Jurnal Mantik
Volume 4, Number 3, Nov 2020, pp. 2063-2070
E-ISSN 2685-4236
https://iocscience.org/ejournal/index.php/mantik

FP-Growth Implementation in Frequent Itemset Mining for Consumer Shopping Pattern Analysis Application

I Made Dwi Putra Asana[1], I Komang Arya Ganda Wiguna[2], Ketut Jaya Atmaja[3], I Putu Anjas Sanjaya[4]

[1,2,3,4]Informatics Department, STMIK STIKOM Indonesia, Denpasar

E-mail: dwiputraasana@stiki-indonesia.ac.id

ARTICLEINFO

Article history:
Received: 01/11/2020
Revised: 08/11/2020
Accepted: 23/11/2020

Keywords:
Data Mining, Association rule mining, frequent itemset, fp-growth, consumer shopping pattern

ABSTRACT

Most retail companies have implemented computer-based information systems for recording sales transaction data. In the implementation of information systems, the data collected in the database is processed limited to making reports such as sales reports and inventory reports. Database generated from computer-based information systems can be further processed to obtain more valuable information. One strategy for using sales transaction data is to analyze consumer spending patterns. Consumer spending patterns can be in the form of associations of items that are often purchased simultaneously. The association between goods can be determined using the frequent itemset search technique. The Fp-growth algorithm is an algorithm that can be used to determine frequent itemsets in a data set. This article describes the results of implementing the FP-Growth algorithm in the consumer shopping pattern analysis application. The resulting shopping pattern is in the form of goods that are often purchased simultaneously by consumers. From the results of the application of the fp-growth algorithm, it was found that the minimum value of support had an effect, namely the smaller the input value of support, the more pairs of items were obtained. The application of the FP-Growth algorithm in determining frequent itemsets in association data mining can find customer spending habits in buying goods simultaneously.

1. Introduction

Every retail company has a challenge to compete with other retail companies. Along with the addition of retail companies, it creates more competitive competition between retail companies. Every retail company must be able to recognize consumer spending habits as a basis for providing services. Information on consumer shopping habits can be the basis for providing attractive promos for consumers. So that retail companies have a strategy in exploring consumer shopping information, which is a must for retail companies. The solution that can be given is data mining. The data mining group that can be used to find customer spending habits is association rule mining.

Ayunadi Swalayan is a retail company located on Jl. Tukad Batanghari, Dauh Puri Klod, West Denpasar, Denpasar City, has been established since 2004 (14 years). This business has used computer-based Point Of Sales (POS). The number of transactions per day averaged 1,120 transactions. The number of items contained in the database is around 15,200 items. In order for this large amount of data to be more useful, it can be reprocessed by applying a data mining method so that it can produce information on customer spending habits.

Data mining, often called knowledge discovery in database (KDD), is an activity that includes collecting and using historical data to find regularities, patterns or relationships in large data sets. The output of data mining can be used to improve future decision making. Data mining is the process of finding meaningful relationships, patterns, and trends by examining large amounts of data stored in storage using recognition techniques such as engineering and mathematics. Data mining has several groups based on the tasks that can be done, one of which is association, the task of association in data mining is to determine the attributes that appear at one time. In the business world it is more commonly called shopping cart analysis[1].

Association rule mining is a data mining technique to find associative rules between a combination of items. An example of an associative rule of purchasing analysis in a super market is to find out how likely a customer is to buy bread along with milk. With this knowledge, super market owners can arrange the placement of their goods or design marketing campaigns using discount coupons for certain combinations of goods[2].

Accredited "Rank 4"(Sinta 4), DIKTI, No. 36/E/KPT/2019, December 13th 2019.

The association between goods is known to use the frequent itemset search technique. Frequent itemset is a collection of one or more items with the appearance of the same or more than the minimum support threshold in the data set[3]. Several studies on retail company transaction data analysis use associated data mining techniques[4] [5]. In this study, a frequent itemset search was performed using the a priori algorithm. The priori algorithm has problems in data processing time[6]One of the alternative algorithms in frequent itemset is FP Growth. The Fp-growth algorithm is an algorithm of the association rule technique that can be used to determine frequent itemset in a data set[7]The FP-Growth algorithm is a development of the Apriori algorithm with a faster frequent itemset search process[8]. Frequent Pattern-Growth (FP-Growth) is an algorithm that can save time and storage media, especially for large databases[9][10].

