

DATA MINING IMPLEMENTATION WITH ASSOCIATION METHOD AND APRIORI ALGORITHM FOR STORE DISPLAY DESIGN IN HOME CENTER INDONESIA

DINA IKRAMINA SETIANI, FAUZAN NASAFI, TUGA MAURITSIUS
EMIL ROBERT KABURUAN AND RIYANTO JAYADI

Information Systems Management Department, BINUS Graduate Program – Master of
Information Systems Management
Bina Nusantara University

Jl. K. H. Syahdan No. 9, Kemanggisian, Palmerah, Jakarta 11480, Indonesia
{ dina.setiani; fauzan.nasafi }@binus.ac.id; { tmauritsus; emil.kaburuan; riyanto.jayadi }@binus.edu

Received April 2021; accepted June 2021

ABSTRACT. *Visual merchandising is a very important tool and has an impact on customer desires in shopping. Data mining is a new method to support the process of analyzing organizational decisions using data in the organization. This research uses Apriori algorithm in processing data mining to be made into information. Data mining has various functions, one of which is association, which is a function used to determine the relationship between existing items. This research explains the data mining design model and the results of this research indicate that data mining can provide solutions to problems that arise related to the appearance of stores or visual merchandising in Informa based on a combination of sales data and by utilizing data mining technology.*

Keywords: Data mining, Association, Apriori algorithm, Visual merchandising, Store display

1. **Introduction.** The presence of shopping centers is one of the most dynamic businesses, especially businesses in the store-based retail sector. This has resulted in the creation of a competitive business climate. In order to survive, the retail business must be able to anticipate market changes and must be responsive in adjusting it so that it is always in accordance with the people's lifestyle. Informa is a retail store that sells furniture products in Indonesia. Informa sells various types and types of furniture, ranging from furniture for residential, office, indoor to outdoor furniture. Informa itself categorizes products sold as household appliances, commercial, electronics, accessories, furniture sets, tables & chairs, mattresses & beds, storage and sofas.

Visual merchandising is a very important way of communicating to customers in order to get the attention of these customers [1]. Visual merchandising has several main elements, such as window display, store design, color & lighting and store layout [1]. Visual merchandising is a very important tool and has an effect on customer desires in shopping [1]. At the merchandise display level, luxury products are often staged in a manner that reflects the iconic aesthetic [2]. And with a good store layout can increase the likelihood that customers will be in the store with a longer period of time [3]. Stress the importance of product staging techniques as a contemporary challenge for retailers wishing to capture an edge in making merchandise standout in-store [2]. Several researchers have demonstrated the influence of visual merchandising on consumers' perceptions and/or purchase intentions [4]. Customer behavior is learning behavior between cluster or organization which have associated with product to respond to the customer and social requirements [5]. Market basket analysis is one of the most popular areas in the field of data-mining,

which is used to analyze customer behaviors with respect to the purchase of commodities [6]. Data mining applied to the field of commercialization allows, among other aspects, to discover patterns of behavior in clients, which companies can use to create marketing strategies addressed to their different types of clients [7]. Data mining has a variety of functions namely association, classification, clustering, forecasting, regression, sequence discovery and visualization [8]. One of them is association, which is a function that is used to determine the relationship between existing items [9]. The four main objectives of the organization that reflect the marketing concept are [10]: orientation to customers in all aspects of the organization and operations, the pursuit of profitable sales volumes in a scientific manner, business integration towards common goals, social responsibility in decision making. Marketing is a social and managerial process where individuals and groups get their needs and desires by creating, offering and exchanging products and values with other individuals and groups [11].

This paper proposes a data mining design model, namely the implementation of data mining with the association method and the Apriori algorithm for shop display design at Indonesian home centers. With this research, several product recommendations will be made with the right layout so that they can increase sales.

In Section 1 we present the motivation of this research followed by discussing some theoretical background that guides the research in Section 2. Section 3 discusses the methodology implemented in this research. In Section 4 we present the results of the research and conclude the paper in Section 5.

2. Theoretical Background.

2.1. Data mining techniques. In the retail world, data mining can extract predictive information hidden from large databases by identifying valuable customers, predicting customer behavior in the future, and making knowledge-based decisions, which move the company to go beyond analyzing past events and allowing companies to be oriented in future. The choice of data mining techniques must be based on data characteristics and business needs [11]. Data mining can be seen as a business movement process that is focused on discovery and continues to use beneficial knowledge from organizational data [12]. This can be used as a guide in making decisions and forecasting effects and decisions. For example, data mining can increase the speed of response of marketing campaigns by segmenting customers into groups with different characteristics and needs. That can predict how it is possible for an existing customer to bring his business to competitors [12].

