JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 2, April 2022, hlm. 159-166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i2.1539
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

ANALYSIS AND IMPLEMENTATION NAÏVE BAYES FOR FAST TRACK STUDENT GRADUATION DIAGNOSIS AT STMIK ROYAL

Andhini Asri Awaliyah Arifin[1], Arridha Zikra Syah[1*], Akmal[1]
[1]Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
email: [*]azsyra@gmail.com

Abstract: Student graduation is one thing that needs to be considered because it is included in the College's Internal Quality Assurance Standards (SPMI). STMIK Royal is one of the universities experiencing problems with student graduation. To achieve the quality of these graduates, a graduation diagnosis with the standards that have been set for students who are still carrying out studies is very necessary so that anticipatory steps can be taken from the beginning to overcome the occurrence of graduation problems in the academic field. This research aims to diagnose fast-track student graduation by using a data mining model with a classification function. The technique used for the classification data mining model is the Naïve Bayes Algorithm. The dataset used as training and testing data is data from the 2021 STMIK Royal fast track students. The criteria used to determine student graduation are Gender, class, credits, GPA, Tuition Fee, Guidance Process, and the KKL Report. The results of the classification modeling with the Naïve Bayes Algorithm produce an accuracy value of 83%.

Keywords: data mining; graduation; naïve Bayes

Abstrak: Kelulusan mahasiswa adalah salah satu hal yang harus diperhatikan karena termasuk ke dalam Standar Penjaminan Mutu Internal (SPMI) perguruan tinggi. STMIK Royal adalah salah satu perguruan tinggi yang mengalami masalah kelulusan mahasiswa. Untuk mencapai kualitas lulusan tersebut, diagnosis kelulusan dengan standar yang telah ditetapkan untuk mahasiswa yang masih menjalankan studi sangat diperlukan sehingga dapat dilakukan langkah antisipasi dari awal untuk menanggulangi terjadinya permasalahan kelulusan dalam bidang akademik. Tujuan diadakannya penelitian ini untuk untuk mendiagnosis kelulusan mahasiswa fast track dengan menggunakan model data mining dengan fungsi klasifikasi. Teknik yang digunakan untuk model data mining klasifikasi adalah Algoritma Naïve Bayes. Dataset yang akan digunakan untuk menjadi data latih dan data uji adalah data mahasiswa fast track STMIK Royal tahun 2021. Kriteria yang digunakan untuk diganosis kelulusan mahasiswa, diantaranya Jenis Kelamin, Kelas, SKS, IPK, Uang Kuliah, Proses Bimbingan, dan Pengumpulan Laporan KKL. Hasil pemodelan klasifikasi dengan Algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 83%.

Kata kunci: data mining; kelulusan; naïve bayes

INTRODUCTION

STMIK Royal is a private university in Asahan Regency that focuses on two study programs: Information Systems and Computer Systems. STMIK Royal student graduation diagnosis information is very important to improve services that can make students feel comfortable, can graduate with a fast track based on the information obtained, and can be used for the campus accreditation process. [1].

Presenting the ups and downs of a student's ability to complete lectures quickly is one element of the campus accreditation assessment [2]. One way to increase the number of students completing their lectures quickly is to promote the fast track program.

In 2020, Information Systems Study Program at STMIK Royal opened a graduation pathway with a fast track program. As many as 60 students enrolled in the route, but only 28 students could complete the graduation with the fast track program. Based on this information, it is necessary to monitor and evaluate the tendency of students to graduate quickly, graduate on time, or graduate late. Monitoring or evaluation of performance will produce useful information to help students, lecturers, administrators, and policymakers [3].

The main factors that greatly influence students' graduation are internal and external. Internal factors come from within, which include student intelligence, understanding of the material in class, and students' independent abilities. In contrast, external factors come from outside themselves, including economic conditions, social conditions, and relationships [4].

To support this research, several complete datasets regarding the graduation of fast track students are needed, such as detailed student data starting from NIM, Name, Class Program, Gender, Total Credits, Cumulative Achievement Index (GPA), Collection of KKL Reports, and UKT that have been completed paid. In this case, the data sample is taken from Information Systems Study Program at STMIK Royal.

Information Systems Study Program at STMIK Royal stores academic data, student data, and other data that is very meaningful when the pattern or knowledge is correctly known to make decisions [5]. Analysis of the available data is also required to consider the information obtained. By utilizing data mining, each data set can provide important knowledge valuable for organizations such as educational institutions [6].

Data mining has several methods, one of which is the Naïve Bayes classification. Naïve Bayes is a simple probabilistic prediction technique based on the application of Bayes' theorem with the assumption of strong dependence. In other words, in Naïve Bayes, the model used is an independent feature model [7].

