



## The Utilization of Naive Bayes and C.45 in Predicting the Timeliness of Students' Graduation

Agung Wibowo<sup>1</sup>, Danny Manongga<sup>2</sup>, Hindriyanto Dwi Purnomo<sup>3</sup>

<sup>1,2,3</sup>Magister Sistem Informasi, Fakultas Teknologi Informasi,  
Universitas Kristen Satya Wacana, Indonesia

<sup>1</sup>agungwibowo@unw.ac.id, <sup>2</sup>danny.manongga@uksw.edu, <sup>3</sup>hindriyanto.purnomo@uksw.edu

### Abstract

An assessment of the success of a college is if the student's graduation rate is on time and high every year. The timeliness of students' graduation can be influenced by several factors. This study aims to determine the profile of the students who graduated both on time and not on time given a certain graduation predicate set by the institution and to know the factors influencing students' graduation. The model used in this study using the NBC to determine the graduation pattern and the Decision tree to determine the influencing factors. In calculating the NBC algorithm using Rapidminer, it was found that the profiles of students who graduated on time and late with the predicate of less satisfactory, satisfactory, very satisfactory and cum laude. In the Decision Tree calculation, the highest gain values are obtained in the IPK3, IPS1, and IPK2 attributes. This research needs to be developed further by increasing the number of attributes and data, and it is necessary to make a system to determine the accuracy of students' graduation from the patterns that have been produced so that it can help universities to increase the level of students' graduation every year.

**Keywords:** Information Systems, Data Mining, UNW.

### 1. INTRODUCTION

Students are the most important part of evaluating the success of the implementation of study programs in tertiary institutions. Some efforts to improve the quality of a tertiary institution can be made in various ways, including increasing the quota of new students, increasing students' achievement both in academic and non-academic achievements, and also increasing the graduation rate of students each year. Ngudi Waluyo University, for example, has low graduation rates in Pharmacy study program. The comparison of new students and graduate students is very high, this is because there are many students who graduate not on time. The following Table 1 is a comparison between the level of graduates and the acceptance of new students in the last four years in the Pharmacy study program.

Table 1. Comparison of New and Graduated Students (Source: BAAK)

Entry Year	New Students	Graduating on time	Graduating not on time
2012	88	38	50
2013	63	33	30
2014	73	47	26
2015	99	63	36

Students' graduation is a very influential part in evaluating academic activities in a college. According to the research, an indicator of the success of a college is if the students' graduation rate is high every year [1].

Some previous studies which examined the graduation rate of students are shown in Table 2.

Table 2. List three of the previous studies

Title	Journals	Results	Limitation	Comment
Student Academic Performance Evaluation Using Naive Bayes Algorithm at Fasilkom Unilak	Journal of Digital Zone Information & Communication Technology	The attributes that had the most significant influence were school origin, first to fifth semesters achievement index and cumulative achievement index (GPA)	<ul style="list-style-type: none"> <li>• Uses NBC method</li> <li>• The population uses 6th semester students</li> </ul>	
Prediction of Student Graduation on Time Based on Age, Gender, and Achievement Index Using the Decision Tree Algorithm	Journal of Information Technology Volume 13 Number 1, January 2017	Performance Index Attributes of fourth semester (IPS-4) gets the highest Gain value and is eligible to become root.	<ul style="list-style-type: none"> <li>• Uses DT method</li> <li>• The population uses first to fourth semesters students</li> </ul>	
Prediction of Timeliness of Graduates with Data Mining Algorithm C4.5	Fountain of Informatics Journal	The highest information gain value is found in the parents' occupational attributes which are used as the root in the study and then followed by the attributes of the area of origin and the type of school origin	<ul style="list-style-type: none"> <li>• Uses Method C.45</li> <li>• The population used students in 2012</li> </ul>	

High and low students' graduation rates have many factors that become a problem in college. In this research, data analysis will be conducted at Ngudi Waluyo University on the Pharmacy study program for the year of 2012, 2013, 2014 and 2015. The analysis carried out is by measuring the graduation level of students through predictions of the timeliness of on time and not on time graduation based on the cumulative students' achievement index during second, third and fourth semesters which will be classified according to the graduation predicate consisting of satisfactory, very satisfactory, cumlaude and several other attributes such as NIM, Name, Gender, scores on mathematics courses, PMB (school enrollment) test scores, origin and place of birth of students, origin of previous schools, and parents / guardians' occupations of students using the C.45 and Naive Bayes algorithm methods that aim to not only compare the highest accuracy of the two algorithms but also to find out what kind of students' profile who can graduate on time? and what are the factors that can influence the timeliness of students' graduation? It aims to provide information for the institution so that it can

increase efforts in encouraging students to accelerate the students' graduation. In addition it is beneficial for students themselves to be able to graduate on time.

