Analyzing the Influence of Socio-Demographic Characteristics on Travel Mode Choice in Banda Aceh City Through Association Rules

[1] Meutia Ayuni Yose, *[2] Hakan Aslan and [3] Zeliha Çağla Kuyumcu
[1, *2 and 3] Faculty of Engineering, Department of Civil Engineering Sakarya University, Turkey

Abstract:

In formulating improvements to integrated transportation planning for the Banda Aceh city which has been urbanizing speedily and has experienced considerable changes in growing economy and motorization, it is necessary to understand the factors that influence community choice of travel modes, one of which is the influence of the socio-demographic characteristics of the community. This is due to the importance of knowing the characteristics and behaviour of modal users in designing transportation facilities and infrastructure to be provided. In this study, association rules are used to analyse the influence of demographic characteristics on the choice of travel modes in the Banda Aceh city. The data used in this study came from survey data conducted by the Aceh Transportation Agency in 2019. After the data analysed through setting up association rules, it shows that role of socio-demographic variables on the choice of travel modes is quite significant in many aspects.

Key words: Mode Choice, Socio-Demographic Characteristics, Association Rules, Apriori Algorithm

1. Introduction

Transportation is a movement from the place of origin to the destination and occurs because of the process of fulfilling needs. The variety of human needs in terms of moving to a place affect the tendency in the choice of mode.

The choice of transportation mode is one of the most important things in transportation planning. This is because it is important to know the characteristics and behaviour of mode users in designing the transportation facilities and infrastructure that will be provided.

Banda Aceh city, the capital of Aceh Province, is one of the medium sized cities with the population around 270,321inhabitants and consists of 9 sub-districts and 70 villages [1]. As with many medium cities in Indonesia, Banda Aceh city has been urbanizing speedily and has experienced considerable changes in growing economy and motorization. In formulating improvements to integrated transportation planning for the city of Banda Aceh, it is necessary to understand the

*Corresponding author: Address: Faculty of Engineering, Department of Civil Engineering Sakarya University, 54187, Sakarya TURKEY. E-mail address: haslan@sakarya.edu.tr, Phone: +905064714574

factors that influence community choice of travel modes which is one of the characteristics of travel behaviour. One of which is the influence of the socio-demographic characteristics of the community i.e. age, gender, income group, etc.

The choice of the transportation modes in urban area is a random process influenced by several factors, one of which is economical social status. The level of one's economical social status is determined by the education, occupation, and income.

2. Materials and Method

2.1. Materials

The data used in this study came from survey data conducted by the Aceh Transportation Agency in 2019 about railway feasibility study in Banda Aceh city. The survey was conducted to see respondents' perceptions if railroad transportation operates in the city of Banda Aceh. However, in this paper the data used are the socio-demographic data of the community and the types of vehicles used today so that the socio-demographic relationship between the community and the choice of the mode of transportation used will be seen through setting up association rules.

2.2. Method

Before setting up association rules, there is a step called data pre-processing, which is one of the most important first steps in data mining. Pre-processing of a data set is required if the data set consists of data that is unnecessary and incomplete (missing), noisy (outlier) and inconsistent data [2]. This step can be seen in Figure 1 below:
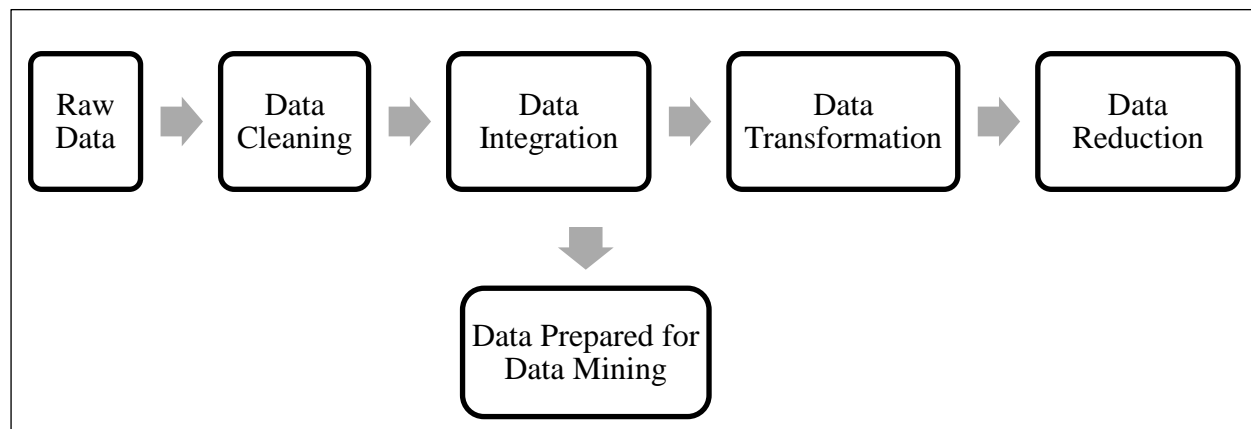

