

**APPLICATION OF LENGTH OF STUDY AND DEGREE OF
EXCELLENCE PREDICTION FOR INFORMATICS DEPARTMENT
STUDENTS OF UMS USING NAÏVE BAYES METHOD**

Papers

Department of Informatics
Faculty of Communications and Informatics



By:

Muh Amin Nurrohmah

Yusuf Sulistyono Nugroho, S.T, M.Eng

**DEPARTMENT OF INFORMATICS
FACULTY OF COMMUNICATIONS AND INFORMATICS
UNIVERSITAS MUHAMMADIYAH SURAKARTA**

JUNE 2015

APPROVAL

Scientific Publications by title:

**APPLICATION OF LENGTH OF STUDY AND DEGREE OF
EXCELLENCE PREDICTION FOR INFORMATICS DEPARTMENT
STUDENTS OF UMS USING NAÏVE BAYES METHOD**

Has been reviewed and approved for examination at:

Day : Saturday

Date : June 20th 2015

Advisor



Yusuf Sulisty Nugroho, S.T, M.Eng.

NIK : 1197

ACCEPTANCE

Scientific Publications by title:

**APPLICATION OF LENGTH OF STUDY AND DEGREE OF
EXCELLENCE PREDICTION FOR INFORMATICS DEPARTMENT
STUDENTS OF UMS USING NAÏVE BAYES METHOD**

Prepared and presented by

Muh Amin Nurrohmat

Has been reviewed and approved for examination at:

Day : Saturday

Date : June 20th 2015

Advisor



Yusuf Sulistyvo Nugroho, S.T, M.Eng

NIK : 1197

This Publication has been accepted as one of requirements

to obtain a bachelor's degree

Date July 6th 2015



Head of Department of Informatics



Dr. Heru Suprivono, M.Sc.

NIK : 970



UNIVERSITAS MUHAMMADIYAH SURAKARTA
FAKULTAS KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI INFORMATIKA

Jl. A Yani Tromol Pos 1 Pabelan Kartasura Telp. (0271)717417, 719483 Fax (0271) 714448
Surakarta 57102 Indonesia. Web: <http://informatika.ums.ac.id>. Email: informatika@fki.ums.ac.id

SURAT KETERANGAN LULUS PLAGIASI

/A.3-IL.3/INF-FKI/VI/2015

Assalamu'alaikum Wr. Wb

Biro Skripsi Program Studi Informatika menerangkan bahwa :

Nama : MUH AMIN NURROHMAT
NIM : L200112010
Judul : APPLICATION OF LENGTH OF STUDY AND DEGREE OF
EXCELLENCE PREDICTION FOR INFORMATICS DEPARTMENT
STUDENTS OF UMS USING NAÏVE BAYES METHOD
Program Studi : Informatika
Status : **Lulus**

Adalah benar-benar sudah lulus pengecekan plagiasi dari Naskah Publikasi Skripsi, dengan menggunakan aplikasi Turnitin.

Demikian surat keterangan ini dibuat agar dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb

Surakarta, 24 Juni 2015

Biro Skripsi
Informatika

Adjie Sapetra, S.Kom

**Turnitin Originality Report**

APPLICATION OF LENGTH OF STUDY
AND DEGREE OF EXCELLENCE
PREDICTION FOR INFORMATICS
DEPARTMENT STUDENTS OF UMS
USING NAÏVE BAYES METHOD by Muh
Amin Nurrohmah

From publikasi september 2015 (publikasi)

Similarity Index

9%**Similarity by Source**

Internet Sources:	3%
Publications:	0%
Student Papers:	9%

Processed on 24-Jun-2015 12:25 WIB

ID: 552412623

Word Count: 2851

sources:**1**

3% match (student papers from 08-Jul-2014)

Class: publikasi maret 2014

Assignment:

Paper ID: [439475766](#)

2

2% match (student papers from 10-Mar-2015)

Class: publikasi

Assignment:

Paper ID: [514619482](#)

3

1% match (Internet from 11-Jun-2015)

<http://eprints.ums.ac.id/view/type/s1/2014.html>

4

1% match (student papers from 07-Jun-2013)

[Submitted to Universitas Muhammadiyah Surakarta on 2013-06-07](#)

