ADI International Conference Series
Vol 4 January 2022

p-ISSN : 2774-9576
e-ISSN : 2747-2981

Prediction of Graduation Time for Students Faculty of Computer Science Amikom University Yogyakarta Using Naïve Bayes Method

Adhy Syahputra[1], Rum Mohamad Andri K. Rasyid[2], Istiningsih[3], Ardi Surwiyanta[4], Haile Qudrat Djojodibroto[5], Edy Anan[6]
Information System AMIKOM University Yogyakarta[1,2], Economics AMIKOM University Yogyakarta[3], Hospitality Yogyakarta Tourism Academy[4], Communication Studies AMIKOM University Yogyakarta[5], Accounting Program AMIKOM University Yogyakarta[6]
Jl Ringroad Utara, Condongcatur, Depok, Sleman, Yogyakarta Indonesia 55283[1,2,3,4,5,6]
Indonesia[1,2,3,4,5,6]
e-mail: adhy.syahputra@students.amikom.ac.id[1]andri@amikom.ac.id[2], Istiningsih@amikom.ac.id[3], surwiyanta@gmail.com[4], haile.djojodibroto@amikom.co.id[5] edyanan@amikom.ac.id[6]

To cite this document:
Syahputra, A., Rasyid, R. M. A. K. ., Istiningsih, I., Surwiyanta, A., Djojodibroto, H. Q., & Anan, E. . (2022). Prediction Of Graduation Time For Students Faculty Of Computer Science Amikom University Yogyakarta Using Naïve Bayes Method. Conference Series, 4(1), 142–151. https://doi.org/10.34306/conferenceseries.v4i1.709
Hash: ABCKyNpfpZZTnbpNTlVLlgZ4YHtOk8Lzl90clXOhjYAvjXeHRqc4cafK0KzOpKNl

Abstract

The graduation rate on time in college cannot be taken lightly. In many cases, it is found that there is an imbalance in the number of students who enter and who have completed their studies so that there is a high accumulation of the number of students in each batch. It is necessary to know the factors that cause students not to graduate on time. Data classification techniques can be used to predict the timeliness of student graduation. The algorithm used is the Naïve Bayes algorithm with data from 122 students from the Informatics Engineering and Information Systems study program at the Faculty of Computer Science, AMIKOM University, Yogyakarta. Based on the results that have been carried out, it can be concluded that the Naïve Bayes algorithm is applied accurately to predict the timeliness of students' graduation. The results of testing the level of prediction accuracy resulting from the trial reached 90.16%.

Keywords: Graduation On-Time, Students, Prediction, Nave Bayes Algorithm, Accuracy

1. Introduction

1.1 Background

AMIKOM University Yogyakarta is a university in the Service Institutions Area of the Directorate of Higher Education Region V, Special Region of Yogyakarta. This college first operated in 1994 under the name of the Academy of Informatics and Computer Management which has two Diploma study programs, namely Informatics Engineering and Informatics Management. In 2002 the Academy of Informatics and Computer Management AMIKOM Yogyakarta changed its form to become the AMIKOM Yogyakarta Faculty of Information and Computer Management which has five programs namely Information Management, Informatics Engineering at the Diploma level, bachelor's degree Informatics Engineering, Master of Informatics Engineering and Information Systems. In 2017 the AMIKOM University Yogyakarta, Faculty of Informatics and Computer Management changed its form to AMIKOM Yogyakarta University, four of the five existing study programs were incorporated in the Faculty of Computer Science except for the Informatics Engineering Study Program at the master's level.

Every year the number of new students from the four study programs above averages about 2200 new students but every year only 2000 graduates so that there is an average of 2000 or about 10% of students per year who cannot graduate on time so that it will be an obstacle in the future. day. This should have been anticipated if known later. This should be anticipated if it is known that students who are expected to graduate are not on time.

Naive Bayes is a statistical classification that can be used to predict the probability of membership of a class. Naive Bayes is based on Bayes theorem which has similar classification capabilities to decision trees and neural networks. Naive Bayes is proven to have high accuracy and speed when applied to databases with large data [1].

