Journal of Technology and Computer (JOTECHCOM)
Vol. 2, No. 1, February 2025, pp. 1~8
ISSN: 3048-0477 (Media Online)

Application of the Chi-Square Automatic Interaction Detection (CHAID) Method in Analysing Variables That Influence Drug Abuse in Province North Sumatera

Muhammad Wira Yuda [1], Rina Widyasari [2], Rima Aprilia [3]
[1,2,3,4] Universitas Islam Negeri Sumatera Utara, Medan
[1] wirayuda0703@gmail.com, [2] rinawidyasari@uinsu.ac.id, [3] rimaaprilia@uinsu.ac.id

Article Info

Article history:

Received February 09, 2025
Revised February 27, 2025
Accepted February 28, 2025

Keywords:

CHAID
Drug Abuse
SPSS
Classification

ABSTRACT

The aim of this research is to classify transmission variables in North Sumatera Province using the Chi-Square Automatic Interaction Detection (CHAID) method. This method is used to determine which factors are the most dominant that cause people in North Sumatra Province to use drugs. This research began with collecting data on drug users in North Sumatra Province taken from BNN PROVSU. After the data was collected, it was processed again using MS Excel and divided into the 6 most dominant factors in the data that trigger drugs. Then, using the SPSS application, calculate which factors are the most dominant that cause people in North Sumatra to use drugs. Based on research conducted by the author on drug abuse in North Sumatra Province using the CHAID method, conclusions can be drawn as follows: The most dominant factor causing people in North Sumatra Province to use drugs is the educational factor. Of the total values that have been examined using the chi-square method, they are X^3 (educational factor) with a value = χ^2 count (13,018) > χ^2 Table (9,488). Education Factors of Drug Abuse in North Sumatra Province Associated with the Ratio of Narcotics Users.

Corresponding Author:

Muhammad Wira Yuda
Universitas Islam Negeri Sumatera Utara
Email: wirayuda0703@gmail.com

1. INTRODUCTION

Narcotics are drugs that can have effects such as stimulating anaesthesia, becoming more enthusiastic, being addictive, and causing a delusional effect or hallucinating. Addiction to using narcotics is one of them, a disease that until now has not been found a way to treat effectively; although narcotics are very necessary in the medical world, there are many people who misuse these narcotics for their own interests and individual satisfaction.

In cases of drug abuse, classification and factor analysis are carried out (i.e., variables), which greatly influence drug abuse in North Sumatra Province using the CHAID method. The Chi-square Automatic Interaction Detection method called CHAID is a tree regression analysis technique that is useful for segmentation. The CHAID method was introduced by Dr. G. V. Kass in 1980 through a piece of writing with the theme "An Exploratory Technique for Investigating Large Quantities of Categorical Data."The CHAID method is a continuation of the method that has appeared in the past, Automatic Interaction Detection (AID). CHAID is a technique in the classification tree that can be used to test the independent variables one by one and arrange them based on the chi-square statistical significance level for the dependent variable.

The objectives to be achieved in this research are as follows: The aim of this research is to classify drug abuse variables in North Sumatra Province using the Chi-Square method Automatic Interaction Detection (CHAID).

2. METHOD

This research method uses data from the North Sumatra Province BNN in 2021 and will be processed using Microsoft Excel and IBM SPSS. As the name suggests, the test statistic that can be used in using the CHAID method is the chi-square test statistic. Statistics in the chi-square test can be used to find independence (freedom) between two variables.

The variables in this research obtained from BNNP-SUMUT are The factors to be classified consist of:

Table 1. Research Variables

Variable	Description	Value of Variable	Measurement Scale
Y	Percentage of drug abuse	Percentage of drug abuse per total number of users in the district/city (people)	Ratio
X_1	Age Factor	1. Adolescent (11-17 years) 2. Adult (> 17 years)	Ratio
X_2	Gender Factor	1. Male 2. Female	Ratio
X_3	Education Factor	1. Graduated from SD 2. Graduated from SMP 3. Graduated from SMA 4. College 5. Not in School	Ratio
X_4	Employment Factor	1. Employed 2. Unemployed	Ratio
X_5	User Status Factor	1. Single 2. Married 3. Widower 4. Widow 5. Divorced	Ratio

The percentage of drug abuse in North Sumatra Province is categorised into two categories, namely the high percentage category of drug abuse in the district/city of North Sumatra Province and the low percentage of drug abuse in the district/city of North Sumatra Province.