5

1% match (student papers from 10-Mar-2015)

Class: publikasi

Assignment:

Paper ID: [514651061](#)

6

1% match (student papers from 04-Feb-2014)

Class: publikasi maret 2014

Assignment:

Paper ID: [393366374](#)

paper text:

APPLICATION OF LENGTH OF STUDY AND DEGREE OF EXCELLENCE PREDICTION FOR
INFORMATICS DEPARTMENT STUDENTS OF UMS USING NAÏVE BAYES METHOD

**1 Papers Department of Informatics Faculty of Communications and
Informatics By:**

Muh Amin Nurrohmah

**APPLICATION OF LENGTH OF STUDY AND DEGREE OF
EXCELLENCE PREDICTION FOR INFORMATICS DEPARTMENT
STUDENTS OF UMS USING NAÏVE BAYES METHOD**

Muh Amin Nurrohmat, Yusuf Sulistyo Nugroho

Department of Informatics, Faculty of Communications and Informatics

Universitas Muhammadiyah Surakarta

Email : amin.nurrohmat@gmail.com

ABSTRACT

Informatics department of Muhammadiyah University of Surakarta has large data. The data are active students and graduate students. Every year, the data becomes larger. On the other hand, the department cannot manage the data well, thus it means that if the data is larger, then the information is smaller. The solution to solve the problem is that the data must be converted into information. This research discusses how to maximize the data into information using *data mining* technique. This research uses *Naïve Bayes* method. It is used to analyze the data, especially in the process of pattern recognition, predicting length of study, and predicting the degree of excellence. After processing the data, the application will display the report, the summary report, and suggestion. Based on the results, the application helps Informatics department to find a solution and take a decision to determine the policy. It is in order to decide where Informatics department will promote its department. Moreover, if Informatics department can recruit good student, it can improve the quality of Informatics department.

Keywords: *data mining, degree of excellence, length of study, Naïve Bayes method*

INTRODUCTION

The development of information technology is very rapid. Now and in the future, information becomes a crucial element in life. Information technology generates large data. Data includes education, economy, industry, and others. These can increase the need for information. However, the need is not offset by presenting sufficient information. The big data cannot be used optimally. Thus, in presenting the information needed require data reprocessing.

The use of *data mining* techniques are expected to provide information where the previous data is only hidden in the data warehouse, it makes the data as valuable information (Huda, 2010).

The implementation of information technology in university generates large data. The data such as student data and graduation data that always grows every year. The data include personal data, the number of graduation and academic score. Based on the data, the hidden information can be found by exploring the data to improve the

competitiveness of university and to support strategic decision-making.

Mauriza (2013), based on data from 342 college students, results show that the samples using *Naïve Bayes* method shows that students will not graduate on time is 256 students or 74.85% of the total sample. This shows that many students who take study more than 8 semesters. Therefore, the department needs to utilize the student data and the graduation data. The data is used to determine student graduation rates by using *data mining* techniques. This research still has weaknesses. The weakness is that the implementation of data mining is still using the existing application. The application was WEKA.

Based on the problems, this research develops application to help the Informatics department. This application predicts the length of study and degree of excellence of Informatics students, Faculty of Communication and Informatics UMS by using *data mining* techniques. Selected techniques are *Naïve Bayes* method. The writer merges the technique with student

data and graduation data. The writer hopes that the use of *data mining* techniques with *Naive Bayes* method can find the information about the length of study and degree of excellence. The use of the application helps Informatics department in finding solutions and policies to improve student achievement, thus students can complete their studies on time.

LITERATURE REVIEW

Nugroho and Setyawan (2014) state in their research titled "*Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi Dan Informatika Universitas Muhammadiyah Surakarta Menggunakan Algoritma C4.5*" that there is a relationship between the classification of the length of study to the data FKI UMS's student who have graduated by using the Decision Tree algorithm C4.5. The writer took 341 data of the 2358 data. The data is the student data who have already graduated. The writer uses some attributes such as school major, gender, school, the average number of credits per

semester, and the role as an lab assistant. The results show that the highest influence variable on the length of study is the average credits per semester. Moreover, the results indicate that the variables that need to be used as consideration for faculty to obtain the effective rate of the length of study is the average of credits.