Figure 1. Preprocessing data [2]

2.2.1. Raw data

In the raw data about railway feasibility study in Banda Aceh city, there is socio-demographic information regarding respondents, namely age, gender, household income, level of education, and mode choice. In addition, there is also information on the character of the trip in the form of transportation costs incurred in a month, the purpose of the trip, the number of trips in a month,

and the type of transportation used. In the data set, there were 2280 respondent data, but because some of the respondent's data were incomplete and some of the respondent's information was not used in this study, the data was cleaned so that there were 2231 respondents.

2.2.2. Data cleaning

In this study, the following arrangements have been made on the data set, considering that it is appropriate to remove the record with the lost data from the database.

Because data such as trip origin, trip destination, travel time and others could not be combined and could not give meaningful relationships, it was removed from the data set. In addition, parts of the data that were left blank and expressed ambiguously belonging to some features were removed from the data set during the data cleaning phase.

There were 49 respondent data that were excluded from the data set because some parts of the information were blank and incorrect (erroneous). Hence, after cleaning the data, there are 2231 respondents' data left. While performing these operations, filter feature in Ms. Excel program was used.

2.2.3. Data arrangement and creating data set

The distributions of the data according to the features in the data set obtained after the data cleaning process are given in Figures 2, 3, 4, 5 and 6.
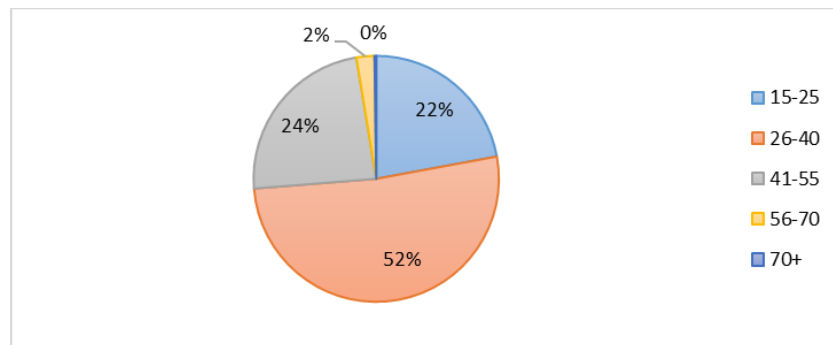

Figure 2. Distribution of data based on the age of the respondents

From the distribution in Figure 2 it can be seen that the largest number of travellers are people aged 26-40 years by 52% and the least travelled is people aged 70+ by 0.2%.
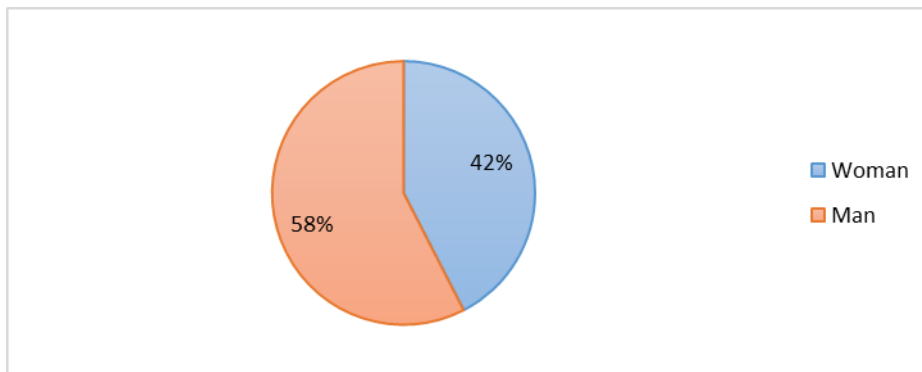
Figure 3. Distribution of data based on gender of respondents

From the distribution in Figure 3, it can be seen that the largest number of travelers are men, which is 58% of the total respondents who traveled.