Forecasting about the probability that a student will graduate on time or not is a difficult thing because often a student who is likely, not able to graduate on time does not show special characteristics or patterns that can be learned, so it is necessary to design a tool that can study the characteristics that can be studied. will then be translated as a pattern.

1.2 Problem Formulation

1. How to predict the punctuality of a student's graduation?
2. To what extent is the accuracy of the Naïve Bayes algorithm able to predict the timeliness of graduates at the Faculty of Computer Science, University of AMIKOM Yogyakarta?

1.3 Limitation of the Problem

1. This research is only limited to graduation at the Faculty of Computer Science and does not include other faculties, considering that up to the time this research was conducted, Faculties other than Computer Science had not yet produced graduates.
2. The subjects in this study were undergraduate students at the Informatics and Information Systems Study Program.
3. Timeliness of graduation is calculated based on the study period for each level that has been previously determined by the stakeholders, namely 4 (four) years.

1.4 Research Purposes and Objectives

The purpose of this study is to determine the accuracy of the Naïve Bayes Algorithm in predicting the accuracy of student graduation.

1.5 Research Benefits

1. As documentation of research work in the form of research reports.
2. As a reference material for future researchers.

2. Research Method

Basic Theory

2.1 2.1 Literature Review

The research entitled the influence of the living environment on student achievement in SMA Negeri 1 Makmur. The purpose of the research conducted by Ernawati and Sabri Yusuf (2014) was to determine the effect of the living environment on student achievement at SMA Negeri 1

Makmur. The approach used is quantitative and the type of research is descriptive and associative. The population in this study were all students of class X MAN Term as many as 241 students and the sample in this study was 15% of the population, namely 35 students. Data collection techniques through questionnaires. The data analysis technique is hypothesis testing (t test). The results of the study obtained that it means that is 4.3106 2,033 concluded that there is a positive influence of the living environment on student achievement in SMA Negeri 1 Makmur.[2]

Research conducted by Soraya et al (2014) aims to determine the effect of education financing by parents on student achievement in class X SMA Negeri 1 Sungai Ambawang. The research method used is associative. The sample in this study amounted to 93 students. The results showed that there was an effect of educational financing by parents on the learning achievement of class X students of SMA Negeri 1 Sungai Ambawang seen from the t count > t table (13.307>1.662) so that Ho was rejected, and Ha was accepted. Simple linear regression calculation obtained the equation Y=68,181+0,205X, meaning that the constant value is 68,181, if education funding by parents is worth 0, then student learning achievement is worth 68,181. The regression coefficient value of the education financing variable by parents is 0.205. This means that every increase in education funding by parents 1, the student's learning achievement will increase by 0.205. While the coefficient of determination in this study shows the contribution of the influence of education financing by parents on learning achievement of 66.1%. [3]

Agus Salim's research (2016) discusses "The Influence of Online Games on Student Learning Behavior in the Department of Islamic Education, Faculty of Tarbiyah and Teacher Training, UIN Alauddin Makassar". There is an influence of online games on the learning behavior of students of the Department of Islamic Education, Faculty of Tarbiyah and Teacher Training at UIN Alauddin Makassar batch 2014.[4]

The study entitled "Naive Bayes Method for Graduation Prediction (Case Study: New College Student Data)" conducted by Syarli et al (2016) used the nave Bayes method in predicting student graduation which resulted in an accuracy rate of 94%. [5].

2.2 Basic Theory

2.2.1 Definition of Expert System

Expert system is a system that seeks to adopt human knowledge to computers, so that computers can solve problems as is usually done by experts, and a good expert system is designed to be able to solve a particular problem by imitating the work of experts (Kusumadewi, 2003: 109).

