Original Article

Modified market basket analysis to identify consumer behavior and promotion programs for supporting shopping tourism

Fajar Stevano Artha (Department of Industrial Engineering, Universitas Atma Jaya Yogyakarta, 225612163@students.uajy.ac.id, Indonesia)
Ririn Diar Astanti (Department of Industrial Engineering, Universitas Atma Jaya Yogyakarta, ririn.astanti@uajy.ac.id, Indonesia)

Abstract

In a retail store located in a tourist destination, tourists can be one of the potential consumers who should be concentrated on. The existence of the retail can support shopping tourism. This research develops a modified market basket analysis (MBA) framework to analyze the relationship between promotional programs and customer behavior in the retail sector. MBA is performed with the apriori algorithm to reveal the relationship between product and product, product and promotion, promotion and customer member, product and customer member. With this analysis, it is hoped that retailers can determine their promotional strategies to retain current consumers and attract new consumers, including tourists. Modified MBA is carried out by adding promotion variables to the point of sales (POS) data that will be analyzed. The framework was tested on a case study of a fashion retail store in the city of Yogyakarta, which is one of the tourism destinations in Indonesia.

Keywords

retail, market basket analysis, customer behaviour, promotional program, shopping tourism

1. Introduction

Retail is a sector that has a large contribution to the economy [Thangavelu, 2019]. Apart from that, retail also plays a role in distributing goods or services from producers to final consumers. With the development of information technology, retailers use computer-based information systems to carry out activities in their business processes, including recording sales transaction. This information system is usually called POS (point of sales). According to Yoseph et al. [2020], retail collects huge volumes of POS data. The data collected by POS is not yet useful for decision making, if it has not been processed. After the data is processed, the data becomes information that can be used to support decision making in retail.

One of the decisions made in retail is determining promotional programs to attract new potential consumers or retain current consumers. One potential type of consumer is tourists, especially if a retailer is in an area that is a tourist destination. According to Geuens et al. [2004]; Correia and Kozak [2016], shopping is one of the activities carried out by a tourist. It creates what is called shopping tourism. Shopping tourism is defined as, "expenditures are spent on goods except for food, drinks and grocery from visitors to a certain area who are not local residents" [Ambagtsheer, 2020]. Shopping tourism contributes to the economy [Timothy, 1995; Leal et al., 2010; Choi et al., 2016, Yiu, 2023]. Therefore, retail in tourist destination areas needs to make efforts to support shopping tourism, for example by carrying out accurate demand forecasts [Silva and Hassani, 2022]. Apart from that, retailers must also understand customer behaviour. This is because most retail purchases are made on impulse purchase [Jeffrey and Hodge, 2007].

Therefore, retailers must have knowledge of what customers might buy if the idea comes into their mind. Several researchers conducted research on market basket analysis (MBA) to analyse consumer behaviour [Hossain et al., 2019; Kavitha and Subbaiah, 2020; Made et al., 2022; Qisman et al., 2021; Ünvan, 2021]. However, that research has not discussed about the relationship between consumer behaviour and promotional programs.

Promotions need to be peformed because, as mentioned above, most retail purchases are made on impulse purchase. Research on promotions in retail was carried out by Kanta [2019] who examined the influence of promotions during the regular season, festive season, and when competitors also carried out promotions. Epstein et al. [2021] emphasizes the influence of promotions on increasing traffic. Sun et al. [2018] discusses the effects of promotions on sales and profits. Khairawati [2019] conducted research to evaluate customer loyalty. Khouja et al. [2020] compared the performance of 5 types of promotions but from the perspective of consumer surplus, average price, utility, and customer supply.

The research presented in this paper attempts to fill the research gap related to shopping tourism and promotional programs in retail which have not been studied by previous research. The proposed framework is developed by using an MBA-based framework that includes promotion types at the data pre-processing stage to obtain customer promotional behaviour that can support the design of promotional programs in retail.

2. Literature review

Research on shopping tourism and retail was carried out by several researchers, for example, Silva et al. [2019]; Sharma et al. [2018]; Piron [2002]; Baruca and Zolfagharian [2012]; Lau et al. [2005]; Nguyen et al. [2016]; Makkonen [2016]; Hadjimarcou et al. [2017]; Boonchai and Freathy [2020]; Lee and Choi [2019] which discusses cross-border shopping tourism. Li et al. [2018]

and Liu et al. [2020] studied about the relationship between retail's prices and shopping tourism. Frago [2021] studied about the COVID-19 pandemic and shopping tourism. Research on retail and shopping tourism was also carried out in connection with demand forecasting in retail and shopping tourism as carried out by Silva and Hassani [2022]. Research on demand forecasting in retail generally examines demand forecasting to support sales in retail. Apart from carrying out demand forecasting to estimate sales, retailers also need to carry out promotional activities to support the achievement of their sales targets.