Figure 4. Distribution of data based on household income of respondents

From the distribution in Figure 4 it can be concluded that the income of most of the people of Banda Aceh city is IDR 20,000.00 - 2,415,000.00 which is 59%. This income value is below the provincial minimum wage set by the Governor of Aceh, which is IDR 3,165,031.00 [2].
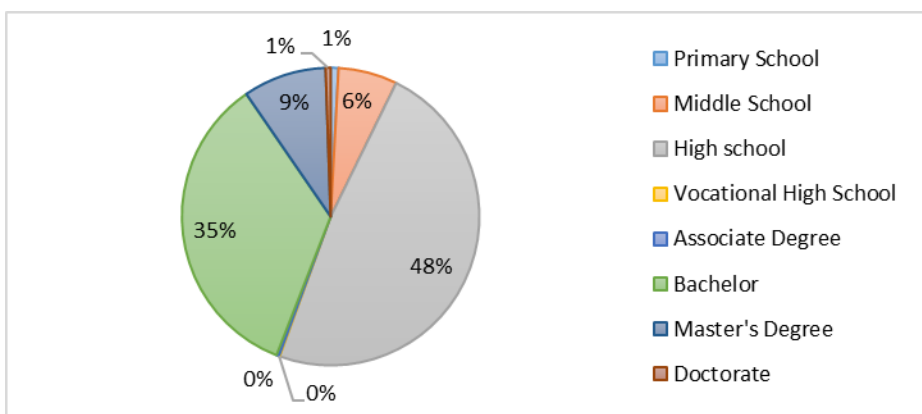
Figure 5. Distribution of data based on level of education of respondents

From the distribution in Figure 5 above, it can be seen that the most dominant education level of

the people of Banda Aceh and its surroundings is high school by 48%, followed by bachelor graduates by 35%. While the lowest education levels are vocational high school, associate degree, and doctorate whose respective percentages do not even reach 1%.
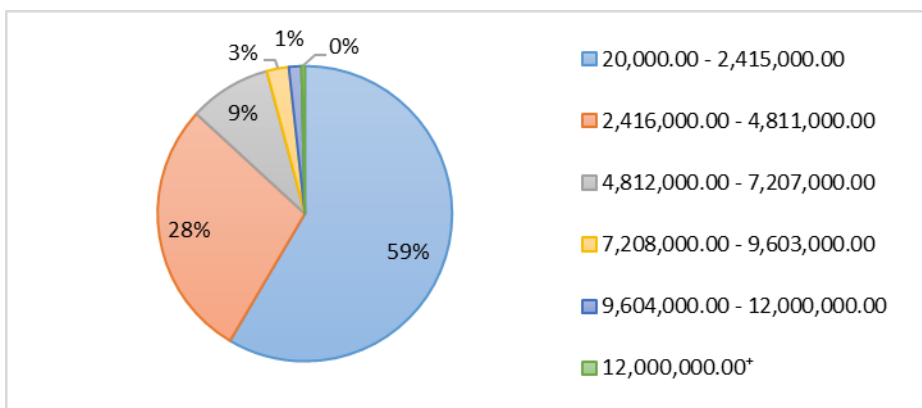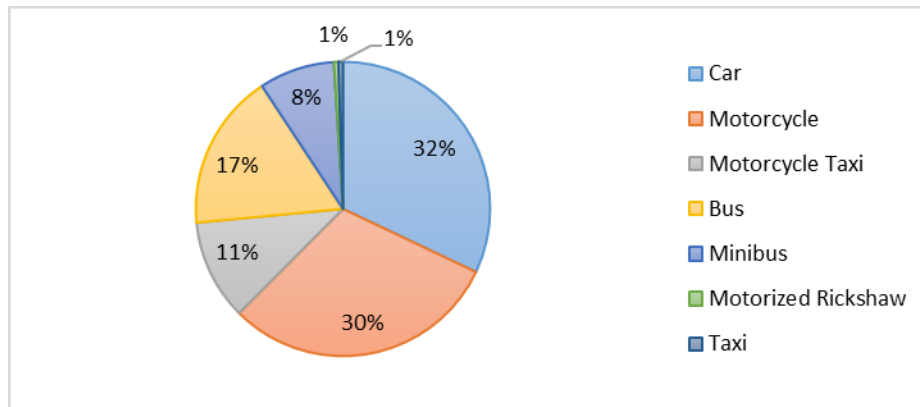

