

## CHAPTER 4

### ANALYSIS AND DESIGN

#### 4.1 Analysis

##### 4.1.1 Research Data

The data used with the CSV extension obtained from Kaggle <https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction> about passenger satisfaction aircraft based on airline facilities. But only 14 attributes that can be used are Inflight wifi service, departure / arrival time convenient, ease of online booking, gate location, food and drink, online boarding, seat comfort, inflight entertainment, on-board service, leg room service, baggage handling, check-in service, inflight service, and cleanliness.

##### 4.1.2 C.45 Algorithm

The following are the stages of using the C4.5 algorithm in data mining

1. Knowing which dataset we will use in test sample data from Airline Passenger Satisfaction csv. The data item start from number 8 until 22 of data number ( 15 data ) in test dataset.

No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	S
1.	5	2	2	2	5	5	5	5	2	2	5	3	3	5	S
2.	2	2	2	2	3	4	4	4	4	4	4	5	4	4	S
3.	4	1	1	1	5	1	5	3	3	4	3	1	3	4	S
4.	2	5	5	5	1	3	4	2	2	2	2	3	2	4	NOD
5.	5	5	5	5	4	5	5	5	5	5	5	5	5	3	S
6.	1	1	4	1	5	5	4	5	5	5	5	3	5	5	S
7.	2	2	2	2	5	5	4	4	4	4	4	3	4	5	S
8.	3	4	0	3	2	0	2	2	4	2	4	4	5	2	NOD
9.	2	5	5	5	2	2	2	2	4	3	3	4	3	2	NOD
10.	5	4	5	3	4	5	4	4	3	5	5	4	5	4	S

11.	3	4	3	1	4	4	5	5	5	3	5	4	5	3	NOD
12.	5	5	5	5	4	5	4	4	4	5	4	5	4	5	S
13.	4	4	4	4	2	4	5	5	5	5	5	5	5	5	S
14.	3	4	3	4	2	3	2	2	4	2	4	4	5	2	NOD
15.	1	2	4	3	4	4	4	4	1	3	2	4	3	4	NOD

*Table 1 : Analysis Data*

Table Indicators in using of C4.5 Implementation :

A1 = Inflight Wifi Service

A2 = Departure / Arrival Time Convenient

A3 = Ease of Online Booking

A4 = Gate Location

A5 = Food and Drink

A6 = Online Boarding

A7 = Seat Comfort

A8 = Inflight Entertainment

A9 = On-board service

A10 = Leg Room Service

A11 = Baggage Handling

A12 = Checkin Service

A13 = Inflight Service

A14 = Cleanliness

S = Satisfaction

In this dataset, there are 17 attributes, but the attributes shown in the image are only attributes that can be used in implementing the C4.5 algorithm.

1. Nomor (int)

2. Id Penumpang (int)
3. Inflight Wifi Service (int)
4. Departure / Arrival time convenient (int)
5. Ease of Online Booking (int)
6. Gate Location (int)
7. Food and Drink (int)
8. Online Boarding (int)
9. Seat Comfort (int)
10. Inflight Entertainment (int)
11. On-board Service (int)
12. Leg Room Service (int)
13. Baggage handling (int)
14. Checkin Service (int)
15. Inflight Service (int)
16. Cleanliness (int)
17. Satisfaction (varchar)

The last attribute called satisfaction is the attribute of the decision result with a value of Satisfied value which means that the passenger is satisfied and Neutral or Dissatisfied which means that the passenger is not satisfied.

Attributes named nomor / number, id will not be mined. Because the attribute value cannot be used to determine the value of passenger satisfaction. However, these attributes can still be displayed along with the attributes that can be used for screening because attributes from number to flight distance can be used as the identity of passengers using the airline.