From previous research it is known that promotions contribute to company revenue and profits, as stated by Epstein et al. [2021]; Khouja et al. [2020]; Phumchusri et al. [2023]. Kanta [2019] where promotions increased profits by around 12-15 % which varied across various product categories. Promotions are also one of the key factors in retail operations [Rooderkerk et al., 2022; Khouja et al., 2020; Phumchusri et al., 2023].

Kim et al. [2019] in their research found a positive relationship between promotions and customer satisfaction. Using the obtained data from point of sales (POS), retail can analyse it to identify about the customer behaviours The technique that is often used is market basket analysis (MBA) [Qisman et al., 2021]. MBA uses data mining to identify recurring relationships between product groups, individual items, or categories. Therefore, MBA can be used in determining customer purchasing behaviour [Ünvan, 2021; Ozgormus and Smith, 2020].

Beside MBA, based on detailed transaction data, it is possible for companies to gain more insight into customer behaviour and preferences by classifying them into meaningful groups to meet their needs more efficiently [Shen, 2021]. Research by Silva et al. [2019] focuses on customer segmentation to allow organizations to understand their clients. Guney et al. [2020]; Moodley et al. [2020]; Shen [2021] carried out segmentation based on their customer profiles, then followed by MBA analysis to identify the customer purchasing behaviour of each segment. Apriori algorithm was introduced by Agrawal [1994] for association rule mining (ARM) based data mining because it is efficient and robust [Reutterer et al., 2017; Zhi-guo et al., 2018].

Apriori algorithm is an analytical method used in data mining to find association rules in transaction data. From previous research it is known that apriori (association rule algorithm) is generally used to determine customer behaviour in MBA [Hossain et al., 2019; Kavitha and Subbaiah, 2020; Made et al., 2022; Qisman et al., 2021; Ünvan, 2021; Verma et al., 2020].

Research on promotional analysis was conducted by Kanta [2019] using OLS regression, autoregressive-distributed lag (ADL) model [Phumchusri et al., 2023], simulation [Khouja et al., 2020], partially profiled least absolute shrinkage and selection operator (partially profiled LASSO) model [Sun et al., 2018], structural equation modeling (SEM) [Khairawati, 2019], gradient boosting (XGBoost) algorithm [Henzel and Sikora, 2020] and conditional on the counterfactual hypothesis baseline prediction (CCFH) [Epstein et al., 2021].

Based on the literature conducted regarding shopping tourism and retail, there has been no research that discusses consumer behaviour to designing promotional programs in retail that support shopping tourism especially using MBA as a basis technique.

3. Proposed framework

In this section, the modified MBA-based framework is presented in Figure 1.

3.1 Step 1: Data collection stage and determination of attribute selection

As stated by Dekimpe [2020], retail can be defined as a big data industry, where big data in retail can be obtained from point of sales (POS). Therefore, in the proposed framework, POS is used as one of the data sources.

If from the POS data, the promotional program for each item cannot be found, it is necessary to find that data from another data source, such as merchandising data. Merchandising data is data that is usually owned by the department responsible for procuring goods and negotiating promotional programs with suppliers. Therefore, description of promotional data is usually provided in the merchandising data.

If from merchandising data, a promotional description of an item cannot be obtained, then data regarding the promotional program for each item can be seen from the merchandise journal. The merchandise journal is a list of promotional programs sent by the merchandising department to each store which contains program information for each brand that will be running.

If the promotional program of each item has been obtained, then the next step is attribute selection. In the proposed framework, preparing the data about description of promotional program for each item to be included in analysis of consumer behaviour, is the contribution of the research presented in the paper. In MBA analysis, usually to analyse consumer behaviour, transaction ID and purchased item attributes per transaction ID are used. Meanwhile, in the framework proposed in this research, apart from transaction ID and purchased items per transaction, the data about promotional programs for each item and the data about the membership status of consumers is used as the attribute to analyse consumer behaviour. After the attribute selection, data that matches the selected attribute is then collected to enter Step 2.

3.2 Step 2: Market basket analysis using apriori-association rule

At this stage, the data that has been collected in Step 1 according to the selected attributes, is then analyzed using MBA. If the data that has been obtained still contains missing data, or the format is not appropriate, data pre-processing can be carried out. Data pre-processing refers to a set of techniques to improve the quality of raw data, such as removing outliers and handling missing values [Fan et al., 2021]. After passing the data pre-processing stage, MBA is performed by utilizing apriori-association rules. As mentioned in the previous section MBA technique was originally found by Agrawal [1994].