## **2. METHODS**

### **2.1. Previous Research**

Many previous studies have examined the graduation prediction analysis using their own method, some of which are shown in Table 2, which are relevant to this study. The research conducted by Nurliana [1] obtains the results that the use of the NBC method has the best accuracy of up to 76.67%, with the most significant attributes being school origin, achievement index in the first to third year and cumulative restoration index (GPA). It is recommended to look for alternative algorithms other than NBC that have good accuracy values and combine the most significant attributes in determining the right class [1].

In a study conducted by Romadhona [2], the results obtained that the highest Information gain value is in the 4th-semester achievement index (IPS-4) with a value of 0.340 and this attribute is eligible as the root. It is recommended to increase the number of training data records in subsequent studies in order to obtain better performance in the results of accuracy [2]. Subsequent research by Indah Puji Astuti [3] obtains the result that the highest information gain value is found in the parents' occupational attribute which is used as root in the study and then continued with the attributes of the region and type of school of origin. From the results of this study, the C4.5 algorithm has an accuracy value of 82%. It is recommended to look for other factors, in addition to student self-data, for example, academic factors, family economic conditions and psychological factors in determining students' graduation [3].

The difference between the previous researches above and the research that will be carried out is the attribute that is used which is the development of previous studies. It was conducted in a different place of study in determining the graduation criteria where in this study, the graduation attribute is added with the graduation predicates of less satisfactory, quite satisfactory, satisfactory, very satisfactory, and cum laude. And data processing methods used Naïve Bayes and decision tree C.45. Where these two algorithms are combined based on their characteristics, Naïve Bayes can predict the future by knowing the graduation pattern of students, decision tree C.45 can find out the most significant attributes in determining graduation.

### **2.2. Prediction of Timeliness of Students' Graduation**

This study has a framework that is based on academic phenomena that occur at Ngudi Waluyo University, namely an imbalance between the number of students entering and graduating in Pharmacy study programs in particular, so it is necessary to find an appropriate solution so that there is an alignment in the academic process that is relevant to evaluate PMB and mathematics exam scores in the first semester and lecture activities in the first 4 semesters of the students supplementing with the attributes of sex, year of birth, type of school of origin, place of origin of students and parents' work. A description of the attributes used is as follows:

### *2.2.1. The Attribute of Gender*

Gender Male (L) and Female (P) are used to determining the level of graduation. According to [4] gender attribute is one of the variables that can be used to determine the level of graduation in students. Learning achievement according to [4] is directly influenced by gender, according to him women have more achievement compared to men.

### *2.2.2. The Attribute of Mathematics Value*

Mathematical value in the Pharmacy study program is used to determine predictions of students' graduation, graduating on time or not. Mathematics grades in first semester are the basis of courses in the Pharmacy study program. The value scales used are A, B, C, D, and E.

### *2.2.3. The Attribute of School Enrollment Test*

The value of student entrance examination is also used to determine the prediction of whether or not the students pass. This value is obtained when the prospective students enroll in the college, the test material is the Academic Potential Test. The range of values given is between 50-90. Based on research [5], the entrance exam results can affect the success of students' studies at a college.

### *2.2.4. The Attributes Of Semester Achievement Index 1-4 And Gpa 2 - Gpa 4*

Semester achievement index and GPA according to [1] is the highest value in determining students' graduation. NBC Achievement Index Algorithm has the highest value in data processing. The higher the semester achievement index value and students' GPA, the higher the opportunities of the students to graduate.

### *2.2.5. The Attribute of the Place of Birth*

The place of birth of students is used in determining students' graduation. Place of birth of students is the origin of the area of students who come from: Java, Lombok, Sumatra, Bali, Kalimantan, Maluku, Riau, East Timor. The attribute of origin of this area is also used in research [6], which from the results of the study shows that regional origin is also a determining variable in predicting graduation.

### *2.2.6. The Attribute of Students' Age*

The age attribute is determined when the students are registered as a student.

