

CHAPTER 4

ANALYSIS AND DESIGN

4.1 Analysis

Sentiment analysis was carried out in this study using 2 methods, namely the Long Short Term Memory method and the Support Vector Machine. First, the data is taken from Twitter as much as 2411 data on the topic of homeschooling. The data is taken using the Crawling method, and the data taken are tweets or comments from people on Twitter social media along with their polarity. Then the data taken from Twitter is stored in CSV format and processed using the Google Collab platform. Then, manual data labeling is carried out which refers to the existing polarity. Researchers divide 2 classes of sentiment, namely positive sentiment, and negative sentiment. Positive sentiment is determined if the polarity is more than 0.0, while negative sentiment is determined if the polarity is less than 0.0. Then to make it easier for the algorithm to work, positive sentiment and negative sentiment are again labeled in the form of numbers 0 and 1. Number 0 for negative sentiment, and number 1 for positive sentiment. After labeling the data manually, the data Pre-Processing process is carried out. The pre-processing of data is related to the lower casing, tokenization, stopword removal, and data cleansing processes. Data pre-processing is done so that the data obtained becomes cleaner information and can reduce the algorithm's obstacles in working. Then, the data is divided into 2 parts, namely training data and test data. After the data is divided into 2 parts, a feature extraction process is carried out to facilitate learning on the model created by converting text data into numeric vectors. Then training data is trained using the existing algorithm to determine the performance of the algorithm. After the training data is run using the existing algorithm, the algorithm will provide data results from the training data that have been trained. After getting the results from the training data, testing is carried out using the test data, then the final results will be obtained from testing the data using the algorithm. The final results obtained are in the form of Accuracy, Precision, Recall, and F1-Score.

The results of the Machine Learning and Deep Learning methods will always change when they are run. Therefore, in this study, the training data and test data were run 30 times to get the best results from the algorithm being tested.

