

Review Article

Study of B40 Schoolchildren Lifestyles and Academic Performance using Association Rule Mining

Puteri NE Nohuddin^{1*}, Zuraini Zainol² and Mohd Hanafi Ahmad Hijazi³

¹Institute of IR4.0, National University of Malaysia, Malaysia

puteri.iv@ukm.edu.my

²Department of Computer, Universiti Pertahanan Nasional Malaysia, Malaysia

zurainiz@upnm.edu.my

³Faculty of Computing and Informatics, Universiti Malaysia Sabah, Malaysia

hanafi@ums.edu.my

*Correspondence: puteri.iv@ukm.edu.my

Received: 17th October 2019; Accepted: 20th December 2020; Published: 20th March 2021

Abstract: B40 community school children experience many underprivileged lifestyles which impacted their academic performance. Through data trends and patterns, the education top management can observe the progress of academic performance and the lifestyle relationships of students in a school or nearby school. It can also help to identify the causes of progress or deterioration of the performance of B40 community students. Therefore, a data analytics framework is essential to help decision-makers to see and analyze the changing trends and patterns of academic progress in data and related lifestyles of B40 community students more effectively and accurately. The objective of this study is to design and develop association rules analysis to deduce the relevance of academic achievement and lifestyle among schoolchildren from the B40 family. The analysis framework is established in several stages that involve data collection and processing and transformation, then the design, application and evaluate of the association rules algorithms. The framework is expected to benefit students, teachers and Education Ministry. This study foresees whether educational programs and healthy lifestyle awareness can be designed specifically for the B40 children so as to improve their academic achievement as desired by the government.

Keywords: *Data mining; Association Rules Mining; B40; Academic Performance; Attribute relationships*

1. Introduction

The government of Malaysia emphasizes the importance of education to all its citizen; this is one of important plans as stated in the 11th RMK. One of the initiatives stresses on education and children especially to the B40 community. To improve lifestyle is to raise the people's economic standards to ensure that the country's economy continues to grow. Therefore, B40 community children must succeed in education or at school or college for better future. Academic monitoring is an important effort that must be made by the school and parents to oversee the children's academic performance. Ministry of Education Malaysia (MOE) has introduced a School Exam Analysis System (SAPS) that maintains the process of school examination analysis. To date there is no system or technique that can help MOE to monitor the progress of students from the B40 community holistically. The introduction of data mining techniques, it allows to analyze patterns and trends of academic progress of schoolchildren and their lifestyles. Through these analyses, the top management are to inspect the progress and academic performance and the lifestyle relationships of students in a school or nearby

school. It also helps to identify the causes of progress or deterioration of the performance of B40 community students. Therefore, the proposed Association Rule based Pattern Analysis framework is essential to help decision-makers to see and analyse the changing trends and patterns of academic progress in data and related lifestyles of B40 community students more effectively and accurately. The objective of this study is to design and develop association rules technique to deduce the relevance of academic achievement and lifestyle among schoolchildren from the B40 family. The framework is developed in several phases: (i) data collection (ii) data processing (iii) data transformation (iv) design and application of the association rules algorithms and (v) evaluation of algorithms.

The rest of the paper is organized as follows. Section 2 discusses a few related topics that are described a group of B40 community schoolchildren studies and data mining techniques literature review. Then Section 3 delivers an explanation of the modules in the projected framework for Association Rule based Pattern Analysis. This is followed by Section 4 which describes the experimental setup of the association rules analysis using a set of survey data and the generated results. Finally, Section 5 is the conclusion section with a summary and future research work.

2. Background and Related Work

In Malaysia, B40 community refers to people or families of households with monthly income of RM3,900 and below. B40 community is growing in numbers, especially in the cities. While earnings per capita has been increasing over the past three years, it should not be overlooked the rising costs caused by inflation and slow wage growth. Precisely, the increase of inflation from year to year (3.5%, 2.1%, 2.09% and 4.1% respectively in 2014, 2015, 2016 and the first half of 2017) thus the ringgit's value could have dropped [1]. A few research findings show B40 children are exposed to many disadvantages. A UNICEF's study [2] disclosed how poverty impacted children who live in low-cost flats in Kuala Lumpur. Among focal findings of this study include (i) 99.7% children in low-cost flats live in relative poverty and absolute poverty, (ii) 15% of these children below 5 years old are underweight, (iii) 22% of the children are undersized and about 23% are either overweight or obese, (iv) whereas most children aged 7 to 17 are schooling, but only 50% of children aged between 5 to 6 year olds attend pre-school, (v) Approximately 1 in 3 households disclose that they did not buy reading materials for their children aged below 18 and (vi) 4 in 10 households do not buy toys for their children aged below 5. In Asia, there are many similar studies initiatives and programs are implanted to assist under privileged children in their academic so as to uplift their economy and indirectly would improve their lifestyle [3, 4, 5].

