

DATA MINING MODEL KLASIFIKASI UNTUK KETEPATAN WAKTU KELULUSAN MAHASISWA

Chetsi Rafelin Patrisia Mk¹, Sylvia J. A. Sumarauw², Vivian E. Regar³
e-mail: chetsiraf@gmail.com

¹²³Universitas Negeri Manado, Sulawesi Utara, Indonesia

Abstrak

Tingkat ketepatan waktu kelulusan mahasiswa merupakan suatu tolok ukur yang memberikan gambaran tentang sejauh mana mahasiswa berhasil menyelesaikan studi yang mereka ikuti di suatu perguruan tinggi. Mahasiswa program studi pendidikan matematika angkatan 2017 yang menyelesaikan studinya tepat waktu hanya sebesar 3.68% dan Mahasiswa angkatan 2019 yang menyelesaikan studi tepat waktu hanya sebesar 2.08%. rendahnya tingkat ketepatan waktu kelulusan akibatnya dapat mempengaruhi kualitas mahasiswa, predikat kelulusan dan juga akreditasi jurusan. Penelitian ini menggunakan metode klasifikasi algoritma naïve bayes untuk mengklasifikasi waktu kelulusan mahasiswa. Metode penelitian yang digunakan adalah jenis penelitian kuantitatif yang mengikuti prosedur data mining berdasarkan prosedur CRISP-DM (Cross Industry Standard Process Model for Data Mining). Pengumpulan data menggunakan kuesioner. Teknik pengolahan data yang diterapkan untuk mengklasifikasikan waktu kelulusan mahasiswa menggunakan metode klasifikasi data mining dengan algoritma Naïve Bayes, yang dibantu oleh *software* RapidMiner Studio Versi 10.1. Penelitian ini menunjukkan nilai akurasi sebesar 75%, yang menunjukkan jika data mining dengan model klasifikasi menggunakan algoritma Naïve Bayes efektif untuk mengklasifikasikan ketepatan waktu kelulusan mahasiswa.

Kata Kunci: Data Mining, Algoritma Naïve Bayes, Kelulusan Mahasiswa

Abstract

The level of timeliness of student graduation is a benchmark that provides an idea of the extent to which students have successfully completed the studies they are taking at a university. Only 3.68% of the 2017 class of mathematics education study program students completed their studies on time and only 2.08% of the 2019 class of students completed their studies on time. The low level of timeliness of graduation can consequently affect the quality of students, the degree of graduation and also the accreditation of departments. This research uses the Naïve Bayes algorithm classification method to classify students' graduation times. The research method used is a type of quantitative research that follows data mining procedures based on the CRISP-DM (Cross Industry Standard Process Model for Data Mining) procedure. Data collection uses a questionnaire. The data processing technique applied to classify students' graduation times uses the data mining classification method with the Naïve Bayes algorithm, which is assisted by RapidMiner Studio Version 10.1 software. This research shows an accuracy value of 75%, which shows that data mining with a classification model using the Naïve Bayes algorithm is effective for classifying students' graduation timeliness.

Keywords: Data Mining, Naïve Bayes Algorithm, Student Graduation

Pendahuluan

Tingkat ketepatan waktu kelulusan merupakan suatu tolok ukur yang memberikan gambaran tentang sejauh mana mahasiswa berhasil menyelesaikan program studi yang mereka ikuti. Lebih dari sekadar angka statistik, tingkat kelulusan mencerminkan kualitas pendidikan yang diberikan oleh perguruan tinggi, efektivitas metode pengajaran, dan kesiapan lembaga dalam membantu mahasiswa mencapai tujuan akademik mereka (Apridiansyah, Veronika, & Putra, 2021).