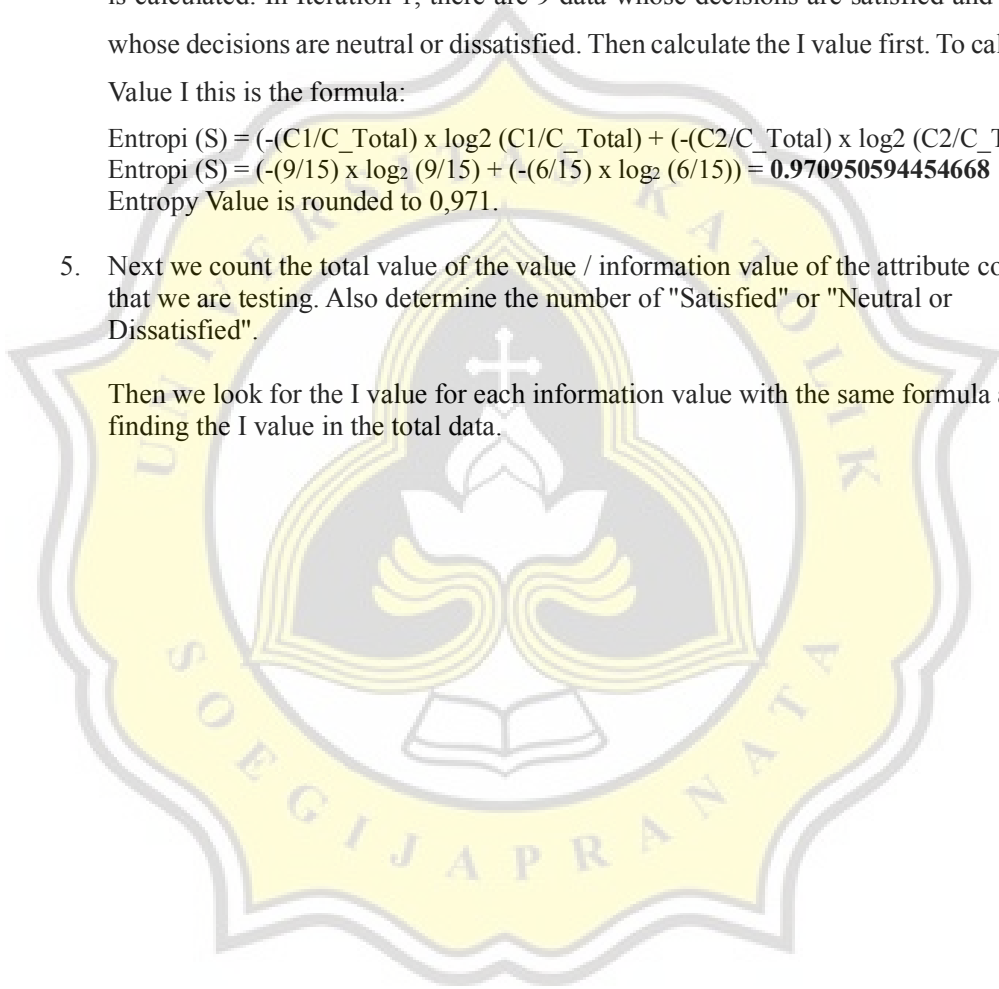
2. Determine the attributes that can be used based on the dataset column, namely Inflight wifi service, departure / arrival time convenient, ease of online booking, gate location, food and drink, online boarding, seat comfort, inflight entertainment, on-board service, leg room service, baggage handling, check-in service, inflight service, and cleanliness.

3. Determine whether the result case is "Satisfied" or "Neutral or Dissatisfied" based on the value of the attribute column used, all of which contain assessments 0, 1, 2, 3, 4, and 5. 0 "berarti tidak memberi nilai", 1 "sangat tidak puas", 2 "tidak puas", 3 "cukup", 4 "puas", 5 "sangat memuaskan".
4. Calculates the decision class, but here we will try the C4.5 algorithm in the libreoffice spreadsheet to solve this dataset. Here I use 15 test data first so that not too much data is calculated. In Iteration 1, there are 9 data whose decisions are satisfied and 6 data whose decisions are neutral or dissatisfied. Then calculate the I value first. To calculate Value I this is the formula:  