This article describes the results of implementing the FP-Growth algorithm in the consumer shopping pattern analysis application. The resulting shopping pattern is in the form of goods that are often purchased simultaneously by consumers. The application built can generate information on customer shopping habits at the Ayunadi Swalayan company. The resulting information can be used by management as a basis for decision support in developing promotional strategies to increase sales.

2. FP-Growth Algorithm

Frequent Pattern Growth (FP-Growth) is an alternative algorithm that can be used to determine frequent itemset in a data set [11], The FP-Growth algorithm is a development of the Apriori algorithm. So that the shortcomings of the Apriori algorithm are fixed by the FP-Growth algorithm.

FP-Growth uses the concept of building a tree in searching for frequent itemsets. This is what causes the FP-Growth algorithm to be faster than the Apriori algorithm. The characteristic of the FP-Growth algorithm is that the data structure used is a tree called the FP-Tree. By using FPTree, the FP-Growth algorithm can directly extract frequent itemset from FPTree.

FP-tree is a compressed data storage structure. The FP-tree is built by mapping each transaction data into each specific path in the FP-tree. Because in every transaction that is mapped, there may be transactions that have the same item, the path is possible to overwrite each other.

The more transaction data that has the same item, the more effective the FP-tree data structure is compressed. Frequent itemset excavation using the FP-Growth algorithm will be carried out by generating a data tree structure (FP-Tree). The FP-Growth method can be divided into 3 main stages, namely as follows:

a) Conditional Pattern Base Generating Stage The Conditional Pattern Base is a subdatabase that contains a prefix path and a pattern suffix. The conditional pattern base generation is obtained through the previously built FP-tree.
b) FP-Tree conditional generation stage At this stage, the support count of each item in each conditional pattern base is added up, then each item that has a number of support counts greater than the minimum support count ξ will be generated with a conditional FP-tree.
c) Frequent itemset search stage If the Conditional FP-tree is a single path, then the frequent itemset is obtained by combining items for each FP-tree conditional. If it is not a single trajectory, then FP-growth will be generated recursively.

FP-growth is an alternative algorithm that can be used to determine the most frequent itemset in a data set. FPgrowth uses a different approach from the paradigm used in the Apriori algorithm (Sensuse, 2012).
FP-Growth is an alternative algorithm that can be used to determine the most frequent data set (frequent item set) in a data set. The FP-Growth algorithm is a development of the Apriori algorithm. FP-growth is a method that is often a mining itemset without a candidate generation. It builds a very dense data structure (FP-tree) to compress the original transaction database[12].

The process of searching for frequent itemsets using the FP-Growth algorithm is carried out by generating a data tree structure or what is called the FP-Tree. The FP-Growth method for generating frequent items through the construction of a decision tree structure is divided into three main stages, namely:

a) Stage of conditional pattern base generation
Conditional Pattern Base is a subdatabase containing the path prefix and pattern suffix. The conditinal pattern base generation is obtained through the FP-Tree that has been built previously.

b) The FP-Tree conditional generation stage
At this stage, the support count of each item in each conditional pattern base is added up, then each item that has a number of support counts greater than or equal to the minimum support count will be generated with the FP-Tree conditional.

c) The frequent itemset search stage.
If the FP-Tree conditional is a single path, then the frequent pattern is obtained by combining items for each FP-Tree conditional. If it is not a single cross, then FP-Growth will be generated recursively.

Accredited "Rank 4" (Sinta 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

3. Research methods

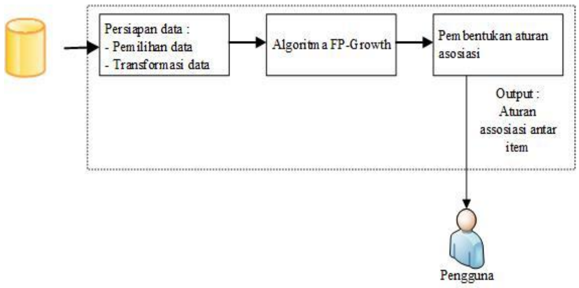

Fig 1. System Overview

The process in the shopping cart analysis application system begins with a data preparation process which includes data selection and transformation. Data selection and data transformation are carried out on POS database input from retail companies (Raw Data). The results from the data preparation stage are then carried out by excavating high frequency patterns (frequent pattern mining) using the FP-Growth algorithm according to the minimum support value entered by the system user. After a high frequency pattern is found, then an association rule is formed based on the minimum confidence value entered by the system user. Then the system will provide an output in the form of association rules between items purchased by customers.

