IAIC Transactions on Sustainable Digital Innovation (ITSDI)
Vol. 4 No. 2 April 2023

p-ISSN: 2686-6285
e-ISSN: 2715-0461

Author Notifications
12 March 2023

Final Revised
23 March 2023

Published
24 March 2023

Use of Data Mining for The Analysis of Consumer Purchase Patterns with The Fpgrowth Algorithm on Motor Spare Part Sales Transactions Data

Tri Ahmad Djabalul Lael[1], *Deskha Akmal Pramudito[2]
Bursa Uludag Universitesi[1], Padjajaran University[2]
Özlüce Görükle Kampüsü, 16059 Nilüfer/Bursa[1]
Jl. Raya Bandung Sumedang KM.21, Hegarmanah, Kec. Jatinangor, Kabupaten Sumedang, Jawa Barat 45363[2]
Turkey[1], Indonesia[2]

e-mail: deskha20001@mail.unpad.ac.id

To cite this document:
Lael, T. A. D., Pramudito, D. A., (2023). Use of Data Mining for The Analysis of Consumer Purchase Patterns with The Fpgrowth Algorithm on Motor Spare Part Sales Transactions Data. IAIC Transactions on Sustainable Digital Innovation (ITSDI),4(2), 119–127. https://doi.org/10.34306/itsdi.v4i2.582

Abstract

This study aims to analyze consumer purchasing patterns for motorcycle parts using data mining methods and FP-Growth algorithms on motorcycle parts sales transaction data. This research aims to obtain helpful information for companies in planning marketing strategies and increasing sales. The data used in this study are motorcycle parts sales transaction data from motorcycle parts stores for one year. The data is then processed using the FP-Growth algorithm to find significant purchasing patterns. The results of this study show that the FP-Growth algorithm can be used to identify substantial consumer purchasing patterns. Some purchase patterns found include a combination of often purchased products, the most active purchase time, and the most purchased product category. Using data mining and the FP-Growth algorithm can assist companies in understanding significant consumer purchasing patterns to improve the effectiveness of marketing strategies and increase sales of motorcycle parts. The novelty of this research lies in using data mining methods and FP-Growth algorithms on motorcycle parts sales transaction data to analyze consumer purchasing patterns. This research also provides valuable information for companies in planning marketing strategies and increasing sales by identifying significant consumer purchasing patterns, such as product combinations often purchased together and the most purchased product categories.

Keywords: FP-Growth Algorithm, Data Mining, and Consumer Buying

1. Introduction

Every company has data stored in its database[1]. The transaction data is getting bigger and bigger day by day. With the increasing amount of data at the company, the role of analysts to analyze data manually needs to be replaced with computer-based applications so that the analysis process can be done precisely and accurately[2][3].

This study applies the FP-Growth algorithm in an application that can determine consumer buying patterns at each different branch with different characteristics. Information will be obtained from the resulting pattern. Association analysis is a data mining technique that forms the basis of various other data mining techniques[4]. Numerous researchers have focused on

 p-ISSN: 2686-6285
e-ISSN: 2715-0461

frequent pattern mining, a stage of association analysis, in order to develop effective algorithms. The significance of a cooperative rule not entirely settled by two boundaries, support (support esteem), to be specific the level of the thing mix in the data set and certainty.

Frequent PatternGrowth (FP-Growth), a variant of the Apriori method that generates a Frequent Pattern Tree (FP-Tree) data structure to identify a data set's most frequently occurring data set (frequent itemset), is the algorithm used in this study. Several things become research issues in light of the described background: 1) How to apply the Information Mining strategy with the FP-Development calculation to applications for dissecting buyer purchasing behaviors? (2) At Sarana Motor, how do customers typically make purchases? 3) How are the resulting purchasing patterns interpreted into information.