$$\text{Entropi (S)} = -(C1/C\_Total) \times \log_2 (C1/C\_Total) + -(C2/C\_Total) \times \log_2 (C2/C\_Total)$$

$$\text{Entropi (S)} = -(9/15) \times \log_2 (9/15) + -(6/15) \times \log_2 (6/15) = \mathbf{0.970950594454668}$$
 Entropy Value is rounded to 0,971.
5. Next we count the total value of the value / information value of the attribute column that we are testing. Also determine the number of "Satisfied" or "Neutral or Dissatisfied".

Then we look for the I value for each information value with the same formula as finding the I value in the total data.



Then count it into this table. Iteration 1

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Inflight Wifi Service	0	0	0	0	0	0	0.4	0.5709
	1	2	1	1	1	0.1333		
	2	4	2	2	1	0.2667		
	3	3	3	0	0	0		
	4	2	0	2	0	0		
	5	4	0	4	0	0		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Departure / Arrival Time Convenient	0	0	0	0	0	0	0.8066	0.1642
	1	2	0	2	0	0		
	2	4	1	3	0.8112	0.2163		
	3	0	0	0	0	0		
	4	5	3	2	0.9709	0.3236		
	5	4	2	2	1	0.2666		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Ease of Online Booking	0	1	1	0	0	0	0.5073	0.4636
	1	1	0	1	0	0		
	2	3	0	3	0	0		
	3	2	2	0	0	0		
	4	3	1	2	0.9182	0.1836		
	5	5	2	3	0.9709	0.3236		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Gate Location	0	0	0	0	0	0	0.7673	0.2036
	1	3	1	2	0.918	0.1836		
	2	3	0	3	0	0		
	3	3	2	1	0.918	0.1836		
	4	2	1	1	1	0.1333		
	5	4	2	2	1	0.2667		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Food and Drink	0	0	0	0	0	0	0.5399	0.4309
	1	1	1	0	0	0		
	2	4	3	1	0.8112	0.2163		
	3	1	0	1	0	0		
	4	5	2	3	0.9709	0.3236		
	5	4	0	4	0	0		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Online Boarding	0	1	1	0	0	0	0.2667	0.7042
	1	1	0	1	0	0		
	2	1	1	0	0	0		
	3	2	2	0	0	0		
	4	4	2	2	1	0.2667		
	5	6	0	6	0	0		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Seat Comfort	0	0	0	0	0	0	0.6434	0.3275
	1	0	0	0	0	0		
	2	3	3	0	0	0		
	3	0	0	0	0	0		
	4	7	2	5	0.8631	0.4027		
	5	5	1	4	0.7219	0.2406		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Inflight Entertainment	0	0	0	0	0	0	0.4812	0.4896
	1	0	0	0	0	0		
	2	4	4	0	0	0		
	3	1	0	1	0	0		
	4	5	1	4	0.7219	0.2406		
	5	5	1	4	0.7219	0.2406		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
On-board Service	0	0	0	0	0	0	0.7496	0.2212
	1	1	1	0	0	0		
	2	2	1	1	1	0.1333		
	3	2	0	2	0	0		
	4	6	3	3	1	0.4		
	5	4	1	3	0.8112	0.2163		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Leg Room Service	0	0	0	0	0	0	0.2163	0.7546
	1	0	0	0	0	0		
	2	4	3	1	0.8112	0.2163		
	3	3	3	0	0	0		
	4	3	0	3	0	0		
	5	5	0	5	0	0		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Baggage Handling	0	0	0	0	0	0	0.7169	0.2539
	1	0	0	0	0	0		
	2	2	2	0	0	0		
	3	2	1	1	1	0.1333		
	4	5	2	3	0.9709	0.3236		
	5	6	1	5	0.6500	0.2600		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Check-in Service	0	0	0	0	0	0	0.4763	0.4946
	1	1	0	1	0	0		
	2	0	0	0	0	0		
	3	4	1	3	0.8112	0.2163		
	4	6	5	1	0.6500	0.2600		
	5	4	0	4	0	0		



KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Inflight Service	0	0	0	0	0	0	0.7264	0.2445
	1	0	0	0	0	0		
	2	1	1	0	0	0		
	3	4	2	2	1	0.2667		
	4	3	0	3	0	0		
	5	7	3	4	0.9852	0.4597		

KOLOM	INFORMASI	JUMLAH	Neutral or Dissatisfied	Satisfied	NILAI I	ENTROPY		GAIN
TOTAL DATA		15	6	9	0.971			
Cleanliness	0	0	0	0	0	0	0.4569	0.5139
	1	0	0	0	0	0		
	2	3	3	0	0	0		
	3	2	1	1	1	0.1333		
	4	5	2	3	0.9709	0.3236		
	5	5	0	5	0	0		

The formula for finding the Entropy value:

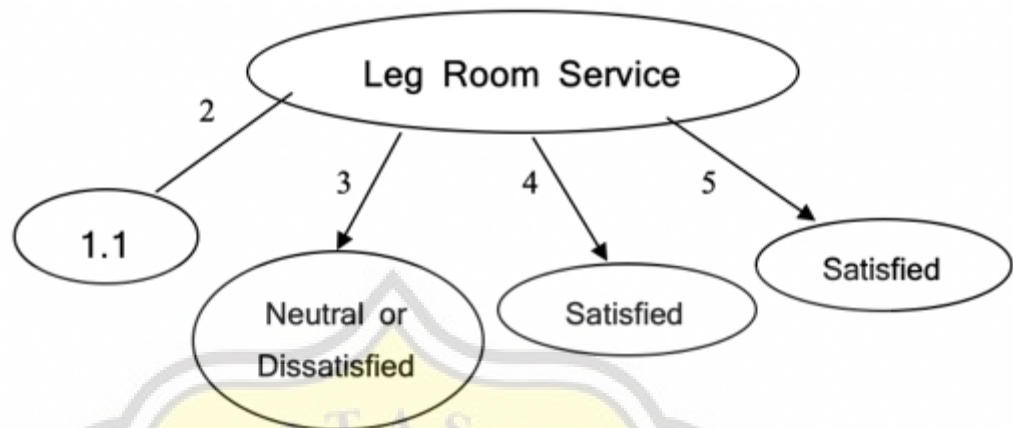
(Column name) = Amount of data per information value / Total number of data x Value I

Then we total the entropy values of each information value into one variable. To find the final entropy / Gain we use the formula:

Value I of total data - The total value of the Entropy in an attribute.

At Iteration 1, the highest Gain value is Leg Room Service with a Gain value of 0.754609761265566. So a total of 15 data 9 "Satisfied" and 6 "Neutral or dissatisfied" produces a decision tree result.

Tree Iteration 1



*Figure 1 : Iteration 1 Calculate*

\*Neutral OD = Neutral or Dissatisfied

\*S = Satisfied

With the following logic

IF Leg Room Service = 3 THEN Satisfaction = Neutral OD

IF Leg Room Service = 4 THEN Satisfaction = Satisfied

IF Leg Room Service = 5 THEN Satisfaction = Satisfied

The values of the attributes of Leg Room Service that still have not made a decision are 2. Then we look for iteration 2. We start from value 1. So it is called tree 1.1 .

6. If iteration 1 is still not complete based on the C4.5 algorithm. Then it will be continued in the next iteration. This iteration will continue repeating until all data have been calculated and produce all decisions on the attribute value being screened.

## 4.2 Desain

The following is a process in the form of a flowchart that shows how the implementation of the C4.5 algorithm program works to determine passenger satisfaction in the form of a decision tree. After going through the process of calculating the C4.5 algorithm through the Libreoffice spreadsheet. This calculation will be implemented in the form of the MYSQL programming language. Because the purpose of this calculation is to know how the C4.5 algorithm works so that it is easier to make the program.