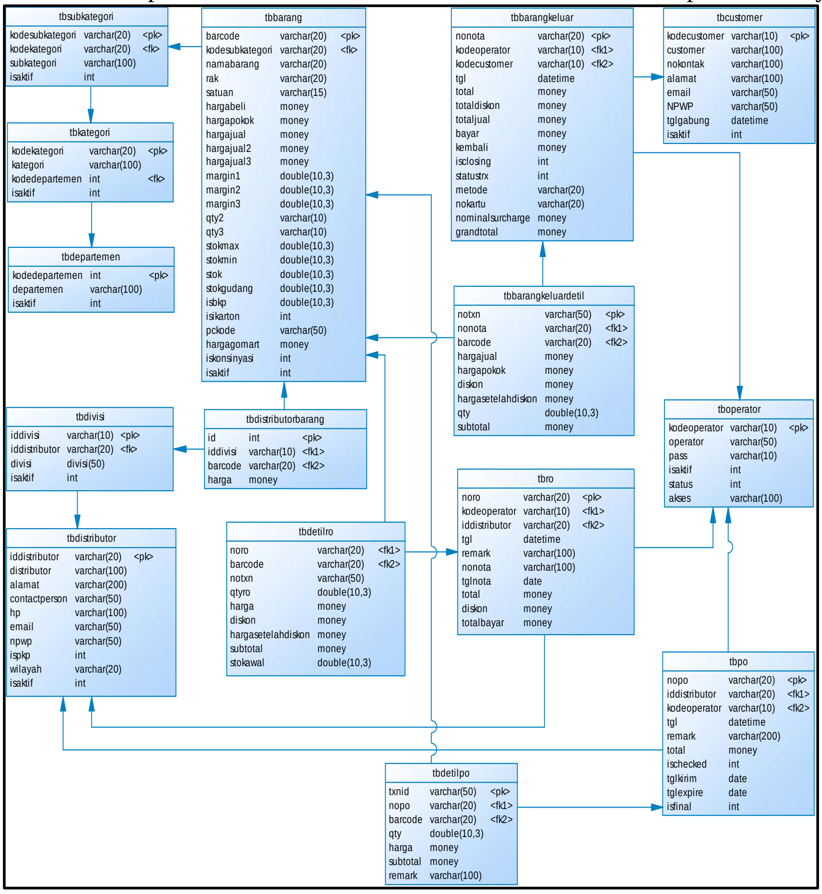

Fig 1 Structure of the AyunadiSwalayan POS Database Table

Figure 2 is a database table structure for the Ayunadi Swalayan POS application which is used as the basis for building a database for data analysis needs. The Database Management System used in Ayunadi Swalayan is SQL Server. In Ayu Nadi Swalayan's research, the focus is on data on sales of household goods, groceries, etc. Figure 2 shows the relationship table that handles data storage of sales and receipt of goods from distributors. To build a datamart, an Extract, Transform, and Load (ETL) process is carried out to produce a more concise table for data analysis.

In the extraction process, the tables and fields are selected in the Ayunadi Swalayan POS database. Based on the needs of the analysis, the tables required are tables of goods, TB of goods out, TB of goods out of detail, and TB of subcategories. Data extraction is carried out based on the join table shown in Figure 3.

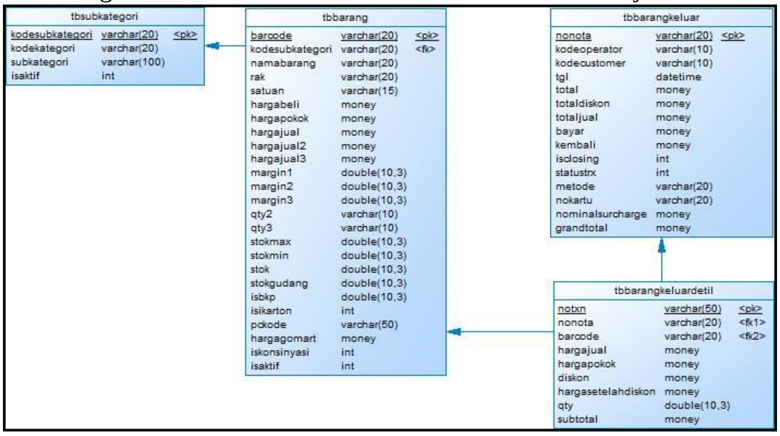

Fig 2 Extraction Table

Based on the table that has been determined in the extraction process, the next step is to transform the data. In the transformation process, data type is determined from the fields used and a joining (merging) of the tables. Joining is done on tbbarang and detailed transactions. Figure 4 shows the results of data transformation for datamart needs.

