

Prediksi Kelulusan Mahasiswa menggunakan Algoritma *Naive Bayes* (Studi Kasus 5 PTS di Banda Aceh)

Munawir¹, Taufiq Iqbal²

AMIK Indonesia

article info

Article history:

Received 19 August 2018

Received in revised form

19 Juni 2019

Accepted 30 September 2019

Available online September 2019

DOI:

<https://doi.org/10.35870/jtik.v3i2.77>

Keywords:

Prediction, Student

Graduation, Naive Bayes

Algorithm.

Kata Kunci:

Prediksi, Kelulusan Mahasiswa,

Algoritma Naive Bayes.

abstract

The e-questionnaire application that researchers built using CodeIgniter and React-Js This study aims to data mining by using rapidminer tools to collect student data from the Feeder application page from the class of 2010-2014 which is assumed that the student class has been declared graduated in 2018. The data was collected from 5 (five) Private Universities in the City Banda Aceh. then by observing the graduation level using data mining can bring a considerable contribution to educational institutions, in an effort to improve curriculum competency in Higher Education, it is expected that the results of data mining can make reference to curriculum standards as a form of graduate competency improvement. The research method uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) which is used as a standard data mining process as well as a research method with stages starting from Business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The results showed that the data mining algorithm for graduation prediction based on the selected pass accuracy attribute revealed that the prediction level was uniform with the algorithm used, Naive Bayes, prediction accuracy was 84%. The data attributes that were found to have significantly influenced the classification process were the GPA and Study Length. The results obtained that students who graduated by 60% are students who are educated in ASM Nusantara and AMIK Indonesia, while in Banda Aceh STIES and Serambi University Mecca the prediction of graduation is 52%. Another thing is different from STIA Iskandar Thani where the prediction of graduating is only 48% and not passing on time is 52%. The results of this prediction can reveal and become a recommendation for prospective students or academics to increase the quantity of graduates and increase student confidence in tertiary institutions.

abstrak

Penelitian ini bertujuan untuk data mining dengan menggunakan tools rapidminer untuk mengumpulkan data mahasiswa dari laman aplikasi Feeder dari angkatan 2010-2014 yang diasumsikan bahwa mahasiswa angkatan tersebut sudah dinyatakan lulus pada tahun 2018. Data tersebut di dikumpulkan dari 5 (lima) Perguruan Tinggi Swasta di Kota Banda Aceh. maka dengan mengamati tingkat kelulusan menggunakan data mining dapat membawa kontribusi yang cukup besar bagi lembaga pendidikan, dalam upaya peningkatan kompetensi kurikulum pada Perguruan Tinggi yang diharapkan hasil data mining dapat menjadikan referensi standar kurikulum sebagai bentuk peningkatan kompetensi lulusan. Metode penelitian menggunakan *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang digunakan sebagai standar proses data mining sekaligus sebagai metode penelitian dengan tahap dimulai dari Business understanding, data understanding, data preparation, modelling, evaluation, dan deployment. Hasil penelitian didapatkan bahwa algoritma data mining untuk prediksi kelulusan berdasarkan atribut ketepatan lulus yang dipilih mengungkapkan bahwa tingkat prediksi seragam dengan algoritma yang digunakan yaitu Naive Bayes, akurasi prediksi sebesar 84%. Atribut data yang ditemukan memiliki signifikan dipengaruhi proses klasifikasi adalah IPK dan Lama Studi. Hasil yang diperoleh bahwa mahasiswa yang lulus sebesar 60% yaitu mahasiswa yang berpendidikan di ASM Nusantara dan AMIK Indonesia, sedangkan pada STIES Banda Aceh dan Universitas Serambi Mekkah sebesar prediksi lulus sebesar 52%. Hal lain berbeda dengan STIA Iskandar Thani dimana prediksi lulus hanya sebesar 48% dan tidak lulus tepat waktu 52%. Hasil prediksi ini bisa mengungkapkan dan menjadi sebuah rekomendasi bagi calon mahasiswa atau pihak akademik agar meningkatkan kuantitas lulusan dan meningkatkan kepercayaan mahasiswa terhadap Perguruan Tinggi.

*Corresponding author. Email: munawir@amikindonesia.ac.id

© E-ISSN: 2580-1643.