One of the initiatives is highlighted in the Eleventh Malaysia Plan for the 2016-2020 is to raise the people's economy to ensure that the country's economy continues to grow. The focus of this initiative is to raise economic standards to reduce poverty and socio-economic imbalances, especially among the lowest 40 percent income households (B40) [6]. This report also placed Malaysia as the 14th country in terms of achieving development from the education sector.

School Examination Analysis System (SAPS) is a system that analyzes school examinations. This system is developed to store, collect and analyze data on student examination results in schools. This system is developed online that can be accessed by all parties including MOE, State Education Department, District Education Office and schools. SAPS is developed to facilitate parents to review their child's test results. The SAPS revision system displays student results slip and scores, thus facilitating parents to monitor their child's academic performance. SAPS measures the achievement of a test and detects students' weakness as SAPS can produce such analyzed data. Data is also shared by all subject teachers as each teacher will enter a score in the system. However, SAPS does not provide a more effective and secure analysis module.

In some developing countries, there are many achievements of different levels on academic performance between schools, which are based on the differences in socio-economic background and student lifestyle that will lead to significant differences in student achievement in school [3]. The B40 community's lifestyle and economic issues have also played an important role. Some parents are less concerned about the management of children's school. Furthermore, total income of parents as a

whole greatly influences in carrying out the duties and responsibilities of a person, as well as the responsibility of parents to their children in the education process. Therefore, lifestyle and academic performance relationships are important [5].

MOE, teachers and parents should monitor and analyze the progress of the B40 community students carefully and deeply to identify the trends and patterns of student performance progress and lifestyle habits. Techniques can be used in this study are data mining techniques which are technologies where new knowledge can be extracted from historical academic data and lifestyle. These techniques are essential to helping teachers and education officers to look and analyze performance data and lifestyle relationships for a change in progress academically in B40 community student data effectively and thoroughly.

Data mining has several techniques that can be used for analysis purposes depending on the type of data used. Among the techniques often used to analyze data are association rules, clustering and also classification techniques. There is a previous study aimed at conducting an analysis of students who obtained low grades in their assessment at Sultan Zainal Abidin University (UniSZA), Terengganu, Malaysia. The study makes comparisons between the three (3) classification algorithms selected; Decision Tree, I Bayes (NB), and Rule Based [7]. Student performance forecasts are more challenging because of the large amount of data in the educational database. Most recently in Malaysia, there is still a lack of existing systems to analyze and monitor student progress and performance. Amirah et al [8] has made comparisons on data mining techniques that have been used to forecast student performance. They also focus on how predictive algorithms can be applied to discover most important attributes in student data. Another study Arora and Badal [12], they applied Association Rule Mining in investigating academic performance of college students. This study will use the relationship regulation analysis technique [9, 10, 11] to look at the pattern of trends that exist in academic achievement among children of the B40 community.

3. The Overview of Association Rule Based Pattern Analysis Framework

In this study, we proposed an Association Rule based Pattern Analysis Framework that is proficient for extracting frequent patterns and rules from a set of data. Figure 1 illustrates the framework of Association Rule based Pattern Analysis (ARPA). Technically, it encompasses of two (2) main components: (i) Association Rule Miner Module and (ii) Ruled based Pattern Analysis as its output. The first component describes the collection of surveys which have been answered by school children and parents among B40 community, and a process of data pre-processing for cleaning and formatting these surveys into a cleaned, formatted, and structured dataset.