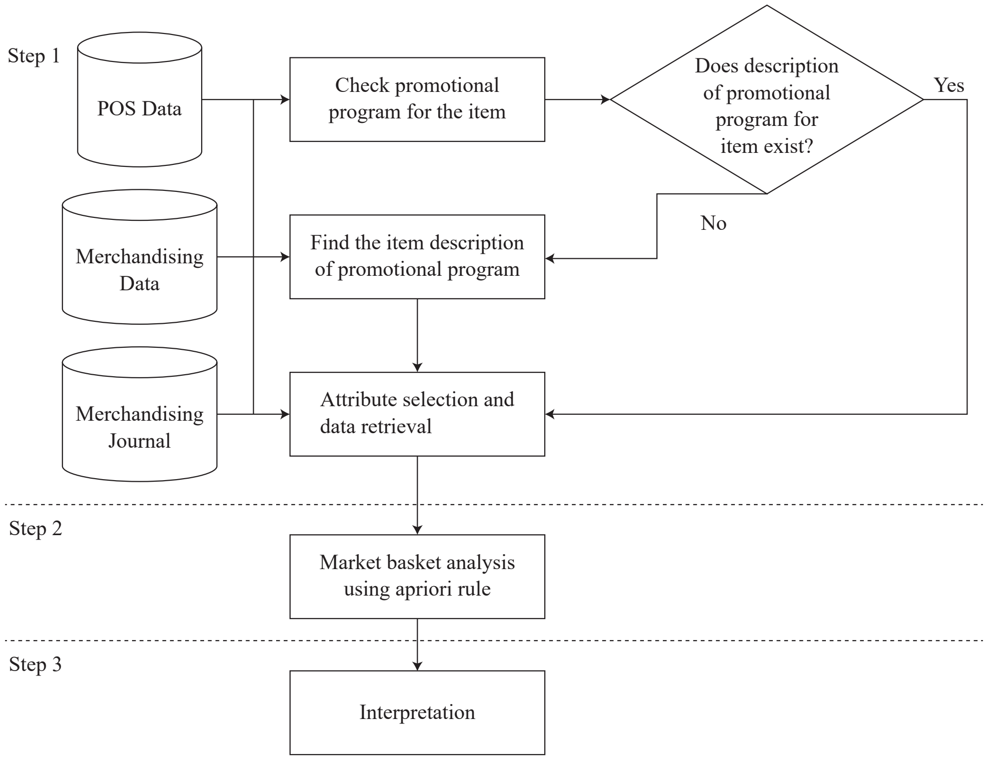

Figure 1: Modified MBA-based framework

However, as explained in Step 1, the MBA analysis proposed in this framework adds the promotional program attribute as a basis for analysing consumer behaviours. The MBA algorithm is performed with R software using library arules.

The MBA technique itself consists of 2 main steps which are: (1) determining frequent set item; (2) generating rule. Step 1 is purposed to find the frequency of occurrence of item or item sets. Support value can be obtained using the following formula [Goedegebuure, 2021]:

$$Support = \frac{\textit{Frequency occurence of item or item sets}}{\textit{Total transaction}}$$

The threshold support value that is used depends on the decision maker. After the support value has been determined, then Step 2 is performed to identify the association rule using confidence value. Confidence Item X → Item Y means how likely item Y is purchased when item X is purchased. Finding confidence value can be obtained using the following formula [Goedegebuure, 2021]:

$$Confidence\ (X \rightarrow Y) = \frac{(Support\ (X,Y)}{(Support\ (X)}$$

In addition, there is another parameter in the apriori-association rule which is "minlen." Minlen describes minimum length of item sets that will be considered for generating association rules [Agrawal et al., 1993].

When determining rules using the apriori-association rule algorithm, the data must be in basket format, consisting of 2 columns, where the first column describes the transaction ID and the second column describes the items purchased and other information such as membership status and shopping bag purchases, as shown in Table 1.

Table 1: Format data for MBA techniqe

Transaction ID	Item
xxxx	Member, Bag, Item 1, Item 2

Based on Table 1, to perform the apriori algorithm for MBA technique, then for each transaction ID, the item purchased, membership status, and shopping bag purchased should be written in one separated by a comma (,).

3.3 Step 3: Interpretation

In this step, the result from MBA is visualized by mapping customer shopping interaction patterns such as:

- Association rule between promotional program and membership status
- Association rule between items
- Association rule between items and membership status

The visualization is performed using R software.

4. Case study to illustrate the applicability of proposed framework

The framework explained in Section 3 is applied to a case study of one fashion retail in the city of Yogyakarta, Indonesia. Yogyakarta is one of the popular tourist destinations in Indonesia [BPS Yogyakarta, 2024]. The tourists visit several tourist attractions such as marinas, nature, museums, and tourist villages [Bappeda Yogyakarta, 2024]. Currently, shopping tourism items are mostly dominated by traditional markets and food souvenirs.