Figure 6. Distribution of data based on mode choice of respondents

From the distribution in the Figure 6, it can be concluded that the most common types of transportation used by the community are private vehicles, namely cars by 32% and motorbikes by 30%. Meanwhile, motorized rickshaws and taxis are the least used, each of them have a percentage of 1%.

Information in the data set for study analysis is classified and can be seen in Table 1. Classification of features in the data set is grouped into classes to build meaningful relationships.

Table 1. Classification of features in the data set

Features	Class	No.
Age	15-25	1
	26-40	2
	41-55	3
	56-70	4
	70+	5
Gender	Woman	0
	Man	1
Household income (IDR)	20,000.00 - 2,415,000.00	1
	2,416,000.00 - 4,811,000.00	2
	4,812,000.00 - 7,207,000.00	3
	7,208,000.00 - 9,603,000.00	4
	9,604,000.00 - 12,000,000.00	5
	12,000,000.00+	6
Level of education	Primary school	1
	Middle school	2
	High school	3
	Vocational high school	4
	Associate degree	5
	Bachelor	6
	Master's degree	7
	Doctorate	8
Mode choice	Car	1
	Motorcycle	2
	Motorcycle taxi	3
	Bus	4
	Minibus	5
	Motorized rickshaw	6
	Taxi	7

3. Data Analysis

The large amount of data, as well as the heterogeneous and versatile features of them, enabled data mining applications to be used as a very effective method in analysis process. In this study the WEKA application is used because it is a good tool in completing various purposes in data mining, where there are four WEKA application interfaces, namely explorer, experimenter, knowledge flow, also simple command line [2].

3.1. Data Analysis Method

In this research, analysis of the relationship with association rules in data mining was carried out. Algorithms are used to scan all available data to reveal large collections of objects. On the first scan, the support level of each object is compared with the initial minimum support level to evaluate whether each object is large. Each subsequent scan starts from the broadly defined objects of the previous scan and a large object cluster is created. The support of this object set is calculated to the end of the data set. This process continues until no other large object set is found [4].

In association rules the concepts of trust level (support), level of support (confidence) and interest (leverage) are criteria that express the strength of the rules.

3.2. Data Analysis

Five different analyses were conducted in this study. The threshold values of the analysis and the variables used are given in Table 2.

Table 2. Details of carried out analyses

Analysis	Variable	The minimum support value	The minimum lift value
Car	Age, gender, household income, and level of education	0.1	1
Motorcycle	Age, gender, household income, and level of education	0.1	1
Motorcycle taxi	Age, gender, household income, and level of education	0.1	1
Bus	Age, gender, household income, and level of education	0.1	1
Minibus	Age, gender, household income, and level of education	0.1	1

The first step carried out in the WEKA application, namely; choose "open file" where the data have been converted into .csv or .arff format. The next step is to select "NumericToNominal" in the filter section and press "apply". In the associate section select "apriori", enter the minimum support and lift value, then press "start" are applied.

3.2.1. Mode choice analysis of car

Best rules found:

 1. Age=2 358 ==> Household Income=2 194   conf:(0.54) < lift:(1.14)> lev:(0.03) [23] conv:(1.14)
 2. Household Income=2 340 ==> Age=2 194   conf:(0.57) < lift:(1.14)> lev:(0.03) [23] conv:(1.15)
 3. Level of Education=6 336 ==> Household Income=2 175   conf:(0.52) < lift:(1.1)> lev:(0.02) [15] conv:(1.09)
 4. Household Income=2 340 ==> Level of Education=6 175   conf:(0.51) < lift:(1.1)> lev:(0.02) [15] conv:(1.09)
 5. Gender=1 444 ==> Age=3 191   conf:(0.43) < lift:(1.09)> lev:(0.02) [16] conv:(1.06)
 6. Age=3 281 ==> Gender=1 191   conf:(0.68) < lift:(1.09)> lev:(0.02) [16] conv:(1.17)
 7. Gender=0 271 ==> Age=2 146   conf:(0.54) < lift:(1.08)> lev:(0.01) [10] conv:(1.07)
 8. Age=2 358 ==> Gender=0 146   conf:(0.41) < lift:(1.08)> lev:(0.01) [10] conv:(1.04)
 9. Gender=1 444 ==> Level of Education=6 218   conf:(0.49) < lift:(1.04)> lev:(0.01) [9] conv:(1.04)
10. Level of Education=6 336 ==> Gender=1 218   conf:(0.65) < lift:(1.04)> lev:(0.01) [9] conv:(1.07)