Setyawan (2014), in his research titled "*Klasifikasi Prestasi Akademik Mahasiswa FKI UMS Menggunakan Metode Decision Tree*" aims to classify students' academic achievement. Student achievement is obtained by using data mining and provides a strategic plan for the Informatics department to find out new students, and to improve student's academic achievement. It increases the accreditation score for informatics department. The writer chooses 342 students as a sample. The classification result shows that male students are students who are predicted to have a satisfactory GPA from Surakarta residency. The classification results predict that the students who are not from science

are the students who have a less satisfactory GPA. The classification results predict that students who have a satisfactory GPA are the students who come from the Surakarta outside the residency. Female students from Surakarta residency are predicted to have a satisfactory GPA. The classification results predict that students who have a less satisfactory GPA are the student majoring in science and social studies. The classification results predict that male students from social studies originating from Surakarta residency, and women students who come from outside residency of Surakarta is a student with a GPA of less than satisfactory. Almost all of the students majoring in science is predicted to have a satisfactory GPA.

Mauriza (2013) in his research entitled "*Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan Metode Naive Bayes*" has the purpose to predict the length of study in the Faculty of Communication and Informatics, Muhammadiyah University of

Surakarta by using Naive Bayes methods. The research is also compared the method with one another. J48 is better than others because of its ability to define and classify each attribute to each class simply. The result shows that there are 342 students as a data. The writer uses *Naive Bayes* method. Apparently, the results show that the students who will graduate on time only 86 students or 25.15%. Students who will not graduate on time 256 students or 74.85%. Based on the results, the faculty needs solutions to improve student achievement. Thus, students can graduate on time and they have satisfactory outcome. Thereby, it can help the faculty to increase the accreditation score.

According to Huda (2010), *data mining* is a mining or the invention of new information by looking for patterns or specific rules. The information is obtained from the big data and is expected to treat the condition. By utilizing student data and graduation data, it is expected to generate information on student graduation rates through *data mining*

techniques. The categories graduation rate is measured from the length of study and GPA. The algorithm is a *apriori* algorithm. The information is presented as the support and the confidence of each category graduation rates. The result is an application to obtain useful information about the graduation rates of students with *data mining* techniques.

RESEARCH METHOD

1. Data Mining Analysis

This research will seek the greatest probability of each attribute.

a. Collecting The Data

This research requires data from all Informatics students either already passed or not passed. This research uses student data and graduation data.

1) Graduation Data

Graduation data are used as training data is the data of students who have passed. Sample data is data student of 2007 to 2010. The attributes of

graduation data are school major, gender, region, school, assistant lab, length of study, and degree of excellence.

2) Student Data

Student data that are used as testing data is the student data which is still active as the students. The writer takes the data randomly as a sample. The attributes of student data are school major, gender, region, school, assistant lab.

b. Data Needs

Determination needs is a necessary in assisting the development of *data mining*. Based on the graduation data attributes and student data attributes, then, it is divided into a variable, such as:

1) School Major: Science, Social and others.

2) Gender: Male and Female.

3) Region: This explains student's home, or students address. The

students are divided based on Indonesia time zone.

- 4) School: This explains where students took the course (Senior high school). It is divided based on Indonesia time zone.
- 5) Lab Assistant: This answer the question "Was the student the lab assistant?"
- 6) Length of Study: This answer the question "Do the students graduate on time?". The criteria for students who graduate on time are if a student graduated ≤ 4 years. The criteria that students do not graduate on time are when students graduate > 4 years.
- 7) Degree of excellence: Predicates graduation are Cumlaude if $GPA > 3.51$, Very Satisfactory when $2.76 < GPA < 3.50$, or Satisfactory if $2.00 < GPA < 2.75$.

c. Cleaning Data

This research needs to perform data cleansing. Thus, the data is relevant to the needs. In addition, to avoid noise and inconsistency data on graduation data attributes and students data attributes in testing the system.

Table 1 Attribute List

Attribute	Content
School Major	IPA, IPS, and LAIN
Gender	PRIA and WANITA
Region	WIB, WITA, and WIT
School	WIB, WITA, and WIT
Lab Assistant	YA and TIDAK
Length of Study	TEPAT and TERLAMBAT
Degree of excellence	CUMLAUDE, SANGAT MEMUASKAN, and MEMUASKAN

d. Data Transformation

Data transformation is a process of changing or merging in the appropriate format which to be processed in *data mining*. Normally, the data that be used in the *data mining* process has a format that cannot be directly used,

therefore the format needs to be changed. The predicted variable is the length of study and degree of excellence. Based on the two categories, it can be developed some categories, which is the combination between two categories.