Table 2. Percentage Grouping of Drug Abuse

Description	Percentage of Drug Abuse	Category
High	≥ 50% of the total	2
Low	≤ 50% of the total	1

In this research, the author carried out several stages to obtain The final result is the conclusion of this research is:

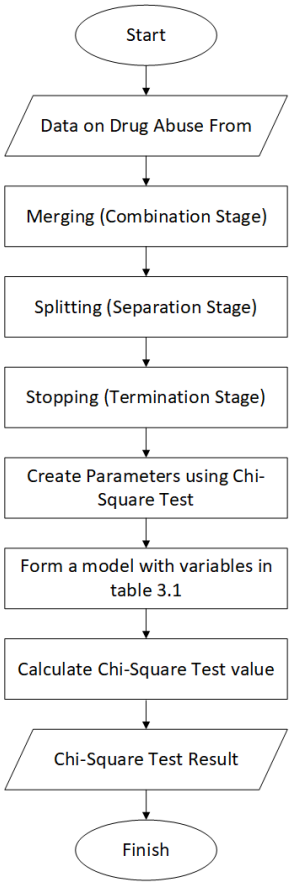

Figure 1. Research Flowchart

3. RESULTS AND DISCUSSION

Data collection was carried out by taking data based on the historical list of drug users in North Sumatra Province TP 2021/2022. This research used Ms. Excel to group variables and SPSS to test the Chi-Square Automatic Interaction Detection (CHAID) method. In applying the method, several variables are needed, namely Variable X and Variable Y. Variable data processing can be done using Ms. Excel and IBM SPSS.

The following is a data table on drug abuse in North Sumatra District/City:

Table 3. Table of data on drug users in North Sumatra Province in 2021.

No.	Regency/City	Number of Cases	Percentage (%)
1	Deli Serdang	139	25.74
2	Asahan	1	0.19
3	Batubara	1	0.19
4	Dairi	1	0.19
5	Medan	360	66.67
6	Binjai	1	0.19
7	Humbang Hasundutan	2	0.37
8	Karo	1	0.19
9	Labuhan Batu	20	3.70
10	Labuhan Batu Selatan	1	0.19
11	Labuhan Batu Utara	0	0.00
12	Langkat	4	0.74
13	Nias	0	0.00
14	Nias Barat	0	0.00
15	Padang Lawas	0	0.00
16	Padang Lawas Utara	0	0.00
17	Serdang Begadai	5	0.93
18	Simalungun	1	0.19

19	Mandailing Natal	1	0.19
20	Nias Selatan	1	0.19
21	Nias Utara	0	0.00
22	Tapanuli Selatan	0	0.00
23	Tapanuli Tengah	0	0.00
24	Tapanuli Utara	2	0.37
25	Toba Samosir	0	0.00
26	Padang Sidempuan	0	0.00
27	Pematang Siantar	0	0.00
28	Sibolga	0	0.00
29	Tanjung Balai	0	0.00
30	Gunung Sitoli	0	0.00
31	Pakpak Barat	0	0.00
32	Samosir	0	0.00
33	Tebing Tinggi	0	0.00
	Total	540	100

Factors influencing drug abuse in North Sumatra are grouped into 5 parts, namely X1: Age Factor, X2: Gender Factor, X3: Education Factor, X4: Job Factor, and X5: User Status Drugs. The following is the amount of data collected in variable X1: Factor age groups have been grouped:

1. Age Grouping (X_1): Teenage Age (11 - 17 Years): 21, Adult Age (> 17 Years): 519
2. Gender Grouping (X_2): Male: 517, Female: 23
3. Education Level of Drug Users (X_3): No Schooling: 29, Elementary School: 78, Junior High School: 124, Senior High School: 269, College: 40
4. Employment Status (X_4): Unemployed: 128, Employed: 412
5. Marital Status (X_5): Married: 248, Single: 259, Widower: 25, Widow: 7, Divorced: 1

Variable X is a variable that influences drug abuse factors, consisting of 5 factors, namely X_1: Age Factor, X_2: Type Factor Gender, X_3: Education Factor, X_4 = Employment Factor, and X_5 = Drug User Status. The following are the results of grouping data on drug users:

Table 4. Dataset of Variable X_1 - Age Factor

X_1 = Age	**Code**
Teenage Age (11-17 years)	1
Adult Age (>17 years)	2

Table 5. Dataset of Variable X_2 - Gender Factor

X_2 = Gender	**Code**
Male	1
Female	2

Table 6. Dataset of Variable X_3 - Education Factor

X_3 = Education Level	**Code**
No Schooling	1
Elementary School (SD)	2
Junior High School (SMP)	3
Senior High School (SMA)	4
College/University	5

Table 7. Dataset of Variable X_4 - Employment Factor

X_4 = Employment Status	**Code**
Unemployed	1
Employed	2

Table 8. Dataset of Variable X_5 - Drug User Status

X_5 = Drug User Status	**Code**
Married	1
Not Married	2

Widower	3
Widow	4
Divorced	5

Dataset Grouping of 540 Drug Users in North Sumatra Province in 2021 on the IBM SPSS application. Testing is carried out by classifying the relationship between the independent variable and the dependent variable with the data. In this case, classification and analysis of factors were carried out which greatly influenced drug abuse in North Sumatra Province. This method forms groups from drug abuse data based on influencing factors. In this phase, incomplete and inconsistent data is cleaned. The following is data for variables X and Y that have been cleaned and grouped.

Table 9. Coding into IBM SPSS

No	Y	X_1	X_2	X_3	X_4	X_5
1	1	2	1	5	2	1
2	2	2	2	4	1	1
3	1	2	1	4	1	3
4	2	1	1	4	1	2
5	2	2	1	3	2	1
6	1	2	1	1	2	2
7	1	2	1	2	1	2
8	1	2	1	3	2	1
9	2	2	1	4	1	2
10	1	2	1	1	2	2
11	2	2	1	2	2	1
12	1	2	1	4	1	2
13	2	2	2	3	1	1
14	2	2	2	4	1	1
15	2	2	1	4	2	1
16	2	2	1	2	2	1
17	2	2	1	4	2	2
18	2	2	1	3	2	2
...	...	...	...	...	...	...
540	2	2	1	4	2	2

3.1. SPSS Manual data processing trial results from educational factors

Percentage of Drug Users ∗ Education Factor. This test uses 540 data points to find out whether there is a relationship between variable X_3 (education factor) and variable Y (high and low drug use).

Table 10. Relationship Between Educational Variables (X_3) * variable Y

No	Y	X_3
1	1	5
2	2	4
3	1	4
4	2	4
5	2	3
6	1	1
7	1	2
8	1	3
9	2	4
10	1	1
11	2	2
12	1	4
13	2	3
14	2	4
15	2	4
16	2	2
17	2	4
18	2	3
...	...	...
540	2	4

Note:
X_3 = Education Factors: 1. No school, 2. Elementary school, 3. Middle school, 4. High school, 5. College.
Y = Drug users: 1. High, 2. Low.

