CHAPTER 5
IMPLEMENTATION AND RESULTS

1.7. Implementation

This research already has a dataset consisting of 1000 data. The first stage is to create a base data and a table to import data. With *the source code* as follows:

```
1. DROP DATABASE IF EXISTS dbProgramScription;
2. CREATE DATABASE dbProgramScription;
3. USE dbProgramScription;
4.
5. create table tblTest
6. (
7. X int not null,
8. Y float not null
9. );
```

From the sql above the first line is namely the command to check and delete if there is a database named dbProgramSscription, the second line command to create a database with the name dbProgramScription, and the line the third uses a database named dbProgramScription. The fifth five contains commands to create tables, rows of commands to create x columns with int data types, namely round numbers and *values* that cannot be empty, and bars the eighth is contains the command to create a column y with a float data type i.e., decimal value or real number and *value* cannot be empty (*not null*).

After the database and tables have been created, the next step is to import the data from the csv file that is imported into the test table in the dbProgramSscript database. With *the source code* as follows:

```
10.Load local data infile 'Documents/ProjectPraProject/thesis/test.csv'
11.Into table tblTest
12.Fields terminated by ','
13.Lines terminated by '\n'
14.Ignore 1 rows;
```

The source code above describes the dataset taken in our local computer from 'Documents/ProjectPraProject/thesis/test.csv and inserted or imported into the tblTest that we have created.

1.8. Results

In this prediction calculation study, there are two build formulas, namely calculations to find the value of the constant a and the coefficient b. With *the source code* as follows:

```
15.Select((sum(y)*sum(x*x))-(sum(x)*sum(x*y)))
16./ ((count(x)*sum(x*x))-(sum(x)*sum(x))) as 'constant a' from tblTest;
17.
18.Select((count(x)*sum(x*y))-(sum(x)*sum(y)))
19./ ((count(x)*sum(x*x))-(sum(x)*sum(x))) as 'coefficient b' from tblTest;
```

Based on the formula of a simple linear regression, the calculation process for the data prediction formula can be carried out. The first step is to calculate the value α and β as follows:

$$\alpha = \frac{(\sum y)(\sum_x 2) - (\sum x)(\sum xy)}{n(\sum_x 2) - (\sum_x)2} = -0.46181061339172225$$

and

$$\beta = \frac{n(\sum xy) - (\sum x)((\sum y)}{n(\sum_x 2) - (\sum_x)2} = 1.014335349349752$$

Based on the calculation results α and β above, the maa linear regression formula for the prediction value can be determined, namely y = -0.46181061339172225 + 1.014335349349752x.

From the build formula above we get a ready-made combined formula to predict the value of Y by entering the value of X that we want using the formula $y = \alpha + \beta x$. With *the source code* as follows:

```
20.Select((sum(y)*sum(x*x))-(sum(x)*sum(x*y)))
21./ ((count(x)*sum(x*x))-(sum(x)*sum(x))) +
22.(((count(x)*sum(x*y))-(sum(x)*sum(y)))
23./ ((count(x)*sum(x*x))-(sum(x)*sum(x))) * 50 as 'Value Y' from tblTest;
```

From the combined formula above we get accurate prediction results and can be used to predict all data that is linear.

The model can be used for variety of values based on the value of X. For example, the prediction value X = 37, so it is obtained result prediction value Y = -0.46181061339172225 + 1.014335349349752(37) = 37.0685973125491 or the prediksi fork value X value = 40 so it obtained the last result of exam value Y = -0.46181061339172225 + 1.014335349349752(50) =. 40.11160337.