Public Transportation Users Segmentation Using CHAID Method

1st Fitri Mudia Sari
Mathematics Department
Universitas Negeri Padang
Padang, Indonesia
fitrimudiasari@yahoo.co.id

2 nd Rahmad Fadhillah
Mathematics Department
Universitas Negeri Padang
Padang, Indonesia
rahmad.fadhill@yahoo.com

3rd Asih Yuhesty
Mathematics Department
Universitas Negeri Padang
Padang, Indonesia
asihyuhesty@gmail.com

4th Siti Hariksa
Mathematics Department
Universitas Negeri Padang
Padang, Indonesia
riksadviero@gmail.com

5th Iva Agustina Sari
Mathematics Department
Universitas Negeri Padang
Padang, Indonesia
ivaagustinasari@yahoo.co.id

6th Irene Simanungkalit
Mathematics Department
Universitas Negeri Padang
Padang, Indonesia
irene_simanungkalit@yahoo.co

Abstract—**Public transportation is one of the transportation media which is frequently used as an alternative if people don't bring or don't have the private vehicle. Nowadays, people in Padang city are frequently used "angkot" and "go-jek". The existence of these two transport gets the different response. The objective of this study was to determine the preference of people in Padang city in choosing public transportation between "angkot" and "go-jek". Classification of public transportation selection seen from cost, service, and timeliness of service.**

Keywords—CHAID, classifying, public interest, public transportation

I. BACKGROUND

The development of urban areas that cover all aspects of community life, such in education, health, industry, infrastructure, and even tourism, has resulted in an increase in community needs for better, comfortable, smooth and affordable transportation services for all levels of society. Nowadays, there is quite a lot of public transportation in the community. One of them is ANGKOT and the recent one is hotly discussed, namely GOJEK. Both of these transportation services provide services to the public in the form of transportation. But the existence of these two public transports has a different response from many people. Based on this, the purpose of this study was to find out which public transportation was more attractive to the public between Angkot and Gojek by using the CHAID method.

The data used in this study is survey data, using interview techniques and direct observation with the help of Questionnaire. The subject of this study is the people of Padang who have used public transportation Angkot and Gojek at least 3 times with a total sample of 50 respondents using Accidental Sampling sampling techniques.

II. DATA ANALYSIS

The analysis used in this study was CHAID analysis (Chi-squared Automatic Interaction Detector). CHAID as a whole work to predict a single variable referred to as the dependent variable, which is based on a number of other variables, referred to as independent variables. CHAID is an iterative technique that tests one by one of the independent variables used in classification and arranges it based on the level of significance of chi-square statistics on the dependent variable [1].

Hypothesis: H_0: $p_{ij} = p_i.p_{.j}$ (There is no relationship between two variables)

H_1: $p_{ij} \neq p_i.p_{.j}$ (There is a relationship between two variables)

The test statistic used is chi-square as follows (Agresti, 2002):

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left[\left(\frac{O_{ij} - e_{ij}}{e_{ij}}\right)^2\right] \quad \ldots\ldots \quad (1)$$

with a degree of freedom: *(r-1)(c-1)*

O_{ij} : observation of line-*i* and column-*j*
e_{ij} : expectation frequency of line-*i* and column-*j*
r : number of lines
c : number of columns

To calculate the expected frequency of each cell used by the formula [2]:

$$e_{ij} = \frac{r_i \ c_j}{n} \ldots\ldots \quad (2)$$

where: r_i : number of line-*i*
c_j : number of column-*j*
n : number of observations

The CHAID analysis steps used are:
1) Categorize dependent variables and independent variables.
2) Create a two-way contingency table for each independent variable with the dependent variable.
3) Calculate the chi-square statistic using equation (1) to examine the significance (having the largest chi-square value and p-value $<\alpha$) of each independent variable on the dependent variable.
4) Test and examine the significance of each independent subcategory in the form of sub 2xd where d are many categories of dependent variables.
5) The insignificant subcategories are combined into a single category and check the significance again.
6) If all independent variables have been significant, Calculate Bonferroni's corrected p-value is based on

contingency tables that have been combined according to each type of variable.

7) Divide the sample into sub-sub-groups according to the categories that have been combined/separated in the best variables based on the smallest corrected bonferroni p-value.

