

## Chi-Square Automatic Interaction Detection (CHAID) Analysis for Home Quality Status Segmentation

Astri Atti<sup>1</sup>, Dominirsep O. Dodo<sup>2</sup>

<sup>1</sup>Faculty of Science and Engineering, Universitas Nusa Cendana, Kupang, Indonesia

<sup>2</sup>Faculty of Public Health, Universitas Nusa Cendana, Kupang, Indonesia

Corresponding Author: [astri\\_atti@yahoo.com](mailto:astri_atti@yahoo.com)

**ABSTRACT:** *Quality of Residence is defined as an assessment of the physical quality of the building and the completeness of the home facility. This study aims to form associations/structured relationships between factors in the classification of observations of the quality of housing eligibility in Kupang Regency by using CHAID method (Chi-square Automatic Interaction Detection). The result of CHAID analysis segmentation shows that the head of household which has house with quality livable there are two groups, that is (1) the head of household group with elementary and junior high school education or equal, working in industry and services as laborers (employees) and (2) the head of household groups with senior high school education and above live in urban areas and work in services. The head of household that have houses with less livable consist of five groups, namely: (1) the head of household educated primary school who lives in urban; (2) the head of household with primary and junior high school education or equivalent who work in agriculture; (3) senior high school headed households who live in rural areas; (4) head of household educated elementary and junior high school or equivalent working in the field of industry and services as well as non-labor status; and (5) high school headed households who live in urban areas and work in industry and agriculture. The quality of non-habitable home category housing (not livable) is owned by head of households who have elementary education who live in rural areas.*

**Keywords:** CHAID, quality of home eligibility, livable

Date of Submission: 05-04-2018

Date of acceptance: 20-04-2018

### I. INTRODUCTION

Housing needs are already a common problem faced by people in this world, both for those in the developed world and in developing countries. According to Suparlan (1985), each need requires a space, in this case the house. The house as a basic necessity in addition to food and clothing is a building used by human beings in an attempt to carry out their life activities, and is an important factor in the formation of human nature and personality. So big the role and meaning of a house, make people always try to improve the quality of his house in order to support all activities.

When the population is not so much, healthy and livable housing is not a problem to worry about. However, over time, increasing population growth, and the growing number of new households, while the area of land for the construction of housing is not increased, causing problems in providing housing and settlements of good quality, such as the quality of homes and the environment, the provision of facilities and better housing and residential infrastructure.

Housing development is one of the national developments in order to improve the quality and human welfare in the hope that all the people of Indonesia are able to occupy a habitable, healthy and quality house. The condition and condition of housing of a society can be one measure of living standards, civilization and personality of the community.

Based on the results of the Statistics of Nusa Tenggara Timur Province (BPS) and National Social Welfare Board (BKSNI) research in 2003, East Nusa Tenggara Province (NTT) is the province that has the largest inhabited households, ie 58.87%. While the remaining 14.89% live in less decent homes and 26.24% live in habitable homes. This may be related to the increasing number of NTT population (BPS 2004). In 2011 the population of 4,776,485 people with a density of 101 souls per km<sup>2</sup>. In 2012 it increased by 2% by 4,871,227 and in 2013 it increased to 4,953,967 people with density of 105 persons per km<sup>2</sup>.

Houses are habitable associated with the location of the potential housing density of houses and the quality of houses and their facilities. According to Darwin (1992), home density is the ratio between the size of the house and the number of occupants. There are 2 indicators that are often used to measure the density of the home floor area and the number of rooms.

The low level of education and lack of knowledge about the habitable home, causing the housing and environment conditions less suitable for habitation. It is based that the level of understanding and knowledge in the management of home and environmental buildings is largely determined by the education that a person has experienced. Sukamdi et.al (1994) states that the educational level of a person is directly related to the level of rationality, the higher the level of community education; the more rational the society will see the world. OeyGardier in Hasan (1993) states that heads of households who have not married and married status better household welfare than those with divorce status.

