CHAPTER 3
RESEARCH METHODOLOGY

3.1 Journal

The number of journals used in this study were 10 journals. The purpose of these journals is to look for the weaknesses of these journals, so that in this study, more detailed research can be carried out than previous journals. The topic of this research is the application of the C4.5 algorithm to classify in predicting the availability of products sold based on Branch, City, Customer Type, Gender, Product Line, Unit Price, Quantity, Tax, Total, Month, Payment, Cogs, Gross Margin Percentage. , Gross Income of the product. The variables used in this study are Branch, City, Customer Type, Gender, Product Line, Unit Price, Quantity, Tax, Total, Month, Payment, Cogs, Gross Margin Percentage, Gross Income. Then there are two predictive attributes, namely High Rating and Low Rating.

3.2 Program Design

The first steps in conducting the research are determining the decision class or predictive attribute, namely High Rating and Low Rating, each attribute is calculated and then added up. Then calculate the gain value from the predicted attribute. After that, calculate the entropy value for each variable Branch, City, Customer Type, Gender, Product Line, Unit Price, Quantity, Tax, Total, Month, Payment, Cogs, Gross Margin Percentage, Gross Income. Then draw the Decision Tree from the highest entropy value, then continue the iteration calculation which then uses the same method as before to get a conclusion from the calculation.

3.3 Coding

In calculating the C4.5 algorithm, the MySQL database is used as a helper program for data processing. Using Mysql can make it easier to calculate the amount of data that is more than doing calculations manually.

3.4 Data

The data will be used to test 1,000 data. The purpose of conducting trials of data is to see the performance of the database, program, time and performance of the C4.5 algorithm in performing calculations.

3.5 Conclusion

After obtaining the results of the research and testing the data, it will be concluded from the calculation of the data using the C4.5 algorithm to see the performance of the C4.5 algorithm in making predictions