiple linear regression in the second model can be seen in Table 3.

| Table 3 Multiple Linear Regression Output Backward-1 |
|-----------------|--------|-------|-------|
| Estimate        | Std.Error | t-value | p-value |
| Intercept       | 7.75135  | 1.39603 | 5.552  | 5.22e-07 |
| Facilities (X_1)| 0.26928  | 0.10911 | 2.468  | 0.0161   |
| Work            | -0.14581 | 0.09832 | -1.483 | 0.1427   |
| Environment (X_2)| 0.31405  | 0.09353 | 3.358  | 0.0013   |

From the results of the analysis, the estimated model is:

\[
\hat{Y} = 7.75135 + 0.26928X_1 - 0.14581X_2 + 0.31405X_4
\] (3)

After doing the elimination, in Table 3 still found p-value > 0.05. Therefore, elimination is carried out on the Work Environment variable (X_2) and multiple linear regression analysis is performed again.

Program Code 3 Multiple Linear Regression Model3

```r
model3 <- lm(MotivasiKerja ~ Fasilitas+Jobdesk, data = DataKuisioner)
summary(model3)
```

In the 3rd model, elimination and multiple linear regression analysis were carried out, the results obtained as shown in Table 4.

| Table 4 Multiple Linear Regression Output Backward-2 |
|-----------------|--------|-------|-------|
| Estimate        | Std.Error | t-value | p-value |
| Intercept       | 7.18280  | 1.35415 | 5.304  | 1.33e-06 |
| Facilities (X_1)| 0.20659  | 0.10147 | 2.036  | 0.0456   |
| Jobdesk (X_4)   | 0.27901  | 0.09129 | 3.056  | 0.0032   |

From the results of the analysis, the estimated model is:

\[
\hat{Y} = 7.18280 + 0.20659X_1 + 0.27901X_4
\] (4)

*name of corresponding author
After eliminating some of the existing variables, the best multiple linear regression model estimation was reached with a p-value < 0.05. So, the significant variables are the Facility (X_1) and Jobdesk (X_4) variables with the equation as above.

Parameter Significance Test
Overall Test (F)

Table 5 Overall Test (F) RStudio

<table>
<thead>
<tr>
<th>p-value</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001897</td>
<td>0.05</td>
</tr>
</tbody>
</table>

The Overall Test (F) performed at Rstudio can be obtained using program code 3. From the output of Program Code 3, the result of the Overall Test (F) is the p-value at the end of the output. Then from the results in Table 5, it is known that the p-value (0.0001897) < (0.05), so it can be concluded that the model is feasible to use or the model is significant.

Partial Test (T)

Table 6 Partial Test (T) RStudio

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>p-value</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>β_1</td>
<td>0.0456</td>
<td>0.05</td>
</tr>
<tr>
<td>β_4</td>
<td>0.0032</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Partial Tests (T) performed at Rstudio can be obtained using Program Code 3 as well. From the output of program code 3, the result of the Overall Test (F) is Pr(>|t|) or p-value contained in each significant variable. Then based on the test results in Table 6, it is known that the p-value of the Facility variable (X_1) is 0.0456 and the Jobdesk variable (X_4) is 0.0032. Both show a value smaller than (0.05), so the conclusion obtained is that these two variables have a significant effect on the model.

Coefficient of Determination Test (R^2)

Table 7 Coefficient of Determination Test (R^2) RStudio

Residual standard error : 1.342 on 66 degrees of freedom
Multiple R-squared 0.2515

The Coefficient of Determination (R^2) performed at Rstudio can be obtained using Program Code 3 as well. The value used is Multiple R-squared of all variables. Then from the results of the analysis in Table 4, the value of the coefficient of determination (R^2) is 0.2515. So it can be concluded that the influence of the independent variables, namely Facilities (X_1), Work Environment (X_2), Salary (X_3) and Jobdesk (X_4) on the dependent variable, namely Work Motivation (Y) is 25.15%. In this case, the remaining 74.85% is influenced by other variables outside of this linear regression model.

Classic Assumption Tests

Normality Test

Program Code 4 Normality Test RStudio

<table>
<thead>
<tr>
<th>1</th>
<th>residual = resid(model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>library(stats)</td>
</tr>
<tr>
<td>3</td>
<td>shapiro.test(residual)</td>
</tr>
</tbody>
</table>

To perform a normality test on Rstudio, a stats package/library is needed in order to run Program Code 4. After the program code is executed, the output is obtained as shown in Table 8.

Table 8 Normality Test

<table>
<thead>
<tr>
<th>p-value</th>
<th>W</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5646</td>
<td>0.9851</td>
<td>0.05</td>
</tr>
</tbody>
</table>

From the results of the normality test in Fig. 8, it is known that the p-value is 0.5646. With a significance level of 5% or = 0.05, it can be concluded that there is no difference in the data or the data is normally distributed.

Autocorrelation Test

*name of corresponding author
To perform the autocorrelation test on Rstudio, we need the lmtest package/library in order to run program code 5. After running the program code, we get the output as shown in Table 9.

**Table 9 Autocorrelation Test**

<table>
<thead>
<tr>
<th>p-value</th>
<th>DW</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6804</td>
<td>2.1247</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Then from the test results obtained, it is known that the p-value (0.6804) > (0.05), so it can be concluded that there is no correlation between the independent and dependent variables or the residual has no autocorrelation.

**Heteroscedasticity Test**

To perform the heteroscedasticity test on Rstudio, we need the lmtest package/library in order to run Program Code 6. After running the program code, the output is as shown in Table 10.

**Table 10 Heteroscedasticity Test**

<table>
<thead>
<tr>
<th>p-value</th>
<th>BP</th>
<th>df</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9856</td>
<td>0.028939</td>
<td>2</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Then from the test results obtained, it is known that the p-value (0.9856) > (0.05), so it can be concluded that the residual variance is different from the measurement data or there is no heteroscedasticity.

**Multicollinearity Test**

To perform a multicollinearity test on RStudio, a package/library car is needed in order to run Program Code 7. After the program code is executed, the output is obtained as shown in Table 11.

**Table 11 Multicollinearity Test**

<table>
<thead>
<tr>
<th>Variabel</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasilitas (X1)</td>
<td>1.115014</td>
</tr>
<tr>
<td>Jobdesk (X4)</td>
<td>1.115014</td>
</tr>
</tbody>
</table>

Then from the test results obtained, the VIF value in the Facility (X1) and Jobdesk (X4) variables is 1.115014, it can be concluded that there is no multicollinearity because the VIF value ≤ 10.

From the overall results of the classical assumption test, it can be concluded that each of the requirements of each test is met.

**DISCUSSIONS**

The four variables were eliminated using the backward elimination method. it is found that the Facilities (X1) variable has a p-value of 0.0456, the Work Environment (X2) has a p-value of 0.1427, Salary (X3) has a p-value of 0.55517 and the jobdesk (X4) has a p-value of 0.0032. with reference to p-value < 0.05, facilities and salary variables become the best model. By using multiple linear regression, it was found that the facility had a positive effect of 20.659% on work motivation and jobdesk had a positive effect of 27.901% on work motivation.

**CONCLUSION**

The results showed that the facility variable had a positive effect on the work motivation variable. So that when the facilities increase, the work motivation variable will also increase by 20.659%. And the Jobdesk variable has a positive effect on the work motivation variable. So that when Jobdesk increases, the Work Motivation variable will also increase by 27.901%.

In this study, the use of larger data and types of companies operating in different fields will give different results. Further research can be done by using other variables to find out other variables that can affect employee motivation.

*name of corresponding author*
REFERENCES