From the formulation of the problem described above, there are several objectives of this study, namely knowing consumer buying patterns in each branch, interpreting the patterns that have been generated into information, designing a good sales strategy and system[5], applying the Data Mining method with the FP-Growth algorithm into the consumer buying pattern analysis application, and testing the patterns that have been implemented successfully or not[6].

2. Research Method

2.1 Method of Collecting Data

This study uses several methods in data collection, namely:

1) Field Research

   The research was conducted directly with Mendi Shopping to obtain primary data. The data collection techniques used are:
   - Observation, namely collecting data by direct observation and systematic recording, to obtain objective data.
   - Direct interview (interview), namely the method of collecting data by conducting a question and answer session directly with the parties concerned in the field studied to obtain the required information.

2) Library Research

   Researchers conducted a literature review to study several books related to this research. Some of these books include books related to data mining and the FP Growth algorithm, books on sales transaction data and other books deemed necessary to support this research.

3) Laboratory Research

   This is the research stage that is carried out using a computer research laboratory to practice directly the results of the analysis, which aims to test the correctness of the designed system.

The theories that form the basis for writing this thesis include data mining theory, the FP Growth Algorithm, KDD and Association Rule.

2.2 Data Mining

Data Mining is not a completely new field. One of the difficulties in defining data mining is the fact that data mining inherits many aspects and techniques from established fields of knowledge[7]. Starting from several disciplines, Data Mining aims to improve on traditional techniques so that they can handle the following:

a. Huge amounts of data

b. High data dimensions
c. Heterogeneous and different data

Grouping Data Mining is divided into several groups, namely:

1) Predictions
Prediction guesses an unknown value and also predicts the value for the future.
2) Description
The description is a way to describe the patterns and trends in your data.
3) Estimation
Estimation is almost the same as classification, except that the target variable is estimated more numerically than categorically. The model is built using a complete record that provides the value of the target variable as a predictive value.
4) Association
The association is tasked with finding the attributes that appear at one time. In the business world, it is more commonly called shopping cart analysis.
5) Clustering
It groups records, observations, or attention and forms a class of similar objects.
6) Classification
In the classification, there is a categorical target variable. For example, income classification can be separated into three categories: high, medium and low.

2.3 FP Growth Algorithm

The FP-Growth algorithm is a viable alternative to determine the most commonly occurring frequent itemset in a dataset, which is an improvement over the Apriori algorithm. FP-Growth addresses the limitations of Apriori by utilizing the concept of building a tree to search for frequent itemsets, resulting in a faster algorithm. The distinctive feature of FP-Growth is that it employs a tree data structure, known as the FP-Tree, which allows for the direct extraction of frequent itemsets from the FP-Tree[9].
The FP-Tree is a condensed storage structure constructed by linking each transaction data to a particular path in the tree. As the transactions may share the same item, the paths can overlap[10]. The effectiveness of the compression process with the FP-Tree data structure is directly proportional to the amount of transaction data that shares the same items. To mine frequent itemsets using the FP-Growth algorithm, a tree data structure (FP-Tree) is generated[11]. The FP-Growth method can be divided into three main stages as follows:

1. The first stage is the generation of the conditional pattern base, which is a sub database consisting of prefix paths and suffix patterns. This is achieved through the use of the previously constructed FP-Tree.
2. The second stage is the FP-Tree conditional generation, in which the support count of each item in every conditional pattern base is aggregated. Items with support count greater than the minimum support count ξ are then generated utilizing the conditional FP-Tree.
3. The final stage is the frequent itemset search. If the conditional FP-Tree is a single path, then frequent itemsets can be obtained by combining items for each conditional FP-Tree. If the conditional FP-Tree is not a single path, the FP-Growth is generated recursively.

2.4 Knowledge Discovery in Database (KDD)

Knowledge discovery in databases (KDD) refers to the process of uncovering potential, implicit, and previously unknown information from a given dataset. This process typically involves the outcomes of data mining, which is the procedure for discovering the inherent tendencies within a data pattern[12][13] The process of KDD involves the challenging task of identifying and detecting patterns in data that are both novel and useful, and subsequently translating these findings into clear and intelligible information. The patterns that are discovered through this process must be valid, informative, innovative, and comprehensible.
[14].