Table 2 Data Transformation

Category	Description
A1	The students complete their course on time. Degree of excellence is Cumlaude.
A2	The students complete their course on time. Degree of excellence is very satisfactory.
A3	The students complete their course on time. Degree of excellence is satisfactory.
B1	The students complete their course lately. Degree of excellence is Cumlaude.
B2	The students complete their course lately. Degree of excellence is very satisfactory.
B3	The students complete their course lately. Degree of excellence is satisfactory.

e. The Use of *Naïve Bayes* Method

Naive Bayes is a method for *data mining* process. It is because this method is easy to implement and does not use numerical optimization, matrix

calculations and others. Bayes is a simple probabilistic-based prediction technique. Bayes is based on the Bayes theorem with strong independence assumption (Prasetyo, 2012).

2. Flowchart

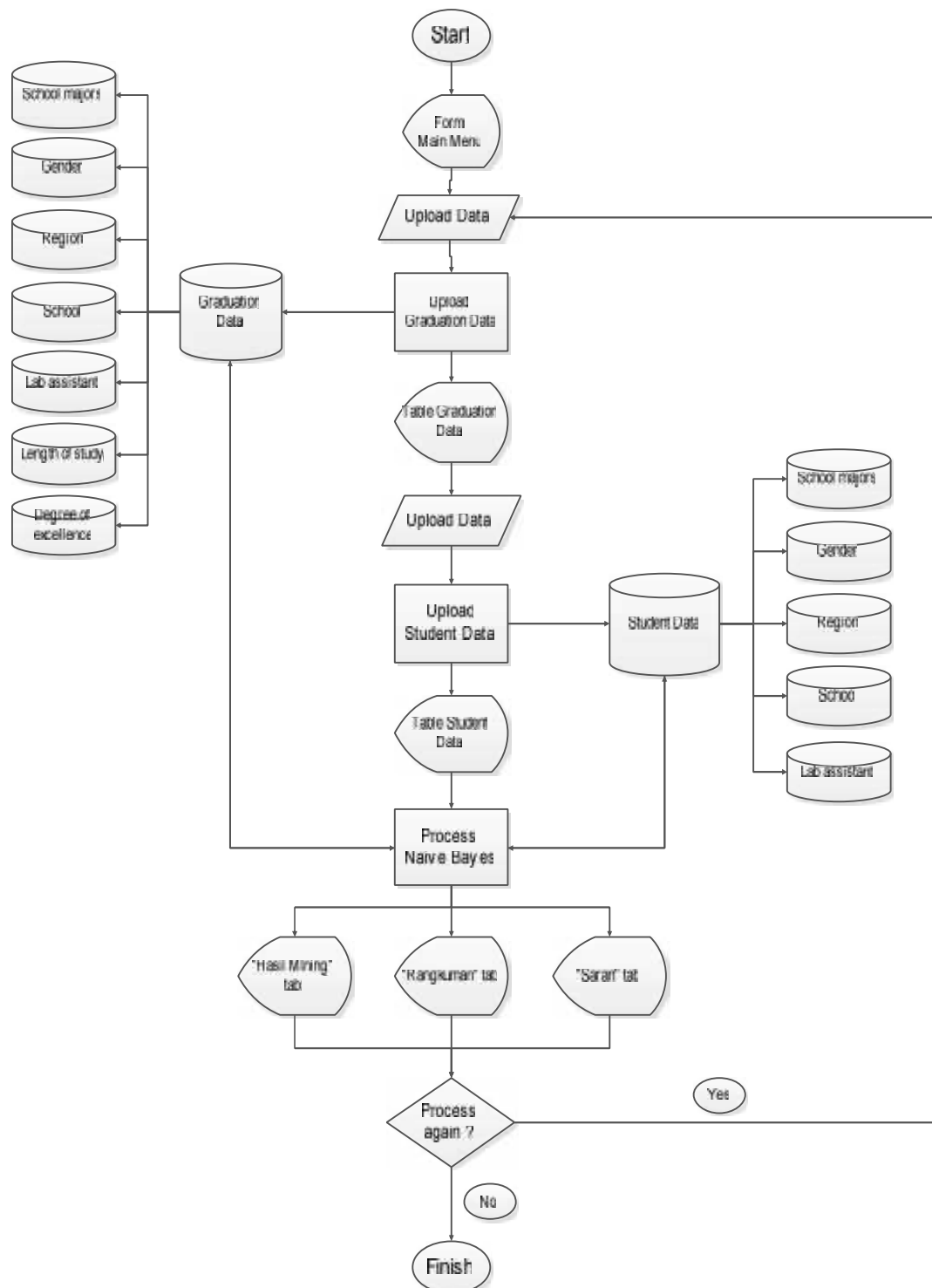


Figure 1 Flowchart of Application

3. Use Case

The Use Case Diagram describe about the interaction between actor and system. Based on the

actor identification and the identification of use case diagram, the User actor has a use case.

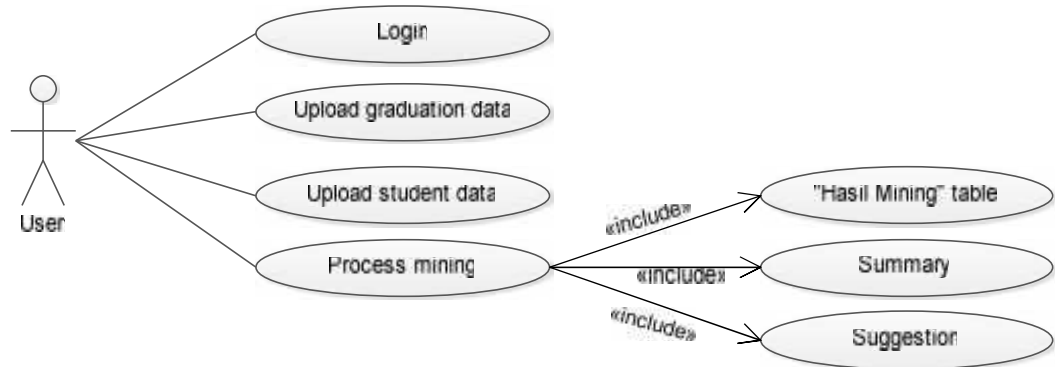


Figure 2 User Use Case

RESULT AND DISCUSSION

