Asia-Pacific Journal of Information Technology and Multimedia 14(1) 2025: 164 – 179
https://doi.org/10.17576/apjitm-2025-1401-10

Mining Association Rules to Determine the Overspending Behavior Among Low Income Households in Malaysia

Petua Persatuan Perlombongan untuk Menentukan Tingkah Laku Berbelanja Terlebih dalam Kalangan Isi Rumah Berpendapatan Rendah di Malaysia

Amira Najiha Azmi, Azuraliza Abu Bakar, Zulaiha Ali Othman, Nor Samsiah Sani

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

**Corresponding author: Azuraliza Abu Bakar (azuraliza@ukm.edu.my)*

Received 11 October 2024
Accepted 20 March 2025, Available online 30 June 2025

ABSTRACT

Overspending behavior in a household can significantly affect the financial burden, debt accumulation, stress, and economic problems. Spending behavior is one of the financial literacy indicators that empowers individuals to make informed financial decisions, budget effectively, and plan for the future. This study proposes an association rules mining approach to investigate the spending behavior among households with income below 40% (B40) in Malaysia. For this purpose, we employ the Apriori algorithm in 2016 and 2019 Malaysia households' income and expenditure survey data obtained from the Department of Statistics Malaysia to discover over-spending items that occurred in household expenditure. The results showed that up to three associated overspending items were discovered based on several support and confidence settings. There are significant changes in spending behavior in the 2016 and 2019 data. Besides food as the main overspending item in 2016 data, other items such as miscellaneous items, restaurants and hotels, and services were overspent in 2019 data. Moreover, three associated items were found only in the 2019 data. This finding benefits the government in improving financial literacy or implementing effective initiatives to improve the nation's living standards.

Keywords: Data Science; Spending Behaviour; Overspending Rules; Apriori Algorithm; Mining Association Rules

ABSTRAK

Tingkah laku berbelanja berlebihan dalam isi rumah boleh menjejaskan beban kewangan, pengumpulan hutang, tekanan dan masalah ekonomi dengan ketara. Tingkah laku berbelanja ialah salah satu petunjuk celik kewangan yang memperkasakan individu untuk membuat keputusan kewangan termaklum, belanjawan dengan berkesan dan merancang untuk masa depan. Kajian ini mencadangkan pendekatan perlombongan petua persatuan untuk menyiasat tingkah laku perbelanjaan dalam kalangan isi rumah berpendapatan di bawah 40% (B40) di Malaysia. Untuk tujuan ini, kami menggunakan algoritma Apriori pada tahun 2016 dan data tinjauan pendapatan dan perbelanjaan isi rumah Malaysia 2019 yang diperoleh daripada Jabatan Perangkaan Malaysia untuk mengetahui item perbelanjaan berlebihan yang berlaku dalam perbelanjaan isi rumah. Keputusan menunjukkan bahawa sehingga tiga item perbelanjaan berlebihan yang berkaitan ditemui berdasarkan beberapa tetapan sokongan dan keyakinan. Terdapat perubahan ketara dalam tingkah laku perbelanjaan dalam data 2016 dan 2019. Selain makanan sebagai item terlebih belanja utama pada data 2016, item lain seperti pelbagai barangan, restoran dan hotel, dan perkhidmatan telah terlebih belanja pada data 2019. Selain itu, tiga item berkaitan hanya ditemui dalam data 2019. Penemuan ini memberi manfaat kepada kerajaan dalam meningkatkan celik kewangan atau melaksanakan inisiatif berkesan untuk meningkatkan taraf hidup negara.

Kata Kunci: Sains Data, Perilaku Perbelanjaan, Petua Terlebih Belanja, Algoritma Apriori, Petua Persatuan Perlombongan

INTRODUCTION

Financial burden is a major challenge for households, especially the lower-income B40 group in Malaysia. Rising living costs, including higher prices for goods and services, have hit this group the hardest. Poor financial management and lifestyle changes often lead to overspending, worsening their financial strain. Spending behavior—how individuals use money, time, and effort—plays a crucial role in this issue (Abdul Rahman et al., 2021; Abdul Shakur et al., 2021). Limited financial knowledge and low income contribute to overspending, compounded by rising education costs and everyday expenses like online shopping, food delivery, and utility bills. Identifying the root causes is crucial for the government to introduce effective guidelines, promote sustainable spending, and enhance national well-being.

Several studies on household income, financial literacy, and spending behavior in Malaysia have used statistical and mathematical models. Methods such as multiple regression (Abu Bakar et al., 2020), OLS and Tobit estimations (Applanaidu et al., 2022), and PLS-SEM (Chih & Sang, 2020) have been applied to examine factors affecting B40 household spending and financial stress. Findings highlight the importance of considering socioeconomic and demographic factors, identifying vulnerable households, and implementing effective policies or stimulus packages to ease the financial burden on B40 households (DOSM, 2017).

Machine Learning (ML) on socio-economic data that used Malaysia household income and expenditure data with various results can be seen in (DOSM, 2020, 2023; Hamid et al., 2021; Han, 2023). (Ho, 2022) use ML algorithms, namely Decision Tree, Naïve Bayes, Neural Network, Support Vector Machines and Nearest Neighbor, to identify attributes that influence excessive spending. The patterns and factors of overspending among household income classes B40, M40, and T20 were identified using 12 attributes, namely the number of households, area, state, strata, race, highest certificate, marital status, gender, housing, income, amount expenses, and categories as attribute classes.