Berdasarkan data yang diperoleh bahwa pada Jurusan Matematika Universitas Negeri Manado persentase mahasiswa angkatan 2017 yang lulus tepat waktu ialah 3,68% dan persentase mahasiswa angkatan 2019 yang lulus tepat waktu ialah 2,08%. Salah satu masalah yang sering dihadapi oleh perguruan tinggi, khususnya Universitas Negeri Manado, adalah ketidakseimbangan antara jumlah mahasiswa yang diterima dan jumlah mahasiswa yang lulus. Meskipun banyak mahasiswa baru diterima, jumlah mahasiswa yang lulus tepat waktu jauh lebih sedikit. Hal ini dapat mempengaruhi kualitas mahasiswa, predikat kelulusan, dan akreditasi jurusan. (Rohmawan, 2018). Oleh karena itu, diperlukan sebuah sistem yang diharapkan dapat mengklasifikasi kelulusan mahasiswa. Sistem ini digunakan untuk mengklasifikasikan apakah mahasiswa akan lulus tepat waktu atau tidak.

Pada dasarnya, data mining sangat berkaitan dengan analisis data untuk mengidentifikasi pola dan kesamaan dalam sekumpulan data dengan penggunaan perangkat lunak (Maskoen & Purnama, 2018). Menurut Yunus dalam Heryana (2019) menyatakan *data mining* memiliki kapasitas untuk memproses data dalam jumlah besar, dan dengan pertumbuhan yang terus meningkat dari tahun ke tahun dalam pengumpulan data, *data mining* dapat mengolah data ini dan mengubahnya menjadi informasi yang berharga. Klasifikasi adalah salah satu peran *data mining* dari lima peran *data mining*. Bansal, Sharma & Goel (2017) mendefinisikan Klasifikasi adalah teknik yang digunakan untuk mengelompokkan data berdasarkan informasi yang sudah ada. Algoritma naïve bayes merupakan salah satu jenis algoritma klasifikasi *data mining*. Bustami (2014) menyatakan algoritma Naïve Bayes adalah metode klasifikasi yang didasarkan pada teorema Bayes. Untuk membantu sistem klasifikasi ketepatan waktu kelulusan mahasiswa dalam menemukan suatu informasi berharga itu diperlukan teknik *data mining* menggunakan peran klasifikasi. Algoritma Naïve Bayes dapat digunakan dalam berbagai bidang, termasuk pendidikan, keuangan, pertanian, ekonomi, asuransi, dan kesehatan. Penelitian pada bidang pendidikan oleh Hartati dan San (2022) menggunakan algoritma naïve bayes untuk prediksi kelulusan mahasiswa dan menunjukkan nilai akurasi sebesar 80%. Penelitian pada bidang ekonomi oleh Wididastuti dkk (2019) tentang pemohon kredit sepeda motor menggunakan algoritma *Naïve Bayes*, hasil menunjukkan angka 93,94% untuk tingkat akurasinya.

Metode

Metode penelitian ini melibatkan penggunaan algoritma Naïve Bayes dalam data mining metode klasifikasi. Menurut Larose, data mining didefinisikan sebagai suatu proses penemuan relasi, pola, dan tren baru yang signifikan dengan menyelidiki dataset besar yang tersimpan di berbagai tempat penyimpanan, dengan menggunakan teknik pengenalan pola seperti statistik dan matematika. (Kamagi & Hansun, 2014). *Data mining* sebagai suatu teknik yang sering digunakan dalam analisis data karena dapat menginspirasi ide-ide baru yang berbeda dari sebelumnya dan bertugas untuk mengekstrak pola-pola menarik dari berbagai dataset yang besar, yang dapat disimpan dalam basis data, data warehouse, atau lokasi penyimpanan lainnya (Sumarauw, 2022).