Figure 7. Associator model for mode choice of car on WEKA application

Figure 7 indicates the results of analysis for the 715 data where the respondents chose car as the mode of transportation in the data set.

- 358 of them are in the age group 2 (26-40 years), and from this age group 194 of them are in the household income group 2 (IDR 2,416,000.00 - 4,811,000.00), where the confidence value is 0.54 and the lift value is 1.14.

- 336 of them are in the level of education group 6 (bachelor degree), and from this level of education 175 of them are in the household income group 2 (IDR 2,416,000.00 - 4,811,000.00), where the confidence value is 0.52 and the lift value is 1.1.

- 444 of them are in the gender group 1 (man), and from this gender group 191 are in the age group 3 (41-55 years old), where the confidence value was 0.43 and the lift value was 1.09.

3.2.2. Mode choice analysis of motorcycle

Best rules found:

 1. Level of Education=3 394 ==> Gender=0 Household Income=1 149   conf:(0.38) < lift:(1.22)> lev:(0.04) [26]
    conv:(1.1)
 2. Gender=0 Household Income=1 211 ==> Level of Education=3 149   conf:(0.71) < lift:(1.22)> lev:(0.04) [26]
    conv:(1.41)
 3. Household Income=1 459 ==> Gender=1 Level of Education=3 151   conf:(0.33) < lift:(1.19)> lev:(0.04) [23]
    conv:(1.07)
 4. Gender=1 Level of Education=3 188 ==> Household Income=1 151   conf:(0.8) < lift:(1.19)> lev:(0.04) [23]
    conv:(1.6)
 5. Gender=0 304 ==> Level of Education=3 206   conf:(0.68) < lift:(1.17)> lev:(0.04) [29] conv:(1.29)
 6. Level of Education=3 394 ==> Gender=0 206   conf:(0.52) < lift:(1.17)> lev:(0.04) [29] conv:(1.15)
 7. Level of Education=6 208 ==> Age=2 138   conf:(0.66) < lift:(1.15)> lev:(0.03) [17] conv:(1.24)
 8. Age=2 392 ==> Level of Education=6 138   conf:(0.35) < lift:(1.15)> lev:(0.03) [17] conv:(1.07)
 9. Household Income=1 459 ==> Level of Education=3 300   conf:(0.65) < lift:(1.13)> lev:(0.05) [33] conv:(1.2)
10. Level of Education=3 394 ==> Household Income=1 300   conf:(0.76) < lift:(1.13)> lev:(0.05) [33]
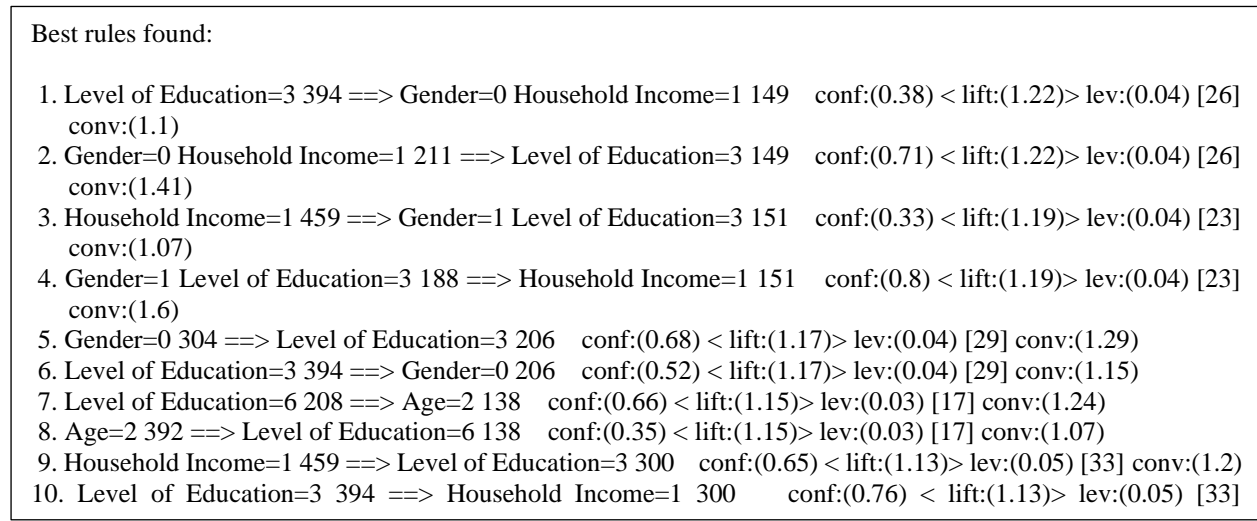
Figure 8. Associator model for mode choice of motorcycle on WEKA application