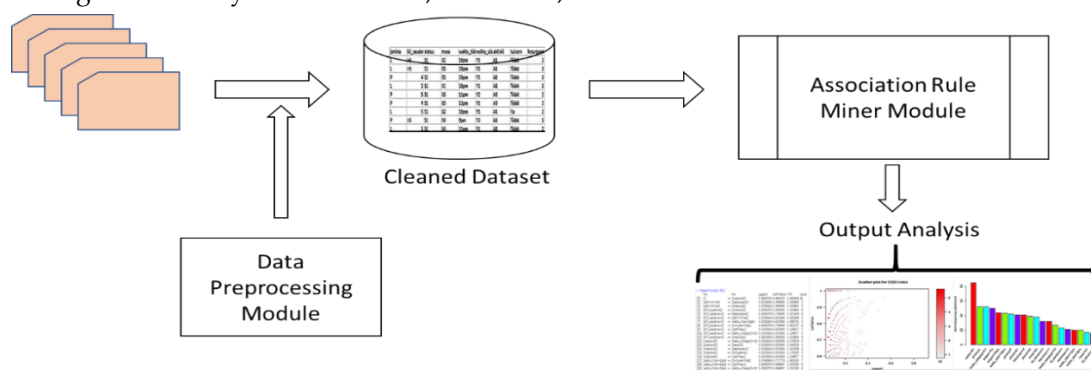


Figure 1. Association Rule based Pattern Analysis Framework

Data preprocessing is designed as a method for discovering inaccurate data from the datasets and substituting, or amending, or removing them into a cleaned data. Therefore, this reduces data redundancy and improves data reliability. In the next stage, Association Rules Mining [13] is applied to investigate the data and to extract the significant rules within the datasets. Then in the output analysis stage, frequent patterns and rules are generated and analyzed.

The generated rules are expected to assist school management and MOE to investigate academic performance in relation to their family's way of living. These patterns and rules are presented in graphs for better insights representations.

Table 1 describes the stages and activities in the proposed ARPA framework. Association Rules Mining is used to uncover the attribute relationships in the data survey.

Table 1. Activities in ARPA

Stage	Activities
1	Surveys are answered by B40 schoolchildren from 2 schools.
2	Data preprocessing is applied to the survey data for cleaning some missing data and standardizing the data format.
3	Frequent patterns and association rules are generated from the cleaned survey data.
4	Frequent patterns and association rules are visualized in graphs for better output analysis

4. Experimental Setup and Results

In this section, the experimental setup and the results are elaborated that include the description of datasets and its schema or data ranges. Then, the results of frequent patterns and association rules are briefly described later in the section.

4.1. Datasets

In this study, the surveys are conducted at two (2) neighboring schools at Sungai Besi, Kuala Lumpur (i) Sekolah Kebangsaan Sg Besi and (ii) Sekolah Kebangsaan Bandar Tasik Selatan. The surveys are conducted in Bahasa Melayu as it is concerned on the school children's understanding on the survey questions. The questions cover on the family status, time spent during after school with families, the activities, extra classes such as tuition and examinations results. After data preprocessing stage, the survey questions have been cleaned and formatted as a cleaned dataset (in Table 2) that consists of 94 records and 34 derived attributes (items) describing student profiles, activities, examination results and sleeping patterns (Figure 3).

Table 2. Parts of Cleaned Survey Dataset

Jantina	Bil_saudara	status	masa	Waktu_tidur	Waktu_ulangkaji	aktiviti	tuisyen	keputusan
L	>5	S1	X2	10pm	Y1	A1	Tidak	2
L	>5	S1	X5	10pm	Y1	A1	Tidak	3
P	4	S1	X5	10pm	Y1	A2	Tidak	3
L	2	S1	X1	10pm	Y1	A3	Tidak	3
P	4	S1	X5	11pm	Y2	A3	Tidak	2

Some attributes are formatted into nominal data to represent the data ranges as shown in Table 3 below.

Table 3. Details of Selected Attributes and Data Ranges

Attribute	Pelajar1	Pelajar2	Pelajar3	Pelajar4	Pelajar5
Masa bersama ibu bapa (sehari) Time spent with parents (daily)	X1= 2-3 jam	X2= 3-4 jam	X3= 4-5 jam	X4= 5-6 jam	X5=>6 jam
Waktu ulangkaji pelajaran Revision time	Y1=1-2jam	Y2= 2-3 jam	Y3= 3-4 jam	Y4= 4-5 jam	Y5=>5 jam
Aktiviti lepas sekolah After school activities	A1=bersukan (sports)	A2=menonton Tv (watch tv)	A3= aktiviti bebas (free activity)	A4= ulang kaji (revision)	<Null>
Status Status	S1=Ibu bapa berkahwin (Parents married)	S2=Ibu bapa bercerai (Parents divorced)	S3=ibu bapa meninggal (Parents passed away)	<Null>	<Null>
Keputusan Akademik Academic result	1=cemerlang (excellent)	2= memuaskan (Satisfactory)	3= Kurang Memuaskan (Less Satisfactory)	<Null>	<Null>

4.2 Association Rules

First stage of the framework, we run the Apriori Algorithm, it extracts out the frequent itemsets from the survey data. The association rules extractor engine is developed using R language. The dataset is rather sparse with a density just above 26.4%, the result shows item = "status" (relationship of the parents) is the most popular item. Among the highest frequencies of selected items are presented in Figure 3.

```
> summary(Danel)
transactions as itemMatrix in sparse format with
94 rows (elements/itemssets/transactions) and
34 columns (items) and a density of 0.2647059

most frequent items:
      status=S1      jantina=L waktu_ulangkaji=Y1      tuisyen=Ya      tuisyen=Tidak      (Other)
           85           52           52           50           44           563
```

Figure 3. A screenshot for item frequency

Figure 4 presents the item frequency histogram that illustrates the occurrence of an item has occurred in the dataset as compared to the other items. The 20 top frequent items are shown in the graph and 3 top frequent items are 'status=berkahwin' (81), 'jantina=L' (52) and 'waktu_ulangkaji=Y1' (52) had the highest frequencies out of the total 94 records in the dataset.