Based on research from a journal entitled "Naïve Bayes Algorithm for Finding Estimated Student Study Time," this study was conducted to estimate student learning time using Naïve Bayes and several characteristics, namely, majors, school origin, area of origin, place of birth, indicators of college outcomes 1 to 5, an indicator of cumulative results, and produces two decision labels, namely passing right and passing late [8].

"Designing a Prediction Model for Student Graduation on Time at UIN Raden Fatah" was conducted to predict

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 2, April 2022, hlm. 159-166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i2.1539
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

student graduation using the Naïve Bayes method and the Rapid Miner application [9].

"Forward Selection Features to Determine Attributes That Affect the Classification of Graduates of the Faculty of Computer Science UNAKI Semarang Using the Naïve Bayes Algorithm" this study focuses on determining the attributes that affect student graduation using the forward selection feature with an accuracy rate of 97,14% [10]. This study aimed to analyze data mining with the Naïve Bayes method in diagnosing fast-track student graduation and to design a web-based application to test the fast-track student graduation diagnosis.

METHOD

This research was carried out during the stages of the data mining process, namely the Cross-Industry Standard Process (CRISP-DM). The stages of the research are shown in Image 1, and then each step is explained.

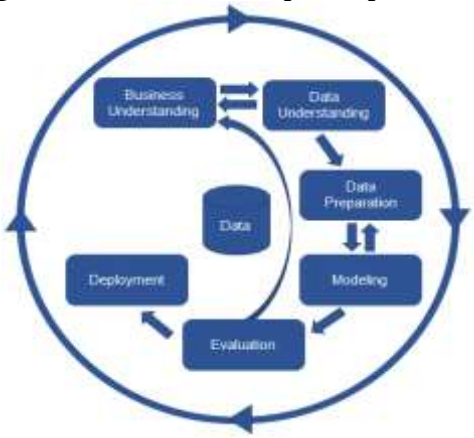

Image 1. Stages of CRISP-DM

Business Understanding

The Fast Track program is a fast-graduate program that can help students accelerate their study period of 3.5 years. In this study, an analysis of the diagnosis of fast-track student graduation at STMIK Royal will be carried out so that students have an idea about the time of the graduate study.

Data Understanding

The secondary data used in this study were obtained directly from interviews with the Head and Secretary of the Information Systems Study Program and the BAAK STMIK Royal, as many as 60 fast track student respondents. The data contains attributes including NIM, Name, Gender, Class, Total Credits, GPA, Tuition, Guidance Process, Collection of KKL Reports, and Decisions.

Data Preparation

There are several processes at the data preparation stage, including selecting data, building data, integrating data, and cleaning data.

Modeling

The classification modeling in this study uses the Naïve Bayes algorithm. The Naïve Bayes method is a simple probabilistic classifier that calculates a set of probabilities by summing the frequencies and combinations of values from a given data set. Naïve Bayes is based on the simplified assumption that attribute values are conditionally independent if given an output value [11].

The equation of Bayes' theorem is:

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

Description:
X: Data with unknown class
H: Hypothesis data is a specific class
P(H|X): Probability of Hypothesis H based on condition X

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 2, April 2022, hlm. 159-166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i2.1539
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

P(H): Hypothesis probability H
P(X|H): Probability of X based on condition on hypothesis H
P(X): Probability X

Evaluation

At this stage, the accuracy of the results and validation were measured using the Confusion Matrix model. The confusion matrix is a useful measurement model for analyzing how well the classifier recognizes tuples of different labels. A confusion matrix is a matrix table consisting of two classes, namely the class considered positive and the class considered negative [12].

Deployment

At this stage, preparation and presentation are carried out based on the knowledge gained through data mining. The presentation of the results of data mining calculations is visualized using a programming language so that it can be understood and read by ordinary people.

RESULT AND DISCUSSION

Based on the results of data collection and studies conducted, the information needed to build a fast-track student graduation diagnosis system is obtained. Furthermore, the data and information will be processed using the Naïve Bayes method. The training data used is the data of 60 students of the Information Systems study program who participated in the fast track program.

Table 1. Sample Data Training

Gender	Class	Total of SKS	GPA	Tuition Fee	Guidance Process	Collection of KKL Reports	Decisions
Female	Regular	Sufficient	Interval 3	Done	Always	On-Time	Fast Track
Female	Non Regular	Sufficient	Interval 2	Done	Always	On-Time	Fast Track
Male	Regular	Sufficient	Interval 1	Done	Often	On-Time	Fast Track
Male	Regular	Sufficient	Interval 2	Not Yet	Always	On -Time	On-Time
Female	Regular	Sufficient	Interval 1	Done	Often	On -Time	On-Time
Female	Regular	Sufficient	Interval 2	Not Yet	Often	On -Time	On-Time
Female	Regular	Sufficient	Interval 1	Not Yet	Seldom	On-Time	Slow Pass
Female	Regular	Sufficient	Interval 3	Done	Seldom	On-Time	Slow Pass
Male	Non-Regular	Sufficient	Interval 2	Done	Seldom	On-Time	Slow Pass