2.2. CRISP-DM. CRISP-DM, which stands for cross-industry standard process for data mining, is a method that has proven to guide the data mining works [7]. To build the mining models, it uses the CRISP-DM methodology. CRISP-DM methodology consists of six major phases [13]: business understanding, data understanding, data preparation, modelling, evaluation, deployment.

2.3. Association & Apriori. Association aims to establish the relationship between items that exist together in a given note [12]. Association rules are one of the essential data and web mining techniques which aim to discover the frequent patterns among itemsets in a transactional database [14]. The Apriori algorithm is a method to discover sets of frequent elements and generates association rules on a set of transaction data. There are some association level metrics, the most popular ones are support, confidence and lift.

It first identifies the frequent individual elements through the transactions and then extends to the increasingly large element sets until the resulting element sets reach a

specified frequency threshold [7]. The Apriori algorithm is usually implemented in investigating the purchases of a store and extracted their association rules [14,17].

2.4. Market basket analysis. Market basket analysis technique is one data mining technique that aims to find products often purchased together from transaction data, here very necessary to determine the rules association. Association rules are a pattern the tendency of buyers to shop goods. Basket size patterns allow marketers to identify light and heavy shoppers in-store (on a given day), in the same way that purchase frequencies identify light and heavy brand and category buyers [15].

3. Research Methodology and Tools. The data mining technique used in this study is the association, also known as market basket analysis. A typical business process is analyzing the sales transaction table by identifying products that are often purchased together by the customer. Existing similarities from purchase data are used to identify groups of similarities from products and habits that occur in the interests of cross-selling. In terms of association, each item is considered information.

The research is organized into 3 steps.

The first step in this research is to collect the data. The data used are data sales of information products generated from daily sales, and then recapitulated every month by the Inventory Planner team with SAP BW (Business Warehouse).

The process of withdrawing data using SAP BW is quite easy and fast, because the tools provided in this module are very complete. Simply drag and drop which is then executed and exported to Excel.

The second step is to use Jupiter Notebook to conduct data preprocessing stage. In this stage, we do the followings:

- 1) Remove leading and trailing whitespace from description column
- 2) Drop rows that do not have invoice numbers
- 3) Create shopping baskets

The third step uses KNIME to conduct the association rule. KNIME is a data analysis, reporting and integration platform including free software. KNIME is used to combine various components and explore information on data mining and machine learning. It can also integrate various data sources, including the pre-processing (ETL: Extract, Transform, Load) stage, the modeling stage, data analysis and visualization without or with programming.

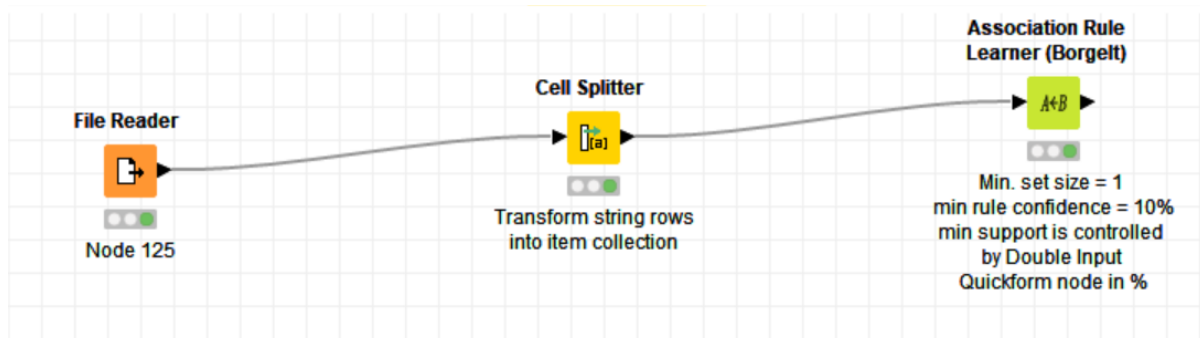


FIGURE 1. Work flow of market basket analysis building association rule

Figures 1 shows the processing flow using KNIME. It contains 3 stages represented by nodes. The followings are detailed information corresponding to each node.

- File Reader: This node is used to read data from ASCII files or URL paths that have been selected previously. Here we use the Data Sales Semester 2 2019 CTC Dept PT Home Center Indonesia.csv file stored in the knime-workspace/informa association directory.