In this experiment, we set the minimum support threshold as 0.05 and the minimum confident threshold as 0.6. As shown in Figure 5, 2300 association rules are generated. These rules are then ranked based on the value of 'support' parameter. Figure 6 illustrates top 20 association rules from 2300 rules generated from the survey datasets. A massive number of rules are generated due to bigger number of columns existed in the datasets. However, not all rules are important nor interesting. Thus, we need to analyse the generated rules based on the combination itemsets, support, confidence and lift values.

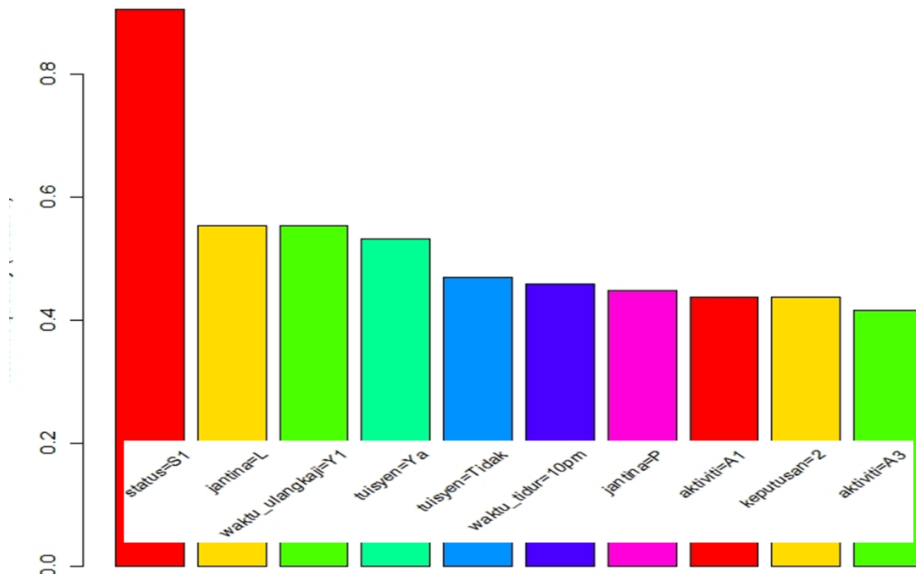


Figure 4. Item frequency histogram

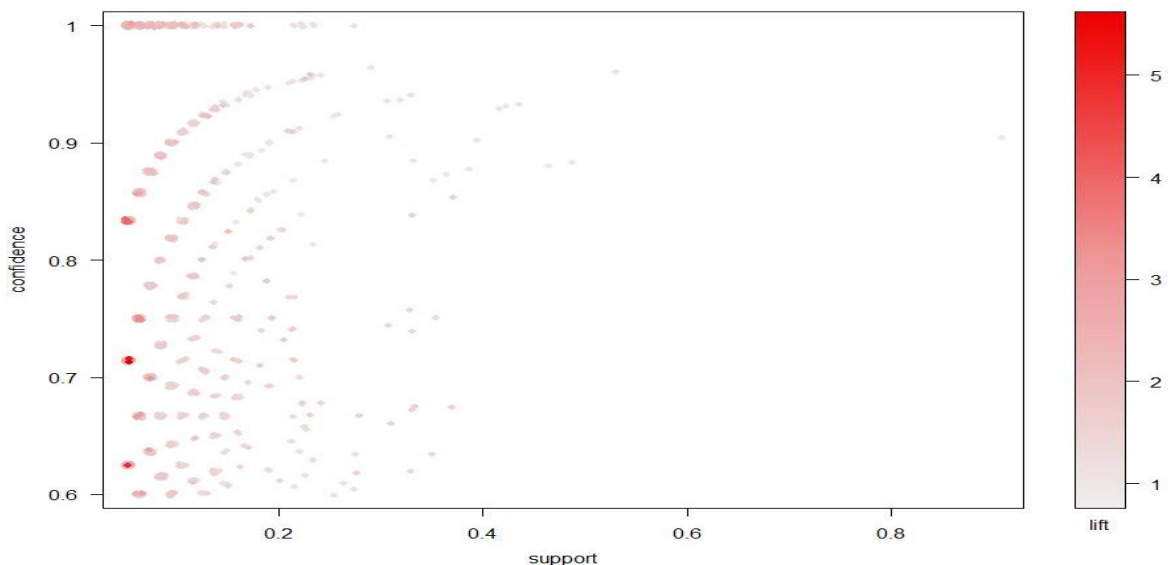


Figure 5. A scatter plot for 2300 association rules

```
> inspect(rules[1:30])
  lhs      rhs      support  confidence  coverage  lift  count
[1] {} => {status=S1} 0.90425532 0.9042553 1.00000000 1.00000000 85
[2] {aktiviti=A4} => {keputusan=2} 0.05319149 1.0000000 0.05319149 2.2926829 5
[3] {aktiviti=A4} => {status=S1} 0.05319149 1.0000000 0.05319149 1.1058824 5
[4] {bil_saudara=1} => {status=S1} 0.06382979 1.0000000 0.06382979 1.1058824 6
[5] {bil_saudara=>5} => {keputusan=3} 0.06382979 0.7500000 0.08510638 3.3571429 6
[6] {bil_saudara=>5} => {aktiviti=A1} 0.05319149 0.6250000 0.08510638 1.4329268 5
[7] {bil_saudara=>5} => {waktu_tidur=10pm} 0.05319149 0.6250000 0.08510638 1.3662791 5
[8] {bil_saudara=>5} => {tuisyen=Tidak} 0.06382979 0.7500000 0.08510638 1.6022727 6
[9] {bil_saudara=>5} => {jantina=L} 0.05319149 0.6250000 0.08510638 1.1298077 5
[10] {bil_saudara=>5} => {waktu_ulangkaji=Y1} 0.05319149 0.6250000 0.08510638 1.1298077 5
[11] {bil_saudara=>5} => {status=S1} 0.08510638 1.0000000 0.08510638 1.1058824 8
[12] {status=S2} => {waktu_ulangkaji=Y2} 0.05319149 0.6250000 0.08510638 2.1759259 5
[13] {status=S2} => {masa=X5} 0.05319149 0.6250000 0.08510638 1.5460526 5
[14] {status=S2} => {keputusan=2} 0.05319149 0.6250000 0.08510638 1.4329268 5