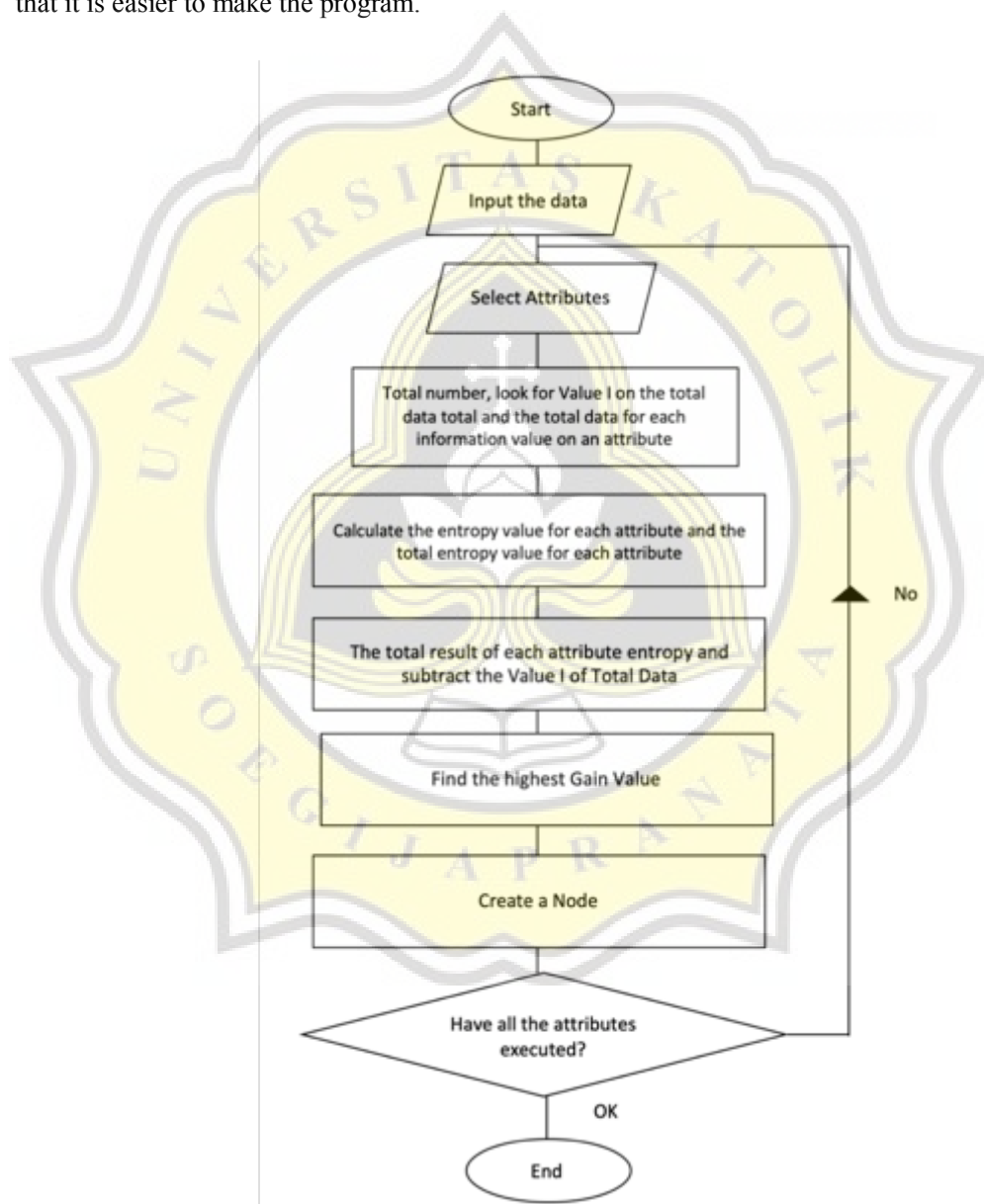


Figure 2 : Flowchart C4.5 Algorithm