The results of the previous study found that there is a significant relationship between air humidity in the home and the density of the inhabitants of the house with the incidence of ARI disease in infants. Unqualified house ventilation is caused due to ventilation and window size not in accordance with building area. The high percentage of households who do not live clean and healthy, high percentage of unhealthy house, and high poverty/low socioeconomic level can also cause ARI. House hygiene is one of the factors that affect the health of its inhabitants, especially in children under five. Thus, home hygiene is a risk factor for the occurrence of ARI in children under five. Therefore it is necessary to improve the quality of residential houses in NTT, especially Kupang Regency, because it belongs to the first priority area of ARI susceptibility in NTT (Atti and Shinta, 2014).

## II. METHOD

This research was conducted in 24 sub-districts in KupangEast Nusa Tenggara by taking samples from each sub-district. Variables to be used in this study can be seen in Table 1 below:

**Table 1.** Variable Operational Concepts and Definitions

Variable	Operational definition	Category in Questionnaire	Categories in Data analysis
Quality of home eligibility	The quality of the dwelling house is an index of physical quality of the building and the completeness of the home facility.		1. Livable Habitation 2. Less Livable 3. Not Livable
Employment	Employment is a field of activity of work / business/company/ office where a person works.	1. Agriculture, plantations, fisheries, livestock, forestry and hunting. 2. Mining and quarrying. 3. Processing industry. 4. Electricity, gas and water. 5. Construction of buildings. 6. Trade, restaurants and accommodation services. 7. Transportation, warehousing and communication. 8. Financial institutions, real estate, leasing and corporate services. 9. Community, social, and individual services.	1. Agriculture (1) 2. Industry (2, 3, 4, 5) 3. Services (6, 7, 8, 9)
Level of education	The level of education in the questionnaire is the highest education completed by looking at the highest certificate / STTB owned.	1. Do not have an elementary certificate 2. Primary school 3. Junior high school 4. Senior high school 5. Senior high school/vocational 6. Diploma I / Diploma II 7. Diploma III 8. Diploma IV / Strata 1 9. Strata 2 / Strata 3	1. Primary school down (1, 2) 2. Junior High School or equivalent (3, 4, 5) 3. Upper Senior high school (6, 7, 8, 9)
Totalof household member	Total of household member is everyone who lives in a household.		1. >5 people 2. 3-5 people 3. 1-2 people

Variable	Operational definition	Category in Questionnaire	Categories in Data analysis
The marital status of householdhead	Marital status is categorized as unmarried and married. Never married are those who do not have a wife/husband at the time of enumeration. Marriage is those who are married at the time of enumeration, either living together or separately.	1. Never married 2. Married 3. Divorce 4. Widowed	1. Never married (1) 2. Married (2) 3. Divorced (3, 4)
Age of household head	Age is the last birthday since the survey was conducted and counted by rounding down.		1. <30 years 2. 30-45 years 3. > 45 years
Employment status of household head	Employment status is the type of position a person in a job.		1. Not laborers 2. Labor
Residential Areas	Residential area is the status of the region where the respondent is domiciled. Residential areas are categorized into two, namely urban and rural areas.	1. Rural 2. Urban	1. Rural 2. Urban

Data were analyzed by using CHAID method. CHAID method (Chi-square Automatic Interaction Detection) is a type of AID (Automatic Interaction Detection) method, which is a method used to analyze structural linkages among variables in a data segment (Fielding, 1977). The CHAID result is a decision tree or dendrogram based on the Chi-square test constructed by repeated group separation into two or more branches beginning with all data. (Ture, et al., 2006). The original CHAID algorithm (Chi-square Automatic Interaction Detection) was introduced by Kass (1980) for the nominal response variable. However CHAID has been extended to the ordinal response variables (Magidson in Magidson&Vermunt, 2006). For more details the stages in the CHAID method are described in the following algorithm:

The stages of CHAID method analysis are as follows (Kass, 1980):

1. For each explanatory variable, cross tabulation is made between the categories of explanatory variables with response categories.
2. From each cross-tabulation obtained, compile a 2xd sub table as possible, d is the number of response variable categories. From the table look for the category pair of explanatory variables that have the smallest test number. If it is not real, combine these two categories into one mixed category. If the number of categories is only two and the test results are not real then the variable does not need to be involved again in the model. Repeat this stage so that the smallest test table of the 2xd sub table pair category (mixed category) explanatory variable goes beyond the critical value.
3. For each mixed category that contains three or more categories of origin, look for binary separation that has the largest test number. If there is to make that separation and go back to stage 2.
4. Calculate the real levels for each new cross-tabulation and note which one has the greatest test number, call it the best-known tabulation. If this number is greater than the critical value, divide the data by that category.
5. Return to stage-1 to do the division based on unselected variables.