The research result is *data mining* application to predict the student length of study and the student degree of excellence for informatics department by using *Naïve Bayes* method. In developing the application, the writer needs two data, such as graduation data and student data with attributes as follow as school major, gender, region, school, lab assistant, length of study, and degree of excellence.

In figure 3, if the user runs the application, application will display login form. User can access the

system by filling username and password. If username and password are correct, then the user will be directed to main form.



Figure 3 Login Form

In figure 4, if the user succeeds to access the system, the

application will display the Main form. This form is used to access all of application commands.

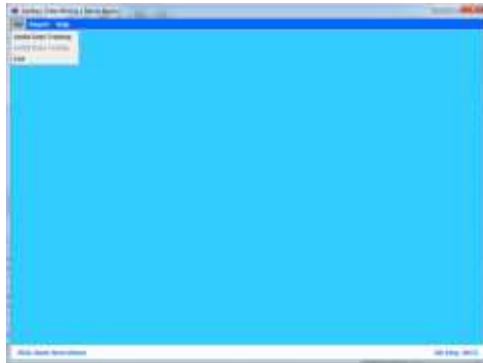


Figure 4 Main Form

In figure 5, if the user succeeds to access the system and click Upload Data Training sub menu on File menu, the application will display Upload Data Training form. The user can take graduation data, and then store the data into table by clicking Upload Data button. If the user succeeds to access Upload Data Training form, the user can process the mining process by clicking Next menu.



Figure 5 Upload Data Training Form

In figure 6, After clicking the next button in Upload Data Training form, the application will display Upload Data Testing form. The user can take student data and save it into table by clicking Upload Data button. If the user succeeds to save the data, then the data is ready to process. The user can click “Proses Mining” button to process the mining process by using *Naïve Bayes* method.