Putri & Rahmawati (2014) mengatakan bahwa dalam metode klasifikasi, terdapat dua tahapan yang dilaksanakan. Pertama, model dibangun dan disimpan sebagai referensi. Kemudian, model tersebut digunakan untuk mengenali, mengklasifikasikan, atau memprediksi data lain dengan tujuan menentukan kelas atau kategori mana di mana objek data tersebut dapat ditempatkan berdasarkan model yang telah tersimpan dalam memori. Menurut Budiman & Ramadani (2015) klasifikasi merupakan suatu proses penemuan model yang mampu menjelaskan dan membedakan kelas data atau konsep yang bertujuan untuk memungkinkan prediksi kelas data atau konsep dari objek yang label kelasnya tidak diketahui.

Menurut Manalu dkk (2017) *naive bayes* adalah metode klasifikasi probabilitas yang sederhana, Probabilitas dihitung dengan menggabungkan kombinasi nilai dari dataset yang tersedia dan frekuensi, dan algoritma ini didasarkan pada asumsi sederhana bahwa nilai atribut secara kondisional saling bebas jika nilai output diberikan. Syukri dkk (2017) mendefinisikan *naive bayes* sebagai metode klasifikasi berdasarkan pada Teorema Bayes, Metode ini mengadopsi probabilitas dan statistik berdasarkan konsep Thomas Bayes, seorang ilmuwan asal Inggris, untuk meramalkan peluang di masa mendatang berdasarkan pengalaman masa lalu. Pendekatan ini dikenal karena asumsi independensi yang sangat kuat terhadap setiap kondisi atau kejadian.

Penelitian ini menggunakan proses *data mining* yang telah dibuat oleh Cross Industry Standard Process Model for *Data Mining* (CRISP-DM) dengan tahapan-tahapan; *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. CRISP-DM menyajikan pedoman prosedur yang standar untuk *data mining* yang dapat digunakan dalam memecahkan masalah- masalah umum di dunia bisnis atau di lingkungan penelitian. CRISP-DM mengungguli metodologi *data mining* lainnya dalam hal kelengkapan dan didokumentasikan dengan baik. Setiap tahapannya telah terstruktur dengan baik dan didefinisikan secara jelas, sehingga dapat diterapkan dengan mudah bahkan oleh pemula (Fadillah, 2015). Subjek penelitian adalah alumni mahasiswa dan mahasiswa aktif Jurusan Matematika, Fakultas Matematika Ilmu Pengetahuan Alam Dan Kebumihan, Universitas Negeri Manado dan teknik pengambilan data menggunakan kuesioner google form, dan pengolahan data berbantuan *software RapidMiner Studio Version 10.1*.

Hasil dan Pembahasan

Hasil Penelitian

1. Business Understanding

Keterlambatan kelulusan mahasiswa sudah menjadi masalah sejak lama bagi jurusan pendidikan matematika Universitas Negeri Manado dibuktikan dengan persentase mahasiswa angkatan 2017 yang lulus tepat waktu hanya sebesar 3,68% dan mahasiswa angkatan 2019 yang lulus tepat waktu hanya sebesar 2,08%. Rendahnya angka kelulusan tepat waktu tentu dipengaruhi oleh banyak faktor baik faktor yang berasal dari luar mahasiswa maupun faktor yang berasal dari dalam diri mahasiswa sendiri. Di jurusan pendidikan matematika unima mahasiswa dikatakan lulus jika telah menyelesaikan minimal sebanyak 144 SKS (Satuan Kredit Semester) dan masa studinya selama 8 semester sampai 15 semester jika lebih dari 15 semester mahasiswa akan dikenakan drop out, masa studi 8 semester dikatakan lulus tepat waktu dan jika lebih dari 8 semester dikatakan lulus tidak tepat waktu. Banyaknya mahasiswa yang tidak dapat menyelesaikan masa studinya selama 8 semester memotivasi agar mengetahui lebih awal waktu kelulusan agar dapat melakukan langkah antisipasi jika akan diklasifikasi lulus tidak tepat waktu.

