Identification of the decline in learning outcomes in statistics courses using the chi-squared automatic interaction detection method

H Prasetyono, A Abdillah, T Anita, A Nurfarkhana and A Sefudin
Indraprasta PGRI University

e-mail: agus.abdillah@unindra.ac.id

Abstract. The decline in learning outcomes in Statistics is a problem that must be immediately solved. Many factors affect learning outcomes such as; intelligence, interest, talent, motivation, learning patterns, learning resources, and so on. The data used in this study are secondary data from TU-FIPPS and BAAK which consists of data on final grades of Statistics students of the Economic Education Study Program, Indraprasta PGRI University. In addition, researchers also use primary data collected through surveys. Data analysis using the Chi-squared Automatic Interaction Detection (CHAID) method. Based on the results of this study, explanatory variables that affect the decline in student learning outcomes Statistics are variables of motivation and group learning frequency. The explanatory variable which has the strongest influence with the decrease in student learning outcomes Statistics is motivation. The value of the accuracy of student classifications produced using CHAID analysis of eighty-five point one percent. The classification tree model of CHAID analysis in this study is better and more accurate for estimating the decline in learning outcomes Statistics in the downward category than in the non-down or very down categories.

1. Introduction

Optimal learning outcomes are the ideals or dreams to be achieved by stakeholders. Both from the students themselves, parents, the business world and the industrial world, as well as institutions where students study. Learning outcomes are changes in behavior after the learning process in the form of changes in student knowledge of a course, from not knowing to know, then changing students' attitudes from bad to good [1]. Learning outcomes of learning are influenced by many factors including internal factors such as; physiological aspects and psychological aspects. Then external factors such as; social environment, non-social environment, and learning approach factors [2]. While [3], factors that influence learning outcomes include; physical factors such as fatigue, disability. Furthermore psychological factors such as intelligence, interests, talents, and motivation.

The fact now is that student learning outcomes in the Statistics subject have decreased dramatically. This is certainly a special concern for us as lecturers or educators. Problems that are considered very serious and must immediately find a solution to the problem quickly and accurately. Many previous studies have examined ways to improve learning outcomes. Most researchers analyze
their research data using multiple linear regression. There are also those who analyze with the path analysis approach. As in research [4], and [5].

One of the statistical methods that can be used to analyze student characteristics toward the decline in learning outcomes in Statistics courses is the Chi-squared Automatic Interaction Detection (CHAID) method. The CHAID method is one of the non-parametric techniques that can be used for selecting variables from large data in determining the most influential variables. The CHAID method will produce a diagram similar to a decision tree diagram and use the Chi-square test for its operation [6]. Relevant studies that use this method include [7], and [8].

The general objective of this research is to find out and analyze the factors that influence the decline in learning outcomes in Statistics courses for even semester students, economic education study programs, Indraprasta PGRI University. Whereas the specific purpose of this study was to classify the characteristics of economic education students who experienced a decline in learning outcomes in the Statistics subject using the CHAID method.

2. Data and sample
The data used are secondary data from the FIPPS Administration and BAAK Indraprasta PGRI University. The data consists of a population of 1,388 regular class students, afternoon classes and extension classes, economic education study program in the even semester. Furthermore, based on the Slovin formula a sample of 94 students was obtained. The data consists of one response variable (Y) which is a decrease in learning outcomes in the Statistics course and twelve explanatory variables, among others; gender (X\textsubscript{1}), regional origin (X\textsubscript{2}), school origin status (X\textsubscript{3}), monthly tuition allowance (X\textsubscript{4}), learning motivation (X\textsubscript{5}), lecture time density (X\textsubscript{6}), task intensity (X\textsubscript{7}), frequency of independent learning (X\textsubscript{8}), group learning frequency (X\textsubscript{9}), learning resources (X\textsubscript{10}), organizational participation (X\textsubscript{11}), and intensity of use of social media (X\textsubscript{12}).