Copyright © 2019. Published by Lembaga Informasi dan Riset (KITA INFO dan Riset), Lembaga KITA (<http://creativecommons.org/licenses/by/4.0/>).

1. Latar Belakang

Pengukuran tingkat kelulusan mahasiswa di suatu Perguruan Tinggi dilakukan setiap tahunnya dengan berbagai cara salah satunya dengan melakukan survei alumni yang menghasilkan suatu data yang menggambarkan kompetensi lulusan tersebut dan keefektifan kurikulum yang diterapkan sehingga dapat digunakan sebagai acuan dalam pengambilan keputusan setiap tahunnya terhadap kebutuhan masyarakat dunia khususnya tingkat daerah. Berbagai model dan analisis dalam memprediksi kompetensi lulusan terhadap kebutuhan masyarakat di dunia pada setiap Perguruan Tinggi telah banyak digunakan dengan berbagai teknik klasifikasi data mining seperti menggunakan algoritma *artificial neural network* (ANN), *naïve bayes* dan *decision tree*. Banyaknya keberhasilan yang ditunjukkan dengan menggunakan teknik dengan berbagai penggunaan dalam memprediksi menggunakan *tools* seperti *weka*, *rapidminer*, dan *orange*. Tetapi, optimalisasi penggunaan *tools* dengan model belum menunjukkan hasil yang tepat dan akurat.

Penelitian sebelumnya telah melakukan prediksi lulusan dengan menggunakan *naïve bayes*. Fithri dan Darmanto (2014) menggunakan beberapa parameter, yaitu jenis kelamin, alamat, umur, status pekerjaan mahasiswa, status pernikahan mahasiswa, rata-rata IPK, jumlah SKS dan status mahasiswa dan menerapkan algoritma *naïve bayes* sehingga dapat menentukan mahasiswa lulus tepat waktu atau terlambat. Murtopo (2016) melakukan prediksi lulusan dengan menggunakan berbagai faktor internal dan eksternal, hasil penelitian yang dilakukan untuk pengukuran akurasi, pengujian dengan memanfaatkan ROC Curva, dan *k-fold cross validation*, pengujian dilakukan sebanyak 10fold, sehingga prediksi kelulusan mahasiswa menunjukkan peningkatan dibanding penelitian sebelumnya.

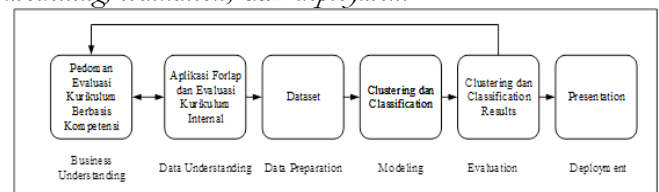
Evaluasi sukses atau tidaknya tergantung pada kinerja akademik terhadap mahasiswa, sehingga dapat dilihat dari ketepatannya dalam menyelesaikan masa studinya di universitas (Nasution, Djahara, dan Zamsuri, 2015), dan diperlukan kajian untuk mendapatkan kondisi yang sebenarnya (Salmu dan Solichin, 2017). Lulusan yang berkualitas tentu saja tidak hanya bergantung dari peran perguruan tinggi semata (Samponu dan Kusri, 2017), akan tetapi

ketepatan lulusan dengan lama studi harus juga dinilai, salah satunya yang bisa melakukan prediksi menggunakan algoritma *Naïve Bayes* (Anuradha dan Velmurugan, 2015).

Penelitian ini bertujuan untuk data mining dengan menggunakan *tools* *rapidminer* untuk mengumpulkan data mahasiswa dari laman aplikasi Feeder dari angkatan 2010-2014 yang diasumsikan bahwa mahasiswa angkatan tersebut sudah dinyatakan lulus pada tahun 2018. Data tersebut di dikumpulkan dari 5 (lima) Perguruan Tinggi Swasta di Kota Banda Aceh. maka dengan mengamati tingkat kelulusan menggunakan data mining dapat membawa kontribusi yang cukup besar bagi lembaga pendidikan, dalam upaya peningkatan kompetensi kurikulum pada Perguruan Tinggi yang diharapkan hasil data mining dapat menjadikan referensi standar kurikulum sebagai bentuk peningkatan kompetensi lulusan.