### *2.2.7. The Attribute of Type of Origin School*

The origin schools are grouped into three types namely high school, vocational high school and MAN (Madrasah Aliyah Negeri). According to [1] the type of school has the most significant influence in increasing the accuracy value of NBC.

### *2.2.8. The Attributes of Parent's Work*

Parents' work attribute is used to describe the economic level of students, whose families have a steady income or not. The attributes of this parent's work consist of: teacher, private employee, civil employee, self-employed, Fisherman, Farmer.

### 2.2.9. The Attributes of Graduating on time or not

Graduation prediction can be divided into two, namely graduating on time where undergraduate students can complete a maximum of study in seventh or eight semester, and graduating not on time is for students who complete the study more than 8th semester (Academic Guide).

Graduates' Attributes are categorized by the title of, among others, unsatisfactory, quite satisfactory, satisfactory, very satisfactory and cum laude. Graduation predicate can be categorized as in Table 3.

### 2.3. Naive Bayes Classifier (NBC)

NBC or also called as Bayesian Classification is an algorithm that classifies statistics based on the Bayes theory which is used to predict the probability of a class membership. The main feature of NBC is a very strong assumption of independence of a condition or an event [8]. NBC has been shown to have high accuracy and speed when applied in large databases [9].

Data mining is a form of process to find a relationship that means the pattern of a large set of data stored in storage with statistical and mathematical techniques [7]. Data mining has several models including NBC or classification models.

The formula of Bayes theory is as follows:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

with the information:

X = data with an unknown class

H = data hypothesis X is a specific class

P (H | X) = H hypothesis probability based on condition X (posteriori probability)

P (H) = probability of hypotheses H (prior probability)

P (X | H) = probability of X based on the conditions in hypothesis H

P (X) = probability of X

### 2.4. Algorithm C.45

C4.5 algorithm is a classification algorithm that uses a decision tree model [3] to determine the attributes that become the root of this decision tree model by looking at the highest gain values of the existing attributes. Entropy and gain calculations are obtained by the following equation:

$$\text{Entropy}(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (2)$$

pi is the number of data classes divided by total data.

Table 3. Graduation Predicate (Academic Guide)

No.	Graduated GPA	Graduation predicate
1	2.00-2.50	Less satisfactory
2	2.51-2.99	Good enough
3	3.00-3.25	Satisfactory
4	3.26-3.50	Very satisfactory
5	3.51-4.00	Cumlaude

### 3. METHODS

#### 3.1. Research Phase

This research has several stages of research that are illustrated as shown Figure 1.

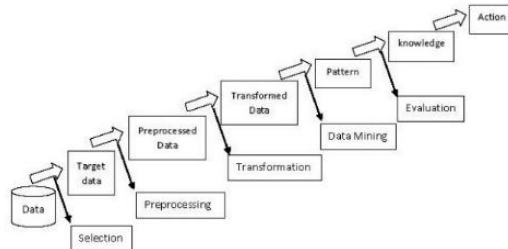


Figure 1. Research Phase

##### 3.1.1. Selection Phase

This study used students' data from 2012 to 2015 at Ngudi Waluyo University in the Pharmacy Study Program, graduating from 2016 to 2019 with a total of 228 students' data.

##### 3.1.2. Preprocessing Phase

Students' data obtained were categorized into two categories, namely personal data and academic data. Personal data consisted of students' biodata and previous educational background. While academic data consisted of data on grades and student achievement indices during the learning process. The attributes of this student data included in Table 4.

Table 4. Types of Attributes in Student Data

No	Attribute Type	Information
1	NIM	Student Identification Number, which contains information on student entry years
2	Name	Student Name Identity
3	Gender	Male or female
4	MK value	Grades in Mathematics Courses in semester 1
5	PMB value	Entrance Examination Selection Score
6	IPS1	1st semester achievement index
7	IPS2	2nd semester achievement index
8	IPK2	Second semester cumulative achievement index
9	IPS3	3rd semester achievement index
10	IPS4	4th semester achievement index
11	IPK3	3rd semester cumulative achievement index
12	IPK4	4th semester cumulative achievement index
13	Place	Student Birthplace
14	Date of birth	Birth Date of student's birth
15	Type of school before	SMA/SMK/MAN/MA
16	Profession	Job parents / guardians of students
17	Passing accuracy	Pass on time or Not
18	Predicate	Not Satisfactory, Quite Satisfactory, Satisfactory, Very Satisfactory or Cumlaude.