[15] {status=S2} => {tuisyen=Ya} 0.05319149 0.6250000 0.08510638 1.1750000 5
[16] {status=S2} => {jantina=L} 0.05319149 0.6250000 0.08510638 1.1298077 5
[17] {waktu_tidur=12pm} => {tuisyen=Tidak} 0.07446809 0.7777778 0.09574468 1.6616162 7
[18] {waktu_tidur=12pm} => {jantina=L} 0.06382979 0.6666667 0.09574468 1.2051282 6
[19] {waktu_tidur=12pm} => {waktu_ulangkaji=Y1} 0.06382979 0.6666667 0.09574468 1.2051282 6
[20] {waktu_tidur=12pm} => {status=S1} 0.08510638 0.8888889 0.09574468 0.9830065 8
```

Figure 6. Examples of 20 rules extracted from 2300 rules generated by Apriori Algorithm

In Figure 7, suppose that we are interested in only rules with the right hand side (RHS) indicating “keputusan” which relates to the academic performance, we apply the following code such as rhs=c(“Keputusan=1”, “keputusan=2”, Keputusan=3”) in appearance to make sure that only =1”, =2” and =3” will appear in the RHS of rules. On the other hand, the LHS of rule is set with default = “LHS”. As a result, based on RHS in relation to item “keputusan={1,2,3}, all other items can appear in the LHS.

```
> rules_2 <- apriori(Dane1, parameter = list(minlen=5, support = 0.04, confidence = 0.75),
+ appearance = list(default = "lhs", rhs = c("keputusan=1","keputusan=2","keputusan=3"),
+ default = "lhs"), control = list(verbose = F))
> rules.sorted <- sort(rules_2, by="lift")
> # display number of rules
> rules.sorted
set of 61 rules
> # display all rules
> inspect(rules.sorted)
  lhs      rhs      support  confidence  coverage  lift  count
[1] {jantina=P, waktu_ulangkaji=Y1, aktiviti=A3, tuisyen=Tidak} => {keputusan=3} 0.04255319 1.0000000 0.04255319 4.476190 4
[2] {jantina=P, masa=X5, waktu_ulangkaji=Y1, aktiviti=A3, tuisyen=Tidak} => {keputusan=3} 0.04255319 1.0000000 0.04255319 4.476190 4
[3] {jantina=P, status=S1, waktu_ulangkaji=Y1, aktiviti=A3, tuisyen=Tidak} => {keputusan=3} 0.04255319 1.0000000 0.04255319 4.476190 4
[4] {jantina=P, status=S1, masa=X5, waktu_ulangkaji=Y1, aktiviti=A3, tuisyen=Tidak} => {keputusan=3} 0.04255319 1.0000000 0.04255319 4.476190 4
```

