CHAPTER 6
CONCLUSION

From the classification process for sentiment analysis using the Support Vector Machine algorithm and the chi – square selection feature and information gain on hotel reviews to increase the accuracy of previous research [1] that I have done, it can be said that:

1. From classification calculations to sentiment analysis with the Support Vector Machine algorithm and using Chi – Square feature selection, it can increase higher accuracy than using the Support Vector Machine algorithm without using feature selection as in previous studies [1]. Where when using a method like the previous research [1] without using feature selection I get an average accuracy of 85.24% while when I use the chi - square feature selection I get an average accuracy of 86.68%. The use of chi-square feature selection resulted in an average accuracy increase of 1.44%. This proves that the use of chi-square feature selection can increase the accuracy of this sentiment analysis. So for case studies on both methods between using the chi-square selection feature and without using the selection feature, the chi-square can increase the accuracy of the previous research method [1].

2. Classification calculations for sentiment analysis with the Support Vector Machine algorithm and Information Gain feature selection can increase accuracy higher than using the Support Vector Machine algorithm without using feature selection as in previous studies [1]. Whereas when not using feature selection as in previous research [1], the average accuracy is 85.24%. When I use feature selection using the Information Gain method, it can increase the average accuracy by 85.78%. The use of feature selection with Information Gain can increase the average accuracy by 0.54%. This proves that using the Information Gain feature selection can increase the accuracy of the previous research method[1] in this sentiment analysis. So for a case study between these two methods, namely with the information gain selection feature and without using the selection feature, information gain can increase the accuracy of the previous research method [1].

3. In the classification results for sentiment analysis in hotel reviews using the Support Vector Machine algorithm, it can be seen that between using the chi-square selection feature and information gain, the accuracy value of each experiment proves that the use of the chi-square selection feature for classification in this hotel review is better. higher than the use of the

information gain selection feature. Whereby using the chi-square selection feature, the average accuracy is 86.68%, while using the information gain selection feature, the average accuracy is 85.78%. From the two results, the average accuracy value has a distance difference of 0.90%. This proves that using the chi-square selection feature gets higher results than using information gain. The best method between information gain and chi-square in the classification in this study is chi-square. So it can be concluded that the use of the chi-square selection feature is better than the use of information gain in this research classification.

Then, as a writer, I give suggestions for further research:

1. Further research can look for datasets with more attributes to see the results of the classification which may increase the accuracy of each method.
2. Added a variety of feature selection methods to compare the process and results of the calculation of accuracy in the classification.
3. Further research can use other classification algorithms such as Naive Bayes, Logistic Regression, and others to see the results of the classification which may increase the accuracy of each algorithm.