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 2, April 2022, hlm. 159-166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i2.1539
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 2. Sample Data Testing

Gender	Class	Total of SKS	GPA	Tuition Fee	Guidance Process	Collection of KKL Reports	Decisions
Female	Regular	Sufficient	Interval 3	Done	Always	On-Time	Fast Track
Female	Regular	Sufficient	Interval 1	Done	Often	On -Time	On-Time
Male	Non-Regular	Sufficient	Interval 2	Done	Seldom	On-Time	Slow Pass

The probability of fast track student graduation data is explained in table 3 – 10:

Table 3. Probability of Fast Track Student Graduation

Decision	Value
Fast Track	0,46667
On-Time	0,36667
Slow Pass	0,16667

Based on the results in the table. 3 shows that the highest probability value is a fast plus 0.46667, and the lowest probability is a slow pass around 0.16667. After getting the probability value of fast track student graduation from table 3, the next step is to calculate the probability of each criterion, namely Gender in table 4, Class in table 5, SKS in table 6, GPA in table 7, Tuition Fee in table 8, Guidance Process in table 9, and KKL Report in table 10.

Table 4. Probability of Gender

Gender	Decision		
	Fast Track	On-Time	Slow Pass
Female	0,86	0,68	0,80
Male	0,14	0,32	0,20

Table 5. Class Probability

Class	Decision		
	Fast Track	Fast Track	Fast Track
Regular	0,71	0,91	0,70
Non Regular	0,29	0,09	0,30

Table 6. Probability of SKS

SKS	Decision		
	Fast Track	Fast Track	Fast Track
Sufficient	0,996	0,995	0,99
Insufficient	0,004	0,005	0,01

Table 7. Probability of GPA

IPC	Decision		
	Fast Track	Fast Track	Fast Track
Interval 1	0,107	0,182	0,400
Interval 2	0,286	0,636	0,400
Interval 3	0,607	0,182	0,200

Table 8. Probability of Tuition Fee

Tuition Fee	Decision		
	Fast Track	Fast Track	Fast Track
Done	0,996	0,818	0,300
Belum	0,004	0,182	0,700

Table 9. Probability of the Guidance Process

Guidance Process	Decision		
	Fast Track	Fast Track	Fast Track
Always	0,677	0,270	0,010
Often	0,320	0,725	0,010
Seldom	0,004	0,005	0,980

Table 10. Probability of KKL Report

Laporan KKL	Decision		
	Fast Track	Fast Track	Fast Track
On-Time	0,996	0,995	0,990
Not On-Time	0,004	0,005	0,010

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 2, April 2022, hlm. 159-166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i2.1539
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Table 11. Data Testing

	Gender	Class	Total of SKS	GPA	Tuition Fee	Guidance Process	Collection of KKL Reports	Deci-sions
	Female	Regular	Sufficient	Interval 3	Done	Always	On-Time	?
Fast Track	0,86	0,71	0,996	0,607	0,996	0,677	0,996	0,9272
On Time	0,68	0,91	0,995	0,182	0,818	0,270	0,995	0,0723
Slow Pass	0,80	0,70	0,99	0,200	0,300	0,010	0,990	0,0004

After getting the probability value of each criterion, the next step is to determine the results of data testing or a data test model by taking the value of the opportunities from the training data.

Based on test data with the criteria of Gender "Female," Class "Regular," Total Credits "Sufficient," GPA "Interval 3", Tuition "Already," Guidance Process "Always," and "On Time" KKL Report Collection are predicted by the Naïve Bayes algorithm with a decision that is "Fast Track" with a probability of 92%.

Based on the test data, students who have such criteria are diagnosed as being able to pass quickly.

After the data is calculated, then the data will be tested. The testing phase is carried out to re-check the accuracy of the calculation.

Table 12. Experimental Results

Data Actual	Data Prediction		
	Fast Track	On-Time	Slow Pass
Fast Track	3	1	0
On Time	1	3	0
Slow Pass	0	0	4

In this experiment, a dataset that has been split randomly obtained 80% training data and 20% testing data which will be classified using the Naïve Bayes Algorithm, so the Confusion Matrix in the form of an accuracy rate of 83% is obtained.

To make it easier for students to make a graduation diagnosis, this research produces a web-based diagnostic application that can be accessed anytime and anywhere. Students must log in to access this application, as shown in Image 2.