3. Data analysis technique
The stages of analysis to be carried out in this study are as follows; conduct data exploration, if the data obtained is incomplete, then estimating the missing data to see from other variables related to incomplete variables, conduct a descriptive analysis of explanatory variables and response variables. Then conduct analysis with the CHAID method which is a statistical method for predicting object membership in non-categorical variable classes. Classification trees are formed through recursive data isolation. [9] The algorithm of the CHAID method is as follows;

a) For each explanatory variable, cross-tabulate the explanatory variable categories with the response variable categories

b) Look for category pairs from explanatory variables with a $2 \times d$ sub-table ($d$ is the number of response variable categories) that have values $x^2$ the smallest. Then compare $x^2$ generated with $x^2_{a}$ predetermined. If $x^2 < x^2_{a}$ combine this pair into a new category.

c) For each combined category containing three or more origin categories, look for binary separations that have $x^2$ value the biggest. If $2* \hat{c}_i \geq x^2_{a}$, then make a new separation and return to stage 2.
d) Compute $2\chi^2_i$ value from the explanatory variable that has been combined with the response variable. Look for $2\chi^2_i$ value the biggest and then compare with $\chi^2_{\alpha i}$. If $2\chi^2_i \geq \chi^2_{\alpha i}$, then divide data by that category.

e) If separation occurs in step 4, then return to step 1 for each section of data resulting from the separation.

Furthermore, If there is a merger in the explanatory variables, then Bonferoni test is performed as a multiplier of $\chi^2_{\alpha i}$. The last stage is performing a tree diagram modeling to facilitate interpretation and conclusion.

4. Results and discussion

Based on the results of descriptive analysis, the percentage distribution of student characteristics is obtained, among others; gender of men as much as 26 or 27.7% and women as much as 68 or 72.3%. Students from Jabodetabek are 84% while the remaining 16% are from outside Jabodetabek. 64.9% came from private schools, while 35.1% came from public schools. For monthly tuition or allowance as much as 71.3% less than Rp. 1,000,000, and monthly tuition allowance above Rp. 1,000,000, as much as 28.7%. Student learning motivation 58.5% comes from self, while the remaining 41.5% motivation comes from other people. The density of lecture time, 73.4% stated loose, 14.9% stated moderate, and 11.7% stated solid. Then for the intensity of the task, as much as 61.7% stated a little, 28.7% enough, and 9.6% a lot. The frequency of independent learning 58.5% of students said they rarely studied independently, 17% were moderate, and 29.8% of students often studied independently. While on the frequency of learning in groups, as many as 50% of students stated rarely, 27.7% were moderate, and 22.3% said they often did group learning. Regarding learning resources, in this case, textbooks as much as 47.9% of students have handbooks and the remaining 52.1% students do not have handbooks. Organizational participation of 23.4% joined the organization, and 76.6% of students did not join the organization. Then the intensity of use of social media, as many as 58.5% rarely use, 23.4% are moderate, and 18.1% of students often use social media.

Furthermore, the percentage decline in learning outcomes in Statistics courses found that 62.8% did not go down, as many as 28.7% categories fell, and 8.5% in the categories fell very much.

4.1. Data exploration

Based on data from a survey of a sample of 94 students of economic education at the University of Indraprasta PGRI, it is known that students who did not experience a reduction in learning outcomes in Statistics subjects as many as 59 people, experienced a decline in learning outcomes with a category down by 27 people, and experienced a decline in learning outcomes in the very category down by 8 people.

4.2. Classification of variables using CHAID analysis

Partial chi-square analysis was carried out to see the freedom between the response variables and each explanatory variable. The chi-square freedom test was carried out at a 5% significance level. Table 1 shows that the status of school origin, motivation to learn, density of lecture hours, intensity of work, frequency of group learning, learning resources, and frequency of use of social media have p-values less than the real level of 0.05, which means that the seven variables have a relationship to variables
decreased learning outcomes in Statistics courses. Students who come from private schools will have a greater opportunity to experience a decrease with a very down category compared to students who come from public schools. Student motivation to learn from other people has the opportunity to experience a decrease in learning outcomes with a very down category. If the density of lecture hours is denser, the percentage of students who experience a decrease in learning outcomes in the very down category is also getting bigger.

Then in the intensity of the assignment, the more assignments were given by students by several lecturers, in the sense that the task intensity is too much, then the chances of students experiencing a decrease with a very down category will be even greater. Students who often do group learning will have a greater chance not to experience a decrease in learning outcomes compared to students who rarely conduct discussions or study groups. Furthermore, students who have learning resources, in this case, Statistics textbooks, have a lower chance of learning outcomes compared to students who do not have Statistics textbooks or handbooks. In addition, students who rarely use social media such as playing Facebook, Instagram, WhatsApp, etc., the chance of not experiencing a decrease in learning outcomes will be even greater.

