Classification of Graduate Profiles Based on Graduate Tracer Study Using Algorithm Naive Bayes Classifier

Junus Sinuraya[1]*, Friendly[2]

Politeknik Negeri Medan
E-mail:Junussinuraya@polmed.ac.id

ABSTRACT

The graduate profile is the role of the graduate of the study program or field of expertise / field of work planned after completing education from the study program. The determination of the profile of study program graduates is generally carried out based on the results of the assessment of stakeholder needs. Based on data from the Ministry of Research, Technology and Higher Education, the IT study program is one of the most majors or study programs in Indonesian universities and the highest number of enthusiasts choose this study program each year. Each year, graduates of IT study programs have a large number of graduates, both vocational and non-vocational colleges. The number of IT graduates is large but they have low graduate competencies, even they do not have competencies in the IT field so that their work is not in accordance with the graduate profile that has been designed. Therefore, it is necessary to conduct research to classify the profile of graduates who have worked based on tracer study data using the Naïve Bayes Classifier method. This study uses attributes, namely study program, value criteria, gender and field of work and the labels used are status (Linear and Non-Linear). The results of the study on the classification of the profile of graduates using the Naïve Bayes Classfier method show that alumni work not according to the profile of graduates by 73% and according to the profile of graduates by 23%, with a data accuracy rate of 87% and are included in the good classification category. it is necessary to conduct research to classify the profile of graduates who have worked based on tracer study data using the Naïve Bayes Classifier method. This study uses attributes, namely study program, value criteria, gender and field of work and the labels used are status (Linear and Non-Linear). The results of the study on the classification of the profile of graduates using the Naïve Bayes Classfier method show that alumni work not according to the profile of graduates by 73% and according to the profile of graduates by 23%, with a data accuracy rate of 87% and are included in the good classification category. it is necessary to conduct research to classify the profile of graduates who have worked based on tracer study data using the Naïve Bayes Classifier method. This study uses attributes, namely study program, value criteria, gender and field of work and the labels used are status (Linear and Non-Linear). The results of the study on the classification of the profile of graduates using the Naïve Bayes Classfier method show that alumni work not according to the profile of graduates by 73% and according to the profile of graduates by 23%, with a data accuracy rate of 87% and are included in the good classification category. gender and field of work and the labels used are status (Linear and Non-Linear). The results of the study on the classification of the profile of graduates using the Naïve Bayes Classfier method show that alumni work not according to the profile of graduates by 73% and according to the profile of graduates by 23%, with a data accuracy rate of 87% and are included in the good classification category. gender and field of work and the labels used are status (Linear and Non-Linear). The results of the study on the classification of the profile of graduates using the Naïve Bayes Classfier method show that alumni work not according to the profile of graduates by 73% and according to the profile of graduates by 23%, with a data accuracy rate of 87% and are included in the good classification category.

INTRODUCTION

The graduate profile is the role of graduates of the study program or field of work that is planned after completing education from the study program. Determination of the profile of graduates of study programs is generally carried out based on the results of studies of stakeholder needs. The graduate profile is prepared based on the competency study material for the graduate curriculum of the study program. Based on data from the Ministry of Research, Technology and Higher Education, the IT study program is one of the most numerous majors or study programs in Indonesian universities and the largest number of enthusiasts choose the study program every year. The needs of stakeholders for IT graduates are increasing every year both from within and outside the country, but stakeholders have difficulty getting quality graduate competencies. Graduates of the IT study program each year have a large number of graduates, both vocational and non-vocational universities. According towww.kompas.comwhich was published on May 15, 2019 explained that 60% of IT graduates are not ready to work, the number of IT graduates is large but they have low graduate competence, even they do not have competence in the IT field.

Medan State Polytechnic (POLMED) is a state vocational college in the city of Medan, has 5 (five) departments, one of which is the Department of Computer Engineering and Information Technology (KI). has various profiles of graduates in the IT field, the KI Department has 2 Study Programs, namely Computer Engineering (CE) which has a graduate profile as a Network Engineer and Network Administrator, Programmer, Embedded System, and Computer Support and the Informatics Management (MI) study program has a graduate profile as Business Analyst, Programmer, Software Engineering and IT Entrepreneur. The KI department graduates every year + 180 alumni with 60% female and 40% male presentations.