2.2.2 Definition of Naive Bayes

The Naive Bayes algorithm is a classification method using probability and statistical methods proposed by British scientist Thomas Bayes. The Naive Bayes algorithm predicts future opportunities based on experience, so it is known as the Bayes theorem. The main characteristic of this Nave Bayes Classifier is a very strong assumption (nave) of the independence of each condition or event.

$$P(c|x) = \frac{P(x|c)\ P(c)}{P(x)}$$

Figure 2.1 Bayes formula

Description:
x : Data with unknown class
c : Hypothesis data is a specific class
P(c|x) : Probability of hypothesis based on condition (posteriori probability)
P(c) : Probability of hypothesis (prior probability)
P(x|c) : Probability based on the conditions in the hypothesis
P(x) : Probability c

2.2.3 Student

A student is someone who's in the process of gaining knowledge or studying and is enrolled currently undergoing education at one form of tertiary institution consisting of academic, polytechnic, high schools, institutes, and universities (Hartaji, 2012: 5).

2.2.4 Cumulative Grade Point Average

Cumulative Grade Points Average (CGPA) is the average credit score which is the final unit of value that describes the value of the teaching and learning process each semester or can also be interpreted as a quantity or number that states the achievement of success in the teaching and learning process of students in a semester.

PK	Predikat
4,00	*Summa Cumlaude*
3,51-3,99	*Cumlaude*
2,76-3,50	Sangat Memuaskan
2,00-2,75	Memuaskan

Table 2.1 Assessment of Cumulative Grade Point Average

2.2.5 System Design

2.2.5.1 Flowchart

According to Jogiyanto (2005), Flowchart or flow chart is a diagram with graphic symbols that state the flow of the algorithm or the process of running the program, flowchart is a diagram in the form of symbols and can show the flow of data and operations that occur in a system. Flowcharts are used as a communication and documentation aid. [6]

2.2.5.2 Data Flow Diagrams (DFD)

According to Kendall (2003). DFD is a data logic model or process that is created to describe where the data comes from and where the data comes out of the system, where the data is stored, what processes produce the data and the interaction between the stored data and the processes imposed on the data. This diagram is used to describe the flow of data on the system in a structured manner.

2.3 System Analysis and Design

2.3.1 Overview

2.3.1.1 Profile of AMIKOM University Yogyakarta

Based on grievance of several entrepreneurs who stated that finding professional workers in the field of Information Technology was difficult, the founders of this Foundation wanted to solve the difficulties in finding experts in the informatics fields are the reason AMIKOM Yogyakarta Foundation was established.

Because the public's trust in AMIKOM is getting higher as evidenced by the number of registrants that has increased significantly from year to year, the institution would feel the need to build a representative campus. On January 20, 2017, based on the Decree of the Ministry of Research, Technology and Higher Education No. 99/KPT/I/2017 STMIK AMIKOM Yogyakarta officially changed its form to AMIKOM Yogyakarta University which has 3 faculties with 1 Master's program, 13 Bachelor programs and 2 Diploma programs.

2.3.2 SWOT

Analysis SWOT analysis is one of the design methods for evaluating strengths.), weaknesses (weaknesses), opportunities (opportunities), and threats (threats). In this application, this analysis is applied and an explanation of each of the SWOT components, including:
Strength: The Information Systems Study Program and the Informatics Study Program are favorite study programs at AMIKOM University Yogyakarta, this is indicated by the number of active students in the study program Information Systems totaling 2970 and 4292 students in

the Informatics Study program according to the results of reporting on the Higher Education Database in 2018. One of the reasons for this is the high demand for computer workers, thereby increasing the interest of high school graduates to continue their education to study programs in the computer sector.

Weakness: The high interest in studying in the computer field study program is not matched by a good understanding of the curriculum so that many students are not ready to take education plus there are things that interfere with student concentration in attending lectures including the factor of availability of funds, residence, marriage, interest in games and so on that hinder student graduation.

Opportunity: It is necessary to build a system that can predict the timeliness of student graduation based on the student's condition before the student has problems with his study period so that both students and universities can take preventive measures so that students do not exceed the specified study period.

Threat: Students who do not have a good understanding of the curriculum and disturbances during the study period are threatened not to graduate on time and even to be returned to their parents (dropping out) without being predicted either by themselves or by the University.

.