Figure 7. Parts of rules query for 61 association rules

The result on the rules query is a set of 61 association rules. The top three rules with respect to the lift measure, a popular measure of rule strength, are listed in Table 4. We can deduce that high frequencies, with approximate 0.043 support, 1.0 confidence and 4.47 lift values, rules on a group of school female children who spend 1-2 hours of studying, spend their spare time with free activities and they don’t attend tuition class are having ‘less satisfactory’ (“kurang memuaskan”) examination results. Also the second highest rule on a group of school female children who spend 1-2 hours of studying, spend their spare time with free activities and they don’t attend tuition class are having not satisfied examination results however spend approximate 6 hours with the family. The third rule is similar but with additional parents’ status = married. We can also assume that these rules are parent-child rules.

Table 4. Top three rules of 61 association rules

Rules	support	confidence	lift
{jantina=P,waktu_ulangkaji=Y1,aktiviti=A3,tuisyen= Tidak} => {keputusan=3}	0.04255	1.00	4.476
{jantina=P,masa=X5,waktu_ulangkaji=Y1,aktiviti=A3,tuisyen=Tidak} => {keputusan=3}	0.04255	1.00	4.476
{jantina=P,status=S1,waktu_ulangkaji=Y1,aktiviti=A3,tuisyen=Tidak} => {keputusan=3}	0.04255	1.00	4.476

Figure 8 shows an interactive plot graph that presents 61 association rules. In the graph, the confidence levels are plotted on the Y axis whereas the support levels on the X axis for each rule. The

details of each rule can be scrutinized by pointing the cursor on the interactive plot graph (which cannot be presented in this paper). As shown in Figure 8, the interactive plot graph recommends that rules that are concentrated in the circle contain about 18 rules (only 2 rules are shown in Figure 8) which are listed below the scatter plot.

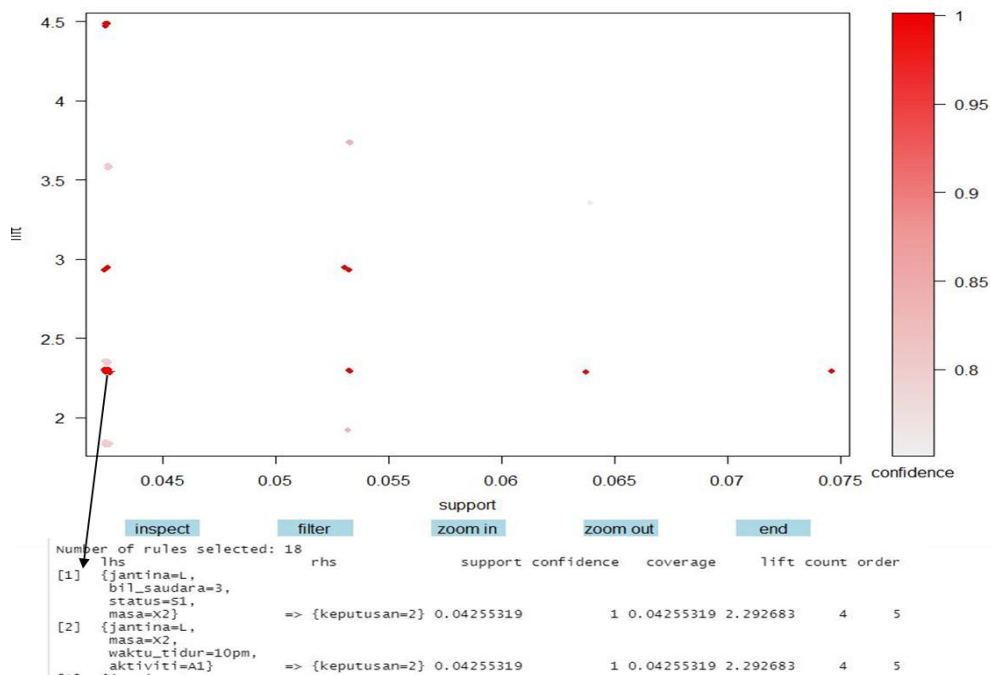


Figure 8. The scatter plot graph

Figure 9 shows an interactive scatter plot that follows these processes:

- Individual rules inspection by choosing each rule and clicking the inspect button.
- Sets of rules inspection by choosing a rectangular region of the plot and clicking the inspect button.
- Selected region inspection (zoom in/zoom out buttons).