In Yogyakarta, there are 9 shopping malls [Niagatour, 2024], therefore in shopping tourism, as defined by Ambagtsheer [2020], namely tourists who spend their money other than for food and drinks can be developed if the shopping mall has the right strategy to attract tourists to shop. One type of retail that can support shopping tourism is fashion retail. Therefore, this case study was conducted at a fashion retail store in one of the shopping malls in Yogyakarta. It is hoped that through the proposed framework, fashion retail managers can understand consumer behaviours related to the items sold with the promotions run. So that this can be used as input for fashion retail managers to be able to determine the right promotional program to attract visitors.

4.1 Step 1: Data collection stage and determination of attribute selection

The data was obtained from POS data stored on the local server of the retail store from July-December 2019 as shown in Table 2. The original data consists of 592,667 rows, 172,334 transaction ID and 13 columns (attribute) as follows: 1. date: transaction date; 2. store: shop code; 3. regnum: register number or POS number where the transaction is carried out; 4. cashier: code for the cashier serving; 5. type: consists of S (for normal sales transactions) and SV (for cancel/return transactions); 6. trx: transaction running number from each POS, this number can be the same at different POS; 7. time: transaction time and hours; 8. item: SKU of the item purchased; 9. descs: description of the item purchased; 10. qty: quantity per item purchased; 11. gross: price before discount including tax; 12. disc: discount amount (if any) including tax; 13. retail: price before discount including tax.

Based on Table 2, it can be seen that in the “disc” column which describes the promotional program, there is no information yet. Therefore, based on the proposed framework, in the merchandising data, it is possible to find details about the ongoing promotional program for each item.

Based on Table 3, the “Brand” column, “Program” column, and “Description” column describe the brand name and the ongoing promotional program respectively. For example, in the “Program” column, “SP” indicates special price while 80 represents the percentage of discount.

In the proposed framework, the attributes used for MBA analysis are transaction ID, purchased item per transaction, promotional program, and membership status of customers. Based on Table 2, the transaction ID is not explicitly visible. Therefore, Table 2 needs to be transformed by combining the columns ‘regnum,” “cashier”, and “trx” as shown in Table 4. The combination is performed using Microsoft Excel using the function “CONCATENATE.”

Table 4: Transaction ID data from POS

regnum	cashier	type	trx	time	TransactionID
31	91262	S	30840	14.54	319126230840
92	90899	S	20522	20.27	929089920522
102	2006	S	31839	19.03	102200631839
102	2006	S	31843	19.12	102200631843

Next, data from Table 2, 3, and 4 are then integrated into 1 table as explained in Section 3 above. The results of the integration of Table 2, 3 and 4 can be seen in Table 5, where Table 5 will later be used at the rule determination stage using the apriori association rule algorithm.

Table 2: Data retrieved from POS

Date	Store	Regnum	Cashier	Type	Trx	Time	Item	Descs	Qty	Gross	Disc	Retail
1-Jul-19	15004	31	91262	S	30840	14.54	10021000	EXECUTIVE LDS 00	1	259900	–	259900
1-Jul-19	15004	92	90899	S	20522	20.27	10021000	EXECUTIVE LDS 00	1	259900	–	259900
1-Jul-19	15004	102	2006	S	31839	19.03	10021000	EXECUTIVE LDS 00	1	399900	–	399900
1-Jul-19	15004	102	2006	S	31843	19.12	10021000	EXECUTIVE LDS 00	1	299900	–	299900

Table 3: Promotional Product for Items

item	descs	gross	disc	retail	Brand	Program	Description
2000001965627	BAG - CHRISTMAS	145.900	116.720	29.180	PRECIOUS ONE	80	PRECIOUS ONE 80
2000002330455	RC CUP & SAUCER FANC	79.000	63.200	15.800	ST YVES	80	ST YVES 80
2000002330455	RC CUP & SAUCER FANC	79.000	63.200	15.800	ST YVES	80	ST YVES 80
2000002330455	RC CUP & SAUCER FANC	79.000	63.200	15.800	ST YVES	80	ST YVES 80
2000002582106	LONG VEST BLACK	199.900	150.000	49.900	CEIL	SP	CEIL SP
2000002636571	PANTS W/ B TERACOTA	199.900	150.000	49.900	CEIL	SP	CEIL SP
2000001965542	CUSHION - OWL DESIGN	145.900	116.720	29.180	PRECIOUS ONE	80	PRECIOUS ONE 80
2000002594260	CUSHION OWL GREEN	129.000	103.200	25.800	CHRYSALIS	80	CHRYSALIS 80
2000002594260	CUSHION OWL GREEN	129.000	103.200	25.800	CHRYSALIS	80	CHRYSALIS 80
2000002594260	CUSHION OWL GREEN	129.000	103.200	25.800	CHRYSALIS	80	CHRYSALIS 80
2000002876243	BOY JOGGER M61 S	99900	0	99900	P&C KIDS	B1G1F	P&C KIDS B1G1F
2000002971719	BOY-JOGGER BLAC S	99900	99900	0	P&C KIDS	B1G1F	P&C KIDS B1G1F

