CHAPTER 6
CONCLUSION

6.1. Conclusion

Homeschooling topic reviews in this study amounted to 2411 with a positive class of 1940 and a negative class of 471. The models made in this sentiment analysis research are Long Short Term Memory and Support Vector Machine. The data from the two models are divided equally, namely 20%, 50%, and 80% of the testing data. The Long Short Term Memory model created is the result of training using the best neurons, namely 600 neurons with a sigmoid activation function, batch size 32, epoch 10, and the distribution of the amount of testing data is 20%. In these parameters, the model was tested 5 times and the results were quite good with an accuracy of 83.8%, precision 88.2%, recall 92.3%, F1-Score 90.2%. Each give good results to determine the performance of the model that has been built. This means, that LSTM is suitable for large data because LSTM is made to complement the shortcomings of RNN which cannot predict words based on previous information stored for a long time. LSTM is able to remember a collection of information that has been stored for a long time, while deleting information that is not needed. LSTM is more efficient in processing, predicting, as well as classifying data on a larger scale.

While the Support Vector Machine model created is the result of training using regularization parameters, namely the lambda parameter of 1 ($\lambda = 1$) and the distribution of the amount of testing data by 20%. In this test, the data used is unbalanced. On these parameters, 10x model testing was carried out and the results obtained were 71.0% Accuracy, 83.9% Precision, 79.6% Recall, and 81.7% F1-Score. The accuracy obtained is only 71%. It is proven that with unbalanced data, less than optimal results are obtained for the Support Vector Machine model. This is because the Support Vector Machine is not optimal for large and unbalanced datasets.

Then the Support Vector Machine test was carried out with balanced data using the undersampling method, and the highest Accuracy results were 94%, 91% Precision, 99% Recall and 95% F1-Score. It is proven that with balanced data the model will produce good results as well. This happens because the model can study the data in a balanced way and is not more

inclined to the majority of data. This proves that Support Vector Machine is good for balanced datasets.

The conclusion is, the Long Short Term Memory method gets a higher accuracy of 83.8% and Support Vector Machine with unbalanced dataset gets 71% accuracy. As for the Support Vector Machine model with balanced data using the Undersampling method the accuracy for this model is 94%.

6.2. Suggestion

In this study, several things can be developed for future research, namely by adding the number of datasets or adding sentiment classes such as normal sentiment. For the Long Short Term Memory method, can add other architectures such as learning rate, dropout layer, or regularization layer for future research. And for the Support Vector Machine method, can try many value for other architectures such as iterations or learning rates.