The outcome of an example interaction is shown in Figure 9. A cursor is moving to the selected bubble, and by clicking the inspect button, the rules are displayed.

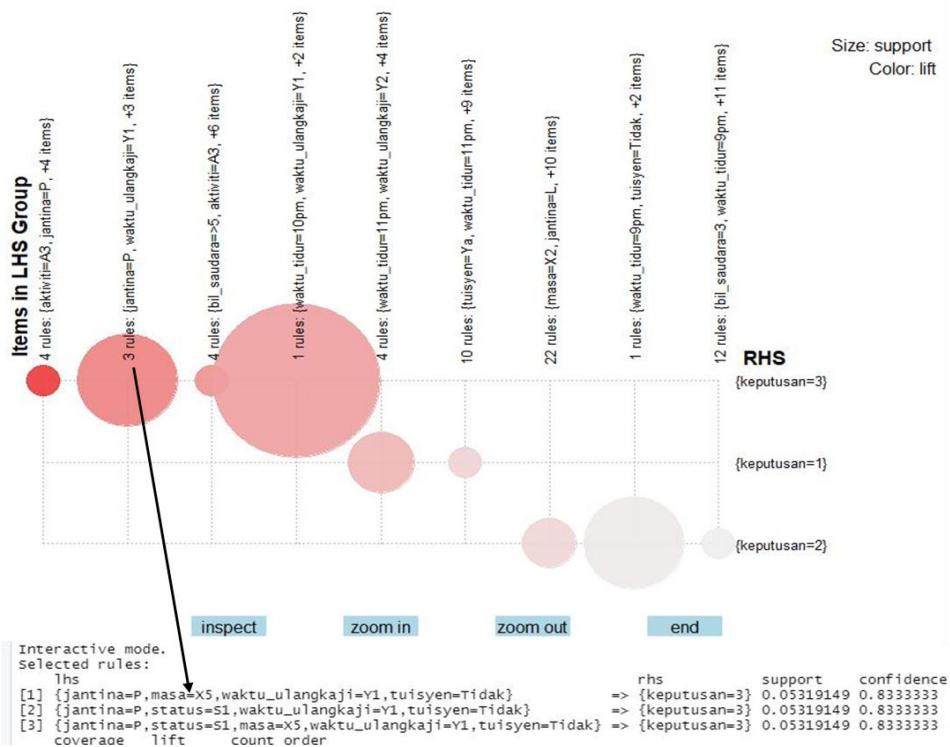


Figure 9. Interactive mode for scatter plot (inspecting rules with high lift).

We also explore further on the rules to investigate on school children with excellent examination results. The rules parameters are defined as in Figure 10 and about 10 rules are generated. However, the parameters on LHS can be improved to extract more accurate and reliable rules.

```
rules_selected <- apriori(Dane1, control = list(verbose=F),
  parameter = list(minlen=2, support=0.001, confidence=0.06),
  appearance = list(default="none", rhs = c("keputusan=1"),
    lhs=c("jantina=L","jantina=P","waktu_ulangkaji=Y1","waktu_ulangkaji=Y2","waktu_ulangkaji=Y3"))

rules.sorted_selected <- sort(rules_selected, by="lift")
summary(rules.sorted_selected)
inspect(rules.sorted_selected)
> inspect(rules.sorted_selected)
```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{jantina=P,waktu_ulangkaji=Y2}	=> {keputusan=1}	0.08510638	0.5714286	0.1489362	1.6785714	8
[2]	{waktu_ulangkaji=Y2}	=> {keputusan=1}	0.14893617	0.5185185	0.2872340	1.5231481	14
[3]	{jantina=L,waktu_ulangkaji=Y2}	=> {keputusan=1}	0.06382979	0.4615385	0.1382979	1.3557692	6
[4]	{jantina=L,waktu_ulangkaji=Y3}	=> {keputusan=1}	0.04255319	0.3636364	0.1170213	1.0681818	4
[5]	{jantina=P}	=> {keputusan=1}	0.15957447	0.3571429	0.4468085	1.0491071	15
[6]	{jantina=L}	=> {keputusan=1}	0.18085106	0.3269231	0.5531915	0.9603365	17
[7]	{jantina=P,waktu_ulangkaji=Y1}	=> {keputusan=1}	0.07446809	0.2916667	0.2553191	0.8567708	7
[8]	{waktu_ulangkaji=Y3}	=> {keputusan=1}	0.04255319	0.2857143	0.1489362	0.8392857	4
[9]	{waktu_ulangkaji=Y1}	=> {keputusan=1}	0.14893617	0.2692308	0.5531915	0.7908654	14
[10]	{jantina=L,waktu_ulangkaji=Y1}	=> {keputusan=1}	0.07446809	0.2500000	0.2978723	0.7343750	7