### 3.1.3. Transformation Phase

Some attributes of students' data had empty values, so they needed to be removed and the value of the attributes was made simpler to facilitate calculations in data mining. Table 5 and Table 6 are the attributes and their values after the transformation stage.

Table 5. Attribute type after going through the transformation stage

No	Jenis Atribut	Keterangan
1	Gender	Male or female (M/F)
2	MK	Grades in Mathematics Courses in semester 1
3	PMB	Entrance Examination Selection Score
4	IPS1	1st-semester achievement index
5	IPS2	2nd-semester achievement index
6	IPK2	Second-semester cumulative achievement index
7	IPS3	3rd-semester achievement index
8	IPS4	4th-semester achievement index
9	IPK3	3rd-semester cumulative achievement index
10	IPK4	4th-semester cumulative achievement index
11	Place	Student Birthplace (Province)
12	Age	Age of entry when registering as a student
13	Type of school before	(SMA(senior high scholl)/SMK(vocational scholl/MAN/MA(islamic senior high scholl))
14	Profession	Job parents/guardians of students (Self-employed, Private employee, Farmer, Civil Servants, Fisherman)
15	Passing accuracy	Pass on time or late Graduated on time A (OA) = Graduated on time with a less satisfactory predicate, Late A (LA) = Graduated Late with the Less Satisfactory predicate, Graduated on time B (OB) = Graduated on time with a Satisfactory predicate, Late B (LB) = Graduated Late with a satisfactory predicate(LB), Graduated on time C (OC) = Graduated on time with Very Satisfactory predicate, Late C (LC) = Graduated Late with a very satisfactory predicate, Graduated on time D (OD) = Graduated on time with cum laude predicate (RD), Late D (LD) = Graduated Late with cum laude predicate

Table 6. Data Training that had undergone a data transformation stage

No	Gender	MK	PMB	Ips1	Ips2	Ips3	Ips4	Ipk2
1	M	B	80	2.89	2.21	2.50	2.25	2.55
2	M	B	80	2.74	2.84	3.00	3.25	2.97
3	F	A	90	3.42	3.05	2.70	2.80	3.29
4	F	B	80	2.11	1.89	1.25	0.00	2.00
5	F	A	90	3.63	3.74	3.40	2.90	3.68
6	F	B	80	3.11	2.79	2.60	3.00	2.95
.....	..	..	....	....	....	....	....	....
.....	..	..	....	....	....	....	....	....
228	M	A	90	3.11	2.68	2.65	2.75	2.89

...	Ipk3	Ipk4	Place	Age	School	Profession	Accuracy
1	2.53	1.83	lombok	19	SMK	Entrepreneur	LB
2	2.86	2.40	Java	18	MAN	Entrepreneur	RC
3	3.05	2.26	Java	18	SMA	private	RC

4	1.74	0.86	lombok	18	SMA	Entrepreneur	LA
5	3.55	2.60	lombok	17	MAN	farmer	RD
6	2.83	2.22	Borneo	18	SMA	private	LC
.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....
228	2.81	2.12	bali	19	SMA	Entrepreneur	LC

### 3.1.4. Data Mining

This study used a data mining classification model, so the data used already had a class target. The target class was to graduate on time or late with the predicates of less satisfactory, satisfactory, very satisfactory, and cum laude contained in the attributes of the graduated target. The algorithms used Naive Bayes and C.45, which will get the most accurate results from the two algorithms. The tool used RapidMiner version 9.5. The Design model using Rapid Miner shown in Figure 2.

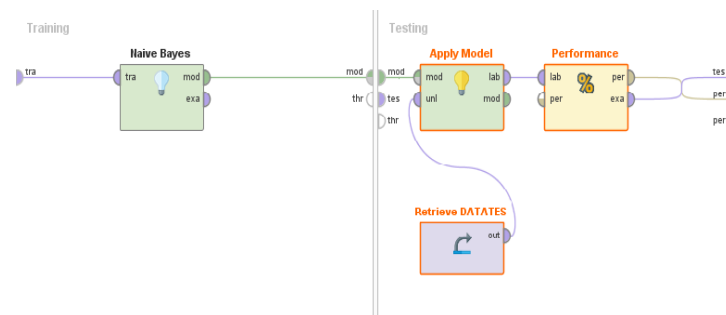


Figure 2. Design Model Using Rapid Miner

#### 3.1.4.1. Naive Bayes Classifier (NBC)

Table 7. Data Testing

Gender	MK	PMB	ips1	ips2	ips3	ips4	ipk2
F	B	80	2,89	2,74	2,70	2,65	2,53
...	ipk3	ipk4	place	age	school	Profession	accuracy
...	2,83	2,60	lombok	19	SMK	Entrepreneur	???