When 679 data where the respondents chose motorcycle as the mode of transportation in the data set were analysed, shown in Figure 8 above; the followings are the reached.

- 394 of them are in the level of education group 3 (high school), and from this level of education group 149 of them are in the gender group 0 (woman) and household income group 1 (IDR 20,000.00 - 2,415,000.00), where the confidence value is 0.38 and the lift value is 1.22.

- 459 of them are in household income group 1 (IDR 20,000.00 - 2,415,000.00), and from this household income group 151 of them are in the gender group 1 (man) and level of education 3 (high school) where the confidence value is 0.33 and the lift value is 1.19.

3.2.3. Mode choice analysis of motorcycle taxi

Best rules found:

 1. Level of Education=6 93 ==> Age=2 64   conf:(0.69) < lift:(1.26)> lev:(0.05) [13] conv:(1.41)
 2. Age=2 134 ==> Level of Education=6 64   conf:(0.48) < lift:(1.26)> lev:(0.05) [13] conv:(1.17)
 3. Age=1 63 ==> Household Income=1 58   conf:(0.92) < lift:(1.23)> lev:(0.04) [10] conv:(2.65)
 4. Household Income=1 184 ==> Age=1 58   conf:(0.32) < lift:(1.23)> lev:(0.04) [10] conv:(1.08)
 5. Household Income=1 184 ==> Level of Education=3 88   conf:(0.48) < lift:(1.23)> lev:(0.07) [16] conv:(1.16)
 6. Level of Education=3 96 ==> Household Income=1 88   conf:(0.92) < lift:(1.23)> lev:(0.07) [16] conv:(2.69)
 7. Gender=0 108 ==> Household Income=1 87   conf:(0.81) < lift:(1.08)> lev:(0.03) [6] conv:(1.24)
 8. Household Income=1 184 ==> Gender=0 87   conf:(0.47) < lift:(1.08)> lev:(0.03) [6] conv:(1.05)
 9. Gender=1 138 ==> Level of Education=6 55   conf:(0.4) < lift:(1.05)> lev:(0.01) [2] conv:(1.02)
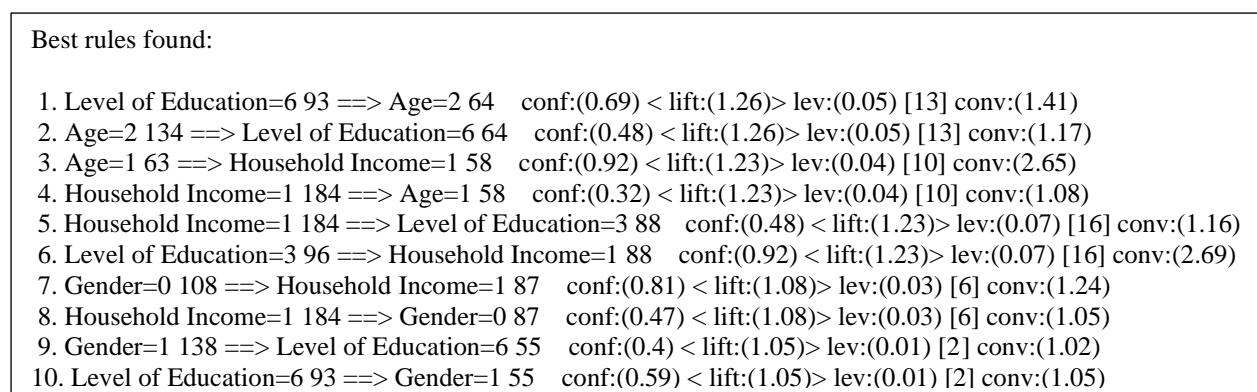10. Level of Education=6 93 ==> Gender=1 55   conf:(0.59) < lift:(1.05)> lev:(0.01) [2] conv:(1.05)