Tracer study is a method that is often used by universities to obtain feedback or satisfaction from alumni/graduates and even stakeholders, with useful feedback making curriculum improvements, curriculum development, improving education system services. Tracer studies can be done online or manually and usually tracer studies are carried out only when accreditation is needed. Polmed conducts tracer studies conducted online and carried out every year. The results of the tracer studies that have been carried out so far have not helped the IP major for improvement, curriculum development and improving the quality of education system services because the results of the tracer study only explain in general working and not working, satisfied and dissatisfied with the education system service, so it cannot be a benchmark for achievement. competence of graduates. The results of the tracer study do not explain that alumni work according to the profile of graduates of each study program and the presentation of each graduate profile, and the grouping of the percentage of graduate profiles based on gender. With the results of a tracer study like this, it is difficult for the IP majors to make improvements in all areas, especially curriculum improvements and academic services.

The Naïve Bayes Classifiers algorithm is one of the methods found in data mining to group or classify easily and has high accuracy.

The purpose of this study, in accordance with Presidential Decree no. 08 of 2012 and Permenristekdikti No. 44 of 2015 concerning KKNI and SN-Dikti. So, the researcher conducted a study on the classification of the profile of graduates who have worked in the Department of Computer Engineering and Informatics at the Medan State Polytechnic based on graduate tracer study data using the nave Bayes classifier algorithm so that it can be a benchmark for achieving graduate competence and improving the curriculum and academic service system.

METHOD

The stages to be carried out in this study are illustrated in Figure 1
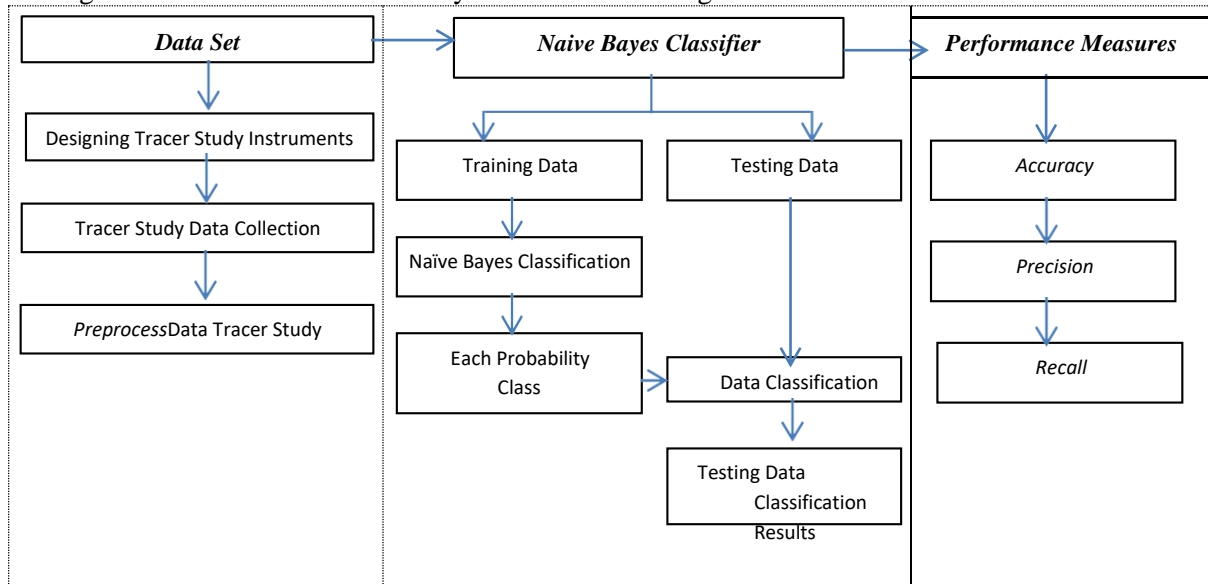