Table 5: Data set for apriori association rule

No Struk	Date	V1
819085319	2019-12-07	DISC_VCR_RP_50.DISC_VCR_RP_50. DISC_VCR_RP_50…
819085320	2019-12-07	DISC_VCR_RP_50.DISC_VCR_RP_50. DISC_VCR_RP_50…
819085321	2019-12-07	REVLON NORMAL.REVLON NORMAL.MAYBELLINE 50…
819085322	2019-12-07	MAX FACTOR 70-1. MAX FACTOR 70-1. MAX FACTOR S
819085323	2019-12-07	MAX FACTOR 70-1.REVLON NORMAL. REVLON NORM…
819085328	2019-12-07	DISC_VCR_RP_50.DISC_VCR_RP_50. DISC_VCR_RP_50…

4.2 Step 2: Market basket analysis using apriori association rule

In the case study presented in this paper, several combinations of minlen, support, and confidence (5 scenarios) to further elaborate the resulting rules are presented in Table 6.

Table 6: Parameters used to generate rules

Scenario	Minlen	Support	Confidence
1	2	0,05	0,5
2	2	0,01	0,5
3	2	0,05	0,1
4	2	0,01	0,1
5	2	0,005	0,5
6	2	0,005	0,1
7	2	0,001	0,5

When the dataset is ready, the next step is to perform association rules using the apriori algorithm. At this stage library arules in R software is used. The results for Scenario 1-5 are presented in Table 7-8, while the result for Scenario 6 and 7 will be visualized in Figure 2-3.

4.3 Step 3: Interpretation

From the results of Table 7-9, it can be seen that members have dominant associations, especially as a “consequence.”

Table 7: Association rules for promotion programs

Minlen; Support; Confidence	Rules	Remarks
2; 0.05; 0.5	0	no rule
2; 0.01; 0.5	1	DV 50k -> Member
2; 0.05; 0.1	0	no rule
2; 0.01; 0.1	2	DV 50k -> Member
		Member -> DV50k
2; 0.005; 0.5	5	PV50k -> DV50k
		PV50k -> Member
		DV 50k -> Member
		DV 50k, PV50k -> Member
		Member, PV50k -> DV50k

Note: Promotional voucher (PV), discount voucher (DV), and membership status for Scenario 1-5.

Table 8: Association rules for promotion programs

Rules	Count
Cosmetics & Fragrance	
{ELIZABETH ARDEN 00} => {Member Customer}	199
{LOREAL 40} => {Member Customer}	186
{LANCOME 10} => {Member Customer}	329
{LOREAL 50} => {Member Customer}	210
{THE SAEM GWP} => {THE SAEM NORMAL}	310
{MAYBELLINE 50} => {Member Customer}	246
{LANCOME NORMAL} => {Member Customer}	326
{SHISEIDO NORMAL} => {Member Customer}	358
{MAKE UP FOREVER NORMAL} => {Member Customer}	564
{C&F PERFUMERY SP, PROMO_VCR_RP_50} => {DISC_VCR_RP_50}	225
{C&F PERFUMERY SP, DISC_VCR_RP_50} => {PROMO_VCR_RP_50}	225
{C&F PERFUMERY SP, DISC_VCR_RP_50} => {Member Customer}	208
Ladies Apparels & intimate	
{CLO B1G1F} => {Member Customer}	196
{CLO 20} => {Member Customer}	246
{PIERRE/C LDS 50-2} => {PIERRE/C LDS 00}	187
{PIERRE/C LDS SP-2} => {PIERRE/C LDS 00}	278
{PIERRE/C LDS 91-1} => {Member Customer}	192
{BRILLIANT GIRL 20} => {Member Customer}	213
{CEIL 20} => {Member Customer}	248
{CEIL B1G1F} => {Member Customer}	326
{ETCETERA 50} => {Member Customer}	308
{SORELLA 50} => {SORELLA 00}	259
{SORELLA 50} => {Member Customer}	219
{YOUNG HEARTS 50} => {YOUNG HEARTS 00}	527
{YOUNG HEARTS 00} => {YOUNG HEARTS 50}	527
{YOUNG HEARTS 91} => {Member Customer}	377
{SORELLA 91-2} => {Member Customer}	374
Home Appliances	
{LENUTA 91} => {Member Customer}	203
{TOYOTERRY SP-1} => {Member Customer}	215
{RHEMA SP} => {Member Customer}	205
{TERRY PALMER 910} => {Member Customer}	240
Mens Apparels	
{PARACHUTE B1G1F} => {Member Customer}	336
{AMBROGIO B1G1F} => {Member Customer}	178
{AMBROGIO UNDERWEAR & SOCKS B1G1F} => {Member Customer}	229
{PIEERE CARDIN SP-50} => {Member Customer}	243
Others	
{P&C KIDS B1G1F} => {Member Customer}	261
{HUSH PUPPIES LDS SHO} => {Member Customer}	586
{HUSH/P UNSX L SP-2, HUSH/P UNSX M SP-2} => {Member Customer}	174

Note: Promotional voucher (PV), discount voucher (DV)) and membership status for Scenario 1-5.