Fig 3 Transformation Table

Load is the final result of the ETL process, load displays data that has gone through the extraction and transformation process. In the Load process the dimensions and facts tables can be seen and processed further to implement. Dimensions and facts tables can be seen in Figure 5.

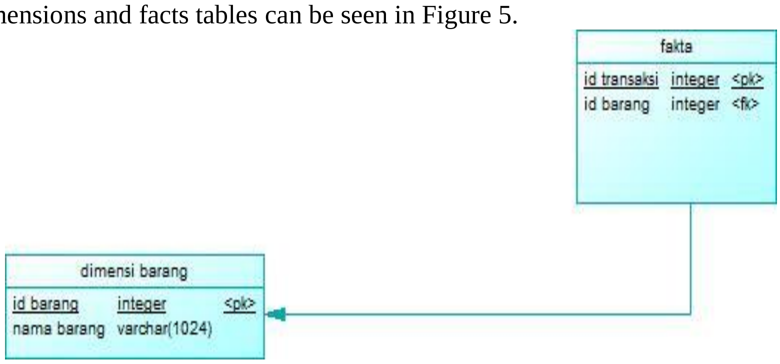

Fig 5 Load Table

4. Results and Discussion

4.1 Home page

The Home page is the page that will be seen for the first time when a customer visits this website. On this page there is a Home menu, Item Data and FP-Growth Algorithm Analysis. Customers can view Item Data and analyze the FP-Growth Algorithm calculations on the FP-Growth Algorithm Analysis menu. Here is the Home page, can be seen in Figure 6.

Accredited "Rank 4" (Sinta 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

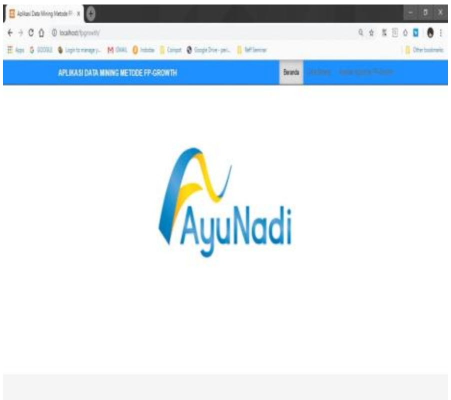

Fig 4 Home page

4.2 Item Data page

Item Data page is a page that displays item data, which consists of the Item Code and Name. This page also has an edit button to edit item data and a delete button to delete item data. Item Data page can be seen in Figure 7.

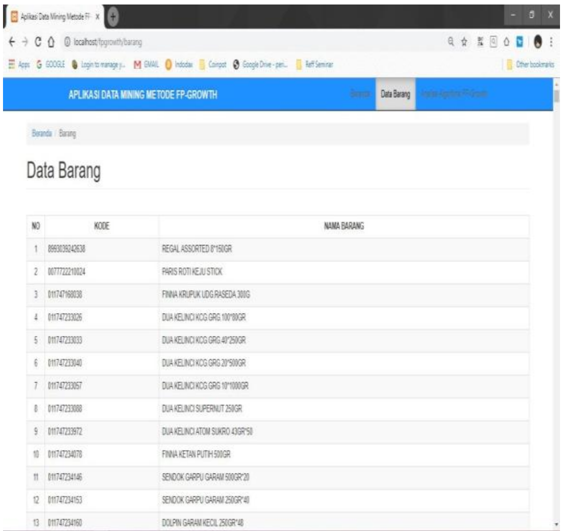

Fig 5 Inventory

4.3 Association Analysis page

The association analysis page is a page that displays the process of item association analysis. The first step in the analysis is that the customer enters a minimum of support, a minimum of confidence, and the number of transactions to be analyzed (Figure 8). After the input process data parameters are entered, then the user presses the process button.