| Category                      | p-value | Category                      | p-value |
|-------------------------------|---------|-------------------------------|---------|
| Gender                        | 0.286   | The intensity of the task     | 0.000   |
| Origin                        | 0.152   | Frequency of independent      | 0.598   |
|                               |         | learning                      |         |
| Status of school origin       | 0.025   | Frequency of group learning   | 0.000   |
| Monthly tuition allowance     | 0.619   | Learning resources            | 0.000   |
| Motivation to learn           | 0.000   | Organizational participation  | 0.605   |
| The density of lecture time   | 0.000   | The intensity of the use of   | 0.004   |
|                               |         | social media                  |         |

**Table 1.** p-value of the chi-square freedom test between the decrease in learning outcomes and each explanatory variable

**Table 2.** Explanatory variables included in the CHAID analysis decision tree diagram

| Depth | Explanatory variable          |
|-------|------------------------------|
| 1     | Motivation to learn          |
| 2     | Frequency of group learning  |

CHAID analysis was performed on 12 explanatory variables with $\alpha = 0.05$ for merging and separation. The tree stopping method used is the minimum observation at the parent node of 10
observations and at the child node of 5 observations. The depth of the decision tree in the analysis of the decline in learning outcomes Statistics courses stopped at depth 2. The decision tree diagram or classification of the CHAID analysis of the decline in learning outcomes of the Statistics course formed in Figure 1 has 5 nodes consisting of 2 inner nodes and 3 terminal nodes so the tree diagram classification analysis CHAID produced 3 classifications. The explanatory variables included in the CHAID analysis tree classification diagram are learning motivation and group learning frequency which can be seen in Table 2. The p-values and chi-square test values of each explanatory variable that have a relationship with the decline in learning outcomes of the Statistics course are according to the CHAID analysis tree classification diagram is shown in Table 3. If decision making is based on the p-value at a real level of 5%, then the two explanatory variables have a relationship with decreasing learning outcomes in the Statistics subject because both of the explanatory p-scores are less than the real level is 0.05.

| Variables             | Chi-square value | p-value |
|-----------------------|------------------|---------|
| Motivation to learn   | 84.749           | 0.000   |
| Frequency of group learning | 6.909           | 0.017   |

The learning motivation variable is the first variable chosen to separate students into two classifications, namely students who have self-motivation and motivation from others, so the learning motivation variable is the variable that has the strongest influence on the decline in learning outcomes in the Statistics course. The decision tree or classification analysis of CHAID shows that students who have the motivation to learn from themselves and who have motivation to learn from others, mostly affect the learning outcomes of Statistics.
Based on Figure 1 above, it can be seen that the first classification of students separated by learning motivation variables is that students who have self-motivation consist of 55 students with 54 details not having decreased learning outcomes in Statistics courses, 1 person has decreased learning outcomes subject Statistics categories go down. The second classification is students who have motivation derived from others, consisting of 39 students with details of 5 people who did not experience a decrease in learning outcomes in Statistics courses, 26 experienced a decrease in learning outcomes in the Statistics category for categories down, and 8 people experienced a decrease in learning outcomes in courses The category statistics are very down. Classifications of students who have the motivation to learn from others are separated by variables of group learning frequency into two classifications. So as to be considered as the third classification, there are at node 3 as many as 31 people or 33% have the characteristics of rarely learning groups will have a chance of 67.7% decreased learning outcomes in Statistics courses with the category down and have a chance of 6.5% to not experience a decrease in results study Statistics courses. The fourth or final classification is in node 4, as many as 8 people or 8.5% have characteristics of the frequency of group learning in the medium category will have a 62.5% chance of experiencing a decrease in learning outcomes in the subject category Statistics dropped and have a chance of 37.5% not to decreased learning outcomes in Statistics courses.

A summary of the characteristics of students who experienced a decline in learning outcomes in the Statistics course in the CHAID analysis classification tree diagram can be seen in the following table 4:

| Classification | Node | Characteristics | Decreased Learning Outcomes | Number of students |
|----------------|------|------------------|-----------------------------|-------------------|
|                | 0    |                  |                             |                   |
| Node 1         |      |                  |                             |                   |
| Node 2         |      |                  |                             |                   |
| Node 3         |      |                  |                             |                   |
| Node 4         |      |                  |                             |                   |

Figure 1. Decision tree

Table 4. Summary of CHAID analysis of learning outcomes in the statistics course
Furthermore, the value of the accuracy of the classification of students produced using CHAID analysis of 85.1%. In table 5 below shows that the percentage for estimating the decline in learning outcomes in the Statistics course in the category did not go down correctly by 91.5%, suspecting the decline in learning outcomes in the Statistics course in the category falling correctly by 96.3%, and in estimating the decline in learning outcomes in the course Statistics by the category really decreased correctly by 0.0%. The CHAID analysis tree classification model in this study is better for estimating the decline in learning outcomes in the Statistics subject in the downward category compared to the non-down or very down category. The CHAID analysis classification tree shows that there is no interaction between explanatory variables. In the classification tree that is formed, there is no cross-influence between explanatory variables because the explanatory variables that separate at each branch are different.