Figure 1. Research Stages

1. *Data Set*

   At this stage, to obtain a data set according to the needs of this research data, the researchers designed a tracer study instrument or questionnaire, after the questionnaire design process the next process was to distribute questionnaires to respondents online, where respondents were alumni of the Department of Computer Engineering and Informatics at the Medan State Polytechnic who were already working. . Variable questionnaires given to respondents in the form of field of work / work skills or according to the profile of graduates of the Department of Computer Engineering and Informatics, Medan State Polytechnic. Alumni questionnaire data will be preprocessed, namely deleting redundancy data and processing graduate profile data to work according to the profile of graduates of the majors (linear) or work that is not in accordance with the profile of majors (non-linear).

2. *Naive Bayes Classifier*

   At this stage the data set that has been processed will be classified using the nave Bayes classifier method. The process of this stage of the data set will be divided into 2, namely training data and data testing. From the training data, the calculation process will be carried out using the nave Bayes classifier to get the probability value of each class. The probability value of the training data will be used for the data testing process so that the classification of the testing data will be obtained.

3. *Performance Measures*

   At this stage, evaluate or measure the accuracy presentation of the results of the classification of testing data. Measuring the classification accuracy of testing data, the calculation of the confusion matrix is carried out, namely accuracy, precision, recall and f-measures.

RESULT AND DISCUSSION

1. Data Tracer Study Alumni

The population of this research is alumni or graduates of the Department of Computer Engineering and Information Technology who are already working or owning a business. The following tables describe the number of samples processed based on the study program, gender, value criteria and area of expertise or occupation.

Table 1. Study program

Study program	Amount	%
Information Management(MI)	116	81%
Computer Engineering(CE)	27	19%
Total	143	100%

Source: Research Data(2020)

Table 2. Gender

Gender	Amount	%
Woman	78	55%
Man	65	45%
Total	143	100%

Source: Research Data(2020)

Table 3. Value Criteria

Value Criteria	Amount	%
Praise	52	36%
Very satisfy	91	64%
Total	143	100%

Source: Research Data(2020)

Table 4. Field of work

Department Profile	Amount	%
Business Analyst	4	3%
Programmer	20	14%
Software Engineering	6	4%
IT Entrepreneur	2	1%
Network Engineer	17	12%
Programmer	0	0%
Computer Support	14	10%
Embedded system	0	0%
Other Fields	80	56%
Total	143	100%

Source: Research data(2020)

2. Data Set

The data set is obtained from a tracer study that has been assigned a status label, namely working in accordance with the graduate profile (Linear) or not in accordance with the graduate profile (Non Linear). The number of data sets used for training data or training data is 128 records. The attributes used as data sets are Study Program, Value Criteria, Gender.

Table 5. Training Data or Training Data

No	Study program	Value Criteria	Field of work	Gender	Status
1	CE	Very satisfy	Programmer	Man	LINEAR
2	CE	Very satisfy	Network Engineer and Network Administrator	Man	LINEAR
3	CE	Praise	And others	Woman	NON LINEAR
4	CE	Very satisfy	Network Engineer and Network Administrator	Woman	LINEAR
5	CE	Very satisfy	Computer Support	Man	LINEAR
6	CE	Very satisfy	Network Engineer and Network Administrator	Man	LINEAR

7	CE	Praise	And others	Man	NON LINEAR
8	CE	Very satisfy	And others	Man	NON LINEAR
9	CE	Very satisfy	Computer Support	Man	LINEAR
10	MI	Praise	And others	Woman	NON LINEAR
11	CE	Very satisfy	Programmer	Man	LINEAR
12	MI	Very satisfy	Computer Support	Man	NON LINEAR
13	CE	Very satisfy	And others	Man	NON LINEAR
14	CE	Very satisfy	And others	Woman	NON LINEAR
15	CE	Very satisfy	And others	Man	NON LINEAR
16	CE	Very satisfy	And others	Woman	NON LINEAR
17	MI	Very satisfy	Network Engineer and Network Administrator	Man	NON LINEAR
18	MI	Praise	Programmer	Man	LINEAR
19	CE	Very satisfy	Network Engineer and Network Administrator	Man	LINEAR
20	MI	Praise	And others	Woman	NON LINEAR
.	.	.	.	.	.
128	MI	Praise	Programmer	Woman	LINEAR