2.3.5 System Design

2.3.5.1 Data Flow Diagram

1. Context Diagram / DFD Level 0

Context diagram is an outline description of the system design to be built, the context diagram will show the entire process of the system being built. Context diagrams show the inputs, processes, and outputs of the designed system. The following context diagram is shown in the figure

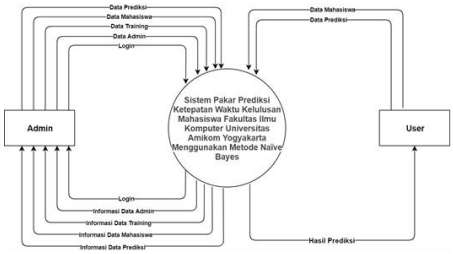

Figure 3.1 Context Diagram

2. DFD level 1

Data flow diagram level is a translation of data from context diagrams. The level 1 data flow diagram shows clearer processes or relationships between the admin and the system and users with the process system can be seen in the figure:

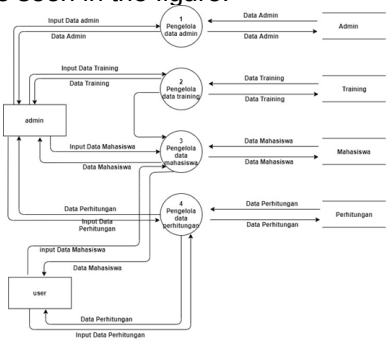

Figure 3.2 DFD Level 1

2.3.5.2 Entity Relationship Diagram (ERD)

Entity Relationship Diagram (ERD) is a concept that describes the relationship between data and is based on the perception of a real world consisting of a set of objects, called entities & the relationships between these objects. Or it can also be called a network model that uses an abstract arrangement of stored data from the system.

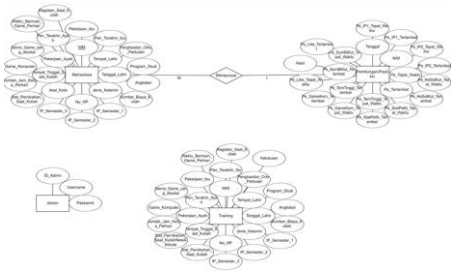

Figure 3.3 Entity Relationship Diagram (ERD)

2.3.5.3 Relationship Table

Figure 3.4 Relationship Table

3. Findings

3.1 Database Implementation

In designing this expert system, the first step is to implement the database system. The following is the implementation of the database that is applied:

3.1.1 Database Creation

This system is designed using a MySQL database using a GUI from Navicat Premium 12. The name of the system database is description.

Table 4.1 Database Creation
CREATE DATABASE description;

3.2 Application Implementation

3.2.1 Main Form

Figure 4.1 Main

Form The Main Form is the display that will appear when the operator data validation on the login form is successful.

3.2.2 Training Data Processing

Form Figure 4.2 Training Data Processing Form

The training data processing form serves to manage training data. the training data form has several processes including: adding training data, saving training data, editing training data, deleting training data. The initial display of this menu all textboxes will be empty and the admin only needs to fill in the training data and when clicking the add button, the data that has been filled in will automatically appear in the table below and a message appears that the data has been successfully added To edit existing data the admin only needs to click on the NIM in the data in the table below then the textbox will be filled with these data then the admin just changes the data that you want to change and finally click save then the data will be automatically updated with a notification message appearing that the data has been updated. then the delete button functions to delete the training data in the table below by clicking on the data you want to delete. Admin can also search for training data based on the NIM of the data.

3.2.4 Student Data Processing

Form Figure 4.3 Student Data Processing Form

The student data processing form functions to manage student data. Student data processing form has several processes including: adding student data, saving student data, editing student data, deleting student data. The initial display of this menu all textboxes will be empty and the admin only needs to fill in the training data and when clicking the add button, the data that has been filled in will automatically appear in the table below and a message appears that the data has been successfully added To edit existing data the admin only needs to click on the NIM in the data in the table below then the textbox will be filled with these data then the admin just changes the data that you want to change and finally click save then the data will be automatically updated with a notification message appearing that the data has been updated. then the delete button functions to delete the training data in the table below by clicking on the data you want to delete. Admin can also search for training data based on the NIM of the data.

3.2.5 Prediction Form

Figure 4.3 Student Data Processing Form

This form serves to predict graduation based on data that has been obtained from student tables and training data tables.