This study (Huang et al., 2017) employed Linear Regression to explore the spending pattern and examine the effects of important socio-economic factors on the consumption pattern of food and non-alcoholic beverages among B40 households in Malaysia. The regression analysis indicates that, apart from semi-skilled occupation, all other linear model factors significantly affect food expenditure. The findings of these researchers suggested that policymakers and stakeholders should draft guidelines for achieving sustainable spending, and the government should take immediate interventions in handling the issues among low-income groups.

Husin (2022) used the Naive Bayes classification algorithm to assess and map the potential of low-income families in Indonesia, aiming to anticipate poverty rates. The study classified 219 low-income families based on 11 attributes—such as income, education, health, and utilities—into categories of "poor" and "very poor," achieving 93% accuracy. The classification was enhanced with geographic data and images of households. Findings suggest that Naive Bayes can effectively support government efforts in identifying and addressing poverty in Bantul Regency.

Ismail et al. (2023) analyzed the relationship between socioeconomic development (SED) indicators and rural harmless sanitary toilet (RHST) penetration in China (2007–2017), finding that optimizing SED indicators by region can improve RHST rates. Källestål et al. (2020) used association rule mining to explore links between socio-demographic factors and chemical exposures in the U.S., highlighting environmental inequalities and aiding public health decisions. Similarly, Li et al. (2022) applied association rule mining to identify factors influencing economic rewards, showing that earnings determinants vary by macro (e.g., self-worth, school practice) and micro perspectives.

Several studies employed statistical analysis, mathematical modeling, and machine learning to identify factors that influence overspending, and some studies focus on expenditure items that are most overspent. However, the current work is limited to the most influences on overspending behavior. The combination of factors has even more impact on overspending behavior severity. Therefore, this study aims to investigate the overspending behavior of the B40 households using the association rules mining method. Association rules mining discovers meaningful relationships among data based on counting the frequent patterns. The frequent patterns of expenditure items that are overspent or adequately spent together will give new insight into data besides other methods introduced previously.

MATERIALS AND METHODS

This study adopted the data analytics modeling methodology consisting of five phases, namely, data preprocessing and preparation, association rules mining (Apriori) algorithm development, rules evaluation, rules visualization, rules interpretation, and insights.

A. Data Preprocessing and Preparation

The Household Income and Expenditure Survey (HIES) dataset for 2016 and 2019 was obtained from the Department of Statistics Malaysia (DOSM). Generally, the original data consists of three databases, including the household's head profiles (approximately 16,000 records and 10 attributes), household's members' profiles (approximately 60,354 records and 12 attributes), and the expenditure items database (approximately 60,354 records and 139 items attributes). The databases were then integrated based on the household's head information, which reduced the number of records. The HIES data must be cleaned, preprocessed, and formatted for analysis. This step involves removing duplicates, handling missing values, converting categorical data into a suitable format, and organizing the data for association rule mining. The three databases are integrated based on the household's head identity number (ID), obtaining 14,551 and 16,355 records for 2016 and 2019, respectively. The B40 household data is extracted based on the defined income category, i.e. total household income less than RM4850 (2016) and RM4850 (2019). Table 1 shows the summary of data description and Classification of Individual Consumption According to Purpose (COICOP) item categories. The 139 spending item attributes are reduced into 12 item categories based on the COICOP published by DOSM. The 12 attributes contain the percentage of item expenses in the categories.

TABLE 1. List of Parameters After Data Preparation

No	Parameter	Description
1	ID	Household identification number
2	Household Size	The number of households
3	Region	1 Peninsula, 2 Sabah and W. P. Labuan and 3 Sarawak
4	State	1 Johor, 2 Kedah, 3 Kelantan, 4 Melaka, 5 State Sembilan, 6 Pahang, 7 Penang, 8 Perak, 9 Perlis, 10 Selangor, 11 Terengganu, 12 Sabah, 13 Sarawak, 14 W.P. Kuala Lumpur, 15 W.P. Labuan and 16 W.P. Putrajaya
5	Strata	1 city; 2 rural areas
6	Type of residential	1 bungalow, 2 semi-D, 3 terraces, series or trips, city houses, 4 longhouses (Sabah & Sarawak only), 5 flats, 6 apartments, 7 condos, 8 shop/office houses, 9 rooms, 10 replacement/temporary huts and 11 other
7	Status	1 owned, 2 rented, 3 squatters owned, 4 squatters rent, 5 quarters and 6 other
8	Sex	1 male; 2 female
9	Age	1 less than 26, 2 26-60, 3 more than 60
10	Race	1 Bumiputera, 2 Chinese, 3 India, 4 Other

11	Marital Status	1 never married, 2 married, 3 widows/widows, 4 divorced, 5 separated
12	Highest Certification	1 Degree/Advanced Diploma, 2 Diploma/Certificate, 3 STPM, 4 SPM/SPMV, 5 PMR/SRP, 6 No Certificate
13	Employment	1 employer, 2 salaried workers, 3 unemployed or unpaid workers
COICOP items category (Data represent the percentage spending of item)		
14	Food	Food and non-alcoholic beverages
15	ATN	ATN tobacco, narcotics (ATN)
16	Clothes	Clothing and Footwear
17	Housing	Housing, water, electricity, gas and other fuels
18	Furnishing	Furnishings, household equipment, and routine household
19	Health	Health
20	Transportation	Transportation
21	Communication	Communication
22	R&C	Recreation and Culture
23	Education	Education
24	R&H	Restaurant and hotel
25	Miscellaneous	Miscellaneous goods and services
26	Overspending	1 Yes, 0 No

B. Labelling the Overspending Items

The 12 categories of spending items are further preprocessed by determining the percentage spending of each item that indicates overspending. The spending percentage of a particular item is represented as 1 for overspending and 0 for adequate spending. The representation is determined based on the reference guide obtained from the Guide to Preparing an Expenditure Plan (Budget), where a personal finance expert. Dave Ramsey's Household Budget Percentages (2023 Edition) uses a combination of income percentages and set figures drawn from national income averages to determine his recommendations. Additionally, his recommended budgeting percentages differ based on household size, the need for childcare, and other variables. (Mansor et al., 2022) suggested the limit of item budget as: Housing costs: 25%, Saving: 15%, Food: 12%, Childcare: 12%, Giving: 10%, Miscellaneous: 5%, Insurance: 4%, Utilities: 4%, Personal spending: 4%, Lifestyle and entertainment: 4%, Transportation: 3%, Health: 2%. In this study, we adopted the threshold suggested by (Munisamy et al., 2022) which the income received can be distributed by item and the suggested percentages are as follows: savings (10%), food (10-15%), utilities (5-10%), housing (25%), transportation (10%), health (5-10%), insurance (10-25%), recreation (5-10%), personal expenses (5-10%), gifts (10%) and Miscellaneous (5-10%). Figure 1 depicts the example of preprocessed dataset.

Num-Household	State	Area	Strata	Ethnic	Total Expenses	Total Income	Food	A&T	Clothes	Housing	Furnishing	Health	Transportation	ommunicatic	R&C	Education	R&C	Miscellaneous
5	4	1	1	1	2482	2839	1	1	0	0	0	0	0	0	0	0	1	1
2	9	1	1	1	4270	4403	0	0	0	1	1	1	1	1	1	0	1	1
1	9	1	2	2	2267	2627	1	1	1	0	0	0	1	0	0	0	1	1
4	4	1	1	1	3001	4808	1	0	1	0	1	0	1	0	1	0	1	0
4	9	1	2	2	3458	2686	1	1	1	1	1	1	1	0	1	0	1	1
4	9	1	1	3	2785	3023	1	0	0	0	0	0	0	1	1	0	0	1
5	9	1	1	1	2930	3351	1	1	0	0	0	0	1	1	1	1	1	1
2	9	1	2	1	2892	3648	1	0	1	0	0	0	0	0	1	0	1	0
5	9	1	1	1	2510	3315	1	1	0	1	0	0	1	1	1	0	0	1
2	4	1	1	1	806	1310	1	1	1	1	0	0	1	0	1	1	1	1
2	4	1	1	2	2840	4372	1	1	1	1	0	0	0	1	1	0	1	1
5	9	1	1	1	2662	4267	1	1	0	1	0	1	0	1	1	0	1	1
3	4	1	1	1	2713	4147	1	0	0	0	0	0	1	0	1	0	1	0
1	9	1	2	1	1617	806	1	1	1	1	1	1	1	1	1	1	1	1
4	9	1	1	1	3205	3861	1	1	1	1	0	0	1	1	1	0	1	1
2	9	1	1	1	2083	2550	1	1	1	1	0	0	1	1	1	0	1	1
4	12	2	1	1	1793	2224	1	1	1	1	0	1	1	1	1	1	1	1
1	9	1	1	2	1770	2398	1	1	0	0	0	0	1	0	1	0	1	1
4	9	1	1	1	2672	3159	1	1	1	1	1	1	1	1	1	1	1	1
2	9	1	2	3	2826	2617	1	1	1	1	0	1	1	1	1	0	1	1
5	4	1	1	1	6193	4771	0	1	1	0	1	0	1	1	0	1	1	1
1	9	1	1	3	1424	1077	1	1	1	1	1	0	1	1	1	1	1	1
2	9	1	2	2	3566	4827	1	0	0	0	0	1	1	0	1	0	1	0
1	4	1	2	1	1928	1959	1	1	1	1	1	0	1	1	1	0	1	1

FIGURE 1. Example of Preprocessed Data

C. Apriori Algorithm Development

The experiment conducted using association rules mining algorithms known as such as Apriori or FP-Growth, on the preprocessed dataset to discover patterns and associations between different transactions or spending behaviors. The algorithms identify frequent item sets and generate association rules that describe relationships between sets of items or transactions. In this study, we employed the Apriori algorithm to search for frequent patterns and association rules. Let X be the spending items, and support (X) is the number of item X occurred in the transaction. Typically, association rules are essential if they satisfy a minimum support and confidence threshold that users or domain experts can set. Additional analysis can be performed to discover interesting statistical correlations between associated items. Figure 2 shows the pseudocode of Apriori algorithm (Othman et al., 2020). The Apriori algorithm generates association rules of spending and overspending behavior of the B40 household in Malaysia.

The algorithms compute the frequent items occurred in the data based on the minimum support threshold. Then, the association rules in the form of X→Y are generated from the frequent items, and the rules with minimum confidence will be considered. In this study, the X and Y can represent the occurrence of either spending or overspending items. The Eq (1) and Eq (2) showed the computation of support and confidence of the association rules. The support for rules X→Y is the frequency of item X and Y occur together.

$$support(X \rightarrow Y) = support(X \cup Y) = P(X \cup Y) \quad (1)$$

The accuracy of the association rules is measured by the confidence of X→Y rules denoted as confidence(X→Y). It is a conditional probability of item Y occurs if item X occurs as in Eq (2).

$$confidence(X \rightarrow Y) = P(X) = \frac{support(X \rightarrow Y)}{support(X)} \quad (2)$$

Support and confidence form the threshold for establishing association rules. However, these measures are still insufficient to filter out worthless association rules. Lift can solve this weakness in association rules. Lift can measure the correlation of association rules and is used to assess whether the sets of items X and Y are independent, positively correlated or negatively correlated. If lift is equal to 1, then the set of items is independent, if lift is less than 1, then the set of items is negatively correlated, if lift is greater than 1, then the set of items is positively correlated as shown in Eq (3).

$$lift\ (X \rightarrow Y) = \frac{confidence\ (X \rightarrow Y)}{support\ (Y)} \quad (3)$$

```
Apriori(T, ε)
  L1 ← {large 1 - itemsets}
  k ← 2
  while Lk−1 is not empty
    Ck ← Apriori_gen(Lk−1, k)
    for transactions t in T
      Dt ← {c in Ck : c ⊆ t}
      for candidates c in Dt
        count[c] ← count[c] + 1

    Lk ← {c in Ck : count[c] ≥ ε}
    k ← k + 1

  return Union(Lk)

Apriori_gen(L, k)
   result ← list()
   for all p ∈ L, q ∈ L where p1 = q1, p2 = q2, ..., pk-2 = qk-2 and pk-1 < qk-1
     c = p ∪ {qk-1}
     if u ∈ L for all u ⊆ c where |u| = k-1
       result.add(c)
   return result
```

FIGURE 2. Apriori Algorithm

D. Rule Evaluation

Analyzing the generated association rules to evaluate their significance and relevance to overspending behavior. This step involves examining the support, confidence, and lift metrics to identify meaningful patterns. High-confidence rules might indicate strong associations between certain spending behaviors and the likelihood of overspending. By setting appropriate thresholds for support and confidence levels in the association rule mining process, patterns related to overspending can be uncovered. For instance, association rules might reveal common associations between specific purchases or spending habits that often lead to overspending.

In this study, we set minimum support values between 0.4 and 0.5 and a minimum confidence of 0.9 to focus on identifying strong and reliable association rules. These higher thresholds help ensure that the extracted rules are both frequent and trustworthy. Additionally, we lowered the minimum support and confidence to 0.08 to observe a broader set of rules, including less frequent associations. This approach allows us to compare how threshold settings influence the number and diversity of rules, helping to identify patterns that might be overlooked at higher thresholds. Then, the number of rules, the minimum and maximum length of antecedent and consequences of rules are recorded.

Rule Visualization: Presenting the discovered patterns and associations through visualizations, such as graphs, charts, or tables, to facilitate better understanding and interpretation of the results would help identify actionable insights for financial planning or interventions to prevent overspending.

Insights and Recommendations: Based on the identified patterns and associations, draw actionable insights and make recommendations. For example, financial institutions or advisors could use these insights to create personalized budgeting advice, recommend alternative spending habits, or develop targeted financial education programs to prevent overspending.

RESULTS

Association rule mining is often an iterative process that requires parameter adjustment, including different variables or segmentation criteria, to uncover more specific or refined overspending patterns. The basis of association rules mining, specifically the Apriori algorithm, is to find the frequent items that occur together. The lower the minimum support, the more association rules are generated. In this study, setting up a larger minimum support causes fewer overspending rules generated. We employed the Apriori algorithm for association rules mining. Several minimum support thresholds at the lowest 0.1 and (0.4-0.8) and minimum confidence of 0.9 were set for the 2016 and 2019 HIES datasets. The value 0.9 of minimum confidence is set to ensure only the most reliable rules are generated. In addition, to investigate more variations of overspending items in the association rules, the minimum support and confidence are set to the smallest value, 0.1.

Table 2 describes the association rules for the 2016 and 2019 datasets with the minimum and maximum rule length range [1 to 5]. The number of all rules (All-Rules) of overspending and adequate spending rules, and the number of overspending rules with the occurrence of one, two, or three overspending items (1-item, 2-item, 3-item) were reported. The number of generated rules decreases when the minimum support value increases. In addition, the overspending rules are extracted from the overall rules generated. At the highest minimum support of 0.8, no overspending rules are generated.

TABLE 2. Description of Association Rules for HIES 2016 and 2019 Dataset

Min-Sup (Min-Conf)	2016				2019			
	All-Rules	Overspending Rules			All-Rules	Overspending Rules		
		1-item	2-item	3-item		1-item	2-item	3-item
0.1-0.3 (0.1-0.9)	1032872	449176	28	0	4066	3907	78	81
0.4 (0.9)	2590	506	0	0	127	51	8	0
0.5 (0.9)	1056	113	0	0	52	18	1	0
0.6 (0.9)	377	24	0	0	22	5	0	0
0.7 (0.9)	127	2	0	0	8	1	0	0
0.8 (0.9)	33	0	0	0	3	0	0	0

A. General Analysis with the Baselines

The study employed the Association Rules Mining method on Malaysian households' income and expenditure data and has yet to be used elsewhere. Experimental comparative analysis could not be done directly with other studies. Generally, we observed our findings on associated items with those on household expenditures reported by the (Redjeki et al., 2015; Sabri et al., 2008; Sabri, 2019). Table 3 depicts the item's ranking of this study (using higher support of 0.5-0.8 and confidence of 0.8-1.0) and the reported expenditure via spending items by (Redjeki et al., 2015; Sabri et al., 2008; Sabri, 2019). It is observed that most overspending items listed in the 2016 and 2019 rules are the necessities items listed as expenditures in the 2016 and 2019 DOSM reports. The R&H (Restaurant and Hotels) has been the third rank over-spending item in 2019 rules. Similarly, R&H is listed in the third rank of 2019 and 2022 reports. The association items obtained from this study have consistent patterns with the ones reported in DOSM.

TABLE 3. General Observation of Items Ranking of Overspending vs Expenditure Items

	Spending Items of Household Expenditure DOSM Reports (Rank by Percentage of Expenditure)						Overspending Items Association Rules (Rank by Highest Support)	
Rank	2016	(%)	2019	(%)	2022	(%)	2016	2019
1	Food	25.5	Food	25.6	House	25.6	Food	Food
2	House	24.7	House	24.2	Food	24.5	Miscell	Miscell
3	Transp	11.8	R&H	12.6	R&H	12.9	Transp	R&H
4	R&H	11.8	Transp	10.8	Transp	10.0	House	Transp
5	Miscell	6.5	Miscell	6.8	Comm	5.6	Comm	Comm
6	Comm	4	Comm	4.2	Miscell	5.4	Clothes	R&C
7	Clothes	3.6	Furnish	3.6	Furnish	4.1	Furnish	House
8	R&C	3.5	Clothes	3.5	Clothes	2.9	ATN	-
9	Furnish	3.2	R&C	3.5	Health	2.8	-	-
10	ATN	2.5	ATN	2.3	ATN	2.1	-	-
11	Health	1.8	Health	2	Service	2.0	-	-
12	Edu	1	Edu	0.9	Edu	0.8	-	-

*Abbreviation (Furnishing: Furnish, Communication: Comm, Transportation: Transp, Miscellaneous: Miscell, Education: Edu)

B. Overspending Association Rules of 2016 and 2019 Dataset

We observe the overall behavior of 2016 and 2019 patterns by representing the rules in the item matrix as depicted in Figure 3 and Figure 4 for 2016 and 2019 data colored in red, respectively. The overspending association items in Figure 3 show that the B40 households' overspending is mainly overspending on food associated with other overspending items: ATN, clothes, housing, furnishing, transportation, communication, R&C, and R&H. On the other hand, overspending on food is also associated with adequate spending on health, education, and miscellaneous. In Figure 4, the 2019 data of B40 households showed significant changes in overspending association rules occur in most items except for adequate spending on health. The 2019 rules showed that not only food was the main overspending item, but associations of other items also occurred, indicating vast changes in spending behavior in three years.

We selected association rules representing the various support and confidence values, which show one-item, two-item and three-item combinations of association rules. Tables 4 - 5 show the list of one and two-item overspending rules for 2016 and 2019 datasets, while Table 6 presents the three-item association rules for 2019 datasets. The rules are represented as Antecedence (antecedence support), Consequence (consequence support), and Sup as the rules support. The lift value is between 1-1.3 indicating the set of items is positively correlated.

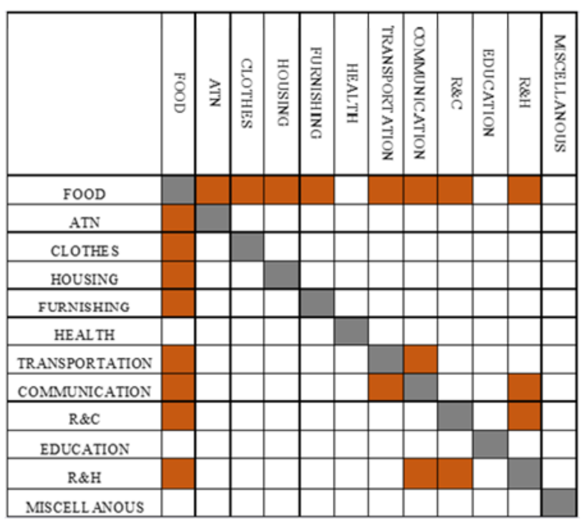

FIGURE 3. The Matrix of B40 Households' Overspending in 2016 Association Rules

In Table 5, we selected rules at support values greater than 0.6 and confidence of 0.9-1.0 since many one-item overspending association rules were generated in the 2016 and 2019 datasets. The rules with the highest confidence, such as 1.0, indicate the most reliable rules. The 2016 rules in Table 4 showed that the households mainly overspent on food, associated with adequate spending on other items such as education, miscellaneous, and health. There are no associations among overspending items at these support values, meaning that around 50%-70% of households overspend solely on food, which is a usual behavior of this income group. The 2019 rules in Table

4 showed significant changes in household overspending items, where the association rules showed that more items were overspent by households in 2019 compared to 2016 households, such as Miscellaneous, R&H, and Transportation. The overspending pattern of the B40 households in 2019 increased and covered more items (Rules 6- Rules 9 of 2019).

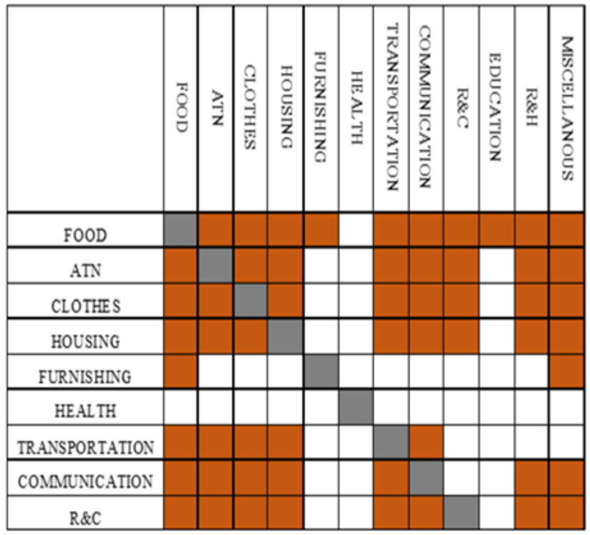

FIGURE 4. The Matrix of B40 Households' Overspending in 2019 Association Rules

Table 5 depicts the 2016 and 2019 two-item association rules. The two-item association rules 2016 occurred at lower rules support (0.1-0.3), indicating only 10-30% of 2016 households overspent in the two-item association. The association rules showed that overspending on items such as clothes, miscellaneous items, transportation, furnishing, housing, ATN, and communication is associated with overspending on food. In contrast, the 2019 association rules of two overspending items occur at higher rules support (0.4 and 0.5). The association rules were generated in two and three lengths. It is observed that 40-50% of 2019 households overspending on two items is associated with adequate spending on other items. The analysis shows the changes in overspending patterns, indicating the lifestyle changes amongst B40 in three years. The changes are evidenced by increased rules' support values in two associated overspending items.

Table 6 presents the selected three-item rules only discovered in the 2019 data. The selected rules represent the support values of 0.3 and 0.4, with the highest confidence value of 0.9. The support values indicate that about 30-40 per cent of the household had overspending on three expenditure items. As discussed earlier, the 2019 rules indicate significant lifestyle changes in which the lower-income group has shown overspending behavior in more than two expenditure items.

TABLE 4. Association Rules of 1-Item Overspending Rules for 2016 and 2019 Dataset

	2016 Overspending Association Rules			2019 Overspending Association Rules		
	Antecedence (0.1-0.7)	Consequence (0.7-1)	Sup	Antecedence (0.5-0.8)	Consequence (0.9-1)	Sup
1	Food=1	Health=0	0.7	Food=1, Education=0	Health=0	0.7
2	Food=1	Education=0	0.7	Education=0, Furnishing=0, Food=1	Health=0	0.6
3	Food=1, Education=0	Health=0	0.7	Food=1, ATN=0	Health=0	0.6
4	Food=1, Health=0	Education=0	0.7	Food=1	Health=0	0.8
5	Food=1	Education=0, Health=0	0.7	Furnishing=0, Food=1	Health=0	0.7
6	Food=1	Miscellaneous=0	0.7	Miscellaneous=1	Health=0	0.6
7	Miscellaneous=0, Food=1	Health=0	0.6	R&H=1	Health=0	0.6
8	Food=1, Health=0	Miscellaneous=0	0.6	Transportation=1	Health=0	0.6
9	Miscellaneous=0, Food=1	Education=0	0.6	Miscellaneous=1, Education=0	Health=0	0.6
10	Food=1, Education=0	Miscellaneous=0	0.6	Health=0, Furnishing=0, Food=1	Education=0	0.6

TABLE 5. Association Rules of 2-Item Overspending Rules for 2016 and 2019 Dataset

	2016 Overspending Association Rules			2019 Overspending Association Rules		
	Antecedence (0.1-0.7)	Consequence (0.7-1)	Sup	Antecedence (0.5-0.8)	Consequence (0.9-1)	Sup
1	Miscellaneous=1	Food=1	0.3	Miscellaneous=1, Food=1	Health=0	0.5
2	Transportation=1	Food=1	0.2	Miscellaneous=1, Food=1, Education=0	Health=0	0.4
3	Housing=1	Food=1	0.2	Food=1, R&H=1, Education=0	Health=0	0.4
4	Communication=1	Food=1	0.2	R&C=1	Food=1	0.5
5	Clothes=1	Food=1	0.1	R&H=1	Food=1	0.5
6	Furnishing=1	Food=1	0.1	Transportation=1	Food=1	0.5
7	ATN=1	Food=1	0.1	Communication=1	Food=1	0.4
8	Transportation=1	Miscellaneous=1	0.1	Clothes=1	Food=1	0.1
9	Communication=1	Miscellaneous=1	0.1	Furnishing=1	Food=1	0.1
10	-	-	-	ATN=1	Food=1	0.1
11	-	-	-	Transportation=1	Miscellaneous=1	0.1
12	-	-	-	Communication=1	Miscellaneous=1	0.1

TABLE 6. Association Rules of 3-Item Overspending Rules for 2019 Dataset

	Antecedence (0.3-0.8)	Consequence (0.6-0.8)	Sup	Conf
1	Miscellaneous=1, Housing= 1	Food=1	0.3	0.9
2	Transportation= 1, R&H=1	Food=1	0.3	0.8
3	Communication= 1, R&H=1	Food=1	0.3	0.8
4	Communication=1, Transportation= 1	Food=1	0.3	0.8
5	Miscellaneous=1, Communication=1	Food=1	0.3	0.8
6	Miscellaneous=1, R&H=1	Food=1	0.4	0.8

7	Food= 1, Housing= 1	Miscellaneous=1	0.3	0.7
8	Food= 1, Communication= 1	Miscellaneous=1	0.3	0.7
9	Food= 1, R&H= 1	Miscellaneous=1	0.4	0.7
10	Food= 1, Communication= 1	R&H=1	0.3	0.7
11	Food= 1, Transportation= 1	R&H=1	0.3	0.7
12	Miscellaneous= 1, Food= 1	R&H=1	0.4	0.7
13	Food= 1, R&H= 1	Transportation=1	0.3	0.7
14	Food= 1, Communication= 1	Transportation=1	0.3	0.7

C. Rules Interpretation

Interpreting the association rules requires a deeper investigation into the dataset and the context in which these associations were derived. It suggests interesting insights or strategies for financial planning, such as identifying certain spending habits that, when occurring together, could lead to more appropriate spending behaviors despite initial patterns of overspending. The association rule suggests an interesting relationship or pattern in the dataset that signifies a potential behavior or trend related to overspending. Several examples of association rules from Tables 6 and 7 are interpreted as follows:

1. Rule 1: Miscellaneous= 1, R&H= 1→Food= 1

if an individual overspends on miscellaneous, and restaurants and hotels (R&H), there is a likelihood of overspending on Food;

2. Rule 2: Transportation= 1, R&H= 1→Food= 1

if an individual overspends on transportation, and restaurants and hotels (R&H), there is a likelihood of overspending on Food;

3. Rule 3: Miscellaneous= 1, Food= 1→Health= 0

an individual overspends on miscellaneous and food maybe due to he has adequate or no spending in health;

4. Rule 4: Transportation=1, Furnishing=0→Health= 0

the individuals or households meeting the conditions of overspending on transportation and having adequate spending on furnishing are likely to also have adequate spending on health-related expenses;

5. Rule 5: Miscellaneous=1, Food=1, Education=0→Health= 0

the individuals or households meeting the conditions of overspending on miscellaneous and food and having adequate spending on education are likely to also have adequate spending on health-related expenses.

Rules 1 and 2 indicate that those with a lower income or specific socio-economic characteristics tend to overspend across multiple categories, including miscellaneous, transportation, dining out, and overall food expenditures. Individuals or households prioritizing spending on miscellaneous transportation and dining out at restaurants may also allocate a significant portion of their budget to food. Rules 3, 4, and 5 indicate that individuals or households prioritizing spending on food, miscellaneous, and transportation may also prioritize their health-related expenditures. The spending behavior might reflect a conscious decision to allocate resources to lifestyle and well-being.

D. General Interpretation of the Overspending Association Rules

We analyze the association rules in general and discuss them concerning socioeconomic aspects. The rule pattern of item1=1 and item2=1→item3 =1, implies a connection between overspending on item1 and overspending on item2, suggesting a correlated behavior which predicts that when this joint overspending behavior occurs, there is an increased likelihood of overspending on item3. Reducing the spending on item1 or item2 would reduce the cost of spending on item3. Two aspects that could be related to the rule's patterns are lifestyle choices and income and spending patterns. Lifestyle choices reflect the individuals or households that prioritize spending on item1 and item2 may also allocate a significant portion of their budget to item3 in general, which might reflect a specific lifestyle choice or preference. The income and spending patterns indicate that the rule could suggest that those with a lower income or specific socioeconomic characteristics tend to overspend across multiple categories, including item1, item2, and item3.

The rule pattern of item1=1 and item2=0→item3 =0 implies a potential socio-economic pattern where a specific spending behavior in item1 and item2 is linked to a corresponding behavior in item3 spending. It could suggest a financial prioritization or lifestyle choice among individuals or households. Two aspects that could be related to the rule's patterns are lifestyle prioritization and financial stability. Lifestyle Prioritization describes the individuals or households prioritizing item1 and item2, which may also prioritize item3 expenditures, which reflect a conscious decision to allocate resources to lifestyle and well-being. Financial stability describes adequate spending on item2, which may indicate financial stability or a certain socio-economic status. The rule suggests that the income group are financially stable enough to afford adequate item2 is also likely to allocate sufficient funds to item3 needs.

CONCLUSION

The study on households' spending behavior using association rule mining provides a data-driven approach to understanding spending patterns, identifying triggers for overspending, and devising strategies to promote better financial management and responsible spending habits. Our findings showed significant changes in the spending behavior of households below 40% income group (B40) between 2016 and 2019 in Malaysia. While in 2016, the household overspent mainly on food, which is a regular pattern, in 2019, changes occurred where the overspending of the B40 group occurred in food, hotels, restaurants and miscellaneous items. The spending behavior changes in 2016 and 2019 showed that regardless of having constraints in income, this group is very much influenced by the lifestyle, especially those living in urban areas. Lifestyle, entertainment, communication, and especially online social media usage are predicted to be the overspending elements in the households of B40. The research finding is vital to the government and several specific ministries to provide initiatives and awareness of financial literacy to the young generation. Besides, this study would open a new research area that introduces different approaches to exploring socioeconomic data and, in the future, could improve financial management, literacy and low-income households' socioeconomic sustainability.

ACKNOWLEDGEMENT

The authors thank the Ministry of Higher Education for the Grant Project Code: LRGS/1/2020/UKM/01/5/2 for funding and supporting this research.

REFERENCES

Abdul Rahman, Mariah, Nor Samsiah Sani, Rusnita Hamdan, Zulaiha Ali Othman, and Azuraliza Abu Bakar. “A Clustering Approach to Identify Multidimensional Poverty Indicators for the Bottom 40 Percent Group.” PLOS ONE 16, no. 8 (August 2, 2021). https://doi.org/10.1371/journal.pone.0255312.

Abdul Shakur, Elia Syarafina, Nor Hayati Sa’at, and Noorhaslinda Kulub Abd. Rashid. “Penentuan Tingkah Laku Pengguna Dalam Mempengaruhi Perbelanjaan Penggunaan Mampan Dalam Kalangan Isi Rumah Melayu.” Kajian Malaysia 39, no. 1 (April 30, 2021): 165–92. https://doi.org/10.21315/km2021.39.1.8.

Abu Bakar, Azuraliza, Rusnita Hamdan, and Nor Samsiah Sani. “Ensemble Learning for Multidimensional Poverty Classification.” Sains Malaysiana 49, no. 2 (February 28, 2020): 447–59. https://doi.org/10.17576/jsm-2020-4902-24.

Applanaidu, Shri-Dewi, Azira Abdul-Adzis, S. J. Jan, and N. Zainal Abidin. “Socio-Economics Factors Affecting B40 Households Food Expenditure in Malaysia.” Journal of Positive School Psychology 6, no. 5 (2022): 4539–53.

Chih, C. Y., and C. M. Sang. “Mining Association Rules between Education, Family Background and Earning.” Journal of Computers, 2020, 56–69. https://doi.org/10.3966/199115992020043102007.

Dosm. Department of Statistics Malaysia, 2017. https://www.dosm.gov.my/portal-main/release-content/report-on-household-expenditure-survey-2016.

Dosm. Department of Statistics Malaysia, 2020. https://www.dosm.gov.my/portal-main/release-content/report-on-household-expenditure-survey-2019.

Dosm. Department of Statistics Malaysia, 2023. https://www.dosm.gov.my/portal-main/release-content/report-on-household-expenditure-survey-2022.

Hamid, N.A., Z. Yaacob, Z.B. Ibrahim, S. Azizan, and N. Rashid. “A Socio-Economic Perspective of B40 Household Budgets in Malaysia.” Journal of Contemporary Issues in Business and Government 27, no. 02 (February 2, 2021). https://doi.org/10.47750/cibg.2021.27.02.003.

Han, Jiawei. Data Mining Concepts and Techniques. Cambridge, MA, USA: Elsevier, 2023.

Ho, Su-Wei. “Top 7 Essential Items the B40 Group Needs and How You Can Help.” MyPF.my, August 26, 2022. https://mypf.my/2022/08/26/top-7-essential-items-the-b40-group-needs-and-how-you-can-help/.

Huang, Hongtai, Rogelio Tornero-Velez, and Timothy M Barzyk. “Associations between Socio-Demographic Characteristics and Chemical Concentrations Contributing to Cumulative Exposures in the United States.” Journal of Exposure Science & Environmental Epidemiology 27, no. 6 (September 13, 2017): 544–50. https://doi.org/10.1038/jes.2017.15.

Husin, Mukhriz Mat. “Perbelanjaan Isi Rumah Meningkat, Rakyat Terus Terbeban.” Sinar Harian, June 5, 2022. https://www.sinarharian.com.my/article/205611/perbelanjaan-isi-rumah-meningkat-rakyat-terus-terbeban.

Ismail, N.A., L. Daud, S. Mohd, N. Samat, and A.R. Ridzuan. “Consumption Pattern Determinants of Low-Income Household: Evidence from Malaysia.” Jurnal Ekonomi Malaysia 57, no. 3 (October 2023). https://doi.org/10.17576/jem-2023-5703-03.

Källestål, Carina, Elmer Zelaya Blandón, Rodolfo Peña, Wilton Peréz, Mariela Contreras, Lars-Åke Persson, Oleg Sysoev, and Katarina Ekholm Selling. “Assessing the Multiple Dimensions of Poverty. Data Mining Approaches to the 2004–14 Health and Demographic Surveillance System in Cuatro Santos, Nicaragua.” Frontiers in Public Health 7 (January 29, 2020). https://doi.org/10.3389/fpubh.2019.00409.

Li, Yong, Shikun Cheng, Jiangshui Cui, Mingjun Gao, Zifu Li, Ling Wang, Cong Chen, Davaa Basandorj, and Tianxin Li. “Mining of the Association Rules between Socio-Economic Development Indicators and Rural Harmless Sanitary Toilet Penetration Rate to Inform Sanitation Improvement in China.” Frontiers in Environmental Science 10 (April 14, 2022). https://doi.org/10.3389/fenvs.2022.817655.

Mansor, Maslina, Mohamad Fazli Sabri, Mustazar Mansur, Muslimah Ithnin, Amirah Shazana Magli, Abd Rahim Husniyah, Nurul Shahnaz Mahdzan, et al. “Analysing the Predictors of Financial Stress and Financial Well-Being among the Bottom 40 Percent (B40) Households in Malaysia.” International Journal of Environmental Research and Public Health 19, no. 19 (September 30, 2022): 12490. https://doi.org/10.3390/ijerph191912490.

Munisamy, Ananthan, Sheerad Sahid, and Muhammad Hussin. “A Financial Literacy Model of Malaysian B40 Households: The Case of Financial Well-Being, Education Level and Socioeconomic Status.” International Academic Symposium of Social Science 2022 46 (September 20, 2022): 64. https://doi.org/10.3390/proceedings2022082064.

Othman, Zulaiha Ali, Azuraliza Abu, Nor Samsiah, and Jamaludin Sallim. “Household Overspending Model amongst B40, M40 and T20 Using Classification Algorithm.” International Journal of Advanced Computer Science and Applications 11, no. 7 (2020). https://doi.org/10.14569/ijacsa.2020.0110751.

Redjeki, Sri, M. Guntara, and Pius Anggoro. “Naive Bayes Classifier Algorithm Approach for Mapping Poor Families Potential.” International Journal of Advanced Research in Artificial Intelligence 4, no. 12 (2015). https://doi.org/10.14569/ijarai.2015.041205.

Sabri, M. F., M. A. Othman, J. Masud, L. Paim, M. McDonald, and T. K. Hira. “Financial Behavior and Problems among College Students in Malaysia: Research and Education Implication.” Universiti Putra Malaysia Institutional Repository, 2008.

Sabri, Mohd Fazli. “Mengurus Bajet 2019 Anda Dan Keluarga.” UTAMA, 2019. https://inspem.upm.edu.my/article/mengurus_bajet_2019_anda_dan_keluarga-47451.