Source: Data Processing(2020)

3. Naive Bayes Classifer

From the training data table, calculations are carried out using the nave Bayes classifier method as follows:

a. Number of Class/Label
   Linear Graduate Profile Class:
   $$P(Y=Linear)=42/128=0.328$$

   Non-linear Graduate Profile Class:
   $$P(Y=Non\ Linear)=86/128=0.672$$
b. The probability of each class/attribute

Table 6. Study Program Probability

Study program	Linear	Non Linear	P(Linear)	P(Non Linear)
Computer Engineering(CE)	15	9	0.357	0.105
Information Management(MI)	27	77	0.643	0.895
Total	42	86		

Source:Data Processing(2020)

Table 7. Probability Criteria Value

Value Criteria	Linear	Non Linear	P(Linear)	P(Non Linear)
Praise	16	31	0.381	0.360
Very satisfy	26	55	0.619	0.640
Total	42	86		

Source: Data Processing(2020)

Table 8. Probability Criteria Gender

Gender	Linear	Non Linear	P(Linear)	P(Non Linear)
Man	26	32	0.619	0.372
Woman	16	54	0.381	0.628
Total	42	86		

Source: Data Processing(2020)

c. Testing data or test data to be classified

Table 9. Testing data or Test Data

No	Study program	Value Criteria	Gender	Status
1	MI	Very satisfy	Woman	??????
2	MI	Praise	Man	??????
3	MI	Praise	Woman	??????
4	MI	Very satisfy	Woman	??????
5	MI	Very satisfy	Woman	??????
6	CE	Very satisfy	Man	??????
7	MI	Very satisfy	Man	??????
8	MI	Praise	Woman	??????
9	CE	Very satisfy	Man	??????
10	MI	Praise	Woman	??????
11	MI	Very satisfy	Man	??????
12	MI	Very satisfy	Man	??????
13	CE	Very satisfy	Man	??????
14	MI	Very satisfy	Woman	??????
15	MI	Praise	Woman	??????

Source: Research Data(2020)

1. Test Testing Data 1
   The probability of being included in the Linear Graduate Profile
   category is Class. Linear = 0.328X0.643X0.619X0.381
   Class.Linear=0.050
   Probability belongs to the category of Non Linear Graduate
   Profile Class. Linear=0.672X0.895X0.640X0.628
   Class.Linear=0.242

2. Test Testing Data 2
   The probability of being included in the Linear Graduate Profile
   category is Class. Linear = 0.328X0.643X0.381X0.619
   Class.Linear=0.050

   The probability of being included in the Non-Linear Graduate Profile category is
   Class.Linear=0.672X0.895X0.360X0.372
   Class.Linear=0.081

3. Test Testing Data 3
   The probability of being included in the Linear Graduate Profile
   category is Linear Class = 0.328X0.643X0.381X0.381
   Class.Linear=0.031
   The probability of being included in the Non-Linear Graduate
   Profile category is
   Class.Linear=0.672X0.895X0.360X0.628
   Class.Linear=0.136

4. Test Testing data 4
   The probability of being included in the Linear Graduate Profile
   category is Class.Linear = 0.328X0.643X0.619X0.381

Class.Linear=0.050

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.640X0.628
Class.Linear=0.242

5. Test Testing Data 5

The probability of being included in the Linear Graduate Profile
category is Class.Linear = 0.328X0.643X0.619X0.381
Class.Linear=0.050

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.640X0.628
Class.Linear=0.242

6. Test testing data 6

The probability of being included in the Linear Graduate Profile
category is Class.Linear = 0.328X0.357X0.619X0.619
Class.Linear=0.045