- Cell Splitter: This node uses predefined delimiter characters to divide the contents of the selected column into sections. Inside this node contains a collection (list or set) of cells with split output from Data Sales Semester 2 2019 CTC Dept PT Home Center Indonesia.csv. The resulting output consists of two columns, namely Row ID and Col0_SplitResultSet.
- Association Learner (Borgelt): The node uses the Apriori algorithm applied by Christian Borgelt. Induction of association and data mining rules is a powerful method for market basket analysis. So its basically logic Apriori and association rule techniques are a technical pair for achieving the level of data accuracy height and data processing techniques arranged systematically and can be mapped computationally algorithmically [16]. The parameters set in this node are
 - Minimum set size: 2
 - Minimum support: 2%
 - Minimum rule confidence: 1%

4. Results and Discussions.

4.1. **Data acquisition and preprocessing.** Sales data is the main input in this research. Due to some limitations, we only collect data from CTC Department during the Semester 2 2019 sales. Table 1 presents the structure of the data.

TABLE 1. Informa sales data

No.	Data name	Data type	Detail
1	AH Dept	Char	Sales department name on Informa
2	AH Commodity	Char	Commodity code based on its designation
3	AH Commodity Desc	Char	Description of the commodity code based on its designation (Commercial/Outdoor/etc.)
4	AH Class	Char	Commodity code based on type
5	AH Class Desc	Char	Description of the commodity code based on type (Table/Chair/Stool/etc.)
6	AH Prd Grp	Char	Commodity group code
7	AH Prd Grp Desc	Char	Descriptions of commodity groups (Basi Chair/Dining Table/Chair/Folding Table/etc.)
8	Article Type	Char	Commodity type code
9	Company	Char	Informa's branch name
10	Company_Code	Numeric	Informa's branch code
11	Karyawan	Char	Employee name & ID
12	Article	Numeric	The article code of the commodity
13	Description	Char	Description of the article code
14	Quantity	Numeric	Number of articles sold
15	Amount	Numeric	Total sales price
16	Transdate	Numeric	Total transaction
17	Receive.No	Numeric	Transaction code
18	Dept	Char	Department code
19	Month	Numeric	Transaction month code
20	Type Transaction	Char	Type of transaction (Cash and Carry/Delivery)

4.2. **Association rule implementation results.** Following are the outputs from the KNIME processing tools. Workflow in the Data Sales Semester 2 2019 CTC Dept PT Home Center Indonesia.csv file consists of 3 nodes, namely Files Reader, Cell Splitter and Apriori algoritihm (setting using KNIME).

▲ Association Rules - 2:110 - Association Rule Learner (Borgelt) (Min. set size = 1)