8) Interpret the results of the tree diagram output.
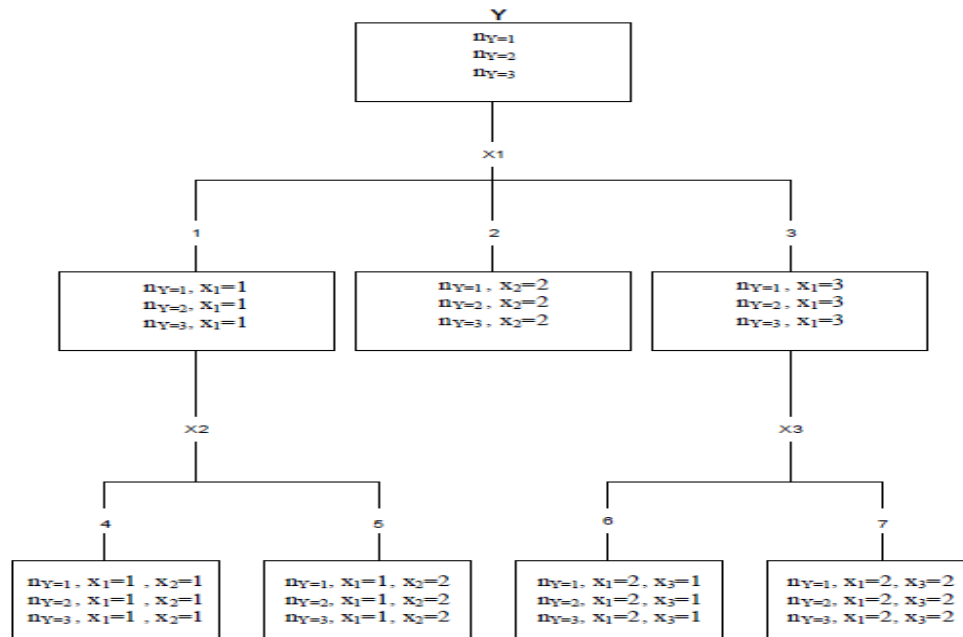
In general, the tree diagram of Chaid is as follows [3]:

Figure 1. Overview of T-Tree diagrams

III. VARIABLES AND DATA STRUCTURES

Determination of variables in this study aims to find information related to consumers who are users of Angkot and Gojek. Consumers have various factors that influence their decision in choosing Angkot or Gojek.

These variables are summarized as follows:

1. Demographic variables
   X1 = Gender
   X2 = Age
   X3 = Last education
   X4 = Job

2. Assessment Variable
   X5 = Suitability of Tariff
   X6 = Ease of Accessing Transportation
   X7 = Services provided by the Driver
   X8 = Convenience of Using Transportation
   X9 = Security Using Transportation
   X10 = Timeliness of service provided to respondent's needs

Based on the research objectives related to the interest of the public to use public transportation for public transportation or gojek, the researcher sets the response variable as follows: Y = The decision to use public transportation for Angkot or Gojek.

IV. RESULTS AND DISCUSSION

A. Determination of significant independent variables

TABLE I.   CHI-SQUARE VALUE OF GOJEK TRANSPORTATION

Variables	df	Chi-Square	Asymp. Sig. (2-sided)

Variables	df	Chi-Square	Asymp. Sig. (2-sided)
Gender	1	2.351	0.125
Age	3	3.615	0.306
Job	3	7.739	0.005
Suitability of Tariff	1	19.121	0.000
Services provided by the Driver	1	2.702	0.100
The convenience of Using Transportation	1	3.712	0.294
Security Using Transportation	1	2.702	0.100
Timeliness of service provided to respondent's needs	1	0.110	0.741

Based on the Chi-Square values in Table 1, the variables that significantly influence the public interest in using public transportation Gojek are: X5 (Tariff), X10 (the suitability of the service time given gojek) and X4 (the type of work that consumers have).

TABLE II.   CHI-SQUARE VALUE OF ANGKOT

Variables	df	Chi-Square	Asymp. Sig. (2-sided)
Gender	1	0.000	1.000
Age	3	2.873	0.412
Job	3	0.884	0.634
Suitability of Tariff	1	1.471	0.225
Services provided by the Driver	1	10.272	0.001
The convenience of Using Transportation	1	1.471	2.225
Security Using Transportation	1	2.371	0.395
Timeliness of service provided to respondent's needs	1	8.420	0.004

Based on the Chi-Square values in Table 2, the variables that significantly influence the public interest in using public transportation Angkot are: X6 (ease of access to transportation), X9 (security using transportation) and X10 (suitability of service time given to respondents' needs).