Table 9: Association rules for product and membership status for Scenario 1-5

Rules	Count
Cosmetics	
{REVLON NORMAL} => {Member Customer}	1.215
{MAYBELLINE NORMAL} => {Member Customer}	1.026
{WARDAH NORMAL} => {Member Customer}	1.002
{ELIZABETH ARDEN 00} => {Member Customer}	199
{LOREAL 40} => {Member Customer}	186
{LANCOME 10} => {Member Customer}	329
{LOREAL 50} => {Member Customer}	210
{LANCOME NORMAL} => {Member Customer}	326
{SHISEIDO NORMAL} => {Member Customer}	358
{MAYBELLINE 50} => {Member Customer}	246
{MAKE UP FOREVER NORMAL} => {Member Customer}	564
Fragrances	
{C&F PERFUMERY SP} => {Member Customer}	1.217
Ladies Apparels	
{CLO B1G1F} => {Member Customer}	196
{CLO 20} => {Member Customer}	246
{BRILLIANT GIRL 20} => {Member Customer}	213
{CEIL 20} => {Member Customer}	248
{CEIL B1G1F} => {Member Customer}	326
{PIERRE/C LDS 91-1} => {Member Customer}	192
{ETCETERA 50} => {Member Customer}	308
{SORELLA 50} => {Member Customer}	219
{SORELLA 91-2} => {Member Customer}	374
{YOUNG HEARTS 91} => {Member Customer}	377
{Member Customer, YOUNG HEARTS 50} => {YOUNG HEARTS 00}	256
{Member Customer, YOUNG HEARTS 00} => {YOUNG HEARTS 50}	256
Home Appliances	
{LENUTA 91} => {Member Customer}	203
{TOYOTERRY SP-1} => {Member Customer}	215
{TERRY PALMER 910} => {Member Customer}	240
{RHEMA SP} => {Member Customer}	205
Mens Apparels	
{PIEERE CARDIN SP-50} => {Member Customer}	243
{PARACHUTE B1G1F} => {Member Customer}	336
{AMBROGIO B1G1F} => {Member Customer}	178
{AMBROGIO UNDERWEAR & SOCKS B1G1F} => {Member Customer}	229
Unisex	
{HUSH/P UNSX M SP-2} => {Member Customer}	1.101
{HUSH/P UNSX L SP-2, HUSH/P UNSX M SP-2} => {Member Customer}	174
Others	
{TAMMIA ACCESS 00} => {Member Customer}	1.012
{P&C KIDS B1G1F} => {Member Customer}	261
{HUSH PUPPIES LDS SHO} => {Member Customer}	586

This indicates that the right promotional program can trigger members to come and shop.

From Table 7, it can be seen that members have a strong association with the combination of DV50 (discount voucher 50 %) and PV50 (promotional voucher 50 %) which can be seen in the associations {Member Customer, PROMO_VCR_RP_50} =>{DISC_VCR_RP_50} and {DISC_VCR_RP_50, Member Customer} => {PROMO_VCR_RP_50}.

From Table 8 and 9, it can be seen that from the cosmetics category members who are Revlon, Maybelline, and Wardah customers still shop even though there are no promotional programs or normal prices and no voucher programs, because only Ultima customers are sensitive to vouchers in cosmetics. There was a large sale of Tammia accessory products even at normal prices, because the placement of the products is near the cashier, allowing impulse buying. For house brand products, there is an indication that members really prefer the B1G1F (Buy 1 Get 1 Free) promotion, especially for mens apparel and ladies apparel products.

From Figure 2 and 3, it can be seen that there is 1 largest node which is a representation of the association {} => {Member Customer} and 1 largest node in Figure 3 which is a representation of {DISC_VCR_RP_50} => {Member Customer}. This indicates that the members are the largest potential customers who must be managed well by retailers, and this largest potential is quite sensitive to voucher programs, especially DV50.