Table 10. Gender Contingency ∗ Drug Users

Crosstab				
Count				
		Gender		Total
		Male	Female	
PN Percentage	Low	171	9	180
	High	346	14	360

Total	517	23	540

Table 11. Calculation of the Frequency of Educational Expectations ∗ DrugUsers

Education Factor * PN Ratio	$F0$	Fh	$(F0 - Fh)^2$	$\frac{(F0 - Fh)^2}{Fh}$
No School - Low	11	9.6	1.96	0.204
No School - High	18	19.3	1.69	0.087
Elementary School - Low	36	26	100	3.846
Elementary School - High	42	52	100	1.923
Middle School - Low	47	41	36	0.878
Middle School - High	76	82.1	37.21	0.453
High School – Low	72	90.3	334.89	3.708
High School - High	199	180.6	338.56	1.874
College - Low	14	13	1	0.007
College - High	25	26	1	0.038
	X^2			**13.018**

3.2. Manual Calculation Results without SPSS

Calculating Expected Frequency (Fh)
Formula:

$$Fh = \frac{Row\ Total\ x\ Column\ Total}{Grand\ Total} \quad (1)$$

Fh (No School – Low) $= \frac{29}{540}\ x\ 180 = 9.6$
Fh (No School – High) $= \frac{29}{540}\ x\ 360 = 19.3$
Fh (Elementary – Low) $= \frac{78}{540}\ x\ 180 = 26$
Fh (Elementary – High) $= \frac{78}{540}\ x\ 360 = 52$
And so on.

Calculating $(F0 - Fh)^2$
$(F0 - Fh)^2$(No School – Low) $= (11 - 9.6)^2 = 1.96$
$(F0 - Fh)^2$(No School – High) $= (18 - 19.3)^2 = 1.69$
$(F0 - Fh)^2$(Elementary – Low) $= (36 - 26)^2 = 100$
$(F0 - Fh)^2$(Elementary – High) $= (42 - 52)^2 = 100$
And so on.

Calculating $\frac{(F0 - Fh)^2}{Fh}$
$\frac{(F0 - Fh)^2}{Fh}$ (No School − Low) $= \frac{1.96}{9.6} = 0.204$
$\frac{(F0 - Fh)^2}{Fh}$ (No School − High) $= \frac{1.69}{19.3} = 0.087$
$\frac{(F0 - Fh)^2}{Fh}$ (Elementary − Low) $= \frac{100}{26} = 3.846$
$\frac{(F0 - Fh)^2}{Fh}$ (Elementary − High) $= \frac{100}{52} = 1.923$
And so on.

Chi-Square Calculation $X^2 = 13.018$

Note: Hypothesis Decision: The PN education factor in North Sumatra Province is significantly associated with the level of drug user ratio.

$X^2_{table\ 0.05\ (db)} = X^2_{table\ 0.05(k-1)(b-1)} = X^2_{table\ 0.05\ (5-1)(2-1)} = X^2_{table\ 0.05(4)(1)} = X^2_{table\ 0.05(4)} = 9.488$
Hypothesis Testing:

$$X^2_{count}(12.018) > X^2_{table}(9.488)$$

3.3. Manual Chi-Square Calculation Results for Education Factors * Drug Users

Table 12. Calculating the p-value of the Education Factor

No	Y	X_3
1	1	5
2	2	4
3	1	4
4	2	4
5	2	3
6	1	1
7	1	2
8	1	3
9	2	4
10	1	1
…	…	…
540	2	4

X^2 count	13.018
db	4
a	0.05
X^2 table	9.488
p-value	0.011
φ	0.156

Search Formula p-value:
p -value = CHISQ.DIST.RT(13,018;4) = 0,011
Above is the calculation of the p-value, obtained from there lationship between variable $X_3 * Y$ whose chi-square value is calculated.
The results of the Chi-Square calculation using SPSS are shown in the picture:

Table 13. Chi-Square Test Education * Drug Users

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	13.059[a]	4	.011
Likelihood Ratio	12.965	4	.011
Linear-by-Linear Association	7.035	1	.008
N of Valid Cases	540		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.67.

Table 14. Results of Phi and Crammer's V Values in SPSS

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.156	.011
	Cramer's V	.156	.011
N of Valid Cases		540	

The results of the Chi-Square independence test indicate that there is no significant association between educational factors and the ratio of drug users in North Sumatra Province, with $\chi^2(4) = 13.018$, $p = 0.011$, $\varphi = 0.156$.

