CHAPTER 3
RESEARCH METHODOLOGY

3.1 Data Collection

For this study, the data will be obtained from company X. The data obtained is data in the form of data from the results of the company's recruitment interview test. The total data obtained is 350 per aspect. Therefore in total, there are 1050 data. Because this study will analyze 3 personality aspects: motivation, work enthusiasm, and self-awareness. All data collected is in Indonesian language.

3.2 Data Preprocessing

After data is collected, go preprocess the data. Before preprocessing the data, first check whether there are typos in the dataset that will be used. Then after that do the preprocess stage of removing punctuation or symbols that are not used, removing stopwords or words that have no meaning, and doing the process of stemming, which is changing words into standard forms.

3.3 Creating Classification Model

After that, go create the model. To make the model, this study will use the LDA algorithm but what is used is the version that uses labels or we can call it L-LDA. From each aspect 3 models will be made that can classify each value per aspect. The value per aspect will be categorized into 5: KS (Kurang Sekali), K (Kurang), C (Cukup), B (Baik), and BS (Baik Sekali). Those 5 categories will later be used as topics in the interview test data grouping. For more details, it will be explained in chapter 4 about the detailed way of making classification models which will later be used to classify interview answers so as to produce whether the person in certain aspects belongs to the category KS, K, C, B, or BS.

To create the model, from each aspect we will use 250 data from the dataset and we will use the other 100 for testing the classification model. We will create the model based on 2 approaches that is already mentioned in chapter 1.

3.4 Predicting Test Data

After finished creating the classification models, we will try the classification models to predict aspect values from each aspect from the testing data. After that we can analyze which approach is better at classifying the testing data. Then we will write the report and show the results.