Based on the analysis using modified MBA by incorporating promotional program and membership status, the result shows that retail can find out how much influence promotional program and membership status has on the purchase of an item. In the case study used in this study, it can be seen that the promotional program and membership status influence the purchase of an item. BIGIF is effective for mens apparel and ladies apparel products but not effective for cosmetics products. An

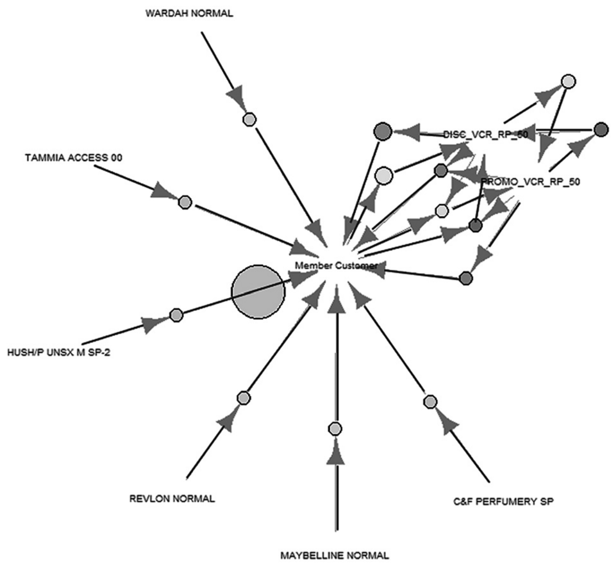

Figure 2: Graph of association rules for Scenario 6

Figure 3: Graph of association rules for Scenario 7

effective promotional program for product cosmetics is promotional vouchers and discount vouchers. In addition, there is a close relationship between membership status and the purchase of items and the use of promotional programs.

By understanding this, to attract potential visitors including tourists, apparel products can be given a B1G1F promotion. In addition, from the results shown in Table 7 and 8 and Figure 2 and 3, the dominance of members in the use of promotional vouchers is also very high. This can be an opportunity for retailers to attract tourists to become members of the retail. Although there are tourists who may not visit the retail very often, with the existence of online retail, it can allow these tourists to make repeat purchases online in the future.

5. Conclusion

Based on the case study, the proposed framework can be applied to fashion retail, where fashion retail already has a point of sales (POS) system to collect transaction data and merchandising data to collect information about the promotional program for each item. This is different to previous research on MBA [Hossain et al., 2019; Kavitha and Subbaiah, 2020; Made et al., 2022; Qisman et al., 2021; Ünvan, 2021; Verma et al., 2020].

In the proposed framework, MBA is performed by adding the variables related to: (1) promotional program of item and (2) membership status. These results can be used by retailers to identify which product's sales are sensitive to promotional programs. Products whose sales are sensitive to promotional programs and have a large sales quantity can be used as an attraction for consumers, including tourists, to visit the retailer.

References

Agrawal, R, Imielinski, T., and Swami, A. (1993) Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 207-216.

Agrawal, R. (1994). Fast algorithms for mining association rules. *Proceedings of the 20th VLDB Conference Santiago, Chile*, pp. 487-499.

BPS Yogyakarta (2024). Jumlah kunjungan wisatawan Nusantara menurut tujuan-tabel statistik. Badan Pusat Statistik Provinsi Di Yogyakarta (Retrieved January 4, 2025 from https://yogyakarta.bps.go.id/id/statistics-table/2/NTE0IzI=/jumlah-kunjungan-wisatawan-nusantara-menurut-tujuan.html).

Bappeda DIY (2024). Aplikasi dataku daerah istimewa Yogyakarta (Retrieved January 4, 2025 from https://bappeda.jogjaprov.go.id/dataku/data_dasar/cetak/214-jenis-objek-wisata).

Dekimpe, M. G. (2020). Retailing and retailing research in the age of big data analytics. *International Journal of Research in Marketing*, Vol. 37, No. 1, pp. 3-14.

Epstein, L. D., Inostroza-Quezada, I. E., Goodstein, R. C., and Choi, S. C. (2021). Dynamic effects of store promotions on purchase conversion: Expanding technology applications with innovative analytics. *Journal of Business Research*, Vol. 128, pp. 279-289.

Fan, C., Chen, M., Wang, X., Wang, J., and Huang, B. (2021). A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, Vol. 9, pp. 1-17.

Goedegebuure, R. (2021). Chapter 5 Methods | Market basket analysis (Retrieved January 5, 2025 from https://bookdown.org/robert_statmind/testing3/methods.html.

Guney, S., Peker, S., and Turhan, C. (2020). A combined approach for customer profiling in video on demand services using clustering and association rule mining. *IEEE Access*, Vol. 8, pp. 84326-84335.

Jeffrey, S. A. and Hodge, R. (2007). Factors influencing impulse buying during an online purchase. *Electronic Commerce Research*, Vol. 7, Nos. 3-4, pp. 367–379.

Henzel, J. & Sikora, M. (2020). Gradient boosting application in forecasting of performance indicators values for measuring the efficiency of promotions in FMCG retail. *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems*, pp. 59-68.

Hossain, M., Sattar, A. H. M. S., and Paul, M. K. (2019). Market basket analysis using apriori and FP growth algorithm. *22nd International Conference on Computer and Information Technology*, pp. 18-20.