File Hilite Navigation View

Table 'default' - Rows: 37 Spec - Columns: 11 Properties Flow Variables

Row ID	[S] Consequent	[L] Antecedent	[I] ItemSe...	[D] Relativ...	[D] RuleCo...	[D] Absolut...	[D] Relativ...	[D] RuleLift	[D] RuleLift%	[D] Absolut...	[D] Relativ...
Row0	X032646 TRINITY 2 TABLE &	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	46.943	4,694.3	1	0.03
Row1	KIEL TABLE TOP 120X60 CM	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	46.943	4,694.3	1	0.03
Row2	DOBSON TABLE 80X80 CM ...	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	23.471	2,347.1	2	0.061
Row3	ROMA-3 TABLE BASE REC...	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	23.471	2,347.1	2	0.061
Row4	DIRMY TABLE DIA 60CM W...	[DACEY 2 CHAIR WHITE]	3	0.091	1.46	205	6.24	12.022	1,202.2	4	0.122
Row5	SAGA CERAMIC TABLE GREY	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	7.824	782.38	6	0.183
Row6	LONDON Z BARCHAIR RED	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	7.824	782.38	6	0.183
Row7	HADANO 2 END TABLE BR...	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	6.706	670.61	7	0.213
Row8	MARVEL BAR STOOL BLACK	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	3.611	361.1	10	0.304
Row9	BATILDA CHAIR BROWN	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	5.216	521.59	9	0.274
Row10	TATILDA BASE Z-018 2 DIA 43	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	5.868	586.79	8	0.243
Row11	MAY BAR STOOL FROSTED...	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	3.009	300.92	12	0.365
Row12	HECTOR BAR STOOL BROWN	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	4.694	469.43	10	0.304
Row13	LONDON Z BARCHAIR BLACK	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	3.912	391.19	12	0.365
Row14	TITANS STOOL WHITE	[TITANS STOOL BLACK]	3	0.091	3.3	91	2.77	7.738	773.78	14	0.426
Row15	LOBITO STOOL STAINLESS...	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	2.257	225.69	16	0.487
Row16	MELINDA2 BARSTOOL BLA...	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	1.641	164.14	22	0.67
Row17	LONDON ELM STOOL IVORY	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	2.257	225.69	16	0.487
Row18	TIMMY STOOL BLACK	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	2.608	260.79	18	0.548
Row19	DERWOOD 2 CHAIR WHITE	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	1.878	187.77	25	0.761
Row20	DACEY 2 CHAIR RED	[DACEY 2 CHAIR BLACK]	2	0.061	2.86	70	2.13	3.477	347.72	27	0.822
Row21	TITANS STOOL YELLOW	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	2.778	277.77	13	0.396
Row22	GREATWALL CHAIR BLACK	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	0.976	97.594	37	1.126
Row23	DACEY 2 CHAIR MILD GREY	[DACEY 2 CHAIR BLACK]	2	0.061	2.86	70	2.13	2.183	218.34	43	1.309
Row24	DACEY 2 CHAIR MILD GREY	[DACEY 2 CHAIR WHITE]	3	0.091	1.46	205	6.24	1.118	111.83	43	1.309
Row25	SIMPSON STOOL BLACK	[DACEY 2 CHAIR BLACK]	1	0.03	1.43	70	2.13	1.067	106.69	44	1.339
Row26	TROY BAR STOOL BLACK	[TITANS STOOL BLACK]	1	0.03	1.1	91	2.77	0.681	68.132	53	1.613
Row27	ALBA CHAIR ALUMINIUM ...	[YOLORADO TABLE + CHAIR BL]	2	0.061	1.9	105	3.2	1.026	102.61	61	1.856
Row28	TITANS STOOL BROWN	[TITANS STOOL BLACK]	7	0.213	7.69	91	2.77	5.878	587.84	43	1.309
Row29	TITANS STOOL BROWN	[X151875 COLORADO TBL +CHAI]	2	0.061	1.44	139	4.23	1.099	109.95	43	1.309
Row30	DACEY 2 CHAIR WHITE	[DACEY 2 CHAIR BLACK]	6	0.183	8.57	70	2.13	1.374	137.39	205	6.239
Row31	DACEY 2 CHAIR BLACK	[DACEY 2 CHAIR WHITE]	6	0.183	2.93	205	6.24	1.374	137.39	70	2.13
Row32	TITANS STOOL LIME GREEN	[TITANS STOOL BLACK]	6	0.183	6.59	91	2.77	5.702	570.16	38	1.156
Row33	TITANS STOOL RED	[TITANS STOOL BLACK]	11	0.335	12.1	91	2.77	9.027	902.75	44	1.339
Row34	TITANS STOOL ORANGE	[TITANS STOOL BLACK]	6	0.183	6.59	91	2.77	6.565	656.54	33	1.004
Row35	TITANS STOOL ORANGE	[X151875 COLORADO TBL +CHAI]	2	0.061	1.44	139	4.23	1.433	143.27	33	1.004
Row36	TITANS STOOL TURQUOISE	[TITANS STOOL BLACK]	9	0.274	9.89	91	2.77	8.552	855.23	38	1.156

FIGURE 2. Result of association rule

Figure 2 presents the result of association rule using Apriori algorithm by applying the above parameters setting.

The screen capture shows 37 rules that meet the above-mentioned parameter setting. It shows, for example from Row0, if someone buys Dacey 2 Chair Black (call it item A) then he or she will likely buy X032646 Trinity 2 Table (call it item B) with confidence 1.43%. The number of 1.43 can be found in column 6. This says that more than 1.43% of time a buyer of item A will also buy item B. The biggest confidence from the figure can be found in Row33 with confidence of 12.1%. This means much of the time (12.1%) if someone buys Titans Stool Black, then she or he will also buy Titans Stool Red.

The strength of the association can also be seen from column 9 (Rule Lift). The output of this process can be used as a reference in finding out which products that usually are bought together. This can be used as a reference in determining the display position of goods in the showroom to increase customers shopping experience.

5. Conclusions. Based on the data mining implementation design by applying the Apriori algorithm and the association method to product sales data, it generates patterns that can help companies determine product sales strategies. The results are as follows.

- a) Recommend the right product selection for display in the store based on a combination of sales data.
- b) Recommend the right products to customers so that they can increase product sales.
- c) Provide a better shopping experience to customers.

With the results of this research, it is hoped that Informa, especially the CTC division, can apply the concept of display based on products that are of interest to customers. So it can help increase sales. In the future, one can do more detailed research on all product categories in Informa, so that not only the CTC division will get the benefits but all divisions in Informa.