2. Data Understanding

Tahapan awal dilakukan dengan pengumpulan data mahasiswa yang telah lulus menggunakan kuesioner penelitian. Dalam pengujian, data dipisahkan menjadi dua bagian: Data latih dan data uji, yang nantinya akan diolah menggunakan algoritma Naive Bayes. Data latih digunakan untuk membangun tabel probabilitas, sementara data uji berfungsi untuk mengevaluasi probabilitas yang telah dibuat. Data latih akan diambil sebanyak 90% dari data yang diperoleh dan data uji sebanyak 10% dari data yang diperoleh.

NIM	JK	AS	JS	JM	Asal	Organisasi	PO	Pekerjaan	Pengulangan MK	Beasiswa	MS	Umur	IPK
19504084	L	SMA	MIA	SB	Bolaang Mongondow	HMJ	M	Tidak	Ya	Ya	9	17	3.35
19504015	P	SMA	IPA	SN	Kota Manado	GMU	M	Ya	Ya	Ya	9	17	3.41
19504010	P	SMA	IPA	SN	Sulawesi Utara	FMIPAK, LP	M	Tidak	Ya	Tidak	10	17	3.4
19504135	L	SMA	MIPA	SB	Minahasa Selatan	EM, SANGGA	M	Tidak	Ya	Tidak	8	17	3.76
19504046	P	SMA	AN 2 Mar	SB	Pinangunian	BEM dan HM	M	Tidak	Ya	Ya	9	16	3.37
19504001	P	SMA	MIPA	SN	Minahasa	GenBI	T	Ya	Tidak	Ya	9	17	3.71
19504031	L	SMA	Ipa	SB	Manado	HMJ	M	Tidak	Ya	Tidak	9	17	3.26

Gambar 1 Sampel Data Alumni Mahasiswa

Keterangan:

NIM: Nomor Induk Mahasiswa, JK: Jenis Kelamin, AS: Asal Sekolah, JS: Jurusan Sekolah,

JM: Jalur Masuk, PO: Penghasilan Orang Tua, MS: Masa Studi, IPK: Indeks Prestasi Kumulatif

Untuk memperjelas data pada Gambar 4.1 yang di atas, berikut ini diuraikan penjelasannya sebagai berikut:

1. NIM, merupakan serangkaian nomor identitas dari setiap mahasiswa yang pernah menempuh bangku perkuliahan dan hanya berlaku selama masa studi perkuliahan tengah berlangsung.
2. JK, merupakan Jenis Kelamin dari setiap mahasiswa dengan laki-laki disimbolkan dengan L dan perempuan disimbolkan dengan P.
3. AS, merupakan Asal Sekolah yang ditempuh mahasiswa sebelum memasuki perguruan tinggi dan terdapat tiga kategori yakni Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK) dan Madrasah Aliyah (MA).
4. JS, merupakan asal jurusan mahasiswa pada bangku sekolah, untuk sekolah menengah atas memiliki tiga jurusan yakni IPA, Bahasa dan IPS sedangkan Sekolah Menengah Kejuruan memiliki banyak jurusan contohnya yaitu: akuntansi, keperawatan, teknik komputer dan jaringan dan sebagainya.
5. JM, merupakan atribut yang menjelaskan mahasiswa masuk perguruan tinggi melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN), Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN), atau jalur mandiri Baku Beking Pande (B2P).
6. Asal Daerah, merupakan atribut yang menjelaskan dari daerah mana mahasiswa tersebut berasal.
7. Organisasi, merupakan atribut yang menjelaskan keikutsertaan mahasiswa dalam organisasi baik organisasi dalam kampus maupun organisasi di luar kampus.
8. PO, merupakan atribut yang berisi penghasilan orang tua mahasiswa atau keluarga mahasiswa yang dapat dilihat pada Tabel 1.