Figure 9. Associator model for mode choice of motorcycle taxi on WEKA application

When 246 data where the respondents chose motorcycle taxi as the mode of transportation in the

data set were analyzed; the followings have been obtained.

- 93 of them are in the level of education group 6 (bachelor degree), and from this level of education group 64 of them are in the age group 2 (26-40), where the confidence value is 0.69 and the lift value is 1.26.

- 163 of them are in the age group 3 (41-55), and from this age group 58 of them are in the household income group 1 (IDR 20,000.00 - 2,415,000.00), where the confidence value is 0.92 and the lift value is 1.23.

3.2.4. Mode choice analysis of bus

```
Best rules found:

 1. Household Income=1 341 ==> Level of Education=3 230   conf:(0.67) < lift:(1.11)> lev:(0.06) [22] conv:(1.19)
 2. Level of Education=3 234 ==> Household Income=1 230   conf:(0.98) < lift:(1.11)> lev:(0.06) [22] conv:(5.24)
 3. Household Income=1 341 ==> Gender=1 Level of Education=3 123    conf:(0.36) < lift:(1.1)> lev:(0.03) [11]
    conv:(1.05)
 4. Gender=1 Level of Education=3 126 ==> Household Income=1 123    conf:(0.98) < lift:(1.1)> lev:(0.03) [11]
    conv:(3.53)
 5. Gender=1 Household Income=1 191 ==> Level of Education=3 123    conf:(0.64) < lift:(1.06)> lev:(0.02) [6]
    conv:(1.08)
 6. Level of Education=3 234 ==> Gender=1 Household Income=1 123    conf:(0.53) < lift:(1.06)> lev:(0.02) [6]
    conv:(1.05)
 7. Age=1 164 ==> Household Income=1 153    conf:(0.93) < lift:(1.05)> lev:(0.02) [7] conv:(1.53)
 8. Household Income=1 341 ==> Age=1 153    conf:(0.45) < lift:(1.05)> lev:(0.02) [7] conv:(1.03)
 9. Household Income=1 341 ==> Gender=0 150    conf:(0.44) < lift:(1.05)> lev:(0.02) [7] conv:(1.03)
10. Gender=0 161 ==> Household Income=1 150    conf:(0.93) < lift:(1.05)> lev:(0.02) [7] conv:(1.5)
```

Figure 10. Associator model for mode choice of bus on WEKA application

When 384 data where the respondents chose bus as the mode of transportation in the data set were analyzed; the below findings are obtained.

- 341 of them are in the household income group 1 (IDR 20,000.00 - 2,415,000.00), and from this household income group 123 of them are in the level of education group 3 (high school) and gender group 1 (man), where the confidence value is 0.36 and the lift value is 1.11.

3.2.5. Mode choice analysis of minibus

Best rules found:

1. Age=2 91 ==> Level of Education=6 40   conf:(0.44) < lift:(1.36)> lev:(0.06) [10] conv:(1.18)
2. Level of Education=6 60 ==> Age=2 40   conf:(0.67) < lift:(1.36)> lev:(0.06) [10] conv:(1.45)
3. Gender=1 Level of Education=3 44 ==> Household Income=1 43   conf:(0.98) < lift:(1.18)> lev:(0.04) [6] conv:(3.81)
4. Household Income=1 153 ==> Gender=1 Level of Education=3 43   conf:(0.28) < lift:(1.18)> lev:(0.04) [6] conv:(1.05)
5. Gender=1 Household Income=1 79 ==> Level of Education=3 43   conf:(0.54) < lift:(1.17)> lev:(0.03) [6] conv:(1.14)
6. Level of Education=3 86 ==> Gender=1 Household Income=1 43   conf:(0.5) < lift:(1.17)> lev:(0.03) [6] conv:(1.12)
7. Age=1 55 ==> Household Income=1 53   conf:(0.96) < lift:(1.17)> lev:(0.04) [7] conv:(3.17)
8. Household Income=1 153 ==> Age=1 53   conf:(0.35) < lift:(1.17)> lev:(0.04) [7] conv:(1.06)
9. Household Income=1 153 ==> Level of Education=3 82   conf:(0.54) < lift:(1.15)> lev:(0.06) [10] conv:(1.14)
10. Level of Education=3 86 ==> Household Income=1 82   conf:(0.95) < lift:(1.15)> lev:(0.06) [10] conv:(2.98)