3. Findings

3.1 Data mining with the FP-Growth Algorithm

In this study, data mining applies association rules with the FP-Growth algorithm in analyzing spare part sales data at TB-Damar to obtain patterns of violations that occur. To extract spare part sales transaction data, the authors use an algorithm that forms the basis of other algorithms, namely frequent pattern growth (FP-Growth)[15]. The frequency of occurrence of each item from transaction data can be seen in the table below:

Table 1. Frequent pattern growth (FP-Growth)

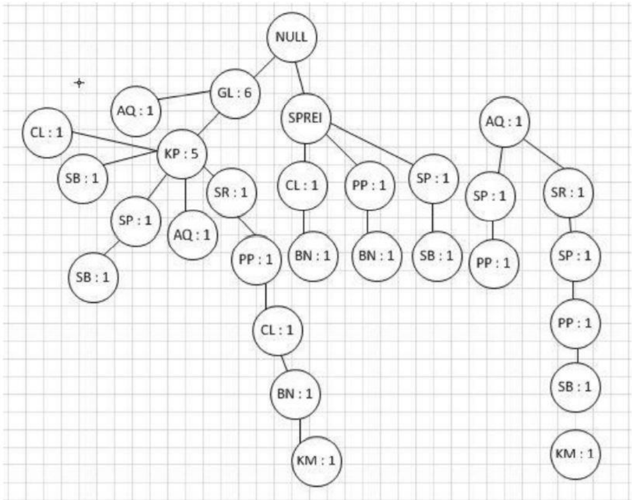
Transaction	Itemset	Frequency
1	SCREW, VALVE ADJUSTING (3C11)	2
2	OIL SEAL (45P1)	6
3	GRAPHIC 1 (UR YB 2006) SCORPIO-Z	5
4	GASKET, CYLINDER (KT) RXS	4
5	CABLE, CLUTCH (3C11)	6
6	BUSH (3C11)	3
7	BRAKE PAD KIT (54P2)	3
8	BATTERY ASSY (12V 3.5A) 3C	3
9	AXLE	

Table 2. Formation of FP-Tree

Transaction	Spare Part Sales Dataset
1	{ i, a, h }
2	{ a, h, g }
3	{ e, a, h }
4	{ f, g }
5	{ a, g, h }
6	{ d, b, a, h }
7	{ f, e, a }
8	{ e, f }
9	{ a, b }
10	{ c, g }

11	{ d, e, a }
12	{ a, d, b, g }
13	{ a, b, c }
14	{ e, h, i }
15	{ a, e, f }

From Table 2 above, the next step must be to form the FP-Tree path. Where the results of the formation of paths 1 to 15 can be seen in the following figure:



Picture 1. FP-Tree line

3.2 Application of FP-Growth

In order to utilize the FP-Growth algorithm for analyzing sales transactions of spare parts, it is necessary to execute a series of steps to identify frequent itemsets[16]. The first step involves constructing an FP-Tree, which can then be used to generate the Conditional Pattern Base, Conditional FP-Tree, and Frequent Itemset. Each of these three stages can be accomplished by referring back to the previously constructed FP-Tree. The creation of the FP-Tree yields the Conditional Pattern Base as the end result, as demonstrated below:

Table 3. Basic Conditional Pattern

Suffix	Basic Conditional Pattern
a	-
b	{(a:4)}
c	{(a,e:2)}
d	{(a,b:2)}
e	{ (a:3),(d:1) }
f	{ (a,e:2), (e:1)}
g	{(a,b,d:1), (a:2), (f:1), (c:1)}

 p-ISSN: 2686-6285
e-ISSN: 2715-0461

h	{ (e:1), (a,b,d:4), (a,e:1) }
i	{ (e,a:1), (a,h:1) }

To find the Conditional FP-Tree by adding up the existing support counts, a larger support count will be generated with the conditional FP-tree. The following table results in Conditional FP-Tree:

Table 4. Constrained FP-Tree

Constrained FP-Tree
-
{(a:4)}
{(a,e:2)}
{(a,b:2)}
{ (a:3),(d:1) }
{ (a,e:2), (e:1)}
{(a,b,d:1), (a:2), (f:1), (c:1)}
{ (e:1), (a,b,d:4), (a,e:1) }
{ (e,a:1), (a,h:1) }

After looking for the Conditional FP-tree, the next step is to find the Frequent Itemset. This stage searches for a single path and then combines it with the items in the Conditional FP-Tree[17]. The following table is the result of the Frequent Itemset:

Table 5. Frequently Used Items

Frequent Itemset
-
(a,b):4
(a,c):2,(b,c):2
(a,d):2,(b,d):2
(a,e):3
(a,f):2,(e,f):3
(a,g):3
(e,h):2,(a,h):5,(b,h):4,(d,h):4
(h,i:2)

p-ISSN: 2686-6285
e-ISSN: 2715-0461

If the minimum support is 40% and the minimum confidence is 60%, then the strong association rule can be seen in the following table:

Table 6. Rule of Strong Association

Item B	Item A	Confidence	Support
Cable clutch	Axle, gasket cylinder	100%	13.33%
Axle	Oil seal, graphic 1	100%	13.33%
Axle	Battery assy	100%	20%
Gasket cylinder	Cable clutch, gasket cylinder	67%	13.33%
Oil seal	Axle, cable cluth	67%	13.33%
Axle	Bush	67%	13.33%
Axle	Brake	67%	13.33%
Cable cluth	Gasket cylinder	75%	20%
Axle	Graphic 1	60%	20%
Axle	Oil seal	83.3%	33.3%
Oil seal	Screw valve adjusting	100%	13.33%

3.3 Implementation and Testing

Implementation of Data Mining in analyzing spare part sales transactions uses association rules with the FP-Growth algorithm[18]. The FPGrowth algorithm, as the main process of association rules, begins with forming itemset candidates, forming FP-Tree, and then determining Conditional Base Patterns, Conditional FP-Tree and Frequent Itemsets. Before testing the data, what was done was to prepare 150 spare art sales transaction data and then input the data into Microsoft Excel[13].

The Association Rules (Text View) results also explain that items a and b are related by explaining them in text. The calculation results above are different from the sample data calculations because the amount of data processed is 150, which is more than the existing data sample, which is 15.

4. Conclusion

The FP-Growth algorithm can determine the sales results of the most widely sold sport motorbike spare parts at Sarana Motor Tangerang. Spare parts that meet minimum support and confidence and are widely sold are screw valve adjusting, oil seals, battery assy, axle, cylinder gasket, and cable clutch. Applying the Data Mining method with the FP-Growth algorithm into an application for analyzing consumer purchasing patterns is very beneficial for the company because TB-Damar will know which spare parts are purchased the most and assist in ordering spare parts at the head office. Implementing the FP-Growth algorithm on the RapidMiner application starts with inputting sales data which will become a database in Ms Excel. The data will be processed using the RapidMiner application, forming support and confidence and producing a strict rule.

From the results of this study, the authors can provide several suggestions that can be considered for research in further development. In implementing this data, the larger the data taken, the more accurate and larger results will be obtained. If the development of this research is carried out, it is better to add samples as well. One of the techniques used in this research is RapidMiner. Besides that, you can also use other data mining applications such as Tanagra or WEKA. The distribution of spare parts must be by brand and type so that consumers are clearer and more accurate in ordering or selling motorbike spare parts. This is useful to make it easier for consumers to purchase and admin in recording sales transaction data.

References

[1] G. N. B. Safrizal and G. N. Budiadyana, “Analysis Application Design Career Development Center In The STMIK Insan Pembangunan and (Case Study: Information Study Program),” *IAIC Trans. Sustain. Digit. Innov.*, vol. 1, no. 1, pp. 66–77, 2019.

[2] J. Yang *et al.*, “Brief introduction of medical database and data mining technology in big data era,” *J. Evidence-Based Med.*, vol. 13, no. 1, pp. 57–69, 2020.

[3] A. Dogan and D. Birant, “Machine learning and data mining in manufacturing,” *Expert Syst. Appl.*, vol. 166, p. 114060, 2021.

[4] S. Wang, J. Cao, and P. Yu, “Deep learning for spatio-temporal data mining: A survey,” *IEEE Trans. Knowl. Data Eng.*, 2020.

[5] L. I. Khalaf, S. A. Aswad, S. R. Ahmed, B. Makki, and M. R. Ahmed, “Survey On Recognition Hand Gesture By Using Data Mining Algorithms,” in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–4.

[6] M. S. Amin, Y. K. Chiam, and K. D. Varathan, “Identification of significant features and data mining techniques in predicting heart disease,” *Telemat. Informatics*, vol. 36, pp. 82–93, 2019.

[7] D. Ahmad, N. Lutfiani, A. D. A. R. Ahmad, U. Rahardja, and Q. Aini, “Blockchain technology immutability framework design in e-government,” *J. Adm. Publik (Public Adm. Journal)*, vol. 11, no. 1, pp. 32–41, 2021.

[8] M. Yusup, P. A. Sunarya, N. Lutfiani, and E. A. Nabila, “A Blockchain-Based Framework Gamification for Securing Learners Activity in Merdeka Belajar-Kampus Merdeka,” in *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 2022, pp. 1–6.

[9] E.-H. A. Rady and A. S. Anwar, “Prediction of kidney disease stages using data mining algorithms,” *Informatics Med. Unlocked*, vol. 15, p. 100178, 2019.

[10] M. A. Ledhem, “Data mining techniques for predicting the financial performance of Islamic banking in Indonesia,” *J. Model. Manag.*, vol. 17, no. 3, pp. 896–915, 2022.

[11] L. Rutkowski, M. Jaworski, and P. Duda, *Stream data mining: algorithms and their probabilistic properties*. Springer, 2020.

[12] G. Ramaswami, T. Susnjak, A. Mathrani, J. Lim, and P. Garcia, “Using educational data mining techniques to increase the prediction accuracy of student academic performance,” *Inf. Learn. Sci.*, vol. 120, no. 7/8, pp. 451–467, 2019.

[13] Sudaryono, U. Rahardja, and N. Lutfiani, “The Strategy of Improving Project Management Using Indicator Measurement Factor Analysis (IMF) Method,” in *Journal of Physics: Conference Series*, 2020, vol. 1477, no. 3, doi: 10.1088/1742-6596/1477/3/032023.

[14] M. E. Lokanan, “Data mining for statistical analysis of money laundering transactions,” *J. Money Laund. Control*, vol. 22, no. 4, pp. 753–763, 2019.

[15] I. Amsyar, E. Christopher, A. Dithi, A. N. Khan, and S. Maulana, “The Challenge of Cryptocurrency in the Era of the Digital Revolution: A Review of Systematic Literature,” *Aptisi Trans. Technopreneursh.*, vol. 2, no. 2, pp. 153–159, 2020.

[16] S. Dewan and L. Singh, “Use of blockchain in designing smart city,” *Smart Sustain. built Environ.*, 2020.

[17] H. Mansour, “How successful countries are in promoting digital transactions during COVID-19,” *J. Econ. Stud.*, vol. 49, no. 3, pp. 435–452, 2022.

[18] D. Gohil and S. V. Thakker, “Blockchain-integrated technologies for solving supply

chain challenges," *Mod. Supply Chain Res. Appl.*, vol. 3, no. 2, pp. 78–97, 2021.