Kanta, S. (2019). Assessing promotional effectiveness in retail industry through estimation of sales components. *Espacios*, Vol. 40, No. 20, pp. 1-7.

Kavitha, M. and Subbaiah, D. S. (2020). Association rule mining using apriori algorithm for extracting product sales patterns in groceries. *International Journal of Engineering Research and Technology*, Vol. 8, No. 3, pp. 5-8.

Khairawati, S. (2019). Effect of customer loyalty program on customer satisfaction and its impact on customer loyalty. *International Journal of Research in Business and Social Science*, Vol. 9, No. 1, pp. 15-23.

Khouja, M., Subramaniam, C., and Vasudev, V. (2020). A comparative analysis of marketing promotions and implications for data analytics. *International Journal of Research in Marketing*, Vol. 37, No. 1, pp. 151-174.

Kim, W. H., Cho, J. L., and Kim, K. S. (2019). The relationships of wine promotion, customer satisfaction, and behavioral intention: The moderating roles of customers' gender and age. *Journal of Hospitality and Tourism Management*, Vol. 39, pp. 212-218.

Made, I., Putra Asana, D., Gede, I., Sudipa, I., Tri, A. A., Mayun, W., Putu, N., Meinarni, S., and Waas, D. V. (2022). *Informal: Informatics Journal*, Vol. 7, No. 1, pp. 38-45.

Moodley, R., Chiclana, F., Caraffini, F., and Carter, J. (2020). A product-centric data mining algorithm for targeted promotions. *Journal of Retailing and Consumer Services*, Vol. 54, No. 101940.

Niagatour (2024). 9 Mall Terbesar di Jogja, yang Suka Belanja Wajib Mampir! Paket Wisata Domestik & Internasional Murah & Terbaik (Retrieved January 4, 2025 from https://niagatour.com/mall-terbesar-di-jogja/).

Ozgormus, E. and Smith, A. E. (2020). A data-driven approach to grocery store block layout. *Computers and Industrial Engineering*, Vol. 139, No. 105562.

Phumchusri, N., Kosawanitchakarn, W., Chawanapranee, S., and Srimook, S. (2023). Evaluating promotional pricing effectiveness using convenience store daily sales data. *Journal of Revenue and Pricing Management*, Vol. 22, pp. 362-373.

Qisman, M., Rosadi, R., and Abdullah, A. S. (2021). Market basket analysis using apriori algorithm to find consumer patterns in buying goods through transaction data (case study of Mizan computer retail stores). *Journal of Physics: Conference Series*, Vol. 1722, No. 1.

Reutterer, T., Hornik, K., March, N., and Gruber, K. (2017). A data mining framework for targeted category promotions. *Journal of Business Economics*, Vol. 87, No. 3, pp. 337-358.

Rooderkerk, R. P., DeHoratius, N., and Musalem, A. (2022). The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners. *Production and Operations Management*, Vol. 31, No. 10, pp. 3727-3748.

Shen, B. (2021). E-commerce customer segmentation via unsupervised machine learning. *ACM International Conference Proceeding Series*, Part F168982.

Silva, J., Varela, N., López, L. A. B., and Millán, R. H. R. (2019). Association rules extraction for customer segmentation in the SMES sector using the apriori algorithm. *Procedia Computer Science*, Vol. 151, pp. 1207-1212.

Sun, L., Zheng, X., Jin, Y., Jiang, M., and Wang, H. (2018). Estimating promotion effects using big data: A partially profiled LASSO model with endogeneity correction (Retrieved November 22, 2018 from https://ssrn.com/abstract=3289112).

Thangavelu, S. M. (2019). Global supply chain in Singapore's services sector: retail value chain and productivity improvements. *Journal of Southeast Asian Economies*, Vol. 36, No. 2, pp. 244-255.

Ünvan, Y. A. (2021). Market basket analysis with association rules. *Communications in Statistics-Theory and Methods*, Vol. 50, No. 7, pp. 1615-1628.

Verma, N., Malhotra, D., and Singh, J. (2020). Big data analytics for retail industry using MapReduce-Apriori framework. *Journal of Management Analytics*, Vol. 7, No. 3, pp. 424-442.

Yoseph, F., Ahamed Hassain Malim, N. H., Heikkilä, M., Brezulianu, A., Geman, O., and Paskhal Rostam, N. A. (2020). The impact of big data market segmentation using data mining and clustering techniques. *Journal of Intelligent and Fuzzy Systems*, Vol. 38, No. 5, pp. 6159-6173.

Zhi-guo, T., Peng-yue, L., and Jie, H. (2018). A regional periodic pattern mining algorithm based on apriori-like principle. *Microelectronics & Computer*, Vol. 35, No. 9, pp. 85-89.

Received: October 21, 2024
Revised: January 6, 2025
Accepted: January 15, 2025