The test scores and the critical values referred to in the above analysis phase are statistics with the criteria of the Chi-square test.

### III. RESULTS AND DISCUSSION

Dendrogram separation results of CHAID analysis can be seen in Figure 1 with the value of  $\alpha$  used by 5%. In the first stage of separation of CHAID, the most realistic association variable in determining a person having a home for Livable, Less Livable or not Livable is educational variables. Based on the educational variables, the head of household is divided into three groups namely the head of household group educated primary school to lower, junior high school or equivalent, and senior high school and above. Of the 570 respondents around 34.7% were head of household groups with lower primary education and most had houses with Less Livable (41.4%) and not Livable (43.4%), and only 15.2% owned Livable houses. For head of household who have junior high school education or equivalent only 25.6% who own Livable house and the rest into the category of house Less Livable and not Livable, while the head of household educated senior high school upwards has about 44.1% who owns Livable house.

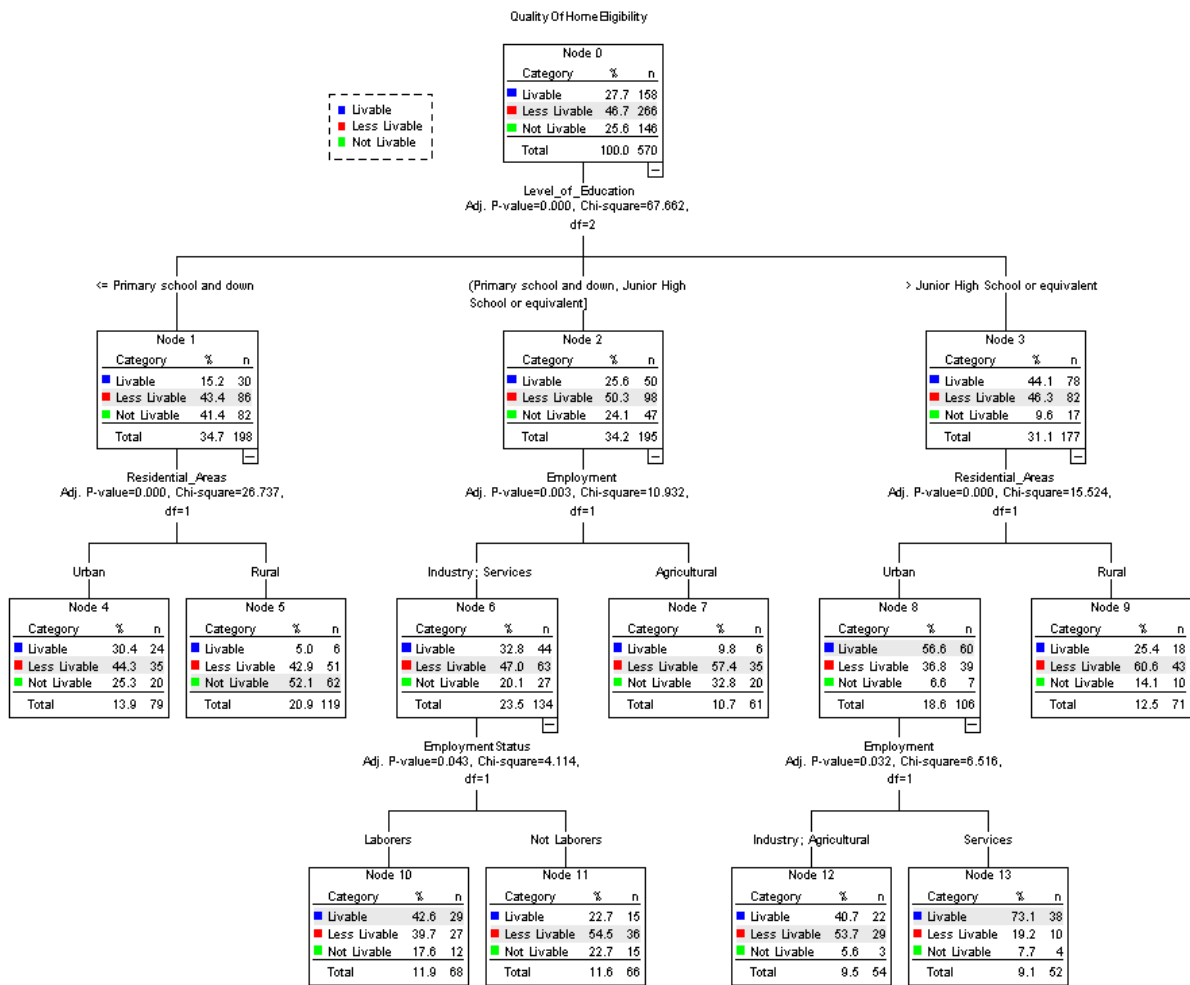


Figure 1. Dendrogram

In head of household educated primary school, there is a structured association between residential area variable with the quality of dwelling house. Of 34.7% of household headed households with under primary education, 13.9% (79 head of households) live in urban areas and have 30.4% livable houses and the rest are less livable and not livable. While 20.9% (119 head of households) who live in rural areas only about 5% who own Livablehouses.

Furthermore, in head of households with junior high school education or equivalent, there is a structured association between work and business (industry and service) variables with residential quality, as well as structured association in business variable (farm) with quality of residential house. Of the 34.2% (195) head of households with junior high school education or equivalent, 23.5% (134) work in industry and services with employment status as laborers (11.9%) and not laborers (11.6%). Of 68 head of households who work as industrial and service laborers, and have junior high school education or equivalent with 42.6% of livable houses, and less livable and not livable are 39.7% and 17.6%, respectively. While head of households whose work status is not more dominant workers who have house less livable (54.5%). For head of households with junior high school education/equivalent and working in agriculture only 9.8% have livable houses, most of them belong to less livable and not livable categories.