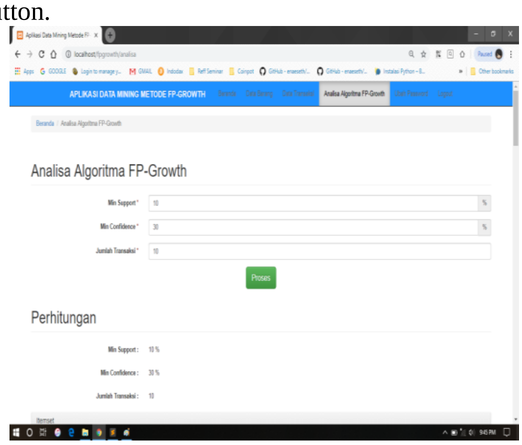

Fig 6 FP-Growth Algorithm Analysis page

4.4 Itemset Support

Figure 9 shows itemset support data which is the initial stage of frequent itemset with the FP-Growth algorithm. Itemset support is the support value of each item to the total transactions being analyzed.

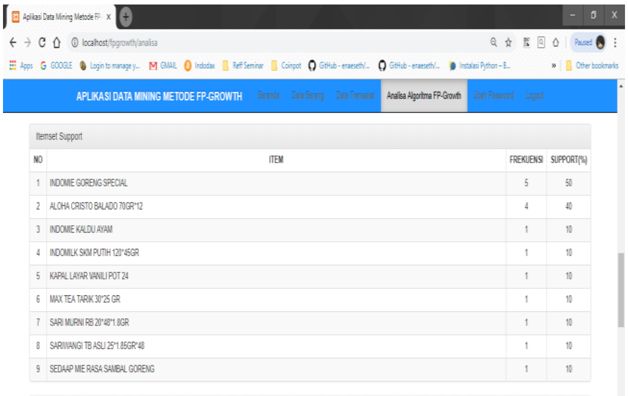

Fig 9 Itemset Support

4.5 Priority Itemset

Itemset priority is the stage after itemset support is found. Priority items are items that have a support value greater than or equal to the support threshold value. The results of the itemset priority are shown in Figure 10.

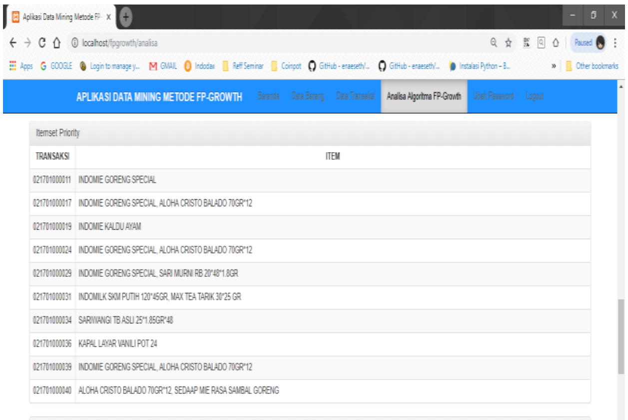

Fig 10 Priority Itemset

4.6 Conditional FP-Tree

Based on the results of the itemset priority, the system then formed a Conditional FP-Tree. The FP-Tree conditional shows the total number of encounters between goods, for example Aloha and Indomie Goreng, shown in Figure 11.

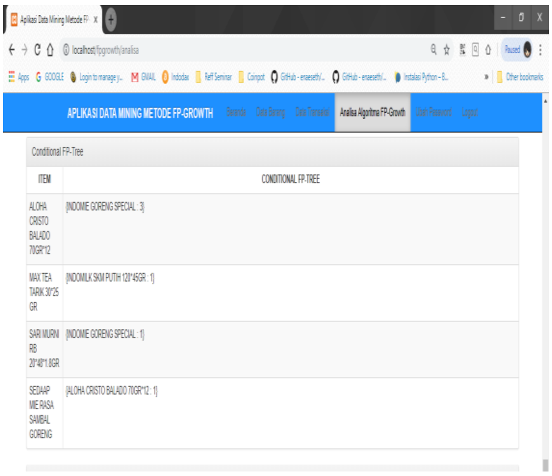

Fig 11 Conditional FP-Tree

Accredited "Rank 4" (Sinta 4),DIKTI, No. 36 / E / KPT / 2019, December 13th 2019.

4.7 Frequent Itemset

Frequent Itemset shows pairs of items that have an appearance value that exceeds the support threshold value. Figure 15 shows the results of the frequent itemset.

Fig 12 Frequent Itemset

4.8 Association Rule

Figure 13 shows the results of association rule mining.*Association Rule* contains rule items that have a value of support and confidence greater than the minimum input support and confidence.

