CHAPTER 4
ANALYSIS AND DESIGN

4.1 Analysis

The principle of this apriori algorithm uses association methods, associations are rules for finding correlations between items in the database. This algorithm has several steps called iterations or passes. Each iteration produces a high frequency pattern.

The following is a sample of transaction data that has been obtained:

Table 4.1: Initialize Nama Barang as Kode Barang

Kode Barang	Nama Barang
B1	Kawat Loket PVC CAB 1/2
B2	Gerobak Arco
B3	Pompa Sepeda Tabung
B4	Paku Beton 70
B5	Selang Dof 5/8
B6	Rel Pintu J3 Classic

Specify minimum support: suppose that the minimum support calculated must be at least 2.

Here is a list of transactions at the hardware store:

Table 4.2: Transaction List

Konsumen	Barang
K1	B1,B2,B5
K2	B2,B4
K3	B2,B3
K4	B1,B2,B4
K5	B1,B3
K6	B2,B3
K7	B1,B3
K8	B1,B2,B3,B5
K9	B1,B2,B3
K10	B6

Iteration 1: Calculate the transaction that contains all items by scanning the database for 1 itemset, after finding it, from the itemset does it meet the minimum support count, if it meets the minimum support value then the itemset will become a pattern.

Table 4.3: Iteration 1

Itemset	Support Count
B1	6
B2	7
B3	6
B4	2
B5	2
B6	1

It is known that the minimum support value is 2, then scan the database if there is a support count that does not meet the data will be deleted.

Table 4.4: Iteration 1: After Eliminated

Itemset	Support Count
B1	6
B2	7
B3	6
B4	2
B5	2

By eliminating B6, all items meet the minimum support.

Iteration 2: Scan the database to find a combination of 2 items and calculate the Support Count

Table 4.5: Iteration 2

Itemset	Support Count
B1, B2	4
B1, B3	4
B1, B4	1
B1, B5	2
B2, B3	4
B2, B4	2
B2, B5	2
B3, B4	0
B3, B5	1
B4, B5	0

Then from the data that has been obtained do a scan again by removing items that do not meet the value of support. So that it becomes like the table below.

Table 4.6: Iteration 2: After Eliminated

Itemset	Support Count
B1, B2	4
B1, B3	4
B1, B5	2
B2, B3	4
B2, B4	2
B2, B5	2

Perform the process for the next iteration until there are no more itemset that meet the minimum support.

Iteration 3: Then, from the transaction data obtained still have 3 item combinations, then the iteration will continue until the last itemset combination. As in the previous step, look for 3 combinations of items in the transaction database then calculate the Support Count.

Table 4.7: Iteration 3

Itemset	Support Count
B1, B2, B3	2
B1, B2, B5	2

Confidence is the strong relationship between items in associative rules. Here is the formula for finding Confidence:

$$Confidence = P(B|A) = \frac{Amount\ of\ Transaction\ contains\ A \wedge B}{Amount\ of\ Transaction\ contains\ A} \; X \; 100\%$$

Table 4.8: Confidence

Itemset	Confidence
If buy B1 then buy B2	4 / 6 = 66%
If buy B1 then buy B3	4 / 6 = 66%
If buy B1 then buy B5	2 / 6 = 33%
If buy B2 then buy B3	4 / 7 = 57%
If buy B2 then buy B4	2 / 7 = 28%
If buy B2 then buy B5	2 / 7 = 28%

Then the value that has been obtained will be calculated the value of Confidence, if the Confidence value exceeds 50% then the itemset includes a high

frequency pattern or combination of items that often appear in the database. Then the percentage level that approaches the largest number can be used as a recommendation item.

So the conclusion of the results in the table above is when buying B1(Kawat Loket PVC CAB ½) then most likely also buy B2(Gerobak Arco) and other possibilities when buying Kawat Loket PVC CAB ½ then also buy B3(Pompa Sepeda Tabung) because the Confidence value exceeds 50% and the percentage is highest. While the frequency of the most items can be used as Best Seller items.

4.2 Design

The amount of data to be processed will continue to be added from tens to hundreds so that it can be known how accurate and how fast the program is running.

The data obtained will be stored in the CSV file. The following are table data from consumer transactions:

	A	B
1	Konsumen	Nama Barang
2	1	Kawat Loket PVC CAB 1/2,Kawat Loket PVC 1/4,Reskam Reyner Putih
3	2	Gerobak Arco,Engsel Plat Amani 4,Amplas Velcro
4	3	Gerobak Arco,Lem Niponplas,Rantai Kapal 3/16
5	4	Gerobak Arco,Tang Potong Keramik,Gembok Jazz 30 pj
6	5	Kawat Loket PVC CAB 1/2,Paku Kalsibot
7	6	Rell J4 Muller,Selang Dof 5/8,Pelampung Radar 99
8	7	KP Top Kuning,Lem Niponplas,Engsel Bras HPP 3
9	8	Lem Dextone 5,Kawat Loket PVC 1/4,Rantai Kapal 3/16
10	9	Gerobak Arco,Tang Potong Keramik,Paku Kalsibot
11	10	Lem Dextone 5,Engsel Nisio 4,Pompa Sepeda Tabung
12	11	Rell J4 Muller,Selang Dof 5/8,Pelampung Radar 99
13	12	Kawat Loket PVC CAB 1/2,Palu Batu Gagang Full 3 LB,Reskam Reyner Putih
14	13	Lem Dextone 5,Palu Batu Gagang Full 3 LB,Selang Benang 1"
15	14	Graji Gorok Mata 18/20,Paku Beton 70,Gembok Jazz 30 pj
16	15	Paku Beton 70,Tang Potong Keramik,Amplas Velcro
17	16	Kawat Loket PVC CAB 1/2,Paku Beton 70,Selang Benang 1"
18	17	KP Top Kuning,Engsel Plat Amani 4,Pompa Sepeda Tabung
19	18	Gerobak Arco,Lem Niponplas,Rantai Kapal 3/16
20	19	Paku Beton 70,Engsel Nisio 4,Engsel Bras HPP 3
21	20	Kawat Loket PVC CAB 1/2,Kawat Loket PVC 1/4,Pompa Sepeda Tabung

Illustration 4.1: Transaction List CSV

The purpose of this amount of data is, because of the many transactions that occur every day, the large amount of transaction data can be utilized so that the results of this apriori algorithm can be more accurate.

Above is the flowchart of the program:

Start the Program after that input CSV files into Program until transaction lists is loaded. Set the minimum support to calculate each transactions. If transaction data count is fulfill Support count then transaction list get into the transaction table. All items will be search for each pattern. After Iteration 1 is created then next Iteration will be created. Result – YES: Confidence Calculated – NO: Input higher Support.
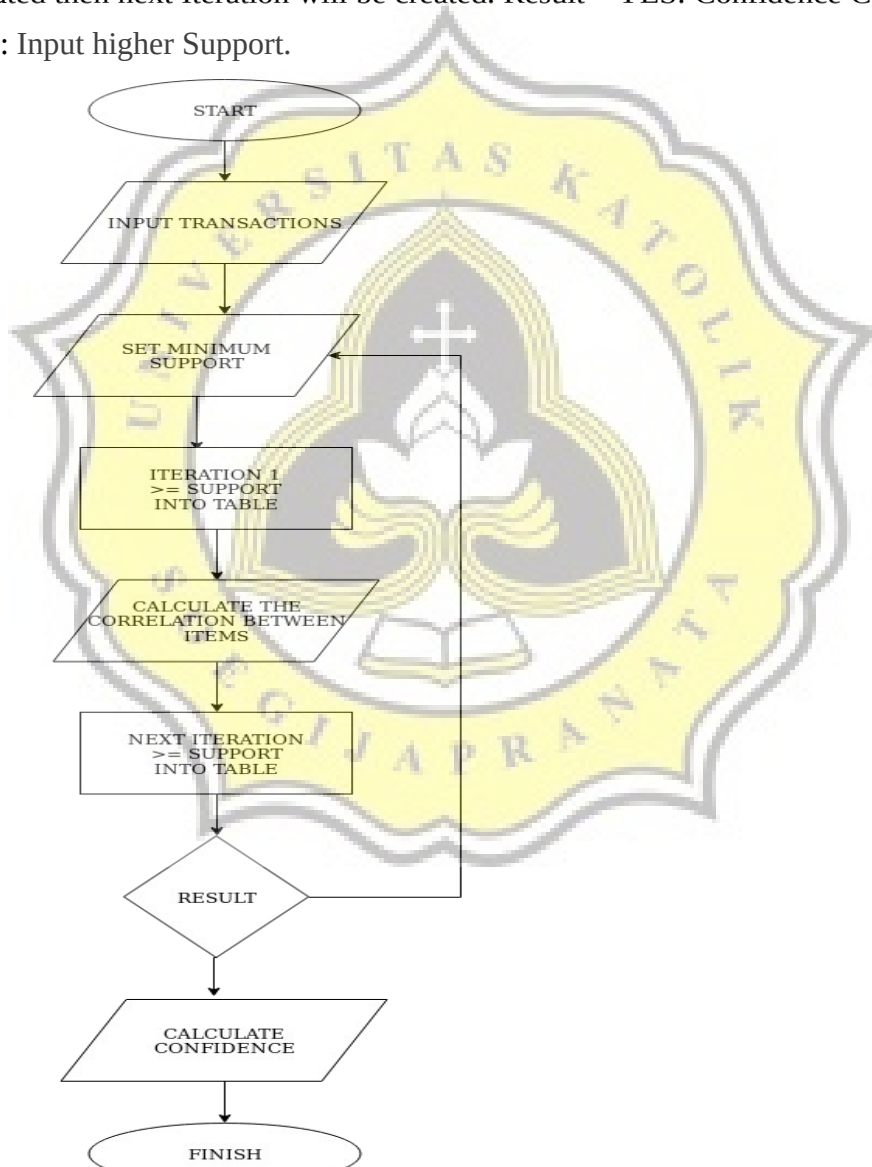

Illustration 4.2: Flowchart