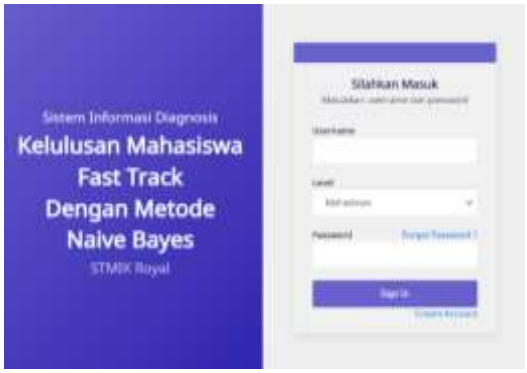

Image 2. Login Page

After successful login, students will be directed to the main page to make a diagnosis, as shown in Picture 3.

Image 3. Diagnostic Questions

Image 3 shows questions for students to make a graduation diagnosis. After the diagnosis process is complete, students will know the diagnosis results using the Naïve Bayes Algorithm shown in Image 4.

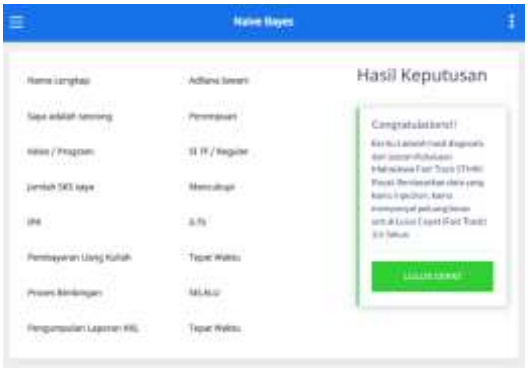

Image 4. Diagnosis of Graduation with Naïve Bayes Algorithm

The diagnosis process using the Naïve Bayes Algorithm can be seen in Image 4, which is the core diagnosis process.

CONCLUSION

This study tested the model using the Naïve Bayes algorithm and fast-track student graduation data. Based on the discussion using the Naïve Bayes Algorithm and the Confusion Matrix Model, an accuracy value of 83% is obtained, which is included in the Good Classification category.

Data Mining with the Naïve Bayes method is a good method and can be applied to diagnose fast-track student graduation based on previous experience.

BIBLIOGRAPHY

[1] M. R. Qisthiano, T. B. Kurniawan, E. S. Negara, and M. Akbar, "Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 987, 2021, doi: 10.30865/mib.v5i3.3030.

[2] L. Abd, R. Hakim, A. A. Rizal, and D. Ratnasari, "Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Student Graduation Prediction Using K-Nearest Neighbor (K-Nn)," *J. Teknol. Inf. dan Multimed.*, vol. 1, no. 1, pp. 30–36, 2019.

[3] E. Erwin, I. Permana, and M. S. Hayat, "Strategi Evaluasi Program Praktikum Fisika Dasar," *J. Penelit. Pembelajaran Fis.*, vol. 9, no. 1, p. 12, 2018, doi: 10.26877/jp2f.v9i1.2308.

[4] P. S. C. Moonallika, K. Q. Fredlina, and I. B. K. Sudiatmika, "Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara)," *J. Ilm. Komput.*, vol. 6, no. 1, pp. 47–56, 2020.

[5] Susriyanti, S. Defit, and N. Istianingsih, "Menentukan Pola Pembelian Produk Dengan Rule Mining Algoritma Apriori Pada Ud. Pelita Kita Padang," vol. 4, no. 3, pp. 100–110, 2021.

[6] N. Yustira, D. Witarsyah, and E. Sutoyo, "Implementasi Algoritma Naïve Bayes Classification Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu (Studi Kasus : Program Studi Sistem Informasi Universitas Telkom)."

[7] M. winny Amelia, A. S. . Lumenta, and A. Jacobus,

"Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naïve Bayes," *J. Tek. Inform.*, vol. 11, no. 1, 2017, doi: 10.35793/jti.11.1.2017.17652.

[8] Y. E. Fadrial, "ALGORITMA NAIVE BAYES UNTUK MENCARI PERKIRAAN WAKTU STUDI MAHASISWA NAIVE," *J. Inf. Technol. Comput. Sci.*, vol. 4, pp. 5–24, 2021.

[9] G. Testiana, "Perancangan model prediksi kelulusan mahasiswa tepat waktu pada UIN Raden Fatah," *JUSIFO (Jurnal Sist. Informasi)*, vol. 4, no. 1, pp. 49–62, 2018.

[10] M. F. Nugroho and S. Wibowo, "Fitur Seleksi Forward Selection Untuk Menetukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes," *J. Inform. Upgris*, vol. 3, no. 1, pp. 63–70, 2017, doi: 10.26877/jiu.v3i1.1669.

[11] E. Manalu, F. A. Sianturi, and M. R. Manalu, "Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries," *J. Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017.

[12] L. A. Utami, "Analisis Sentimen Opini Publik Berita Kebakaran Hutan Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization," *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 103–112, 2017.