.3 Discussion

4.3.1 Application of Naïve Bayes to Data Testing

4.3.1 Application of Naïve Bayes to Data Testing

1. Counting the number per class or label

Here the researcher uses 2 classes or labels, namely the TIMELY class and the LATE class. Then count the number of classes or labels that are on time and late respectively based on the training data.

On Time	Late
140	27

2. Counting the number of cases or variables per class

Here the researcher only uses 7 cases or variables in the training data, namely variables:

a. Housing accommodation
b. Tuition Fees
c. GPA Semester 1
d. GPA Semester 2

e. Activities during study
f. Marital Status
g. Computer Games

Then count the number of cases or variables per class. The case or variable that will be calculated is based on the data to be predicted in the table above.

Housing accommodation (Kost) & On Time	Tuition Fees (Parents) & On Time	GPA Semester 1 (Good (3 to 3.5)) & On Time	GPA Semester 2 (Good (3 To 3.5)) & CorrectTime	Activities during study (Only College) & On Time	Marital Status (Not Married) & On Time	Computer Games (Very Likes) & On Time
81	123	59	68	92	137	28

Place of Residence During College (Kost) & Late	Source of Tuition Fees (Parents)) & Late	IP Semester 1 (Good (3 to 3.5)) & Late	G.P. Semester 2 (Good (3 To 3.5)) & Late	Activities During College (Lecture Only) & Late	Marital Status (Not Married) & Late	Computer Game (Very Likes) & Late
18	26	13	14	16	26	5

3. Multiplying all cases or variables per class

From the above data, calculations can be made by multiplying all cases or variables per class.

a. On-time Class

The formula would be:

$$\frac{\text{Tempat Tingal Saat Kuliah}}{\text{Data Yang Tepat Waktu}} \text{ x } \frac{\text{Sumber Biaya Kuliah}}{\text{Data Yang Tepat Waktu}} \text{ x } \frac{\text{IP Semester 1}}{\text{Data Yang Tepat Waktu}} \text{ x}$$

$$\frac{\text{IP Semester 2}}{\text{Data Yang Tepat Waktu}} \text{ x}$$

$$\frac{\text{Kegiatan Saat Kuliah}}{\text{Data Yang Tepat Waktu}} \text{ x } \frac{\text{Status Pernikahan}}{\text{Data Yang Tepat Waktu}} \text{ x } \frac{\text{Game Komputer}}{\text{Data Yang Tepat Waktu}} \text{ x}$$

$$\frac{\text{kelas itu sendiri}}{\text{jumlah data keseluruhan}}$$

Based on the data above it could be calculated:

$$\frac{81}{140}\frac{81}{140} \text{ x } \frac{123}{140}\frac{123}{140} \text{ x } \frac{59}{140}\frac{59}{140} \text{ x } \frac{68}{140}\frac{68}{140} \text{ x } \frac{92}{140}\frac{92}{140} \text{ x } \frac{137}{140}\frac{137}{140} \text{ x } \frac{28}{140}\frac{28}{140} \text{ x } \frac{140}{167}\frac{140}{167}$$

0.8786 x 0.4214 x 0.4857 x 0.6571 x 0.9786 x 0.2000 x 0.8383

Result = 0.01122

b. Delayed Class

The formula would be:

$$\frac{\text{Tempat Tingal Saat Kuliah}}{\text{Data Yang Terlambat}} \text{ x } \frac{\text{Sumber Biaya Kuliah}}{\text{Data Yang Terlambat}} \text{ x } \frac{\text{IP Semester 1}}{\text{Data Yang Terlambat}} \text{ x}$$

$$\frac{\text{IP Semester 2}}{\text{Data Yang Terlambat}} \text{ x}$$

$$\frac{\text{Kegiatan Saat Kuliah}}{\text{Data Yang Terlambat}} \text{ x } \frac{\text{Status Pernikahan}}{\text{Data Yang Terlambat}} \text{ x } \frac{\text{Game Komputer}}{\text{Data Yang Terlambat}} \text{ x}$$

$$\frac{\text{kelas itu sendiri}}{\text{jumlah data keseluruhan}}$$

Based on the data above it can be calculated:

$\frac{1818}{2727} \times \frac{2626}{2727} \times \frac{1313}{2727} \times \frac{1414}{2727} \times \frac{1616}{2727} \times \frac{2626}{2727} \times \frac{5\ 5}{2727} \times \frac{27\ 27}{167\ 167}$

0.6667 x 0.9630 x 0.4815 x 0.5185 x 0.5926 x 0.9630 x 0.1852 x 0.1617
Results = 0.00274

4. Comparing results per class

The results of class calculations are on time	The results of the calculation of the late class
0.01122	0.00274

It can be concluded that the data that has been tried to predict has a result that is on time.

The accuracy of the system built is 90.16%. It was obtained from 122 predicted data, as many as 110 data were in accordance with the predicted results, while 12 data results were not in accordance with the predictions.

$\frac{110\ 110}{122\ 122}$ x 100% = 90.16%

4. Conclusion

4.1 Conclusions

Expert System Prediction Timeliness of Graduation Students of the Faculty of Computer Science, AMIKOM University Yogyakarta using the Naïve Bayes method was built using Visual Basic.Net language using MySQL database. This system has been tested and is running.
Based on the results of research that has been carried out by the author, the following conclusions can be drawn:

1. An Expert System for Predicting Timeliness of Graduation for Students of the Faculty of Computer Science, AMIKOM University Yogyakarta Using the Naïve Bayes Method has been successfully built through the following steps: Analysis, Design / Design, Implementation, and finally Testing.
2. Able to help predict the timeliness of student graduation with existing characteristics with an accuracy rate of 90.16%.

4.2 Suggestions

Based on the analysis, authors in this case realizes that there are still many shortcomings in this study and considerations for improving performance. Expert System Prediction of Timely Graduation of Students of the Faculty of Computer Science, AMIKOM University Yogyakarta Using the Naïve Bayes Method, so criticism and suggestions from readers are highly expected. The author's suggestions in developing the system to be taken into consideration are as follows:
1. Develop a program so that it can be used directly by students.
2. Explore new parameters of student activities that affect the timeliness of graduation.

References

[1] Nugroho, Yuda Septian (2014), Data Mining Using the Naïve Bayes Algorithm for Classification of Graduates at Dian Nuswantoro University, Department of Information Systems, Faculty of Computer Science, Dian Nuswantor University, http://mahasiswa.dinus.ac.id /docs/skripsi/jurnal/13789.pdf, accessed on 24 November 2019.

[2] Ernawati Ernawati, Sabri Yusuf (2014), The Effect of Living Environment on Student Achievement at SMA Negeri 1 Makmur, Journal of Economics and Education (JSEE), http://jfkip.umuslim.ac.id/index.php/jsee/ article/view/261, accessed on September 30, 2019.

[3] Soraya, Nuraini Asriati, Husni Syahrudin, The Effect of Education Financing by Parents on the Learning Achievement of Class X State Senior High School Students, Economics Education Study Program FKIP Untan, https://media .neliti.com/media/publications/215660-elektron-funding-education-by-oran.pdf, accessed on September 30, 2019.

[4] Agus Salim, 2016, The Effect of Online Games on Student Learning Behavior in the Department of Islamic Education, Faculty of Tarbiyah and Teacher Training at UIN Alauddin Makassar, http://repositori.uin-alauddin.ac.id/11338/1/Pengaruh%20Game%20Online%20terhadap%20Behaviour%20Belajar%20Mahasiswa%20Jurusan%20Education%20Religion%20Islam%20Faculty%20Tarbiyah%20dan%20Keguruan%20UIN%20Alauddin%20Makassar.pdf, accessed on 30 September ber 2019.

[5] Syarli, Asrul Ashari Muin (2016), Naive Bayes Method for Graduation Prediction (Case Study: New College Student Data) http://ejournal.fikom-unasman.ac.id/index.php/jikom/article/view/76/52, accessed on November 24, 2019.

[6] http://sir.stikom.edu/1961/4/BAB_III.pdf, accessed on November 30, 2019.