Figure 11. Associator model for mode choice of bus on Weka application

When 185 data where the respondents chose minibus as the mode of transportation in the data set were analyzed;

- 91 of them are in the age group 2 (26-40), and from this age group 40 of them are in the level of education 6 (bachelor degree), where the confidence value is 0.44 and the lift value is 1.36.

- 44 of them are in the gender group 1 (man) and level of education group 3 (high school). From this gender group and level of education group 43 of them are in the household income group 1 (IDR 20,000.00 - 2,415,000.00), where the confidence value is 0.98 and the lift value is 1.18.

are the findings attained.

4. Discussion

In order to reveal the factors affecting the mode choice, the apriori algorithm was used in this study, and analyses were made by determining appropriate thresholds for support and lift values. In the respondents' data, mode choices were classified according to the respondents' choice which are car, motorcycle, motorcycle taxi, bus and minibus. Following, relevant analyses were carried out on the basis of this classification. Analysis of the choice of mode for motorized rickshaw and taxi was not carried out due to the small percentages available. As a result of these analyses using WEKA application;

Table 3. The association rules result between mode choice and other variables with the largest object

Variable	Mode choice (group; total respondents)				
	1; 715	2; 679	3; 246	4; 384	5; 185
Age (group; total respondents)	2; 358*	2; 392	2; 134	1; 164	2; 91
Gender (group; total respondents)	1; 444	0; 304	0; 108	1; 191	1; 79
Household income (group; total respondents)	2; 340	1; 495	1; 184	1; 341	1; 153
Level of education (group; total respondents)	6; 336	3; 394	6; 93	3; 234	6; 60

* In the mode choice group 1 (car), age group 2 (26-40 years) is the most dominant age group with a total of 358 respondents (50%), gender group 1 (man) is the most dominant gender group with a total of 444 respondents (62%), and so on.

In the mode choice group 4 (bus) and 5 (minibus) which are public transportation, the dominant gender group is gender group 1 (man). Another interesting finding is that the household income group 1 (IDR 20,000.00 - 2,415,000.00) dominates in the mode choice 2 (motorcycle) group. This can be caused by many things, one of which is the ease of owning a private vehicle, especially a motorcycle, and inadequate public transportation facilities in Banda Aceh city [5].

Conclusions

From the analyses it can be expressed that socio-demographic factors play an important role in the choice of transportation modes, especially in the city of Banda Aceh which can be seen that there are more private vehicle users than public transportation users influenced by age, gender, level of education, and household income. This can lead to various transportation problems such as congestion, increasing CO_2 emissions, noise pollution, traffic clutter, and others. By looking at the socio-demographic analysis in this study, it is hoped that it will provide input to stage holders in making policies in the city of Banda Aceh. One effort that can be done is to improve the public transportation system so that people who use private vehicles want to switch to public transportation.

References

[1]   Statistical Year Book of Banda Aceh, Bureau of Statistics, Government of Aceh, 2019.
[2]   Srivastava S. Weka: A Tool for Data pre-processing, Classification, Ensemble, Clustering and Association Rule Mining. International Journal of Computer Applications (0975-8887) 2014: Vol. 88 No. 10.
[3]   Aceh Provincial Government. https://humas.acehprov.go.id/. Accessed on 09th October 2020.
[4]   Silahtaroğlu G. Veri madenciliği kavram ve algoritmaları. 3. Basım. İstanbul: Papatya; 2016.

[5]   Saleh SM, Sugiarto S, Anggraini R. Analysis on Public's Response Toward Bus Reform Policy in Indonesia Considering Latent Variables. The Open Transportation Journal 2019:13:17-24.