Figure 10. Association rules in which keputusan = 1 (excellent)

5. Conclusion

In conclusion, the Association Rule based Pattern Analysis Framework was established to generate frequent patterns and rules from the survey dataset in order to find out the relationship between B40 community school children lifestyle and academic performance. The framework is developed in several phases: (i) data collection (ii) data processing (iii) data transformation (iv) design and application of the association rules algorithms and (v) evaluation of algorithms. In line with the objective of this study is to design and develop association rules technique to deduce the relevance of academic achievement and lifestyle among schoolchildren from the B40 family. The findings support that if B40 families spend less attention to having extra hours of studying after school and attending class tuition undeniably affects their children academic performance. In future, this research is expected to develop a clustering technique for grouping similar rules in clusters. Therefore, it can assist the decision makers in ministry and schools to interpret the output analysis in more effective manner especially in providing support on extra classes and awareness on family lifestyle in relation to the importance of their children academic progress.

Acknowledgement

This research is supported by Geran Universiti Penyelidikan Universiti Kebangsaan Malaysia with grant number GUP-2018-076.

References

- [1] CompareHero. Golongan Isi Rumah B40, M40 dan T20 di Malaysia Berdasarkan Pendapatan. Available: <https://www.comparehero.my/budgets-tax/articles/klasifikasi-pendapatan-di-malaysia-t20-m40-dan-b40>. 2020. [Accessed online on 1 Aug 2019].
- [2] UNICEF Malaysia & DM Analytics. Children without: A study of urban child poverty in low cost flats in Kuala Lumpur. February 2018. Available: <https://www.unicef.org/malaysia/reports/children-without> [Accessed online on 1 Aug 2019].
- [3] J. Kohyama. Self-Reported Academic Performance and Lifestyle Habits of School Children in Japan. *International Journal of Child Health and Nutrition*, pg 90-97, 2017.
- [4] P. Banerjee. A systematic Review of Factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*. 2016.
- [5] Khairunnisa, F. Sabrian, and Safri. Hubungan Gaya Hidup Dengan Prestasi Akademik Mahasiswa Keperawatan Universitas Riau. *Jurnal Online Mahasiswa*, Vol 2 No 2, Oktober 2015.
- [6] A. Zainol. Elevating B40 Households Towards a Middle-Class Society. *Economic Planning Unit, Prime Minister Department*. 2016.
- [7] F. Ahmad, N. H. Ismail dan A. Abdul Aziz. The Prediction of Students: Academic Performance Using Classification Data Mining Techniques. *Applied Mathematical Sciences*, Vol. 9, no. 129, pages 6415-6426. 2015.

- [8] A. Mohamed Shahiri, W. Husain dan N. Abdul Rashid. A Review on Predicting Student`s Performance Using Data Mining Techniques. *Procedia Computer Science*, volume 72, pages 414-422, 2015.
- [9] P. Nohuddin, C. Frans, R. Christley and C. Setzkorn. Finding interesting trends in social networks using frequent pattern mining and self organizing maps. *Knowledge-Based Systems*, volume 29, pages 104-113. 2012.
- [10] B. Kumar and K. Rukmani. Implementation of Web Usage Mining Using APRIORI and FP Growth Algorithms. *Int. J. of Advanced networking and Applications*, 404(6), 400-404. Retrieved from <http://ijana.in/papers/6.11.pdf>. 2010.
- [11] Z. Zainol, M. T.H. Jaymes and P. Nohuddin. VisualUrText: A Text Analytics Tool for Unstructured Textual Data. *Journal of Physics: Conference Series*, Volume 1018, 2018.
- [12] R. Kumar Arora and D. Badal. Mining Association Rules to Improve Academic Performance. *International Journal of Computer Science and Mobile Computing*. Vol. 3, 2014.
- [13] I. H. Witten, I. H., E. Frank, M. Hall and C. Pal, C. *Data Mining (4th Edition) Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. 2016.



© 2020 by the author(s). Published by Annals of Emerging Technologies in Computing (AETiC), under the terms and conditions of the Creative Commons Attribution (CC BY) license which can be accessed at <http://creativecommons.org/licenses/by/4.0>.