Acknowledgment. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] Y. A. Soomro, S. A. Kaimkhani and J. Iqbal, Effect of visual merchandising elements of retail store on consumer attention, *Journal of Business Strategies*, vol.11, no.1, pp.21-40, DOI: 10.29270/JB-S.11.1(17).002, 2017.
- [2] M. Logkizidou, P. Bottomley, R. Angell and H. Evanschitzky, Why museological merchandise displays enhance luxury product evaluations: An extended art infusion effect, *J. Retail.*, vol.95, no.1, pp.67-82, DOI: 10.1016/j.jretai.2018.11.001, 2019.
- [3] M. Garaus, U. Wagner and C. Kummer, Cognitive fit, retail shopper confusion, and shopping value: Empirical investigation, *Journal of Business Research*, vol.68, no.5, pp.1003-1011, DOI: 10.1016/j.jbusres.2014.10.002, 2015.
- [4] M. R. Kpossa and E. Lick, Visual merchandising of pastries in foodscapes: The influence of plate colours on consumers' flavour expectations and perceptions, *J. Retail. Consum. Serv.*, vol.52, no.2, DOI: 10.1016/j.jretconser.2018.10.001, 2020.
- [5] P. Ponyiam and S. Arch-Int, Customer behavior analysis using data mining techniques, *Proc. of 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life (iSemantic2018)*, pp.549-554, DOI: 10.1109/ISEMANTIC.2018.8549803, 2018.
- [6] H. Li, Y. J. Wu and Y. Chen, Time is money: Dynamic-model-based time series data-mining for correlation analysis of commodity sales Hailin, *J. Comput. Appl. Math.*, DOI: 10.1016/j.cam.2019.112659, 2019.
- [7] J. Silva, N. Varela, L. A. B. López and R. H. R. Millán, Association rules extraction for customer segmentation in the SMES sector using the Apriori algorithm, *Procedia Comput. Sci.*, vol.151, pp.1207-1212, DOI: 10.1016/j.procs.2019.04.173, 2019.
- [8] E. W. T. Ngai, L. Xiu and D. C. K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, *Expert Syst. Appl.*, 2009.
- [9] W. H. Au and K. C. C. Chan, Mining fuzzy association rules in a bank-account database, *IEEE Trans. Fuzzy Syst.*, 2003.
- [10] M. A. Pasigai, Pentingnya konsep dan strategi pemasaran dalam menghadapi persaingan bisnis (The importance of marketing concepts and strategies in dealing with business competition), *Jurnal Ekonomi Balance*, vol.6, no.2, pp.51-56, 2010.
- [11] D. T. Anggraeni and I. K. Nurcahya, Pengaruh retail mix terhadap keputusan pembelian pada bali cenik kids apparel di denpasar (The effect of retail mix on purchase decisions at bali cenik kids apparel in denpasar), Ph.D. Thesis, Udayana University, 2014.
- [12] M. M. Maghfirah, T. B. Adji and N. A. Setiawan, Appropriate data mining techniques and algorithms for using in analysis of customer relationship management (CRM) in bank industry, *Seminar Nasional Aplikasi Teknologi Informasi (SNATi2015)*, Yogyakarta, 2015.
- [13] S. Palaniappan and R. Awang, Intelligent heart disease prediction system using data mining techniques, *2008 IEEE/ACS International Conference on Computer Systems and Applications*, pp.108-115, DOI: 10.1109/AICCSA.2008.4493524, 2008.
- [14] N. Karimtabar and M. J. S. Fard, An extension of the Apriori algorithm for finding frequent items, *2020 6th Int. Conf. Web Res. (ICWR2020)*, pp.330-334, DOI: 10.1109/ICWR49608.2020.9122282, 2020.
- [15] J. Martin et al., Fundamental basket size patterns and their relation to retailer performance, *J. Retail. Consum. Serv.*, vol.54, no.9, DOI: 10.1016/j.jretconser.2020.102032, 2020.
- [16] W. Ningsih, Implementasi data mining untuk penentuan posisi barang pada rak menggunakan metode Apriori pada PT Midi Utama Indonesia, (Data mining implementation for positioning merchandise on shelves using the Apriori method at PT Midi Utama Indonesia), *Techno. Com.*, vol.15, no.4, pp.335-342, 2016.
- [17] T. M. Hossain, J. Watada, Z. Jian, H. Sakai, S. Rahman and I. A. Aziz, Missing well log data handling in complex lithology prediction: An NIS Apriori algorithm approach, *International Journal of Innovative Computing, Information and Control*, vol.16, no.3, pp.1077-1091, 2020.