2. Metode Penelitian

Metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM) digunakan sebagai standar proses data mining sekaligus sebagai metode penelitian dengan tahap dimulai dari *Business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*.



Gambar 1. Desain Penelitian

Objek penelitian adalah lulusan Angkatan tahun 2010 sampai dengan 2014 yang tersebar dari 5 (lima) Perguruan Tinggi Swasta yaitu; 1) AMIK Indonesia Program Studi Manajemen Informatika, 2) ASM Nusantara Program Studi Sekretaris, 3) STIE Sabang Banda Aceh Program Studi Manajemen, 4) STIA Iskandar Thani Program Studi Ilmu Administrasi Negara, dan 5) Universitas Serambi Mekkah Program Studi Manajemen. Sedangkan alat analisis yang digunakan adalah *Rapidmer*.

Untuk teknik modeling metode yang digunakan adalah *Naïve Bayes* yang merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema atau aturan *bayes* dengan asumsi

independensi yang kuat pada fitur, artinya bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Nugroho, 2014). Klasifikasi Bayesian adalah klasifikasi statistik yang bisa memprediksi probabilitas sebuah *class*. Adapun klasifikasi Teorema Bayes sebagai berikut :

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} . PH \dots \dots \dots [NBC - 01]$$

Keterangan

- X : Data dengan class yang belum diketahui
- H : Hipotesis data merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- P(H) : Probabilitas hipotesis H (prior probabilitas)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Teorema Bayes sering pula dikembangkan mengingat berlakunya hukum probabilitas total, menjadi seperti berikut:

$$P(H|X) = \frac{P(H|X)P(H)}{\sum_{i=0}^n P(H_i|X)} . PH \dots \dots NBC - 02]$$

Keterangan

- i : 1,2,3, ... , n jumlah data Hipotesis (prior probabilitas)
- S : Probabilitas total H

Untuk menjelaskan Teorema Naïve Bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, Teorema Bayes di atas disesuaikan sebagai berikut:

$$P(C|F_1, \dots F_n) = \frac{P(F_1, \dots F_n|C)}{P(F_1, \dots F_n)} . P . C \dots NBC - 03]$$

Di mana Variabel C merepresentasikan kelas, sementara variabel F1 ... Fn merepresentasikan karakteristik petunjuk yang dibutuhkan untuk

melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga evidence).

3. Hasil Analisis dan Pembahasan

Dataset

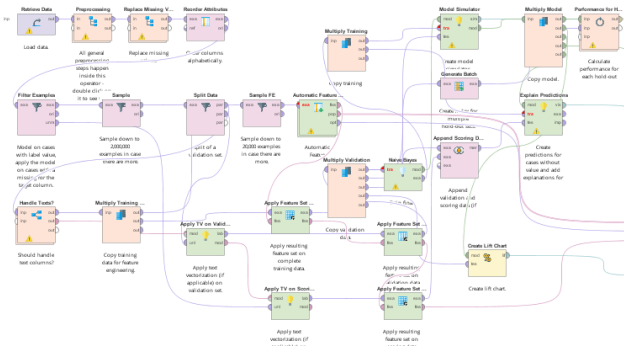
Dataset mahasiswa diambil berdasarkan data kelulusan mahasiswa pada aplikasi Forlap di 5 (lima) Perguruan Tinggi Swasta (PTS) di Kota Banda Aceh. *Dataset* digunakan untuk mengevaluasi dan memprediksi kelulusan tepat waktu di masing-masing PTS. Selain itu, atribut disusun dan mengacu pada standar kelulusan Perguruan Tinggi yang terdiri dari IPK, lama studi, dan ketepatan lulus dari Angkatan 2010–2014 PTS. Data yang diperoleh sebanyak 500 dengan target “Tepat Waktu” dengan *view* data di Rapidminer.