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.105X0.640X0.372
Class.Linear=0.017

7. Test testing data 7

The probability of being included in the Linear Graduate
Profile category is Class.Linear =
0.328X0.643X0.619X0.619
Class.Linear= 0.081

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.640X0.372
Class.Linear= 0.143

8. Test testing Data 8

The probability of being included in the Linear Graduate
Profile category is Linear Class =
0.328X0.643X0.381X0.381
Class.Linear=0.031

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.360X0.628
Class.Linear=0.136

9. Test data testing 9

The probability of being included in the Linear Graduate
Profile category is Class.Linear =
0.328X0.357X0.619X0.619
Class.Linear=0.045

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.105X0.640X0.372
Class.Linear=0.017

10. Test Testing data 10

The probability of being included in the Linear Graduate
Profile category is Linear Class =
0.328X0.643X0.381X0.381
Class.Linear=0.031

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.360X0.628
Class.Linear=0.136

11. Test data testing 11

The probability of being included in the Linear Graduate
Profile category is Class.Linear =
0.328X0.643X0.619X0.619
Class.Linear= 0.081

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.640X0.372
Class.Linear= 0.143

12. Test testing data 12

The probability of being included in the Linear Graduate
Profile category is Linear Class =
0.328X0.643X0.381X0.381
Class.Linear=0.031

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.360X0.628
Class.Linear=0.136

13. Test Testing data 13

The probability of being included in the Linear Graduate
Profile category is Class.Linear =
0.328X0.357X0.619X0.619
Class.Linear=0.045

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.105X0.640X0.372
Class.Linear=0.017

14. Test Testing data 14

The probability of being included in the Linear Graduate
Profile category is Class.Linear =
0.328X0.643X0.619X0.381
Class.Linear=0.050

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.640X0.628
Class.Linear=0.242

15. Test Testing data 15

The probability of being included in the Linear Graduate
Profile category is Linear Class =
0.328X0.643X0.381X0.381
Class.Linear=0.031

The probability of being included in the Non-Linear Graduate
Profile category is
Class.Linear=0.672X0.895X0.360X0.628
Class.Linear=0.136

The results of the classification of testing data using the nave Bayes method with the number of test

data 15 records as follows:

Table 10 . Results of Graduate Profile Classification using nave Bayes classifier

No	Study Program	Value Criteria	Gender	Class Linear	Class NonLinear	Predicated Class	actual Class
1	MI	Very Entertaining	Woman	0.050	0.242	Non Linear	Non Linear
2	MI	Praise	Man	0.050	0.081	Non Linear	Non Linear
3	MI	Praise	Woman	0.031	0.136	Linear	Non Linear
4	MI	Very satisfy	Woman	0.050	0.242	Non Linear	Non Linear
5	MI	Very satisfy	Woman	0.050	0.242	Non Linear	Non Linear
6	CE	Very satisfy	Man	0.045	0.017	Linear	Linear
7	MI	Very satisfy	Man	0.081	0.143	Non Linear	Linear
8	MI	Praise	Woman	0.031	0.136	Non Linear	Non Linear
9	CE	Very satisfy	Man	0.045	0.017	Linear	Linear
10	MI	Praise	Woman	0.031	0.136	Non Linear	Non Linear
11	MI	Very satisfy	Man	0.081	0.143	Non Linear	Non Linear
12	MI	Very satisfy	Man	0.031	0.136	Non Linear	Non Linear
13	kindergarten	Very satisfy	Man	0.045	0.017	Linear	Linear
14	MI	Very satisfy	Woman	0.050	0.242	Non Linear	Non Linear
15	MI	Praise	Woman	0.031	0.136	Non Linear	Non Linear

Source: Data Processing(2020)

4. *Performance Measures*

The performance of nave Bayes classification can be measured using a confusion matrix table, to determine the amount of test data that is correctly classified and the test data is incorrectly classified.