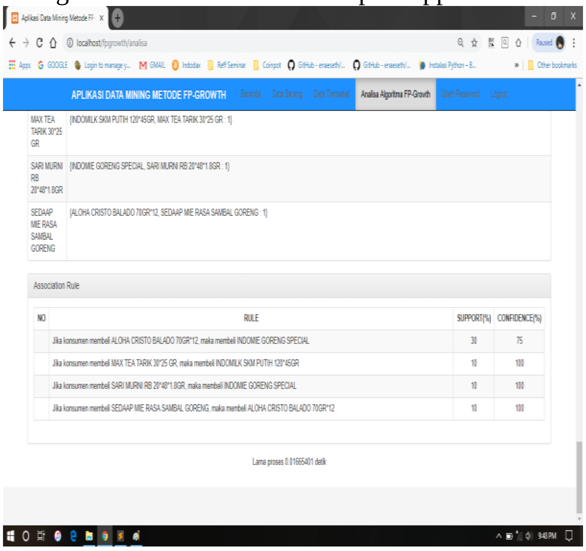

Fig 13 Association Rule

5. Conclusion

The conclusions that can be drawn from the application of the fp-growth algorithm in determining the association between goods in retail companies are as follows:

a) The smaller the input value, the more pairs of items you will get.

b) The application of the FP-Growth algorithm in determining frequent itemsets in association data mining can find customer spending habits in buying goods simultaneously.

c) The benefit of the consumer shopping pattern analysis application with FP-Growth in determining associations between goods is that companies can find out what items are often purchased simultaneously by customers, so that they can adjust the location of items that are often purchased together on one shelf, and keep stock of one of the items that are frequently purchased. often bought together.

Acknowledgment

This research was supported and financed by theministry of technology research and high education in the Republic of Indonesia with a contract number 07/04 / LPPM / PE / IV / 2019. The author also thanks STMIK STIKOM Indonesia for support in the implementation of this research.

6. References

[1] D. T. Larose, *Discovering Knowledge in Data : An Introduction to Data Mining*. New Jersey: John Willey & Sons,Inc, 2005.
[2] K. dan E. T. Lutfi, *Algoritma Data Mining*. Yogyakarta: Andi, 2009.
[3] K. Chavan, P. Kulkarni, P. Ghodekar, and S. N. Patil, "Frequent itemset mining for Big data," in *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, 2016.
[4] A. W. O. Gama, I. K. G. D. Putra, and I. P. A. Bayupati, "IMPLEMENTASI ALGORITMA APRIORI UNTUK MENEMUKAN FREQUENT ITEMSET DALAM KERANJANG BELANJA," *Teknol. Elektro*, vol. 15, no. 2, 2016.
[5] A. Ag. B. Ariana and I. M. D. P. Asana, "ANALISIS KERANJANG BELANJA DENGAN ALGORITMA APRIORI PADA PERUSAHAAN RETAIL," 2013, pp. 2–4.
[6] W. Choiriah, "Penggunaan Algorithma Apriori Data Mining Untuk Mengetahui Tingkatkesetiaan Konsumen (Brand Loyality) Terhadap Merek Kenderaan Bermotor (Studi Kasus Dealer Honda Rumbai)," *J. Teknol. Inf. Komun. Digit. Zo.*, vol. 7, no. 1, p. 44, 2016.
[7] R. N. Arifin, "Implementasi Algoritma Frequent Pattern Growth (FP-Growth) Menentukan Asosiasi antar Produk (Studi Kasus: Nadiamart)," *J. Tek. ITS*, pp. 68–76, 2015.
[8] D. Widiastuti and N. Sofi, "Analisis Perbandingan Algoritma Apriori dan FP-GROWTH Pada Transaksi Koperasi," *UG J.*, vol. 8, no. 1, 2014.
[9] Y. Liu and Y. Guan, "FP-Growth Algorithm for Application in Research of Market Basket Analysis," in *2008 IEEE International Conference on Computational Cybernetics*, 2008, pp. 269–272.
[10] M. . Mythili and A. R. M. Shavanas, "Performance Evaluation of Apriori and FP-Growth Algorithms," *Int. J. Comput. Appl.*, vol. 79, no. 10, 2013.
[11] A. Sijabat, "Penerapan Data Mining untuk Pengolahan Data Siswa dengan Menggunakan Metode Decision Tree," *J. Inf. dan Teknol. Ilm.*, vol. 5, no. 3. ISSN : 2339-210X, 2015.
[12] Ririanti, "Implementasi Algoritma FP-Growth pada Aplikasi Prediksi Persediaan Sepeda Motor (Studi Kasus PT. Pilar Deli Labumas)," *Pelita Inform. Budi Darma*, vol. 6, pp. 139–144, 2014.