International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

Association Rules Analysis on FP-Growth Method in Predicting Sales

Supiyandi[1], Mochammad Iswan Perangin-angin[2], Andre Hasudungan Lubis[3], Ali Ikhwan[4], Mesran[5], Andysah Putera Utama Siahaan[6]

[1,2,6]Faculty of Computer Science, Universitas Pembangunan Panca Budi, Medan, Indonesia
[3] Faculty of Engineering, Universitas Medan Area, Medan, Indonesia
[4]Faculty of Science and Technology, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
[5]Department of Informatics, STMIK Budi Darma, Medan, Indonesia
[4,6]School of Computer and Communication Engineering, Universiti Malaysia Perlis, Pauh, Malaysia

Abstract - Sales transaction data on a company will continue to increase day by day. Large amounts of data can be problematic for a company if it is not managed properly. Data mining is a field of science that unifies techniques from machine learning, pattern processing, statistics, databases, and visualization to handle the problem of retrieving information from large databases. The relationship sought in data mining can be a relationship between two or more in one dimension. The algorithm included in association rules in data mining is the Frequent Pattern Growth (FP-Growth) algorithm is one of the alternatives that can be used to determine the most frequent itemset in a data set.

Keywords: Data Mining, Association Rules, FP-Growth.

I. INTRODUCTION

Promotional items are a strategy on marketing economics to market a product or even service to increase a company's sales quantity [11][12]. Purchasing and selling is an activity that affects the amount of stock of goods. Then the information generated will help managers in deciding a promotional activity for the introduction of an item to the public.

Data mining is a field of science that unifies techniques from machine learning, pattern processing, statistics, databases, and visualization to handle the problem of retrieving information from large databases [2]. The relationship sought in data mining can be a relationship between two or more in one dimension. For example, in product dimension, there is the relation of purchasing a product with other product. Also, the relationship can also be seen between two or more attributes and two or more objects.

Data mining can analyze large data into information in the form of patterns that have to mean for decision supporters. However, the number of existing data is not always followed by the knowledge that can be generated by the data that many, so in the end, the data only become a less useful. Data mining is the process of searching for patterns of interest and hidden pattern of a large data set stored in a database, such as data warehouse and other data storage [1][3][4].

One important factor in data mining is the existence of rules to find a high-frequency pattern among the itemsets. It is called the Association Rule. FP-Growth is an algorithm included in association rules. Frequent Pattern Growth (FP-Growth) is one of the alternative algorithms that can be used to determine the most common set of data in a dataset. The algorithm searches for the Association Rule by using the value parameters of support and confidence.

II. THEORIES

2.1 Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and as the extraction of useful information and related knowledge from large databases. Several different approaches are classified as information/knowledge seeking techniques in KDD [5][7]. There are quantitative approaches, such as probabilistic approaches such as inductive logic, pattern search, and decision tree analysis. Other approaches include deviation, trend analysis, genetic algorithms, artificial neural networks, and mixed approaches of two or more of several approaches. There are several elements in information search techniques/knowledge in KDD are:

- Working large amounts of data
- Requires efficiency related to the volume of data
- Priority determination/accuracy
- Requires high-level language usage
- Use some form of automatic learning
- Produce interesting things

A company needs business intelligence to develop business processes, monitor time, quality cost, and control about the improvement of a company's quality. In this definition presented that business intelligence is the process of converting data into information. From the collection of existing information will be taken into a pattern of knowledge [6][8]. The purpose of business intelligence is to convert very much data into business value through an analytical report. Data Mining is a series of processes to explore the added value of a data set that produces knowledge that has not been known manually.

2.2 Data Mining Step

Data mining is part of Knowledge Discovery Database (KDD). Data mining is not as a stand-alone technology. It is an important step in the KDD mainly relating to the extraction and calculation of the patterns of the data being sliced [9][10]. Some stages in the process of data mining:

- A Precise statement of the problem. For example, the decision if the bank wants to know whether a customer has the potential to have bad credit, or identify a customer whether to move to the competitor of our business and so forth. After finding the business questions that need to be answered by data mining, then determine the type of task to answer the business questions. The basic tasks that form the basis of data mining algorithms are classification, regression, segmentation, association and sequence analysis.