There were X data that were not yet known its class as in Table 7.

How the Naive Bayes algorithm worked [10]:

- 1) First, read the training data in Table 6.
- 2) Second, calculate the mean and standard deviation of the predictor attributes in each class. The result of this stage shown in Table 8.

Table 8. Calculation of Number of classes

Accuracy	RC	RD	LA	LB	LC
Value	0,6	0,05	0,05	0,15	0,15



- 3) Third, counted the same number of cases in the same class. The result of this stage shown in Table 9.

Table 9. Calculation of the number of cases in each class

Class	Nilai	RC	LA	LB	LC	RD
Gender	F	0.83	1	0.67	0.67	1
MTK	B	0.83	1	0.67	0.67	1
PMB	80	0.83	1	1	0.67	-
IPS1	2.89	0.17	-	0.33	-	-
IPS2	2.74	0.17	-	-	-	-
IPS3	2.70	0.08	-	-	-	-
IPS4	2.65	0.08	-	-	-	-
IPK2	2.53	0.08	-	-	-	-
IPK3	2.83	0.33	-	-	-	-
IPK4	2.60	-	-	-	-	1
Place	Lombok	0.16	-	-	-	-
Age	19	0.16	-	0.50	0.33	-
School	SMK	0.16	-	0.50	0.33	-
Profession	Entrepreneur	0.42	1	1	0.33	-

- 4) Fourth, calculated the results of the multiplication of proper attributes A, B, C, and D and late A, B, C, and D. The result of this stage shown in Table 10.

Table 10. Result of the Multiplication of Proper Attributes

Accuracy	RC	LA	LB	LC	RD
Value	$0,72 \times 10^{-8}$	0	0	0	0

- 5) Fifth, compared the results of multiplication of the four on-time groups and four late groups.

Because the result of P (On-time C) is greater, then the decision is Graduated on time with the title of very satisfactory.

### 3.1.4.2. Algorithm C.45

How the C.45 Decision tree algorithm worked [11]:

- 1) First, read the training data in Table 6.
- 2) Second, did the calculation of the total entropy value of the accuracy label

$$\sum_{i=1}^n \left( -\frac{12}{20} * \log_2 \left( \frac{12}{20} \right) \right) + \left( -\frac{1}{20} * \log_2 \left( \frac{1}{20} \right) \right) + \left( -\frac{1}{20} * \log_2 \left( \frac{1}{20} \right) \right) + \left( -\frac{3}{20} * \log_2 \left( \frac{3}{20} \right) \right) + \left( -\frac{3}{20} * \log_2 \left( \frac{3}{20} \right) \right) = 1,695461844238$$

- 3) Third, we looked for information gain values for each attribute and chose the largest value from the calculation results. Entropy and Gain calculation results are shown in Table 11.

Table 11. Entropy and Gain Calculation Results

Ranking	Attribute	Entropy	Value	Gain
1	IPS2	IPS	0	1,70
		IPK2	0	1,69
	IPK3	IPK	0	1,69
			3,40	1
		IPS	0	
			1,83	1
		IPK	0	
			3,00	1
			2,95	1

Ranking	Attribute	Entropy	Value	Gain
	IPS		0	
	2,89		0,92	
	3,11		0,92	
	IPS		0	
	Lombok		1,84	
	Jawa		0,47	
	Bali		0	
	Kalimantan		0	
	Entrepreneur		1,68	
	Private		0,97	
	Civil Servants		0	
	Farmer		0	
	17		0	
	18		1,42	
	19		1,5	
	SMA		1,48	
	SMK		1,5	
	MAN		1	
	A		1,5	
	B		1,5	
	80		1,5	
	90		1,5	
	Male		1,5	
	Female		1,67	

- 4) Fourth, made the node / root node decision from the highest gain value. The calculation results obtained IPS2 attributes with a gain value of 1.70, GPA2 and GPA 3 with a gain value of 1.69 which had the largest gain value, then the attribute would be the root node.
- 5) Fifth, created subdivisions below the root node from the order of high to low gain values and trimmed / eliminated attributes with low values.