Structured associations are also seen in education variables, residential areas, and business fields with residential quality. Of 570 head of households, there are 9.5% of head of households who have senior high school education and live in urban areas with industrial and agricultural fields, while in the service sector there is 9.1%. Of these, the home of the head of households working in the field of industry and services is still dominated by the livable house, while those working in the field of services in general (73.1%) already have livable houses. A summary of the CHAID dendrogram can be seen in Table 2.

**Table 2.** CHAID Segmentation

Livable	Less Livable	Not Livable
1. Primary and Junior High School or equivalent, Industry and services, Laborers (employees) (Node 10)	1. Primary school Down, Urban (Node 4)	1. Primary School Down, Rural (Node 5)
2. Upper School, Urban, Service (Node 13)	2. Primary and Junior High School or equivalent, Agriculture (Node 7)	
	3. High School to Top, Rural (Node 9)	
	4. Primary and Junior High School or equivalent, Industry and Services, Not Laborers (Node 11)	
	5. Senior High Schools, Urban, Industrial and Agriculture (Node 12)	

Another study ever conducted by Atmaja (2004) states that the relationship between the type of work with the physical condition of the house is negative or very small is -39%, meaning that if the type of work is improved then the physical condition of the house is not getting healthier. Similarly, the number of family members with the physical condition of the house has a low level of relationship with negative coefficient, meaning that the more number of family members do not make the physical condition of the house become healthier (Atmaja 2004). This is in line with the results of Anwar's research (1999 in Astuti 2002) which states that the large number of family members is not related to home conditions. While for educational level factor, there is correlation between educations with physical condition of house although very low that is 0,125 but positive meaning that if education is increased one unit hence will influence to house physical condition (Atmaja 2004).

#### IV. CONCLUSION

The results of CHAID analysis segmentation showed that households with houses with livable qualities were two groups, namely groups (1) head of household with primary and junior high school education or equivalent, working in industry and services as laborers and groups (2) head of household who has senior high school education and above lives in urban areas and works in services. Head of household that have houses with less livable quality consist of five groups, namely: (1) head of household educated elementary school who live in urban areas, (2) household heads who have primary and junior high school education or equivalent, and work in agriculture, (3) ) head of household educated senior high school who live in rural, (4) head of households educated elementary and junior high school or equivalent working in the field of industry and services as well as non-labor status, (5) high school headed households who live in urban areas and work in the field industry and agriculture. The quality of not livable category home residence is owned by head of household who have primary education down and live in the rural.

