Sales Data Analysis Using Apriori and Single Exponential Smoothing Algorithm in Pharmaceutical Companies

Ridho Almuqorobin[1], Onny Marleen[2]

[1]Postgraduate Student, Gunadarma University, Jakarta, Indonesia, 10430
[2]Lecturers, Gunadarma University, Jakarta, Indonesia, 10430
Email address: [1]ridho.almuqorobin @ gmail.com, [2]onny_marleen @ staff.gunadarma.ac.id

Abstract— *One of the pharmaceutical industry companies in processing sales data using computers and storing sales data using a database, so that transaction data can be processed into useful information for the company. Sales transaction data every day will always increase, so that the volume of data becomes large, but if left alone the sales data will only become meaningless junk. One way to manage sales transaction data is to use Data Mining techniques. The purpose of this study is to look for patterns of relationships between products in a sales transaction using a priori algorithm and predict the number of sales in the next period using a single exponential smoothing. After the sales data is realized, it produces 5 association rules with a minimum support value of 10% and a minimum trust of 70% with a positive lift ratio value of more than 1. While for forecasting the next period, namely November 2019 based on data from January 2019 to October 2019, it produces estimated value 4,616 with the value of Mean Absolute Deviation = 238,945, Mean Square Error = 2,175,942,257 and Mean Absolute Percentage Error = 0.734 with alpha value = 0.1.*

Keywords— *Data Mining, Apriori, Single Exponential Smoothing.*

I. INTRODUCTION

Over time a company will produce a very large variety of data both in number and size so that the data must be processed first into knowledge, the more knowledge generated from the data, the greater the value or benefits. The process of processing data into knowledge is referred to as Data Mining. Data Mining can help with decision making and allows companies to manage the information contained in sales transaction data into new knowledge. Based on the knowledge gained, the company can arrange strategies to increase its profits, so that in the future the company can be more competitive.

One of the pharmaceutical industry companies in East Jakarta precisely located in the *Jakarta Industrial Estate Pulogadung* (JIEP) area is a company that has customers in almost all regions of Indonesia. One type of medicines that is sold is the type of *Over the counter* (OTC) is a type of medicines used without using a doctor's prescription, the type of *Over the counter* (OTC) medicines at generally sold in pharmacies. Based on data on sales transactions from January 2019 to October 2019, the Surabaya region has high sales compared to other regions throughout Indonesia. The pharmaceutical industry company in computerized sales data processing and stores sales data using a database, so that the transaction data can be processed into profitable information

for the company. Sales transaction data every day will always increase so that the database has a large volume, although until now service activities and sales transactions have not experienced problems, but if left alone, the transaction data will only become meaningless garbage. One way to manage sales transaction data is by utilizing Data Mining techniques.

In research [1] entitled the implementation of Association rule mining to determine food menus with a priori algorithm. In this study using the Association Rule Mining method and a priori algorithm in mining the requested sales transaction data on payment receipts purchased by the cash register. In this study, the data used were 20 transactions with a minimum support of 30% and a Minimum Confidence of 50% so as to produce 4 food combinations that are most often bought by consumers. In a study [2] entitled the analysis of forecasting (estimated) sales of services at the warnet bulian city in muara bulian. In this study, sales data are used to predict warnet bulian city service sales in August 2018 by using a single exponential Alignment method for forecasting calculations and the average calculation error Mean Absolute Deviation is 1,248,462 for August estimates and August 2018 forecasting is Rp. 10,064,421 estimated average value is smaller than July 2018 income of Rp. 16,222,000.

Based on the description above, this research is to analyze sales data using the association rule mining method with a priori algorithm to determine the pattern of medicines sales and calculate the lift ratio to determine whether the association rules produced are valid or invalid. In addition, this study also predicts the number of medicines sold using Single Exponential Smoothing to predict sales in the next period based on previous transaction data. Thus the results obtained can be used to assist decision making in determining marketing strategies.

II. REVIEW OF LITERATURE

Data Mining is a process that employs one or more machine learning techniques (machine learning) to automatically analyze and extract knowledge. Data Mining is an iterative and interactive process to find new patterns or models that are valid (perfect), useful and can be understood in a large database. [3].

Findings in the Database (KDD) are all processes that are collected activities, the use of historical data to find regularities, patterns or relationships with data sets that are

increasing [4]. Process in KDD is a process that consists of a series of repetitive processes.

1. Data Cleaning
   This process can be used to eliminate or eliminate noise that is not completed.
2. Data Integration
   This process is used to complete data from different source platform formats which can then be integrated in the database.
3. Data Selection
   The data contained in the database are then selected or reduced with various techniques so that the data taken is relevant to the data analysis task from the database.
4. Data Transformation
   This process moves the transformation or transfer of data into a form suitable for excavation through summary or aggregation operations.
5. Data Mining
   Data that has been selected or transformed is weighed with various techniques, an essential process for extracting patterns from data with intelligent methods. Data Mining Process is the process of finding interesting patterns or information in data using certain functions.
6. Evaluate Patterns and Presentations
   Knowledge (Pattern Evaluation and Knowledge Presentation) This stage is part of the process of finding knowledge that complements the patterns, patterns of information or information found that are related to previous facts or hypotheses.

III. METHOD

The object of this research is a company in the pharmaceutical industry in the *Jakarta Industrial Estate Pulogadung* (JIEP) area, using January 2019 sales transaction data until October 2019. Sales transactions used are transactions in the Surabaya region, while the type of product chosen is medicine OTC (Over the Counter) which can be purchased freely at pharmacies.

A. Primary Data

Primary data obtained from a company engaged in the pharmaceutical sector in the form of excel, so that data obtained from January to October 2019 in the Surabaya area. Surabaya is one of the main areas of high sales in Indonesia and there are 3,239 sales data records. The sales data is then cleared with the provisions of deleting duplicate data and deleting transactions that are of type return, so that the remaining 2,836 records. After the transaction data is cleared then the data is grouped so that there are 386 transactions and 20 products.

B. Secondary Data

Secondary data is a source of research data obtained through intermediary media or indirectly in the form of books, journals, articles, or archives both published and not publicly published. The advantage of secondary data is the time and

cost needed for research to classify problems and evaluate data, relatively less compared to primary data collection.

C. Association Rule, Apriori Algorithm, Lift Ratio

Data mining techniques will be applied using the association rule using apriori algorithm then calculating the lift ratio.

formula *support* 1 itemset:

$$Support\ (A) = \frac{Jumlah\ Transaksi\ A}{Total\ Transaksi} \quad [5]$$

formula *support* 2 itemset:

$$Support\ (A, B) = \frac{Jumlah\ Transaksi\ A\ dan\ B}{Total\ Transaksi} \quad [5]$$

formula *support* 3 itemset:

$$Support\ (A, B, C) = \frac{Jumlah\ Transaksi\ A, B\ dan\ C}{Total\ Transaksi} \quad [5]$$

formula *confidence association rules* A → B:

$$Confidence\ (A \rightarrow B) = \frac{\Sigma\ Transaksi\ A\ dan\ B}{\Sigma\ Transaksi\ A} \quad [5]$$

formula *benchmark confidence* :

$$Benchmark\ Confidence = \frac{\Sigma\ Transaksi\ item\ dalam\ consequent}{\Sigma\ Transaksi} \quad [6]$$

formula *lift ratio:*

$$Lift\ Ratio = \frac{Confidence}{Benchmark\ Confidence} \quad [6]$$

D. Forecasting

Make forecast using Single Exponential Smoothing which is used to predictions sales in the next period.

Formula forecast next period

$$Ft + 1 = aXt + (1 - a)Ft \quad [7]$$

Formula *Mean Absolute Deviation (MAD)*

$$MAD = \frac{\Sigma\ |At - Ft|}{n} \quad [2]$$

Formula *Mean Square Error (MSE)*

$$MSE = \frac{\Sigma\ |At - Ft|^2}{n} \quad [2]$$

Formula *Mean absolute percentage error (MAPE)*

$$MAPE = \frac{\Sigma\ |At - Ft|/At}{n} 100 \quad [2]$$

IV. RESULT

A. Association Rule, Apriori Algorithm, Lift Ratio

In this study the researcher determined the minimum support value of 10% and a minimum confidence of 70%. So that the itemset used is an item that has a value of support $> =$ 10% and has a value of confidence $> =$ 70%.

Apriori algorithm calculation begins by defining each itemset and sales data is transformed into data that is ready for mining. Sales data is changed in the form of boolean data in a table, each transaction that occurs on a product item with a category will be represented in binary form, where number 1 indicates the transaction and 0 indicates no transaction. Based on the amount of data consisting of 2,836 records then the data are grouped with 386 transactions and 20 products, then converted into boolean data in a table and calculated products

that are often purchased by customers.

TABLE 1. Products Frequently Purchased by Customers

Itemset	Count	Itemset	Count
I-01	264	I-05	34
I-02	208	I-11	32
I-09	84	I-16	30
I-18	74	I-12	30
I-04	68	I-14	29
I-03	68	I-08	27
I-06	45	I-13	20
I-15	39	I-17	12
I-10	37	I-20	4
I-07	35	I-19	1

After getting a product that is often bought by a customer then calculates support (the percentage of purchases for an item in the whole transaction) for each item.

TABLE 2. Itemset with Minimum Support

Itemset	Count	Support	Itemset	Count	Support
I-01	264	68%	I-05	34	9%
I-02	208	54%	I-11	32	8%
I-09	84	22%	I-16	30	8%
I-18	74	19%	I-12	30	8%
I-04	68	18%	I-14	29	8%
I-03	68	18%	I-08	27	7%
I-06	45	12%	I-13	20	5%
I-15	39	10%	I-17	12	3%
I-10	37	10%	I-20	4	1%
I-07	35	9%	I-19	1	0%

After calculating the support value of 20 itemsset, the minimum support (min-sup) is determined at 10%, if there are items that have support below the min-sup 10% will be removed (prune). Then the remaining items are then merged as many as 2 items (k = 2) to form candidate C2. Of the 20 itemset, the support value is found and 9 itemsset meet the minimum support of 10% and 9 items set that met the minimum support, 36 combinations were produced. the combination of 2 items set that produces 36 items is then searched for the value of support and found there are 10 items that meet the minimum support (min-sup) 10%.

TABLE 3. Combination of 2 items support >= 10%

Itemset	Count	Support
I-02, I-01	155	40%
I-09, I-01	66	17%
I-03, I-01	63	16%
I-02, I-09	61	16%
I-04, I-01	55	14%
I-02, I-03	54	14%
I-02, I-04	49	13%
I-01, I-18	49	13%
I-02, I-18	48	12%
I-04, I-09	37	10%

After getting an itemset that meets the minimum support of 10% then do a combination of 3 itemset. After obtaining a product that is often purchased by customers with 3 itemset combinations, then the value of support with 3 combinations is calculated. After getting a combination of 3 itemsset if there are still items that have support below the min-sup 10% will be removed (prune). Then if the remaining itemset can still be

combined again then a 4 item (k = 4) join will be formed C4 candidate, but if it cannot be combined again it means that the maximum combination of the itemset is only up to candidate C3 (k = 3).

TABLE 4. Combination of 2 items support >= 10%

Itemset	Count	Support
I-02, I-04, I-01	46	12%

The next step is to find association rules that meet the minimum requirements by calculating the confidence association rules A → B, in this study the minimum confidence is 70%.

TABLE 5. Association Rules minimum confidence is 70%

Itemset (A → B)	Count AB	Count A	Count B	Support	Confidence
I-02, I-04 → I-01	46	49	264	12%	94%
I-04, I-01 → I-02	46	55	208	12%	84%
I-04 → I-01	55	68	264	14%	81%
I-02 → I-01	155	208	264	40%	75%
I-04 → I-02	49	68	208	13%	72%

After find for support and confidence values for each itemset, then the association rules used are itemset that have a confidence value of more than 70% and then an elevator ratio is calculated to determine the strength of the association rules. Lift ratio is a measure to determine the strength of association rules that have been formed. The lift ratio value is usually used as a determinant of whether the association rules are valid or invalid.

TABLE 6. Association Rules and Lift Ratio

Itemset (A → B)	Count B	Supp	Conf	Brac- Conf	Lift Ratio
I-02, I-04 → I-01	264	12%	94%	0.68394	1.372604
I-04, I-01 → I-02	208	12%	84%	0.53886	1.552098
I-04 → I-01	264	14%	81%	0.68394	1.182598
I-02 → I-01	264	40%	75%	0.68394	1.089561
I-04 → I-02	208	13%	72%	0.53886	1.337245

B. Forecasting

Forecasting is to predict a business condition in the future by using a forecasting method that will be used to measure and estimate business conditions in the future. In this study forecasting is used to predict the number of products sold in the next period using the Single Exponential Smoothing method. The data used to forecasting are sales data from January 2019 to October 2019.

TABLE 7. Monthly Sales

Month	1	2	3	4	5
product sold	2,466	4,945	4,773	4,289	6,035

Month	6	7	8	9	10
product sold	4,060	4,164	2,971	5,721	4,523

Single Exponential Smoothing method parameter values used are alpha (α). Testing the value of α using values (0.1), (0.2), (0.3), (0.4), (0.5), (0.6), (0.7), (0.8), and (0.9). Testing the influence of the value of α is carried out to determine the effect of whether or not the value of α on the value of the error prediction error when the parameter is changed.

Forecast results from January 2019 to November 2019 with an alpha value of 0.1.

TABLE 8. Forecasting and alpha 0.1

α = 0.1					
Month	Actual	Forcast	MAD	MSE	MAPE
1	2,466	2466.000	0.000	0.000	0.000
2	4,945	2466.000	2479.000	6145441.000	50.131
3	4,773	4697.100	75.900	5760.810	1.590
4	4,289	4765.410	-476.410	226966.488	-11.108
5	6,035	4336.641	1698.359	2884423.293	28.142
6	4,060	5865.164	-1805.164	3258617.428	-44.462
7	4,164	4240.516	-76.516	5854.761	-1.838
8	2,971	4171.652	-1200.652	1441564.363	-40.412
9	5,721	3091.065	2629.935	6916557.241	45.970
10	4,523	5458.007	-935.007	874237.186	-20.672
11		4616.501			
Total			2389.445	21759422.570	7.341
MAD			238.945		
MSE				2175942.257	
MAPE					0.734

Based on the forecast using alpha values 0.1, 0.2, 0.3 it can be seen that the November 2019 forecast is 4,616 the number of products sold with the smallest MAPE value is 0.734 which states the percentage of error forecasting results against the actual value on November sales with alpha value 0.1.

V. CONCLUSION

The conclusion that sales data in Surabaya in January 2019 to October 2019 at companies engaged in pharmacy that there are 3,239 sales data records, the data is then cleared with the provisions of deleting duplicate data and deleting transaction records that are of type return so that there are 2,836 records, after Transaction data is cleared then data is grouped so that there are 386 transactions and 20 products. After the sales data have been analyzed it produces 5 association rules with a minimum support value of 10% and a minimum confidence of 70% with a positive lift ratio value. While for the next forecasting period, which is November based on data from January to October produces forecasting value 4,616,501 or 4,616 with Mean Absolute Deviation = 238,945, Mean Square Error = 2,175,942,257 and Mean Absolute Percentage Error = 0.734 with alpha value = 0.1.

The results of the association's rules can make a promotional package, such as a discount or promo buy 2 items get 3 items such as if buying item I-02 and item I-04 then get item I-01 discount, if buying item I-04 and item I-01 then get item I-02. The company is advised to increase the stock of this product and maintain the availability of this product.

REFERENCES

[1] Pandiangan, Henri. "Implementasi association rule mining untuk menentukan menu makanan dengan algoritma apriori," *Jurnal Mantik Penusa,* Vol. 3, No. 1, Juni 2019, 31-36, 2019.
[2] Fauziah., Ningsih, Yulia Istia., Setiarini, Eva. "Analisis peramalan *(forecasting)* penjualan jasa pada warnet bulian city di muara bulian," *Jurnal Ilmiah Ekonomi dan Bisnis,* 10(1), 61-67, 2018.
[3] Listriani, Dewi., Setyaningrum, Anif Hanifa., Eka M. A, Fenty. "Penerapan metode asosiasi menggunakan algoritma apriori pada aplikasi analisa pola belanja konsumen (studi kasus toko buku gramedia bintaro)," *Jurnal Teknik Informatika,* Vol. 9, No. 2, 120-127, 2016.
[4] Henando, Luki. "Algoritma apriori dan fp-growth untuk analisa perbandingan data penjualan leptop berdasarkan merk yang diminati konsumen (studi kasus : indocomputer payakumbuh)," *Jurnal Sistem Informasi Dan Manajemen Informatika,* Vol. 6, No. 1, 1-17, 2019.
[5] Baetulloh, Uci., Gufroni, Acep Irham., Rianto. "Penerapan metode *association rule mining* pada data transaksi penjualan produk kartu perdana kuota internet menggunakan algoritma apriori," *Jurnal Simetris,* Vol. 10, No. 1, 173-188, 2019.
[6] Despitaria, Sujaini Herry., Tursina. "Analisis asosiasi pada transaksi obat menggunakan data mining dengan algoritma apriori," *Jurnal Sistem dan Teknologi Informasi (JUSTIN,)* Vol. 1, No. 1, 1-6, 2016
[7] Gustriansyah, Rendra., Nadia, Wilza., Sofiana, Mitha. "Komparasi Metode Peramalan Jumlah Permintaan Kamar Hotel," *Jurnal Ilmiah Informatika Global,* Volume 09, No. 2, 95-100, 2018.