

## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1. Data Collection

In collecting datasets, I use websites that provide various kinds of datasets. For this research I used data from <https://www.kaggle.com/binaryjoker/airline-passenger-satisfaction>. Data with the file name `airline_passengeer_satisfaction.csv` has a file size of 14.34MB. I downloaded this data on September 20, 2021. To download it you are required to Sign In first (Register if you don't have an account). The downloaded file will be a zip file, so it must be extracted to get the csv file. The total data obtained were 129,880 with 23 measuring columns and 1 response column.

#### 3.2. Algorithm

In choosing the algorithm, I consulted my supervisor. During the consultation, my lecturer informed and suggested the Learning Vector Quantization (LVQ) algorithm. This algorithm has not been used very often. Therefore, I use this LVQ algorithm. After using LVQ I looked for another algorithm to use as a comparison. Then I chose Naive Bayes because this algorithm is an algorithm that is often used, easy and has good accuracy. I use these two algorithms to classify supervised learning data about airline passenger satisfaction that has been obtained previously. In addition to knowing which algorithm is better in accuracy.

#### 3.3. Coding and Design

In this step, the MySQL tools will be used. MySQL is used because the existing dataset is in the form of 2-dimensional data (columns and rows) the same as the MySQL database table. In addition, the installation of MySQL is very easy. By downloading xampp through the website <https://www.apachefriends.org/download.html>. Xampp already provides several versions for Windows, Linux and OS X operating systems. Here I use Linux. After MySQL is installed, the data will be preprocessed. Continuous data such as age and distance will be changed first to make it easier to classify.

### 3.4. Analysis

In analyzing, I will do 5 tests as follows :

**Table 3.1.** Analysis

	<b>Training Data</b>	<b>Testing Data</b>
I	90%	10%
II	75%	25%
III	50%	50%
IV	25%	75%
V	10%	90%

In this analysis, it is divided into 5 stages to determine whether the amount of training has an effect. Influence on Naive Bayes accuracy and on LVQ accuracy.

### 3.5. Make a Report

In making the report, I wrote chapters 1-4 first. After chapter 4 finished, I started the coding stage for program development. Then the results that have been carried out during the coding stage will be recorded in the chapter 5 report. And finally, conclusions will be drawn from the results of the coding stage which will be written in chapter 6.