#### V. SUGGESTION

To reduce the number of homes that less livable and notlivable then need to increase health promotion efforts to increase public knowledge about the quality of the feasibility of dwelling house, especially for the community Kab. Kupang Nusa Tenggara Timur.

#### REFERENCES

- [1]. AttiandShinta. 2014. *Modeling of Acute Respiratory Infection Cases (ISPA) for Case Control in East Nusa Tenggara Province* [Report on Research Results].
- [2]. Astuti, Y. 2002. *Factors Affecting the Quality of the Coastal Housing Residential Environment*. [Thesis]. Padang: Univ. Andalas.
- [3]. Atmaja J. 2004. *Hubungan Faktor Sosial Ekonomidengan Kondisi Fisik Bangunan Rumah tidak Sehat di Kecamatan Lubuk Alung*. *Jurnal Ilmiah R & B* 2004;(4):2.
- [4]. Badan Pusat Statistik Propinsi NTT. 2008. *Statistik Sosial dan Kependudukan Nusa Tenggara Timur 2008*. NTT: BPS.
- [5]. Badan Pusat Statistik dan Badan Kesejahteraan Sosial Nasional. 2004. *Analisis Deskriptif Penyandang Masalah Kesejahteraan Sosial (PMKS) Tahun 2003*. Jakarta : BPS.
- [6]. Bungsu, R. 2008. *Pengaruh Karakteristik Individu, Pengetahuan, Sikap dan Peran Petugasterhadap Penerapan Rumah Sehat di Kecamatan Peureulak Kabupaten Aceh Timur*. [Thesis]. Medan: Univ. Sumatera Utara; 2008.
- [7]. Darwin, Muhadjir. 1992. *Konsep Program Pemukiman untuk Pengembangan Sumber Daya Manusia*. Yogyakarta: Pusat Penelitian Universitas Gajah Mada.
- [8]. Fielding A. 1977. *Binary Segmentation: The Automatic Interaction Detector and Related Technique for Exploring Data Structure* (in O' Muirheartaigh, C. A., and C. Payne. 1977. *The Analysis of Survey Data Vol. I. Exploring Data Structure*). London, New York, Sidney, Toronto: John Wiley & Sons
- [9]. Hasan, Sulaiman. 1993. *Kondisi Sosial Demografidan Tipe Perumahan yang ditempati Penghuni pada Perumnas Kodya Ujung Pandang* (Tesis). Yogyakarta: Universitas Gajah Mada.
- [10]. Kass GV. 1980. *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. *App. Statist* 29(2):119-127.
- [11]. Kemenkes RI. 2014. *Profil Kesehatan Indonesia Tahun 2013*. Jakarta: Kemenkes RI.
- [12]. Kuswantojo, dkk. 2005. *Perumahan dan Permukiman di Indonesia: Upaya Membuat Perkembangan Kehidupan yang berkelanjutan*.

- Bandung: ITB Press.
- [13]. Magidson J, Vermunt JK. 2006. *An Extension of CHAID Tree-based Segmentation Algorithm to Multiple Dependent Variables*. Departement of Methodology and Statistic, Tilburg University. Netherlands.
- [14]. Sukamdi, Riturahardoyo. *et. al.* 1994. *Evaluasi Lingkungan Perumahan di Pedesaan Kabupaten Bantul*. Yogyakarta: Universitas Gajah Mada.
- [15]. Suparlan, parsudi. 1985. *Manusia, Kebudayaan, dan Lingkungannya*. Jakarta : Rajawali Press.
- [16]. Ture M, Akturk Z, Kurt I, Dagdeviren N. 2006. *The effect of Health Status, Nutrition and Some Other Factors on Low School Performance Using Induction Technique*. TrakyaUniversitesi Tip FakultesiDergisi.

Astri Atti "Chi-Square Automatic Interaction Detection (Chaid) Analysis for Home Quality Status Segmentation" American Journal of Engineering Research (AJER), vol. 7, no. 4, 2018, pp.183-188.