Figure 6 Upload Data Testing Form

In figure 7, “Hasil Mining” tab is used to display the results of mining process by using *Naïve Bayes* method. The table contains some attributes, such as the attribute of School major, gender, region, school, lab assistant that is taken from student data and length of study attribute, and degree of excellence attribute that is taken from mining process that use *Naïve Bayes* method

in graduation data. There is an Export excel button that is used to save the result into MS Excel.

The screenshot shows a table with multiple columns and rows of data. The data is organized into a grid with alternating green and yellow rows. At the bottom of the table, there is a legend with colored squares corresponding to different categories.

Figure 7 “Hasil Mining” Table

In figure 8, “Rangkuman” Tab is used to display the summary of mining process. The summary is displayed as the category of data transformation process.

The screenshot shows a report titled 'Rangkuman' with a list of categories and their corresponding values. The categories are listed on the left, and the values are listed on the right. The report is displayed in a table-like format.

Figure 8 “Rangkuman” Report

“View Grafik” button in “Rangkuman” tab is to see the summary graphically. It displays in figure 9.

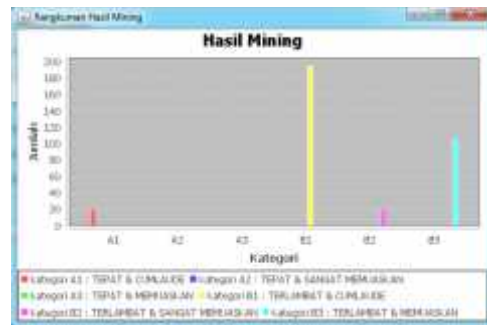


Figure 9 “Rangkuman” Graphic

In figure 10, “Saran” tab is used to display suggestion to informatics department. This suggestion is useful to take a decision, find a solution, and give a policy to improve student achievement. Thus, the student can complete the study on time.

The screenshot shows a report titled 'Saran' with two columns of suggestions. Each suggestion is listed with a category and a value. The suggestions are organized into two columns, each with a title and a list of items.

Figure 10 “Saran” Report

“View Grafik” button in “Saran” tab is to see the summary graphically. It displays in figure 11.



Figure 11 “Saran” Graphic

In figure 12, About menu contains writer photo, name, student number, department, and email.



Figure 12 About Form

In figure 13, “Petunjuk” menu contains how to use the application.



Figure 13 “Petunjuk” Form

CONCLUSION

Based on the result, the writer concludes that:

1. This application can display the information about the prediction result of length of study and degree of excellence in the table form by using graduation data and student data.
2. Graduation data and student data that is processed by this application are school major, gender, region, school, assistant lab, length of study, and degree of excellence.
3. After mining the data, the method to take a decision is to see the suggestion level which is suggested to Informatics department of UMS. It can be a reference in order to maximize promotion of department.

BIBLIOGRAPHY

- Huda, Nuqson Masykur. 2010. *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*, Skripsi. Semarang: Fakultas MIPA Universitas Diponegoro.
- Mauriza, Ahmad Fikri. 2013. *Implementasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika UMS Menggunakan Metode Naive Bayes*, Skripsi. Surakarta: Fakultas Komunikasi dan Informatika UMS.
- Nugroho, Yusuf Sulistyو dan Setyawan. 2014. *Klasifikasi Masa Studi Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta Menggunakan Algoritma C4.5*, Jurnal KomuniTi, Vol. VI, No. 1 Maret 2014. Surakarta: Fakultas Komunikasi dan informatika UMS.
- Prasetyo, Eko. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Setyawan. 2014. *Klasifikasi Prestasi Akademik Mahasiswa FKI UMS Menggunakan Metode Decision Tree*, Skripsi. Surakarta: Fakultas Komunikasi dan Informatika UMS.

Biography

Name : Muh Amin Nurrohmat
NIM : L200112010
Place of Birth : Sragen
Date of Birth : June 16th 1993
Gender : Male
Religion : Islam
Education : S1
Department / Faculty : Department of Informatics / Communications and Informatics
Higher Education : Universitas Muhammadiyah Surakarta
Home Address : Pedak RT. 02/01, Karangwaru, Plupuh, Sragen
Phone number : 085728559827
Email : amin.nurrohmat@gmail.com