*The development of science and technology in the era of the Industrial Revolution 4.0 is growing rapidly. The Industrial Revolution 4.0 itself began to occur through the form of intelligence engineering and the internet of things as movement and connectivity between humans and machines [1]. Thus, a combination of digital and internet technology was obtained with conventional industry, which aims to significantly increase productivity, efficiency and customer service [2].*

*Machine learning can be defined as computer applications and mathematical algorithms adopted by learning models that are sourced from data and can issue predictions for the future [9].*

*If there are more than two variables, then the type of linear relationship can be proven in the multivariate linear regression equation quoted in equation 1 and equation 2 as follows:*

*In this study, a descriptive qualitative method was used using a multivariate linear regression model with the Python programming language. Descriptive research is a research method used to gain the widest possible knowledge of the object of research. This descriptive study provides a detailed picture of a particular condition. Descriptive research aims to describe a situation as it is and interpret objects according to events, or anything related to variables that can be explained with numbers or words.*

*This stage can be referred to as Data Preprocessing. The method used in Data Preprocessing in this model is Data Cleaning.*

*At the stage of the Data Mining & Knowledge process, the process carried out is the selection of the type of method that is in accordance with the character of the data or known as Modeling.*

*The model feasibility test, also known as the F test, is a test carried out to determine whether the independent variables used in viewing the equation have a significant influence on the value of the dependent variable [36]. F-test or ANOVA (Analysis of Variance) on multivariate linear regression can be used to determine whether the model made can work better than a simpler model.*

*The regression coefficient test or known as the t-test has the aim of testing the significant level of the independent variable constants in the regression equation individually [36]. The existing hypothesis is: H0 = The independent variable has no significant effect. H1 = The independent variable has a significant effect.*

*The conclusion is the independent variable (independent) JobLevel with the dependent variable (dependent) that is MonthlyIncome.*

*R-square can be interpreted as a coefficient of determination. R-square shows the proportion of influence between the independent variable (independent) on the dependent variable (dependent)[37]. Based on Figure 10, the coefficient of determination (R-square) is 0.909 (90.9%). So, MonthlyIncome is influenced by Age and YearsAtCompany factors of 0.909 (90.9%). The residual value of the coefficient of determination of 0.091 (9.1%) is influenced by other unknown factors.*

*In Figure 14, the scatter plot is spread around the diagonal line, so it can be assumed that there is a linear relationship between the independent (independent) and dependent (dependent) variables.*

*Multicollinearity test was used to show the correlation between predictor variables. If there is multicollinearity in the regression model, it causes the resulting regression parameter to have a very large error value. The criteria used to determine the existence of multicollinearity between predictor variables is to use the value of Variance Inflation Factors (VIF). If the VIF value is greater than 10, then a multicollinearity problem is found. The VIF value can be obtained by regressing the independent variables [40].*

*From the results of the assumptions above, it can be said that the predictors used in the regression are correlated with each other.*

*Autocorrelation is a test that aims to test the multivariate linear regression model whether there is a correlation between the confounding error in period t and the error in period t1 (previous). If there is a correlation, it is called autocorrelation. A good regression model is a model that is free from autocorrelation [38]. In this step, the Durbin-Watson score is calculated using the durbin\_watson() function from the created statsmodel, then scores with the following conditions:*

*If the Durbin-Watson score is less than 1.5 then there is a positive autocorrelation and the assumption is not met.*

*If the Durbin-Watson score is between 1.5 – 2.5 then there is no autocorrelation and the assumption is satisfied.*

*If the Durbin-Watson score is more than 2.5 then there is a negative autocorrelation and the assumption is not met.*

*Homoscedasticity is a test that aims to test whether the multivariate linear regression model has heteroscedasticity symptoms or not by looking at the presence or absence of certain patterns on the scatterplot graph. The regression model can be said to be a good model if the model has homoscedasticity or does not occur heteroscedasticity [41].*

*From the scatterplot graph (figure 22), it can be seen that the residual points spread randomly, and are spread both above and below the number 0 (zero) on the Y axis. Thus, it can be concluded that there is no symptom of heteroscedasticity in the regression model used.*