B. Grouping people's interest in using public transportation

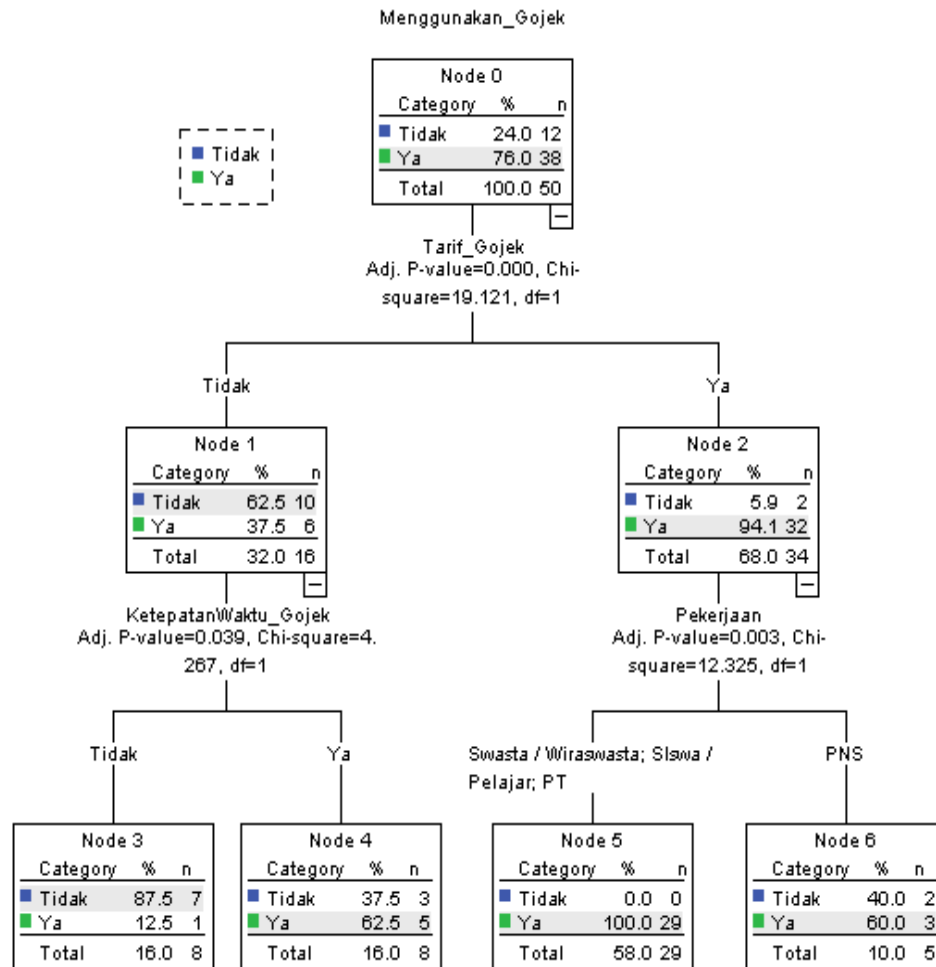Based on the CHAID method analysis results obtained T-tree diagram as follows:

Figure 2. T-Tree Diagram of Gojek

Based on Figure 2 were obtained 4 groups:

1. Group I

The desire of the people to use public transportation GO-JEK is influenced by GO-JEK Tariff in accordance with the economy of the community where the timeliness of GO-JEK shuttle is not in accordance with the wishes of the community.

2. Group II

The desire of the people to use public transportation GO-JEK is influenced by GO-JEK Tariff in accordance with the economy of the community where the timekeeping of GO-JEK is in accordance with the wishes of the community.

3. Group III

The people's desire to use public transportation GO-JEK is influenced by GO-JEK Tariff in accordance with the economy of the community where the work is Private, Self-Employed, and Students / Students.

4. Group IV

The people's desire to use public transportation GO-JEK is influenced by GO-JEK Tariff in accordance with the economy of the community where the work of civil servants is.

While the analysis of community decisions in deciding to use or not use Angkot is obtained T-tree diagram as follows:

ATLANTIS
PRESS

Menggunakan_Angkot

Node 0		
Category	%	n
■ Tidak	50.0	25
■ Ya	50.0	25
Total	100.0	50

Legend:
■ Tidak
■ Ya

Layanan_Akses_Angkot
Adj. P-value=0.001, Chi-square=10.272, df=1

Ya — Tidak

Node 1		
Category	%	n
■ Tidak	32.3	10
■ Ya	67.7	21
Total	62.0	31

Node 2		
Category	%	n
■ Tidak	78.9	15
■ Ya	21.1	4
Total	38.0	19

Keamanan_Angkot
Adj. P-value=0.014, Chi-square=6.039, df=1

YA — Tidak

Node 3		
Category	%	n
■ Tidak	0.0	0
■ Ya	100.0	9
Total	18.0	9

Node 4		
Category	%	n
■ Tidak	45.5	10
■ Ya	54.5	12
Total	44.0	22

Ketepatan\Waktu_Angkot
Adj. P-value=0.020, Chi-square=5.392, df=1

Tidak — Ya

Node 5		
Category	%	n
■ Tidak	58.8	10
■ Ya	41.2	7
Total	34.0	17

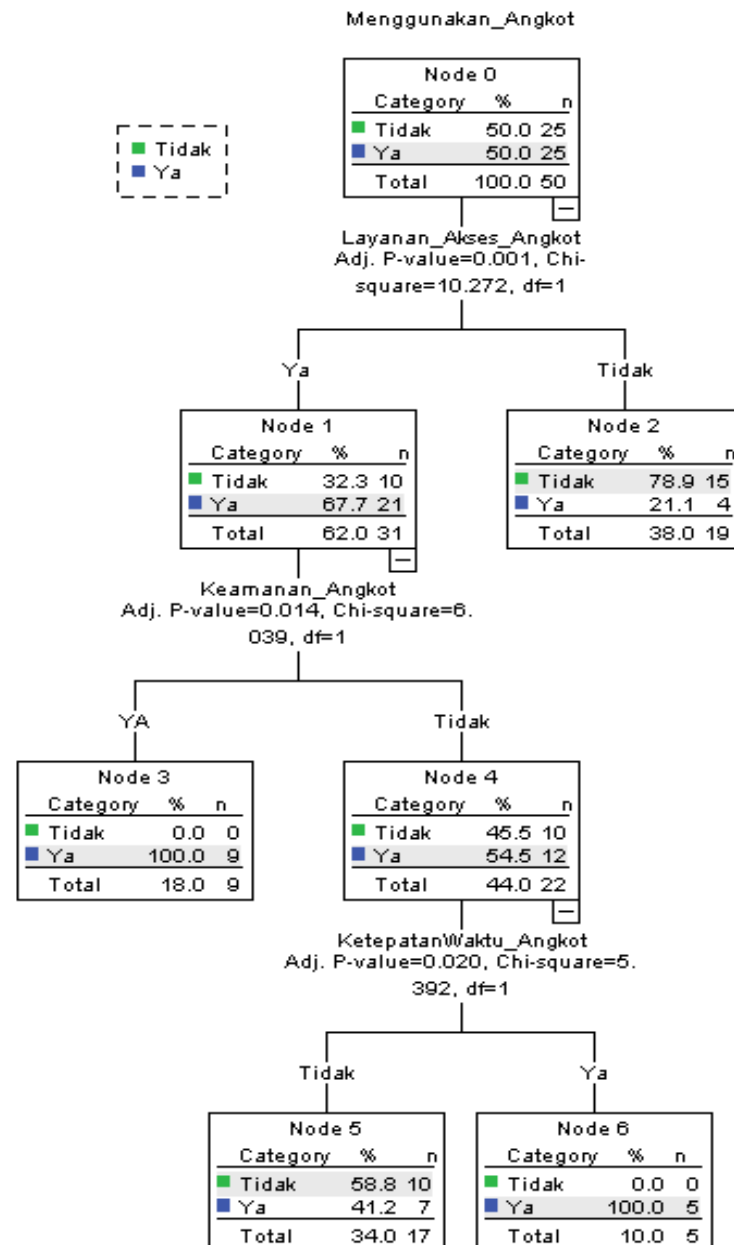| Node 6		
Category	%	n
■ Tidak	0.0	0
■ Ya	100.0	5
Total	10.0	5

Figure 3. T-Tree Diagram of Angkot

Based on Figure 3 were obtained 4 groups:

1. Group I

The people's desire to use public transportation is influenced by public transportation services where people get services from public transportation and people feel safe in using public transportation.

2. Group II

The desire of the public to use public transportation angkot is influenced by public transportation services where people get services from public transportation, people do not feel safe in using public transportation and the timeliness of public transportation is not in accordance with the wishes of the community.

3. Group III

The desire of the public to use public transportation is influenced by public transportation services where people get services from public transportation, people do not feel safe in using public transportation and the timeliness of

transportation of public transportation in accordance with the wishes of the community.

4. Group IV

The people's desire to use public transportation is influenced by public transportation services where people get services from public transportation and people feel safe in using public transportation.

attention and correct their respective services to the types of public transportation needs expected. After finding out what things that affect people's interest in using public transportation, public transportation or gojek is expected that both parties can improve service in the sector for the smooth running of a business and maximum satisfaction can be felt by consumers.

REFERENCES

[1]   Gallagher, C.A., "An Iterative Approach to Classification Analysis". www.casact.org/library/ratemaking/90dp237.pdf, 2000.

[2]   Agresti. A, "Categorical Data Analysis Second Edition". New York : John Wiley & Sons, Inc, 2002.

[3]   Lehmann.T, and Eherler. D, "Responder Profiling with CHAID and Dependency Analysis", www.informatik.uni-freiburg.de/~ml/ecmlpkdd/WS-Proceedings/w10/lehmann.pdf, 2001.

[4]   Simonoff. J.S, "Analyzing Categorical Data", New York : Springer, 2003.

[5]   Muers. J.H, "Segmentation and Positioning for Strategic Marketing Decisions", Chicago: American Marketing Association, 1996.