CHAPTER 3
RESEARCH METHODOLOGY

Linear Regression, generally used in Data Mining to predict a value. In Linear Regression, data is modeled in the form of a graph in the form of two-dimensional lines, so it takes the variables X and Y.

In Linear Regression, the variable Y is referred *to as the response variable* while X is referred to as *the predictor variable*. The two variables are formulated statically with the formula $.y = \alpha + \beta x$

The value of y in the above formulation is considered to be the constant, while the value of α and β is the regression coefficient that affects the delineation of data in a graph of two-dimensional.

The value α and β can be searched using the *least square* method which serves to minimize the error value between the actual data and the data of the predicate result. Given the sample value of the data S with dots (x1, y1), (x2, y2), ... (x3, y3), then the *regression coefficient* can be searched using the following formula :

Constant(α) :

$$\alpha = \frac{(\sum y)\,(\sum_x 2) - (\sum x)(\sum xy)}{n(\sum_x 2) - (\sum_x)2}$$

Coefficient (β)

$$\beta = \frac{n(\sum xy) - (\sum x)((\sum y)}{n(\sum_x 2) - (\sum_x)2}$$

where x is the average of x1, x2, ... xs and y are the averages of y1, y2, ... Ys.

The following example is an example data of mid-semester and final exam scores for the "Algorithms and Computer Programming" course for 12 Informatics Engineering students.

Table 1 Test Scores for Algorithms and Programming Courses

Mid Semester Exam	End of Semester Exams
71	82
50	65
80	78
72	79
93	89
85	74
58	50
82	78
66	76
35	51
89	77
84	92

The above score data can be depicted in the form of a two-dimensional linear graph where the mid and final exam score data are in the form of linear lines. Point x is the data for mid-term exam scores while point y is the data for the end-of-semester exams.

Based on the formula of linear regression, the calculation process for the data prediction formula can be carried out. The first step is to calculate the average values for x and y. Based on the value presented, the value of x is 72.08 and the value of y is 74.25.

Both values can be substituted into a formula to obtain α and β value and as follows:

a=74.25-(0.613)(72k.08)=30.05

Based on the calculation results and above, maa the linear regression formula for the prediction value can be determined, which is y=30.05+0.613x

The model can be used for variety of grades based on mid-semester grades. For example, the prediction of last score for mid test scores = 86, maa is obtained the last exam of prediction result y=30.05+0.613 (86) = 82,768 or prediksi for mid test scores = 47 so it is obtained the last exam of prediction result Y=30.05+0.613 (47) = 58.861.