Tepat Waktu (class)	Angkatan (Year)	IPK (GPA)	Lama Studi (Study Duration)	PTS (College)
NO	2011	3,0	4,500	AMIK Indonesia
YES	2011	3,6	3	AMIK Indonesia
YES	2011	3,7	3	AMIK Indonesia
YES	2011	3,6	3	AMIK Indonesia
YES	2011	3,7	3	AMIK Indonesia
NO	2011	3,3	4	AMIK Indonesia
YES	2011	3,9	3	AMIK Indonesia
NO	2011	3,0	4	AMIK Indonesia

Gambar 1. Dataset

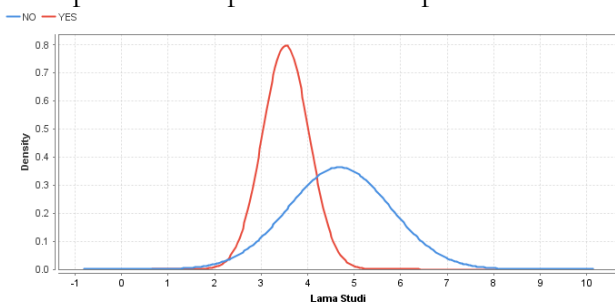
Hasil Eksperimental dan Pengamatan

Tujuan utama dari penelitian ini adalah untuk mengeksplorasi jika memungkinkan untuk memprediksi kelulusan mahasiswa (*output*) berdasarkan berbagai variabel penjelas (*input*) yang dipertahankan dalam model yang digunakan. Klasifikasi model dibangun menggunakan beberapa algoritma Naïve Bayes dan aplikasi Rapidminer digunakan pada tahap ini. Setiap *classifier* diterapkan untuk dua opsi pengujian *cross validation* seperti terlihat pada gambar 2 berikut.



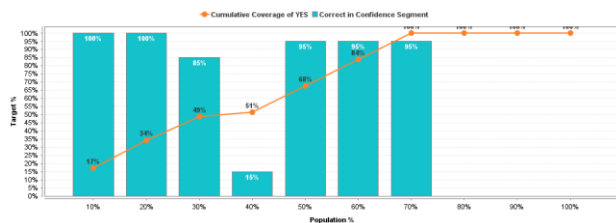
Gambar 2. Model yang digunakan

Model yang diterapkan menghasilkan sebuah informasi dari model yang digunakan (gambar 3), simulator percobaan setiap atribut, *performance*, *lift chart*, *features sets*, *prediction*, *dataset*, *weights*, dan *correlation*. Model yang diterapkan merupakan hasil selanjutnya di-representasi dalam bentuk grafis yang difokuskan pada atribut lama studi untuk mendapatkan hasil prediksi lulus tepat waktu.



Gambar 3. Model pada Atribut Lama Studi

Pada gambar 4 ditunjukkan efektivitas model dengan menghitung rasio antara hasil yang diperoleh dengan model dan hasil yang diperoleh tanpa model. Hanya dibuat untuk masalah dua *class*.



Gambar 4. Lift Chart Naïve Bayes

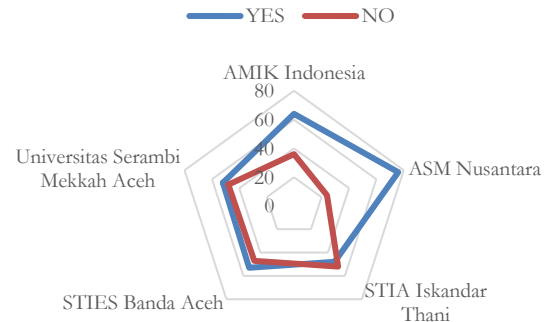
Pada gambar 5, akurasi penggunaan Naïve Bayes pada penelitian ini menunjukkan persentase 84%.
accuracy: 88.84% +/- 3.70% (micro average: 88.81%)

	true NO	true YES	class precision
pred. NO	43	0	100.00%
pred. YES	16	84	84.00%
class recall	72.88%	100.00%	

Gambar 5. Akurasi Naïve Bayes

Pembahasan

Berdasarkan analisis dan pengujian *dataset* dari 5 (lima) Perguruan Tinggi Swasta (PTS) maka dapat diketahui bahwa, AMIK Indonesia memiliki persentase kelulusan mahasiswa sebesar 64%, ASM Nusantara 76%, STIA Iskandar Thani 48%, STIES Banda Aceh 53%, Universitas Serambi Mekkah sebesar 52%, hasil persentase tingkat kelulusan mahasiswa dapat dilihat pada gambar 6.



Gambar 6. Perbandingan Kelulusan Mahasiswa 5 (lima) PTS

Jika dilihat lebih lanjut, maka ASM Nusantara memiliki tingkat kelulusan lebih tinggi daripada 4 (empat) PTS lainnya. Hal tersebut terlihat jelas bahwa IPK lulus mahasiswa ASM Nusantara dengan rata-rata 3.2 dan masa studi 3 tahun. Selanjutnya tingkat kelulusan pada STIES Banda Aceh dan Universitas Serambi Mekkah pada persentase 53% dengan lama studi 4 tahun dan IPK rata-rata 3. Sedangkan pada tingkat lulusan STIA Iskandar Thani lebih rendah bahwa lama studi 4 tahun dan IPK 3.2 tingkat kelulusan sebesar 48%. Sedangkan pada AMIK Indonesia dengan lama studi 3 tahun dan IPK 3.4 tingkat kelulusan sebesar 64%.

4. Kesimpulan

Hasil algoritma data mining untuk prediksi kelulusan berdasarkan atribut ketepatan lulus yang dipilih mengungkapkan bahwa tingkat prediksi seragam dengan algoritma yang digunakan yaitu Naïve Bayes, akurasi prediksi sebesar 84%. Atribut data yang ditemukan memiliki signifikan dipengaruhi proses klasifikasi adalah IPK dan Lama Studi. Hasil yang diperoleh bahwa mahasiswa yang lulus sebesar 60% yaitu mahasiswa yang berpendidikan di ASM Nusantara dan AMIK Indonesia, sedangkan pada STIES Banda Aceh dan Universitas Serambi Mekkah

sebesar prediksi lulus sebesar 52%. Hal lain berbeda dengan STIA Iskandar Thani dimana prediksi lulus hanya sebesar 48% dan tidak lulus tepat waktu 52%. Hasil prediksi ini bisa mengungkapkan dan menjadi sebuah rekomendasi bagi calon mahasiswa atau pihak akademik agar meningkatkan kuantitas lulusan dan meningkatkan kepercayaan mahasiswa terhadap Perguruan Tinggi. Penelitian lebih lanjut perlu dikaji kembali mengenai jumlah SKS, lingkungan, mata kuliah, dosen, dan lain sebagainya agar dapat menjadi suatu kesimpulan yang dapat digunakan bagi Institusi lainnya.

5. Ucapan Terima Kasih

Ucapan terima kasih sebesar-besarnya kepada Ditjen Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi sebagai penyandang dana penelitian pada skema Penelitian Dosen Pemula (PDP) Tahun 2019, dan tidak lupa pula kepada LPPM AMIK Indonesia dan para Dosen, rekan-rekan sejawat, para mahasiswa/i, Alumni serta tim peneliti yang telah membantu menyelesaikan penelitian ini sesuai dengan harapan.

6. Daftar Pustaka

- Anuradha, C. and Velmurugan, T., 2015. A comparative analysis on the evaluation of classification algorithms in the prediction of students performance. *Indian Journal of Science and Technology*, 8(15), pp.1-12.
- Fithri, D.L. and Darmanto, E., 2014. Sistem Pendukung Keputusan Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Nave Bayes. In *Seminar Nasional Teknologi dan Informatika 2014*. Muria Kudus University.
- Murtopo, A.A., 2016. Prediksi Kelulusan Tepat Waktu Mahasiswa STMIK YMI Tegal Menggunakan Algoritma Naïve Bayes. *CSRID (Computer Science Research and Its Development Journal)*, 7(3), pp.145-154.
- Nasution, N., Djahara, K. and Zamsuri, A., 2015. Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes (Studi Kasus: Fasilkom Unilak). *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, 6(2), pp.1-11.
- Nugroho, Y.S., 2014. Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. Dian Nuswantoro Fakultas Ilmu Komputer Skripsi.
- Salmu, S. and Solichin, A., 2017, April. Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta. In *Prosiding Seminar Nasional Multidisiplin Ilmu Universitas Budi Luhur*, ISSN (pp. 2087-0930).
- Samponu, Y.B. and Kusriani, K., 2017. Optimasi Algoritma Naive Bayes Menggunakan Metode Cross Validation Untuk Meningkatkan Akurasi Prediksi Tingkat Kelulusan Tepat Waktu. *Jurnal ELTIKOM*, 1(2), pp.56-63.