4. CONCLUSION

In the educational classification, it can be seen that the percentage of drug users has a low category value: 33.3 and a high category: 66.7 with p-value: 0.012 and chi-square value: 11.204 degrees of freedom: 1. And the College, Middle School, No School, Elementary School category has a low category score: 40.1 and high category: 59.9. And the high school category has a low category score: 26.6 and high category: 73.4.

Based on the results of data analysis using the Chi-Square test analysis, the Education Factor value was obtained with a Chi-Square value = 13.018, phi = 0.156, p-value = 0.011. Based on this value, it can be concluded that the Education Factor for Drug Users in North Sumatra Province is significantly associated with the level of Drug User Ratio.

REFERENCES

[1] R. Aprilia, "Fuzzy Multiple Attribute Decision Making in Hotel Selection," *Jurnal Sains, Matematika dan Terapan*, 2019.
[2] H. Banjar, *Narkoba Dari Masa Ke Masa*, Medan: Mitra Medan, 2019.
[3] F. Ferri, *Ferri's Clinical Advisor*, Elsevier, 2019.
[4] O. Fitriani, S. Handayani, and N. Asiah, "Determinants of Drug Abuse among Adolescents at SMAN 24 Jakarta," *Arkesmas*, vol. 2, no. 1, pp. 135-143, 2017.
[5] I. Islamy and D. Dini, "Penerapan Metode CHAID (Chi-Squared Automatic Interaction Detection) Pada Kelanjutan Siswa yang Mengikuti Lembaga Bimbingan Belajar," *Prosiding Pendidikan Matematika dan Matematika*, vol. 4, no. 1, pp. 1-7, 2021.
[6] A. Majid, *Bahaya Penyalahgunaan Narkoba*, Semarang: Alprin, 2020.
[7] J. Nugraha, *Introduction to Categorical Data Analysis*, Yogyakarta: Deep Ublish, 2016.
[8] M. Pakpahan, D. Siregar, A. Susilawaty, T. Mustar, R. Ramdany, E. I. Manurung, and Maisyarah, *Promosi Kesehatan dan Perilaku Kesehatan*, Jakarta: Yayasan Kita Menulis, 2021.
[9] A. Rizkina and R. Widyasari, "Analisis Persediaan Menggunakan Metode Min Max dan Economic Order Quantity," Medan: Indonesia, 2022.
[10] K. K. Sahu and S. Sahu, "Substance Abuse Cause and Consequences," *Bangabasi Academic Journal*, vol. 9, no. 1, pp. 52-59, 2016.
[11] K. Siti, *Analisis Dinamik Model Penyebaran Penyalahgunaan Narkoba Kelompok Individu yang Direhabilitasi*, Semarang: Universitas Negeri Semarang, 2020.
[12] V. Sulviana, A. H. Wigena, and Indahwati, "Method Implementation CHAID (Chi-Squared Automatic Interaction Detection)," *Segmentation Trend of Sales of Soft Drinks in Indonesia*, vol. 2, no. 2, pp. 24-31, 2018.
[13] S. Syahdan, A. Arif, and M. Megawati, "Analysis of the Factors that Cause Dengue Hemorrhagic Fever (DHF) Using Chi-Square Automatic Interaction Detection (CHAID)," *International Journal of Nature Science and Engineering*, vol. 5, no. 3, pp. 104-113, 2021.
[14] R. Syuhada, H. Mawengkang, and M. S. Lydia, "Analysis of Performance K-Nearest Neighbor for Regulate Learning," *IOP Conference Series: Materials Science and Engineering*, vol. 725, no. 3, 2020. doi: 10.1088/1757-899X/725/1/012132.
[15] D. Wijayanti, *Revolusi Mental Stop Penyalahgunaan Narkoba*, Yogyakarta: Indoliterasi, 2016.
[16] S. Yamin, *Statistical Tutorial: SPSS, LISREL, WARPPLS and JASP*, Depok: PT Dewangga Energi Internasional, 2021.