Table 11. Confusion matrix test

15 Rows	Linear	Non Linear	
Actual : Linear	3(TP)	1(FP)	4
Actual : Non Linear	1(FN)	10(TN)	11
	4	11	

Source: Data Processing(2020)

Based on the confusion matrix test table, it can be calculated recall, precision, f-measures and accuracy.

Table 12. Classification Performance Measurement

Measure	Formula	Result	Percent(%)
Recall	$\dfrac{TP}{TP+FN}$	$\dfrac{3}{3+1}$	75%
Precision	$\dfrac{TP}{TP+FP}$	$\dfrac{3}{3+1}$	75%
F-measure	$\dfrac{2XRXP}{R+P}$	$\dfrac{2X75X75}{75+75}$	75%
Accuracy	$\dfrac{TP+TN}{TP+TN+FP+FN}$	$\dfrac{3+10}{3+10+1+1}$	87%

Source: Data Processing(2020)

Based on the classification performance measurement table, the test accuracy level of 15 records testing data is 87%, so the classification accuracy level for this study is high.

In the confusion matrix test table there are 15 data as test data for classifying graduate profiles using nave Bayes, it is found that graduates who work according to the graduate profile are 3/15 or 20% while graduates who do not work according to the graduate profile are 10/15 or 67% then It can be concluded that the alumni of the Department of Computer Engineering and Informatics at the Medan State Polytechnic work according to the profile of graduates of the curriculum by 20%.

CONCLUSION

From the test results using a confusion matrix table, the data accuracy level for the classification of graduate profiles using the nave Bayes classifier is 87% or is included in the good classification category, Recall is 75%, Precision is 75% and f-measures is 75%. Based on the results of the classification can be used as input for decision makers. Based on training data as many as 128 records and test data as many as 15 records produced graduates who worked according to the graduate profile as many as 4 alumni or 27% and the results of the classification of alumni working according to the graduate profile were 11 alumni or 73%.

REFERENCE

[1.] Regulation of the Minister of Education, No. 73 of 2013 concerning the preparation of learning outcomes (CP) for graduates of study programs at universities. Jakarta : Government of the Republic of Indonesia

[2.] Presidential Regulation, No. 8 of 2012 concerning the preparation of a higher education curriculum based on the Indonesian National Qualifications Framework (KKNI). Jakarta : President of the Republic of Indonesia.

[3.] Jananto, A. (2013). Naive Bayes Algorithm for Finding Estimated Student Study Time. Dynamic, 18(1).

[4.] Handayani, F., & Pribadi, FS (2015). Implementation of the Naive Bayes Classifier Algorithm in Automatic Text Classification of Complaints and Community Reporting through Call Center Services 110. Journal of Electrical Engineering, 7(1), 19-24.

[5.] Saleh, A. (2015). Implementation of the naive bayes classification method in predicting the amount of household electricity use. Creative Information Technology Journal, 2(3), 207-217.

[6.] Nursalim, S., & Himawan, H. (2014). Classification of graduate work fields using the k-nearest neighbor algorithm. J Information Technologist, 10(1), 31-43.

[7.] Asikin, MF, Kurniawaty, D., Sari, SK, & Cholissodin, I. (2016). Implementation of Naïve Bayes Classifier Method for Selection of Practicum Assistants in Hadoop Multinode Cluster Simulation. Journal of Information Technology and Computer Science, 3(4), 273-278.

[8.] Septiari, D. (2016). Implementation of the Naive Bayes Classifier Method in the Eligibility Classification of Prospective Blood Donors (Case Study: PMI Demak Regency). Dian Nuswantoro University, Semarang.

[9.] Bisri, MH Implementation of the Naïve Bayes Algorithm to Predict Student Majors at SMA Kesatrian 1 Semarang.

[10.] Rifqo, MH, & Wijaya, A. (2017). Implementation of the Naive Bayes Algorithm in Determining Credit Provisions. Pseudocode, 4(2), 120-128.

[11.] Hemali, Daniela, Chris J. Hinde, and Roger G. Stone. Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages. (2009)