### 3.1.5. Evaluation

After the data mining process was completed, an accurate model/pattern would be obtained in predicting the timeliness of students' graduation using two algorithms, NBC and C.45. The results of calculations with Naive Bayes can produce graduation patterns based on the attributes used, while C.45 decision trees can find out the most significant attributes in determining students' graduation.

## 4. RESULT AND DISCUSSION

The data used in this study were the data of the students who graduated on time with a total of 147 graduates and who graduated late who were 81 graduates. The number of training data was 228 graduates, the testing data were used by five graduates to determine the accuracy of the NBC and C.45 algorithms. The attributes used as parameters were 15 attributes, of which 14 were predictors and 1 was the result.

#### 4.1. Naive Bayes

Table 12. NBC calculation results

True	LB	OC	LA	OD	LC	LD	OB
LB	29	13	4	0	12	0	2
OC	2	84	0	4	10	0	0
LA	3	0	2	0	0	0	0
OD	0	10	0	21	1	3	0
LC	6	10	1	0	8	0	0
LD	0	0	0	3	0	0	0
OB	0	0	0	0	0	0	0

The results of experiments using NBC (in Table 12) could be known for its result on the label of on-time C (graduated on time with a predicate of very satisfactory) higher than other labels. Table 13, Shown students' profiles based on NBC calculation.

Table 13. Students' profiles based on NBC calculations

No	Pattern	Gender	MTK	PMB	Ips1	Ips2	Ips3	Ips4
1	OB	F	C	65	2.27	2.20	1.96	2.14
2	OC	F	B	75	2.80	2.98	2.98	3.02
3	OD	F	A	86	3.47	3.54	3.61	3.35
4	LA	M	D	65	0.70	0.78	0.50	0.83
5	LB	F	C	67.5	2.26	2.42	2.25	2.43
6	LC	F	B	75	2.65	2.64	2.64	2.55
7	LD	F	A	87	3.68	3.66	3.76	3.40

No	Ipk2	Ipk3	Ipk4	Place	Age	Scholl	Profession
1	2.16	2.08	2.10	Java	17	SMA	Entrepreneur
2	2.98	2.90	2.78	Java	18	SMA	Civil servants
3	3.54	3.53	3.28	Java	18	SMA	Entrepreneur
4	0.45	0.62	0.43	Borneo	18	MAN	Farmer
5	2.31	2.25	2.26	Jawa	18	SMA	Entrepreneur
6	2.60	2.62	2.37	Jawa	18	SMA	Entrepreneur
7	3.66	3.69	3.08	Lombok	19	SMA	Civil servants

#### 4.2. Desicion Tree C.45

Table 14. Decision Tree C.45 calculation results

True	LB	OC	LA	OD	LC	LD	OB
LB	2	0	0	0	0	0	0
OC	37	114	6	8	30	0	2
LA	1	0	1	0	0	0	0
OD	0	3	0	17	1	3	0
LC	0	0	0	0	0	0	0
LD	0	0	0	3	0	0	0
OB	0	0	0	0	0	0	0

The following resultants from Table 14 are produced in the form of trees as shown Figure 3.

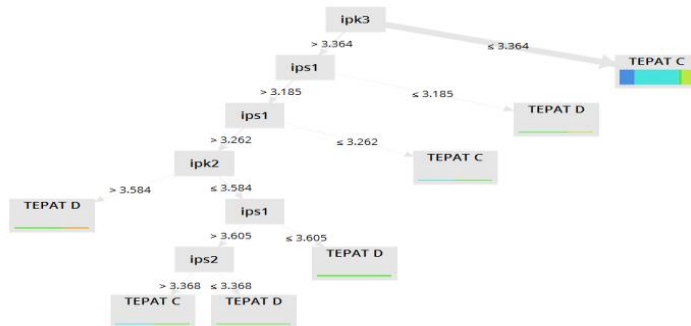


Figure 3. Result Decision Tree

In the decision tree in Figure 3, the selected attributes were IPK3, IPK2, and IPS1, while the other attributes were directly trimmed from the decision tree. It could be seen that with the amount of data and types of data that already existed, only a few attributes were needed to get the output class from the dataset.

Calculation of decision tree algorithm as in Table 14, the root node (root) was selected based on the highest gain value. In the calculation using Rapidminer the highest gain value was obtained in the IPK3 attribute followed by the IPS1, and IPK2 attributes followed by the status of the students graduating on time or late with the title of A, B, C, or D. GPA was obtained from the cumulative study results of students in third semester, IPS1 was obtained from the results of the students in first semester, and GPA was the result of the cumulative study of students in second semester. Based on the results of the decision tree calculation in Figure 3, it could be seen that the attributes that appear in the decision tree were IPK3, IPS1 and IPK2, other attributes (gender, mathematical value, PMB value, origin, school, work) were not displayed or trimmed because they were not selected as the attributes in the decision tree. The three attributes that had the highest gain in calculations using the decision tree algorithm were the factors that affected the student's graduation timeliness. While in the calculation with Naïve Bayes algorithm in finding the status of students' graduation on time with the testing data in Table 14, it could be seen that the students graduated on time with a very satisfactory predicate.

Naïve Bayes algorithm calculation using Rapidminer got the students' profiles as shown in Table 14 Students' profiles can be used as a pattern to find out whether students can graduate on time or late with the title of A, B, C, or D. In this study, the use of two Naïve Bayes and Decision tree algorithms have more complete results compared to studies previous. Naïve Bayes is used for future predictions with patterns generated in graduates and decision tree C.45 is used to determine the attributes that most play a role in students' graduation.

## 5. CONCLUSION

Students' graduation can be predicted in an on-time or late manner and can identify the factors influencing it. By using the Naive Bayes Algorithm and Decision Tree Algorithm, this study finds the factors that most influential in the graduation which is the cumulative achievement index in third and fourth semesters as well as the semester performance index of students in the first semester. Graduation patterns can be known from the beginning of the students entering the school until the second year. With this pattern, the study program can try to prepare the students to graduate on time with the specified graduation predicates. And with the attributes that play a role in graduation, the study program can improve students' competence early so that graduation rates become higher each year. The results of this study need to be further developed by increasing the number of attributes and data, and it is necessary to make a system to determine the accuracy of students' graduation from the patterns that have been produced in order to help the university to increase the level of students' graduation each year.

## 6. REFERENCES

- [1] Nasution, N., Djahara, K. and Zamsuri, A. (2015). Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes (Studi Kasus: Fasilkom Unilak). *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, 6(2), 1-11
- [2] Romadhona, A., Suprapedi, & Himawan, H. (2017). Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree . *Jurnal Teknologi Informasi*, 13(1), 69-83.
- [3] Astuti, I. P. (2017). Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5. *Fountain of Informatics*, 2(2), 41-45.
- [4] Andri, Kunang, Y.N., & Murniati, S. (2013). Implementasi Teknik Data Mining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Pada Universitas Bina Darma Palembang. *Seminar Nasional Informatika* . Yogyakarta.
- [5] Erdogan, S. M. (2005). A Data Mining Application In A Student Database. *Journal Of Aeronautics And Space Technologies*, 2(2), 53-57.
- [6] Sidik, M., Rasminto, H., Iriani, A., & Manongga, D. (2017). Implementasi Data Mining Untuk Prediksi Kelulusan Menggunakan Metode Klasifikasi Naive Bayes. *Jurnal Teknologi Informasi dan Komunikasi*, 8(2), 13-20.
- [7] Daniel, L. T. (2006). *Data Mining Methods dan Models*. John Wiley & Sons, Inc Publication.
- [8] Cahyaningtyas, C., Purnomo, H. D., & Kristianto, B. (2019). The Use of Naive Bayes for Broiler Digestive Tract Disease Detection. *JITCE (Journal of Information Technology and Computer Engineering)*, 03, 1-7.
- [9] Kusriani, & E. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- [10] Saputra, M. F., Widiyaningtyas, T., and Wibawa, A. P. (2018). Illiteracy Classification Using K Means - Naive Bayes Algorithm. *International Journal On Informatics Visualization*, 2 (3), 153.
- [11] Lungu, I., & Pirjan, A. (2010). Research Issues Concerning Algorithms Used For Optimizing The Data Mining Process. *IDEAS*.

- [12] Mujib Ridwan, H. S. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Jurnal EECCIS*, 59-64.
- [13] Garcia, E. P. (2011). Model Prediction of Academic Performance for First Year Students. *IEEE Computer Society*.
- [14] Nugroho, Y.S. (2014). Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta Menggunakan Algoritma C4.5. *KomuniTi*, 84-91.
- [15] Zahroh, F. (2016). Pengaruh Gender Terhadap Motivasi Memilih Sekolah dan Prestasi Belajar. *Journal of Accounting and Business Education* .