

CHAPTER 3

RESEARCH METHODS

3.1 Stages of Data Processing

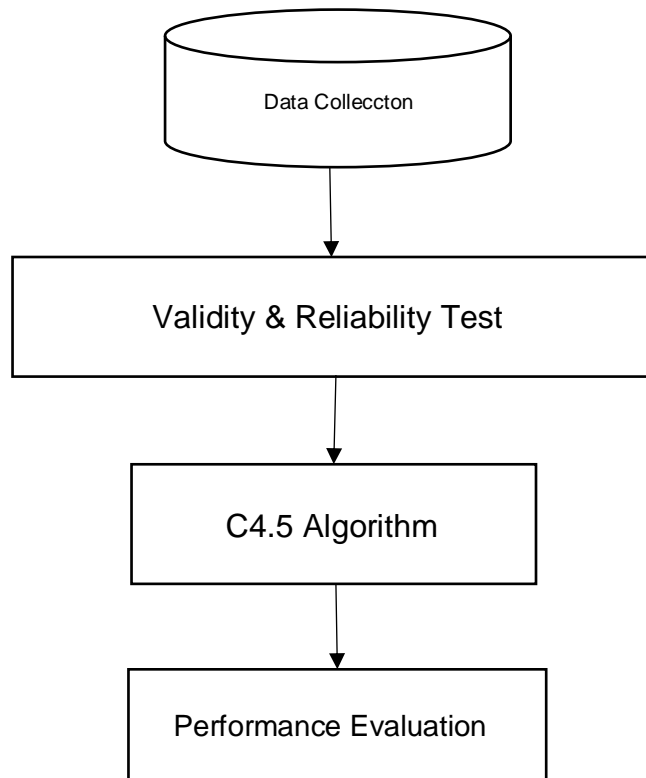


Figure 3.1 Research Stage

Based on figure 3.1, this research begins by collecting data by distributing questionnaires to the respondent. The collected data will be tested for feasibility as a research instrument with validity and reliability test. For generate a decision tree using the C4.5 Algorithm to calculate entropy and gain value. Then perform an accuracy test to determine the accuracy of the C4.5 algorithm model with the confusion matrix.

3.2 Data collection

Data was obtained through a questionnaire given to respondents, students from the Department of the public health 2019 generation. To measure the

instrument for the attitude scale with the Likert scale, use four categories of answers in table 3.2. The list of the questionnaire in table 3.1 :

Table 3.1 Questionnaire Question

NO	QUESTION	ATTRIBUTE
1	The lecturer explains the lecture contract & RPS at the beginning of the lecture	REABILITY
2	The lecturer conveys the lecture material very clearly	
3	Lecturer Teaching according to the field of expertise	
4	Textbooks, workbooks and/or handouts are well provided	
5	Lecturers start and end lectures according to the learning schedule	
6	Lecturers explain and facilitate learning activities well	
7	Assessment of tasks objectively	
8	Lecturers inspire and motivate students	
9	Information about the Department of academic clearly	
10	There is insurance for students	
11	Administration services in Department are easy and fast	
12	Academic services according to sop	
13	University Accreditation	ASSURANCE
14	Department of Accreditation	
15	Politeness of Educational Staff	
16	Campus security and vehicle parking	
17	The campus atmosphere is conducive to study	
18	Freedom to carry out scientific activities on campus	
19	Ease of getting information on campus	
20	Department understands the needs and interests of students	EMPATHY
21	Departments are open and cooperative with students	
22	Lecturers serve students outside of working hours	
23	Lecturers are easy to contact in academic consultations	
24	Friendly lecturers in serving	
25	Classrooms are kept clean and tidy	TANGIBLE
26	Complete facilities in the learning process are available in the classroom (computer, LCD projector, whiteboard)	
27	Laboratory/workshop facilities meet the needs of students in the learning process	
28	The library provides adequate reference books	
29	There are internet facilities that are easily accessible in the campus environment	
30	Availability of adequate and clean toilet facilities	

Table 3.2 Category Scale and Score

Category	Answer Score
Very satisfied	4
Satisfied	3
Not satisfied	2
Very Dissatisfied	1

Where 4 is a very satisfied, 3 is a satisfied, 2 is a not satisfied, and 1 is a very dissatisfied.

3.3 Validity and Reliability test

A validity test is carried out to measure the variables of parameters are valid or not, and invalid variables will be eliminated. Validity test using equation 3.1 :

$$r_{xy} = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{[(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)]}} \quad (3.1)$$

Where r_{xy} is a correlation coefficient, n is the number of samples, $\sum x$ is a total score of the item, $\sum x^2$ is a number of item scores squared, $\sum y$ is a total score, $\sum y^2$ is a total score squared, and $\sum xy$ is the number of multiplication of item scores and total scores.

A reliability test is a tool to measure whether a questionnaire/parameters are reliable or not. If the parameters are not reliable so they can't be used. The parameters are said to be reliable if it gives a Croancbach Alpha value > 0.60 , can be seen in table 3.3. And the coefficient categories that can be used are quite reliable, reliable, and very reliable. Reliability test using equation 3.2 :

$$r_{11} = \left[\frac{K}{K-1} \right] \times \left[1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right] \quad (3.2)$$

Where r_{11} is the reliability of the instrument, K is a number of items in the instrument, $\sum \sigma_i^2$ is a number of item variances, and $\sum \sigma_t^2$ is a total variance.

Table 3.3 Cronbach Alpha Coefficient Category

Cronbach alpha value range	Coefficient category
0,00 – 0,20	Less reliable
0,21 – 0,40	Somewhat reliable
0,41 – 0,60	Quite reliable
0,61 – 0,80	Reliable
0,81 – 1,00	very reliable

3.4 C4.5 Algorithm

To generate the model of the decision tree using the C4.5 Algorithm need to calculate entropy and gain value for each variable/attribute. Entropy is the value of information that states the measure of uncertainty of the attributes from data objects in bits. After calculating the entropy value, attribute selection is carried out using Information Gain. Information Gain is a measure of the effectiveness of an attribute in classifying the data used to determine the order of the attributes in which the has the largest Information Gain value selected.

The formula for calculating entropy using equation 3.3 :

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i \quad (3.3)$$

Where S is the case set, n is the number of partitions of S, and p_i is the probability obtained from the number of classes divided by the total cases.

Formula calculating gain using equation 3.4 :

$$Gain(S,A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3.4)$$

Where S is a case set, A is attribute n is the number of attributes, |S_i| is a number of partitions to –I, and |S| is a number of cases in S.

3.5 Performance Evaluation

Tests were carried out to determine the performance of the C4.5 algorithm model in classifying a predetermined class. In this test, data is divided into 70% training data and 30% data testing. The value generated through the confusion matrix method is in the form of evaluation accuracy, precision, and recall. And the formula for calculate accuracy using equation 3.5, for calculating precision using equation 3.6, and for calculating recall using equation 3.7 :

$$\text{Accurate} = \frac{TP+TN}{TP+FP+FN} 100\% \quad (3.5)$$

Where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative.

$$\text{Precision} = \frac{TP}{TP+FP} 100\% \quad (3.6)$$

Where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative.

$$\text{Recall} = \frac{TP}{TP+FN} 100\% \quad (3.7)$$

Where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative.