Implementation of Association Rules to Manage Cross-Selling and Up-Selling for IT Shop

I Wayan Andis Indrawan[1*], Komang Oka Saputra[2], Linawati[3]
[1]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[2.3]Department of Electrical and Computer Engineering, Udayana University
[*]andisindrawan@gmail.com

Abstract - Nowadays sales marketing strategies in showrooms are increasingly innovative. Utilizing information technology, especially data mining to obtain customer habits in shopping. Customer's shopping habits are recorded in each transaction and stored in a data warehouse. This transaction data is actually very valuable and can be elaborated to be used to manage cross selling and up selling in sales stores, including IT showroom stores. IT Showroom stores usually have unusual or unexpected customers shopping for IT needs. Therefore this potential can be exploited for analysis to get behavior from customers by analyzing transactions in a certain period of time using the Rapidminer application. Rapidminer is a fit application to solve this problem because it is equipped with various algorithms, one of which is association rules. Association rules are algorithms that can be used to analyze data and produce recommendations for cross selling and up selling.

***Index of Terms*- Data Mining, Association Rule Algorithm, Cross Selling, Up Selling, IT Shop.**

I. INTRODUCTION

The retail trade sector is very fast developing. The sales method by displaying the physical goods sold is the advantage of this sale, when compared to online sales [1]. Retail sales have many special tricks in developing their business [2]. Starting from cross selling, up selling and so forth [3]. Cross selling is a method of selling by associating one selling item with another related item [4]. So that buyers are interested in buying these two items together. Cross selling is also an option to develop a retail business [5]. By giving discounts on goods whose sales have a good trend, but on brands that have a nominal more expensive than other similar brands. Both of these methods are only a small part of sales innovation in the world of retail sales.

The retail sales sector is categorized based on goods sold, there are many types ranging from sales of food, medicines, fashion, electronics, to IT equipment. These various items target consumers at the final level or personal consumers, not for businesses. Retail is targeting sales to meet household needs [6] [7] and also has the capability to increase the prestige of the community by elaborating products [8] based on the brand name of the item. As retail growth is increasingly threatened by online businesses, retail businesses need ways to attract shoppers to continue to shop at retail stores rather than online. IT stores are one of the most visible impacts. Online stores easily expose the specifications of an item and are traded online. But apparently that advantage has many gaps that can be exploited from retail stores. Goods sold in online stores sometimes do not match the specifications displayed with the physical arrivals. Many goods sold experience defects in production and also defects during the shipping process. That weakness must be exploited by retail traders by prioritizing the physical quality of goods sold through showrooms and intense promotions. Promotion requires clever trading tricks, one of which is cross selling and up selling as explained earlier [9]. In this era of increasingly rapid information technology, data has an important role to make decisions.

Every IT retail store certainly has a data warehouse that stores sales transactions every day even as a whole since the beginning of the sale [10] [11]. Armed with this transaction data, the seller can process the data with data mining [12]. In the world of data mining there are many data processing applications that can be used, among others, Weka and Rapidminer [13]. Rapidminer is an application that is very easy to use if seen from how to operate it . With [14] which has many problem solving algorithms, trade tricks utilize the FP-Growth (Frequent-Pattern growth) algorithm and the Association Rule which is used to obtain the most frequent sales data that is often bought by customers and the relationship between items purchased by customers. So that retailers can place goods with the highest level of purchase correlation will be placed side by side (cross selling) and goods with the highest frequency level will be given a discount, but on expensive brands (up selling) [5]. So that it can increase sales from the IT Retail store. This idea will be realized in this research.

II. LITERATURE REVIEW

A. Data Mining

Data Mining is the process of analyzing data using software to find patterns and rules of a dataset [12]. Data Mining has the advantage of being able to analyze data with even a large degree of complexity [15]. The results of data mining analysis can help to make decisions based on accurate analysis [16]. Data mining runs by extracting information from data collections [17] using the algorithms needed in accordance with the desired conditions. The results of data mining can involve statistics, database management systems and machine learning [18].

B. Association Rule

Association analysis or association rule mining is a data mining technique for finding associative rules between a combination of items based on support and confidence values. Association analysis is also often referred to as market basket analysis [19].

Association Rule Mining is an algorithm to find the relationship between one item with other items in a dataset. The first step is to look for frequent itemset. In this process, a combination of items that meet the minimum requirements of the support value in the database will be searched. To get the support value of item A can be obtained using the following formula [20] :

$$Support\ (A) = \frac{Jumlah\ transaksi\ \ yang\ mengandung\ \ Item\ A}{Total\ Transaksi}$$

Next to get the support value of 2 items by using the following formula :

$$Support\ (A, B) = P(A \cap B) = \frac{Jumlah\ Transaksi\ \ yang\ mengandung\ \ A\ dan\ B}{Total\ Transaksi}$$

The next process after all frequent items and large item sets obtained then need to look for minimum confidence requirements (mincof) using this formula:

$$Confidence\ (A \rightarrow B) = P(A|B) = \frac{Jumlah\ Transaksi\ \ yang\ mengandung\ \ A\ dan\ B}{Jumlah\ Transaksi\ \ yang\ mengandung\ \ A}$$

C. FP-Growth Algorithm

There are so many algorithms in data mining. One of the highlights is FP-Growth [21]. This algorithm is an alternative algorithm that functions to determine the set of data that most often appears in a data warehouse. This FP-Growth has a slightly different reasoning with Apriori algorithm because it uses a method that takes the set items that often appear, mining without generation of candidates. This builds a very dense data structure (FP-Tree) to compress the transaction database from the original transaction [22].

*D.*RapidMiner

Rapidminer is software that contains an integrated environment for processing data, machine learning, deep learning and predictive analysis [14]. Rapidminer is usually used for commercial to business, research, training, education, rapid prototyping and developing applications and even can display model validation, visualization and optimization results. In addition not only having a paid license, RapidMiner also provides a free license for education [23].

III. METHODOLOGY

This research is elaborated into several steps [24]. Among others :

1. Data from Credible Sources
During this research, the case study used was a computer shop in Denpasar named Agusta Computer. Transaction data used is data from January 2020 to April 2020.

2. Literature Review
This study also sought literature review from similar sources from various disciplines related to this research from scientific papers, journals and a number of online sources.

3. Implementation of FP-Growth and Association Rule
In the third phase of this study will be examined and tested to analyze a number of transactions from the data that has been prepared using the FP-Growth and Association Rule Algorithm. Both of these algorithms have been prepared to process the dataset using an application called RapidMiner. RapidMiner is a suitable application for this research [14].

4. Implementation Results
The results of the implementation of this research are sales data with the greatest intensity level and sales with the most intense level of correlation between goods at a predetermined time period.

5. Conclusions
After obtaining the results of the research implementation, will be followed by a follow-up of the results data obtained for the case study.

IV. RESULT

A. Customer Transaction Data

The following table presents a sample of the five highest customer transaction data at Agusta Computer stores in the period of January 2020 to April 2020.

TABLE I
FIVE HIGHEST TRANSACTION DATA
JANUARY - APRIL 2020

ID Transaksi	Item Transaksi
101	Printer, Tinta Printer
102	Connector RJ45, Crimping Set
103	Printer, Connector RJ45, Crimping Set
104	Printer, Keyboard, Tinta Printer
105	Keyboard, Connector RJ45, Crimping Set
106	Printer, Flashdisk, Tinta Printer
107	Flashdisk, Connector RJ45, Crimping Set
108	Printer, Flashdisk,Connector RJ45, Tinta Printer
109	Keyboard, Connector RJ45, Tinta Printer, Crimping Set
110	Printer, Flashdisk, Keyboard, Crimping Set
111	Printer, Connector RJ45, Crimping Set
112	Printer, Flashdisk, Connector RJ45, Tinta Printer, Crimping Set
113	Printer, Tinta Printer
114	Connector RJ45, Crimping Set
115	Printer, Keyboard, Tinta Printer, Crimping Set
116	Printer, Keyboard, Tinta Printer, Crimping Set
117	Connector RJ45, Crimping Set
118	Printer, Keyboard, Connector RJ45, Tinta Printer, Crimping Set
119	Printer, Tinta Printer
120	Flashdisk, Connector RJ45, Crimping Set

A. Tabular Data

The following table is tabular data from the transaction data sourced from raw data. The tabular data contains the five highest transactions each month from January to April 2020. One example is in the transaction 101 buyers bought printers, and printer ink.

TABLE II
TABULAR TRANSACTION
DATA SET

Row No.	Id$_{Transaksi}$	Printer	Flashdisk	Keyboard	Connector RJ45	Tinta Printer	Crimping Set
1	true	true	false	false	false	true	false
2	true	false	false	false	true	false	true
3	true	true	false	false	true	false	true
4	true	true	false	true	false	true	false
5	true	false	false	true	true	false	true
6	true	true	true	false	false	true	false
7	true	false	true	false	true	false	true
8	true	true	true	false	true	true	false
9	true	false	false	true	true	true	true
10	true	true	true	true	false	false	true
11	true	true	false	false	true	false	true
12	true	true	true	false	true	true	true
13	true	true	false	false	false	true	false
14	true	false	false	false	true	false	true

B. FP-Growth Algorithm Process

The process of FP-Growth algorithm utilizes the concept of tree development, or often referred to as FP-Tree, when searching for frequent item sets. Unlike the Apriori algorithm which uses generate candidate. With this concept, the FP-Growth algorithm is a much faster process compared to the Apriori algorithm. The stages of the FP-Growth algorithm are divided into 3 phases, such as [21] :

1. The generation phase for conditional from pattern base,
2. The generation phase for conditional FP-Tree,
3. Phase for frequent searches of itemset.

Fig. 1 describe the process of applying the FP-Growth algorithm to the dataset implemented using the RapidMiner application.

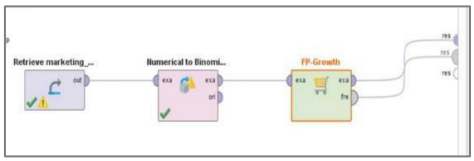


Fig 1. The FP-Growth Algorithm Process

The process of the RapidMiner application in Fig. 1 begins by placing a dataset on the process sheet, which then the dataset must be transformed into binominal data from numerical so that it can be processed by the FP-Growth algorithm. These three processes are linked and processed one by one starting from left to right. The results obtained are in the form of a table as described in the table below.

TABLE III
FP-GROWTH ALGORITH RESULT

Size	Support	Item 1	Item 2
1	0.700	Crimping Set	
1	0.650	Printer	
1	0.600	Connector RJ45	
1	0.550	Tinta Printer	
1	0.350	Keyboard	
1	0.300	Flashdisk	
2	0.350	Crimping Set	Printer
2	0.550	Crimping Set	Connector RJ45
2	0.250	Crimping Set	Tinta Printer
2	0.300	Crimping Set	Keyboard
2	0.250	Printer	Connector RJ45
2	0.500	Printer	Tinta Printer
2	0.250	Printer	Keyboard
2	0.250	Tinta Printer	Keyboard

From the results of Algortima FP-Growth, it can be concluded that the product with the highest sales intensity is an item with the name Crimping Set where this item gets a support value of 0.700, followed by an item that is a printer with a support value of 0.650.

C. Association Algorithm Process

Association rule is a data mining process that has a function to determine all associative rules which can meet the minimum requirements for suppot (minsup) and confidence (minconf) in a dataset. The two conditions will later be used for interesting association rules compared to predetermined limits, namely minsup and minconf. Berikut merupakan implementasi dari Association rule yang diimplementasikan pada aplikasi RapidMiner. This process is a continuation of the previously completed FP-Growth algorithm process.

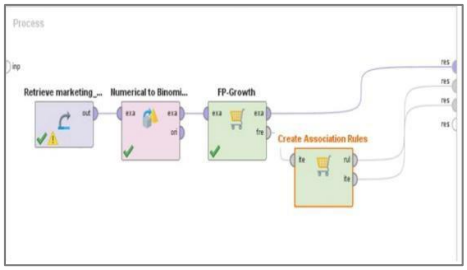
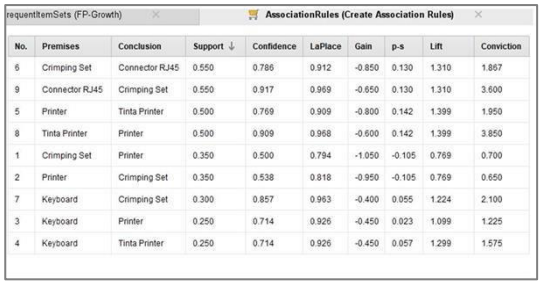


Fig 2. Association Algorithm Process

The process described in Fig. 2 is an advanced process of Fig. 1, where this process is continued by elaborating the

results of data sets that have been processed using the FP-Growth Algorithm followed by processing the data using the Association Algorithm Process. This process is the final process until later we get a final dataset and can be analyzed to make a decision from the company.

TABLE III
ASSOCIATION ALGORITHM RESULT



FrequentItemSets (FP-Growth) | AssociationRules (Create Association Rules)

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s	Lift	Conviction
6	Crimping Set	Connector RJ45	0.550	0.786	0.912	-0.850	0.130	1.310	1.867
9	Connector RJ45	Crimping Set	0.550	0.917	0.969	-0.650	0.130	1.310	3.600
5	Printer	Tinta Printer	0.500	0.769	0.909	-0.800	0.142	1.399	1.950
8	Tinta Printer	Printer	0.500	0.909	0.968	-0.600	0.142	1.399	3.850
1	Crimping Set	Printer	0.350	0.500	0.794	-1.050	-0.105	0.769	0.700
2	Printer	Crimping Set	0.350	0.538	0.818	-0.950	-0.105	0.769	0.650
7	Keyboard	Crimping Set	0.300	0.857	0.963	-0.400	0.055	1.224	2.100
3	Keyboard	Printer	0.250	0.714	0.926	-0.450	0.023	1.099	1.225
4	Keyboard	Tinta Printer	0.250	0.714	0.926	-0.450	0.057	1.299	1.575

The results of processing the dataset can be seen in Table III where the data obtained is a combination of the sale of item sets with the highest level of probability of being bought simultaneously. In the first place the most frequently purchased item combination is Crimping Set and RJ45 Connector with a support value of 0.550 and a Confidence value of 0.785. The combination of item sets with the second highest level of sales with different types of item sets results in a combination of Prnter and TInta Printer sales with a support value of 0.500 and a confidence value of 0.769. So that these two itemset combinations deserve important attention for sellers to manage marketing strategies in the following month.

V. CONCLUSION

Association Rule can be used as an algorithm to analyze Cross-Selling and Up-Selling sales. The Association Rule Algorithm can be interpreted using the RapidMiner Tools in combination with the FP-Growth algorithm. RapidMiner can be used as a transaction data processing machine as a marketing strategy for both Cross-Selling and Up-Selling. RapidMiner can be used as a transaction data processing machine as a marketing strategy for both cross-selling and up-selling. The implementation of this technology can be seen from this research.

REFERENCES

[1] L. P. Dumais, J. R. Tumiwa, and M. P. Sam, "A Comparative Analysis of Consumer Preferences Between Online and Offline Store (Case Study on Fashion Product)," *J. EMBA J. Ris. Ekon. Manajemen, Bisnis dan Akunt.*, vol. 5, no. 3, pp. 3282–3291, 2017.

[2] R. Sarkar, "Online Shopping vs Offline Shopping : A Comparative Study," *Int. J. Sci. Res. Sci. Technol.*, vol. 3, no. 1, pp. 424–431, Jan. 2017.

[3] J. S. Johnson and S. B. Friend, "Contingent cross-selling and up-selling relationships with performance and job satisfaction: an MOA-theoretic examination," *J. Pers. Sell. Sales Manag.*, vol. 35, no. 1, pp. 51–71, Jan. 2015.

[4] P. Surya Sumartha, F. Samopa, P. S. Sumartha, and F. Samopa, "Cross Selling Product Bundling Based on Customer Satisfaction Study Case Meat," *Int. J. Educ. Res.*, vol. 5, no. 1, pp. 241–252, 2017.

[5] E. Miranda and N. Elfreida, "Data Warehouse, Data Mining Dan Konsep Cross-Selling Pada Analisis Penjualan Produk," *ComTech Comput. Math. Eng. Appl.*, vol. 1, no. 2, p. 344, 2010.

[6] M. P. A. Ariawan, P. B. I. S. Putra, and I. M. Sudarma, "Analysis of Enterprise Architecture Design Using TOGAF Framework : A Case Study at Archival Unit of Faculty of Agricultural Technology of," vol. 2, no. 2, pp. 52–57, 2017.

[7] D. Electrical, "Designing a Virtual Data Warehouse in Supporting Sales Information Needs," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, pp. 2–5, 2019.

[8] R. L. Rahardian and M. Sudarma, "Application of Neural Network Overview In Data Mining," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, pp. 1–19, Sep. 2017.

[9] D. Ardiada, P. A. Ariawan, and M. Sudarma, "Evaluation of Supporting Work Quality Using K-Means Algorithm," *Int. J. Eng. Emerg. Technol.*, vol. 3, no. 1, pp. 52–55, 2018.

[10] K. A. Shobirin, A. P. S. Iskandar, and I. B. A. Swamardika, "Data Warehouse Schemas using Multidimensional Data Model for Retail," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, p. 84, 2017.

[11] I. N. A. Prabawa, D. Agung, K. Arimbawa, and I. G. N. Janardana, "Analysis and Design Data Warehouse For E-Travel Business Optimization," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, 2019.

[12] P. Novenando, M. Weking, I. G. Ngurah, W. Partha, and A. I. Weking, "Application of Data Mining with Support Vector Machine (SVM) in Selling Prediction Trend of Spiritual Goods (Case Study : PT . X Bali)," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, pp. 20–24, 2019.

[13] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018.

[14] M. Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," *Jurteksi*, vol. 4, no. 2, pp. 123–132, 2018.

[15] I. W. S. Pramana, P. R. Iswardani, N. Wayan, and S. Aryani, "Application of Data Mining in Optimization of Hotel ' s Food and Beverage Costs," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, pp. 3–7, 2019.

[16] S. I. Murpratiwi, A. A. N. Narendra, and M. Sudarma, "Mapping Patterns Achievement Based on CRISP-DM and Self Organizing Maps (SOM) Methods," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, pp. 1–6, 2017.

[17] K. S. Utami, I. G. W. Darma, N. Wayan, and S. Aryani, "Stock management using K-means method and Time Series method as Stock Order," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, pp. 60–64, 2019.

[18] W. Wahyudin, I. P. A. Wijaya, and I. B. A. Swamardika, "Data Mining for Clustering Revenue Plan Expense Area (APBD) by using K-Means Algorithm," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, p. 87, 2017.

[19] A. P. S. Iskandar, K. A. Shobirin, and K. Oka Saputra, "Analysis of Shopping Cart At Drugs Store By Using An Apriori Algorithm," *Int. J. Eng. Emerg. Technol.*, vol. 2, no. 1, p. 97, 2017.

[20] D. Agung, K. Arimbawa, I. N. A. Prabawa, and P. A. Mertasana, "Implementation of Apriori Algorithm in Determining Tourism Visit Patterns to Bali," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, pp. 10–14, 2019.

[21] Y. D. Lestari, "Penerapan Data Mining Menggunakan Algoritma Fp-Tree Dan Fp-Growth Pada Data Transaksi Penjualan Obat," in *Seminar Nasional Teknologi Informasi dan Komunikasi (SNASTIKOM 2015)*, 2015, no. Snastikom, pp. 60–65.

[22] A. Ikhwan, D. Nofriansyah, and Sriani, "Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)," *Saintikom*, vol. 14, no. 3, pp. 211–226, 2015.

[23] S. Wahyuni, K. S. S. Mochammad, and I. Perangin-Angin, "Implementasi Rapidminer dalam Menganalisa Data Mahasiswa Drop Out," *J. Abdi Ilmu*, vol. 10, no. 2, pp. 2013–2016, 2017.

[24] A. Agung, G. Oka, K. Adnyana, and K. O. Saputra, "Design of Data Warehouse for University Library using Kimball and Ross 9 Steps Methodology," *Int. J. Eng. Emerg. Technol.*, vol. 4, no. 1, 2019.