- Initial Exploration. It is to prepare data to be a source for data mining including data cleaning to learn the pattern.
- After finding the definition of the problem, the next step is to look for data that supports the problem definition. Determine the portion of data used to training data mining based on data mining algorithms that have been made. After the data preparation is completed, the next step is to provide some data into the data mining algorithm.
- Model building and validation. Validate whether data mining provides accurate predictions. After the data training is completed, the data mining needs to be "tested" or validated accurately to the data testing.
- Deployment. This stage selects the right application against data mining to make predictions.

III. RESULTS AND DISCUSSION

In this study, the author uses FP-Growth method that will be utilized in solving the problem of fruit sales analysis which will be used to know a promotion strategy itemset in improving the quality of sales. In this chapter will discuss the analysis of sales data that affect the sales of fruit by using frequent pattern growth.

Table 1. Sales transaction data

No.	Fruit	Date	Week	Day
1	Grape	1-Mar-17	I	Wednesday
2	Apple	1-Mar-17	I	Wednesday
3	Kiwi	1-Mar-17	I	Wednesday
4	Mango	1-Mar-17	I	Wednesday
5	Dragonfruit	1-Mar-17	I	Wednesday
6	Mango	3-Mar-17	I	Friday
7	Orange	3-Mar-17	I	Friday
8	Lemon	3-Mar-17	I	Friday
9	Pear	3-Mar-17	I	Friday
10	Kiwi	3-Mar-17	I	Friday
11	Dragonfruit	5-Mar-17	I	Sunday
12	Sugar Palm Fruit	5-Mar-17	I	Sunday
13	Watermelon	5-Mar-17	I	Sunday
14	Apple	5-Mar-17	I	Sunday
15	Banana	5-Mar-17	I	Sunday
16	Melon	6-Mar-17	II	Monday
17	Snakefruit	6-Mar-17	II	Monday
18	Papaya	6-Mar-17	II	Monday
19	Pear	6-Mar-17	II	Monday
20	Pear	6-Mar-17	II	Monday
21	Banana	9-Mar-17	II	Thursday
22	Dragonfruit	9-Mar-17	II	Thursday
23	Litchi	9-Mar-17	II	Thursday
24	Dates	9-Mar-17	II	Thursday
25	Watermelon	9-Mar-17	II	Thursday
26	Pear	9-Mar-17	II	Thursday
27	Apple	9-Mar-17	II	Thursday
28	Orange	9-Mar-17	II	Thursday
29	Pineapple	11-Mar-17	III	Saturday
30	Melon	11-Mar-17	III	Saturday
31	Banana	11-Mar-17	III	Saturday
32	Orange	11-Mar-17	III	Saturday

33	Banana	11-Mar-17	III	Saturday
34	Pear	11-Mar-17	III	Saturday
35	Tomato	11-Mar-17	III	Saturday
36	Pumpkin	12-Mar-17	III	Sunday
37	Banana	12-Mar-17	III	Sunday
38	Blueberry	12-Mar-17	III	Sunday
39	Orange	12-Mar-17	III	Sunday
40	Avocado	12-Mar-17	III	Sunday
41	Lemon	12-Mar-17	III	Sunday
42	Apple	14-Mar-17	III	Tuesday
43	Pear	14-Mar-17	III	Tuesday
44	Mango	14-Mar-17	III	Tuesday
45	Orange	14-Mar-17	III	Tuesday
46	Lemon	14-Mar-17	III	Tuesday
47	Banana	14-Mar-17	III	Tuesday
48	Apple	14-Mar-17	III	Tuesday
49	Mango	16-Mar-17	III	Thursday
50	Orange	16-Mar-17	III	Thursday
51	Watermelon	16-Mar-17	III	Thursday
52	Pineapple	16-Mar-17	III	Thursday
53	Papaya	16-Mar-17	III	Thursday
54	Snakefruit	16-Mar-17	III	Thursday
55	Mango	16-Mar-17	III	Thursday
56	Grape	16-Mar-17	III	Thursday
57	Litchi	16-Mar-17	III	Thursday
58	Pear	17-Mar-17	III	Friday
59	Tomato	17-Mar-17	III	Friday
60	Banana	17-Mar-17	III	Friday
61	Lemon	17-Mar-17	III	Friday
62	Apple	17-Mar-17	III	Friday
63	Orange	17-Mar-17	III	Friday
64	Kiwi	17-Mar-17	III	Friday
65	Dragonfruit	17-Mar-17	III	Friday
66	Melon	17-Mar-17	III	Friday
67	Pear	17-Mar-17	III	Friday
68	Melon	20-Mar-17	IV	Monday
69	Sugar Palm Fruit	20-Mar-17	IV	Monday
70	Lemon	20-Mar-17	IV	Monday
71	Mango	20-Mar-17	IV	Monday
72	Pear	20-Mar-17	IV	Monday
73	Dates	20-Mar-17	IV	Monday

74	Orange	20-Mar-17	IV	Monday
75	Blueberry	20-Mar-17	IV	Monday
76	Tomato	20-Mar-17	IV	Monday
77	Pumpkin	20-Mar-17	IV	Monday
78	Lemon	22-Mar-17	IV	Wednesday
79	Pear	22-Mar-17	IV	Wednesday
80	Dragonfruit	22-Mar-17	IV	Wednesday
81	Pear	22-Mar-17	IV	Wednesday
82	Papaya	22-Mar-17	IV	Wednesday
83	Melon	22-Mar-17	IV	Wednesday
84	Dates	22-Mar-17	IV	Wednesday
85	Orange	22-Mar-17	IV	Wednesday
86	Apple	22-Mar-17	IV	Wednesday
87	Mango	26-Mar-17	IV	Sunday
88	Banana	26-Mar-17	IV	Sunday
89	Melon	26-Mar-17	IV	Sunday
90	Lemon	26-Mar-17	IV	Sunday
91	Blueberry	26-Mar-17	IV	Sunday
92	Orange	26-Mar-17	IV	Sunday
93	Banana	26-Mar-17	IV	Sunday
94	Pineapple	26-Mar-17	IV	Sunday
95	Sugar Palm Fruit	26-Mar-17	IV	Sunday
96	Lemon	26-Mar-17	IV	Sunday
97	Lemon	26-Mar-17	IV	Sunday
98	Kiwi	28-Mar-17	V	Tuesday
99	Grape	28-Mar-17	V	Tuesday
100	Litchi	28-Mar-17	V	Tuesday
101	Pear	28-Mar-17	V	Tuesday
102	Papaya	28-Mar-17	V	Tuesday
103	Dragonfruit	28-Mar-17	V	Tuesday
104	Orange	28-Mar-17	V	Tuesday
105	Pear	31-Mar-17	V	Friday
106	Lemon	31-Mar-17	V	Friday
107	Tomato	31-Mar-17	V	Friday
108	Mango	31-Mar-17	V	Friday
109	Pear	31-Mar-17	V	Friday
110	Avocado	31-Mar-17	V	Friday
111	Blueberry	31-Mar-17	V	Friday
112	Kiwi	31-Mar-17	V	Friday
113	Dragonfruit	31-Mar-17	V	Friday
114	Banana	31-Mar-17	V	Friday

Table 2 shows an example of some types of fruit taken for experimental material on FP-Growth method calculations.

Table 2. Range per week

No.	Fruit	Initial	Transaction	Range	Remark
1	Snakefruit	SF	2	0,4	II, III
2	Grape	GR	3	0,6	I, III, V
3	Apple	AP	7	0,8	I, II, III, IV
4	Banna	BN	10	1	I, II, III, IV, V
5	Kiwi	KW	5	0,6	I, III, V
6	Mango	MG	8	0,8	I, III, IV, V
7	Papaya	PY	4	0,8	II, III, IV, V
8	Melon	ML	6	0,6	II, III, V
9	Orange	OR	11	1	I, II, III, IV, V
10	Pear	PR	14	1	I, II, III, IV, V

The sample data is based on one-month sales data. The following table will display sample transaction data within one month.

Table 3. Sales rate

Week	1	2	3	4	5	6	7	8	9	10
I		GR	AP	BN	KW	MG			OR	PR
II	SL		AP	BN			PY	ML	OR	PR
II	SL	GR	AP	BN	KW	MG	PY	ML	OR	PR
IV			AP	BN		MG	PY	ML	OR	PR
V		GR		BN	KW	MG	PY		OR	PR

The sample transaction data in the previous table is the sample data to be processed using FP-Growth. The frequency of occurrence of each item from the transaction data can be seen in the following table.

Table 4. Sales item occurrence

No.	Initial	Frequency	Support Frequency	Support
1	SF	2	2/5	40%
2	GR	3	3/5	60%
3	AP	4	4/5	80%
4	BN	5	5/5	100%
5	KW	3	3/5	60%
6	MG	4	4/5	80%
7	PY	4	4/5	80%
8	ML	3	3/5	60%
9	OR	5	5/5	100%
10	PR	5	5/5	100%

Table 4 describes the results of the frequency of occurrence of items of sale through transactions each week. This result will determine support. This study takes the value of support count $\xi = 75\%$ and the value of support count will affect the items that will be analyzed phases of making FP-Tree. It determines the highest support value and that meets the 75% support account. It is the value that a company's agency needs in looking at the average percentage of transactions. The frequency of transactions in a day divided by the number of attributes to find the percentage value of support. It will find an item that meets the frequency, then the item to be used for next is an item that meets the frequency of $\geq 75\%$ as shown in table 5.

Table 5. Support and frequency

No.	Initial	Frequency	Support Frequency	Support
1	AP	4	4/5	80%
2	BN	5	5/5	100%
3	MG	4	4/5	80%
4	PY	4	4/5	80%
5	OR	5	5/5	100%
6	PR	5	5/5	100%

The data will be scanned based on the data frequency occurrence table. Other items that do not meet the frequency of occurrence of support account that is SF with support value 40%, GR with support value 60%, KW with support value 60% and ML with support value 60% each does not meet minimum value support 75% are not scanned. After the process of scanning the data between the frequency table of the emergence of data and the frequency table emergence of support accounts, then the process of data mining can be seen in the following table.

Table 6. Support account occurrence

Week	Transaction
I	AP, BN, MG, OR, PR
II	AP, BN, PY, OR, PR
III	AP, BN, MG, PY, OR, PR
IV	AP, BN, MG, PY, OR, PR
V	BN, MG, PY, OR, PR

IV. CONCLUSION

Based on testing done in the previous stage, it can be concluded that the data mining can be implemented by using a sales database to find trends in pattern itemsets combination so it can be used as valuable information in making decisions to prepare the stock of goods and marketing strategies required later. The application of the FP-Growth algorithm to marketing technique has a very efficient ease and can accelerate the process of formation of the trend of sales item combination pattern. It can be done by promoting items that do not enter the expected confidence.

REFERENCES

I. Fadlina, L. T. Sianturi, A. Karim, and Mesran, "Best Student Selection Using Extended Promethee II Method," International Journal of Recent Trends in Engineering & Research, vol. 3, pp. 21-29, 2017.

II. W. Fitriani, "Comparison Between WEKA and Salford System in Data Mining Software," International Journal

of Mobile Computing & Application, vol. 3, pp. 01-04, 2016.

III. H. A. Hasibuan, and R. B. Purba, "Productivity Assessment (Performance, Motivation, and Job Training) using Profile Matching," International Journal of Economics and Management Studies, vol. 3, pp. 73-77, 2016.

IV. Khairul, and M. Simaremare, "Decision Support System in Selecting the Appropriate Laptop Using Simple Additive Weighting," International Journal of Recent Trends in Engineering & Research, vol. 2, pp. 215-222, 2016.

V. Mesran, K. Tampubolon, R. D. Sianturi, F. T. Waruwu, and A. P. U. Siahaan, "Determination of Education Scholarship Recipients Using Preference Selection Index," International Journal of Scientific Research in Science and Technology, vol. 3, pp. 230-234, 2017.

VI. M. I. Perangin-angin, W. Fitriani, and N. Mayasari, "Tuition Reduction Determination Using Fuzzy Tsukamoto," International Journal of Engineering Science Invention, vol. 5, pp. 68-72, 2016.

VII. A. P. U. Siahaan, "Fuzzification of College Adviser Proficiency Based on Specific Knowledge," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, pp. 164-168, 2016.

VIII. Z. Tharo, "Profile Matching in Solving Rank Problem," IOSR J ournal of Electronics and Communication Engineering, vol. 11, pp. 73-76, 2016.

IX. L. Marlina, and Muslim, "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)," Inte rnational Journal of Engineering Trends and Technology, vol. 38, pp. 380-383, 2016.

X. I. Sumartono, D. Arisandi, and Mesran, "Expert System of Catfish Disease Determinant Using Certainty Factor Method," International Journal of Recent Trends in Engineering & Research, vol. 3, pp. 202-209, 2017.

XI. R. B. Purba, "The Analysis of Good Profitability Governance Using Markov Chain inSmall Business," International Journal for Innovative Research in Multidisciplinary Field, vol. 3, pp. 01-06, 2017.