**Table 5. Accuracy in the classification of CHAID analysis**

| Observed          | Predicted       | Percent correct |
|-------------------|-----------------|-----------------|
|                   | Not down        | Down | Very down |               |
| Not down          | 54              | 5    | 0         | 91.5%          |
| Down              | 1               | 26   | 0         | 96.3%          |
| Very down         | 0               | 8    | 0         | 0.0%           |
| Overall percentage| 58.5%           | 41.5%| 0.0%      | 85.1%          |

5. Conclusions
Most of the students experienced a decrease in learning outcomes in the Statistics subject category by 41.5%. The explanatory variables included in the CHAID analysis tree diagram are the variables of learning motivation and group learning frequency. The explanatory variables that have the strongest influence with decreasing student learning outcomes in Statistics students are learning motivation. The CHAID analysis tree classification shows that there is no interaction between explanatory variables because there is no cross-influence between explanatory variables. The highest classification opportunities that classify the decline in learning outcomes of the subject of Statistics students with the category down are students who have the characteristics of group learning rarely. The value of the accuracy of student classifications produced using a CHAID analysis of 85.1%. The CHAID analysis tree classification model in this study is better for estimating the decline in learning outcomes in the Statistics course in the downward category compared to the non-down or very down category.

Based on the results and discussion in this study, it shows that the lack of student motivation that arises from other people (external motivation) cannot be used as an excuse so as not to experience a decline in learning outcomes in the Statistics course category, because students who do not have the motivation to learn in this study, still has certain characteristics based on other explanatory variables. In addition, group learning activities have an important role for students of economic education at Indraprasta PGRI University. So that the opportunity for learning outcomes in Statistics courses is smaller. Future studies can be carried out using other methods such as network analysis, CART analysis to compare the results obtained between the CHAID method and other methods.

Acknowledgments
Research and writing of this paper are funded by Kemristek Dikti DIPA: Directorate of Research and Community Service No. Contract: 0403/SKP.LT/LPPM/UNINDRA/III/2019 with "PKPT" scheme. Therefore, the authors thank the infinite. This is a pride for the author and an opportunity for authors to develop capacity. The authors express their sincere appreciation to all structural officials, administrative staff, lecturers, and all students who have been pleased and participated until the research is completed.

References
[1] Sudjana N 2009 Penilaian Hasil Proses Belajar Mengajar (Bandung: PT Remaja Rosdakarya)
[2] Syah M 2010 Psikologi Pendidikan dengan Pendekatan Baru (Bandung: PT Remaja Rosdakarya)
[3] Slameto 2010 Belajar dan Faktor-faktor yang Mempengaruhi (Jakarta: Rineka Cipta Mmmm)
[4] Abdillah A 2017 Efektivitas Media Pembelajaran dan Minat Belajar Pengaruhnya Terhadap Hasil Belajar Akuntansi Dengan Motivasi Belajar Sebagai Variabel Intervening Pada Siswa Kelas XI SMK Negeri dan Swasta Di Jakarta Timur, Jupeko 1 vol 2 pp 11-25
[5] H Prasetyono, A Abdillah, T Widiyarto and H Sriyono 2018 Character-based economic learning implementation and teacher’s reinforcement on student’s affective competence in minimizing hoax, Cakrawala Pendidikan 37(3):426-435
[6] Miftahuddin 2012 Penggunaan Metode Chaid (Chi Square- Automatic Interaction Detection) Pada Pohon Klasifikasi Menggunakan Satu Peubah Respon Dengan Perbandingan Taraf Nyata, Jurnal Matematika, Statistik & Komputasi vol 9 pp 11-22
[7] Kresna O 2017 Implementasi Metode Chi-Squared Automatic Interaction Detection pada Klasifikasi Indeks Prestasi Kumulatif Mahasiswa FMIPA UNIROW, Technology Science and Engineering Journavl 1 pp 21-26
[8] Rindy A P, Indahwati, and Farit M A 2013 Analisis CHAID untuk identifikasi ketepatan waktu lulus berdasarkan karakteristik mahasiswa, Xplore vol 2 pp 1-5
[9] Ritschard G 2010 CHAID and Earlier Supervised Tree Methods (Switzerland: Dept of
Econometrics, University of Geneva)