Tabel 1 Penghasilan Orangtua

No	Penghasilan	Keterangan
1	Penghasilan <Rp. 3.000.000	Rendah
2	Penghasilan Rp. 3.000.000 – Rp. 10.000.000	Menengah
3	Penghasilan > Rp. 10.000.000	Tinggi

9. Pekerjaan, merupakan atribut yang menjelaskan apakah selama berkuliah mahasiswa yang bersangkutan memiliki pekerjaan.
10. Pengulangan Mk, merupakan atribut yang menjelaskan apakah selama berkuliah mahasiswa yang bersangkutan memiliki mata kuliah yang dikontrak lebih dari satu kali.
11. Beasiswa, merupakan atribut yang menjelaskan apakah selama berkuliah mahasiswa yang bersangkutan menerima beasiswa berupa bantuan Bidikmisi, KIP Kuliah ataupun Beasiswa Bank Indonesia.
12. MS, merupakan singkatan Masa Studi mahasiswa dalam menyelesaikan program Strata 1 (S1).
13. Umur, merupakan atribut yang menjelaskan umur saat mahasiswa baru memasuki perguruan tinggi.
14. IPK, merupakan singkatan Indeks Prestasi Kumulatif yang diperoleh oleh setiap mahasiswa.

3. Data Preparation

Pada tahap ini, dilakukan proses pembersihan, seleksi, dan transformasi data. Pembersihan data dilakukan dengan membersihkan missing value dan inkonsistensi pada data set, kemudian data akan diseleksi berdasarkan kebutuhan penelitian dan data ditransformasi untuk mempermudah proses data mining.

Tabel 2 Sampel Transformasi Data Latih

JK	AS	JS	JM	A	O	PO	P	MK	B	Umur	IPK	MS
L	SMA	S	SB	DK	A	M	TK	MG	B	Umur<18	3.00<IPK<3.60	Studi>8
P	SMA	S	SN	DK	A	M	K	TM	B	Umur<18	3.00<IPK<3.60	Studi>8
P	SMA	S	SN	DK	A	M	TK	MG	NB	Umur<18	3.00<IPK<3.60	Studi>8
L	SMA	S	SB	DK	A	M	TK	MG	NB	Umur<18	IPK>=3.60	Studi<=8
P	SMA		SB	DK	A	M	TK	MG	B	Umur<18	3.00<IPK<3.60	Studi>8
P	SMA	S	SN	DK	A	T	K	TM	B	Umur<18	IPK>=3.60	Studi>8

Keterangan:

JK: Jenis Kelamin, AS: Asal Sekolah, JS: Jurusan Sekolah, JM: Jalur Masuk, AD: Asal Daerah, O: Organisasi, PO: Penghasilan Orang Tua, P: Pekerjaan, MK: Mata Kuliah Mengulang, B: Beasiswa, IPK: Indeks Prestasi Kumulatif, MS: Masa Studi

Berikut merupakan penjelasan tentang data yang telah ditransformasi:

1. Atribut jenis kelamin memiliki dua kategori yaitu Perempuan disimbolkan dengan P dan Laki-laki disimbolkan dengan L.
2. Atribut asal sekolah memiliki tiga kategori yaitu SMA, SMK dan MA.
3. Jurusan Sekolah memiliki dua kategori yaitu Sejalur disimbolkan dengan S, jurusan saat bersekolah yang masuk dalam kategori sejalur dengan jurusan matematika adalah Matematika dan Ilmu Pengetahuan Alam (MIPA) dan akuntansi dan Tidak Sejalur disimbolkan dengan TS, yang masuk dalam kategori tidak sejalur ialah Ilmu Pengetahuan Sosial (IPS), Keperawatan, Farmasi, Administrasi Perkantoran, TKJ dan sebagainya.
4. Jalur Masuk memiliki 3 kategori yaitu SNMPTN disimbolkan dengan SN, SBMPTN disimbolkan dengan SB dan Jalur Mandiri B2P disimbolkan dengan B2P
5. Asal Daerah dikategorikan menjadi dua yakni dalam kota yaitu yang berasal dari Sulawesi utara disimbolkan dengan DK dan luar kota yaitu yang berasal dari luar Sulawesi utara disimbolkan dengan LK.
6. Organisasi memiliki dua kategori yaitu Aktif disimbolkan dengan A yaitu mahasiswa yang mengikuti satu atau lebih organisasi baik organisasi dalam kampus ataupun organisasi di luar kampus dan Tidak Aktif disimbolkan dengan TA yaitu mahasiswa yang tidak mengikuti satupun organisasi.
7. Penghasilan Orangtua memiliki tiga kategori yaitu rendah disimbolkan dengan R, menengah yang disimbolkan dengan M dan tinggi yang disimbolkan dengan T.
8. Pekerjaan saat berkuliah memiliki dua kategori yaitu Kerja (K) yaitu mahasiswa yang memiliki pekerjaan sampingan saat masih berkuliah dan Tidak Kerja (TK) yaitu mahasiswa yang tidak memiliki pekerjaan sampingan saat berkuliah.
9. Atribut pengulangan mata kuliah memiliki dua kategori yaitu Mengulang yang disimbolkan dengan MG dan Tidak Mengulang yang disimbolkan dengan TM
10. Beasiswa memiliki dua kategori yaitu menerima beasiswa disimbolkan dengan B dan tidak menerima beasiswa disimbolkan dengan NB.
11. Umur saat memasuki perguruan tinggi ditransformasi menjadi 2 kategori yaitu Umur<18 dan Umur>=18 karena umur normal saat memasuki perguruan tinggi ialah 18 tahun.
12. IPK atau Indeks Prestasi Kumulatif memiliki rentang 1-4 dan ditransformasikan menjadi 3 kategori yaitu IPK<3.00, 3.00<IPK<3.60 dan IPK>=3.60.
13. MS atau lama masa studi yang menjadi label atau kelas target ditransformasi menjadi dua kategori yaitu Studi<=8 untuk mahasiswa yang menyelesaikan masa studinya kurang dari atau sama dengan 8 semester atau tepat waktu dan Studi>8 untuk mahasiswa yang menyelesaikan masa studinya lebih dari 8 semester atau tidak tepat waktu.

4. Modeling

Pada tahap ini data yang telah dibersihkan dari *missing value* dan inkonsistensi dan telah ditransformasi akan diolah menggunakan metode klasifikasi Naïve Bayes. Cara pengolahan data menggunakan software RapidMiner Studio Version 10.1

Hasil akurasi *performance vektor* setelah dilakukan proses data mining menggunakan klasifikasi Naïve Bayes mendapat nilai akurasi sebesar 75%.

Tabel 3 Performance Vector

Accuracy: 75.00%

	true Studi<=8	true Studi>8	class precision
pred. Studi<=8	1	0	100.00%
pred. Studi>8	2	5	71.43%
class recall	33.33%	100.00%	

5. Evaluation

Tahapan setelah *modeling* yaitu mengevaluasi proses data mining yang telah dilaksanakan dari tahap *business understanding* hingga tahap *modeling* telah dilakukan sesuai tahapan proses data mining CRISP-DM dan diperoleh tingkat akurasi sebesar 75% menggunakan *confusion matrix*.

6. Deployment

Hasil dari pengklasifikasian mahasiswa aktif dalam bentuk Jurnal dan Skripsi akan dilaporkan kepada program studi Pendidikan Matematika UNIMA.

Pembahasan

Tujuan dari penelitian ini adalah untuk menilai keefektifan data mining dengan menerapkan model klasifikasi menggunakan algoritma Naïve Bayes dalam mengklasifikasikan apakah mahasiswa lulus tepat waktu di Jurusan Matematika Universitas Negeri Manado. Alumni mahasiswa dan mahasiswa aktif menjadi subjek dalam penelitian ini. Peneliti menggunakan kuesioner google form sebagai teknik pengambilan data.

Berdasarkan penelitian oleh Etriyanti dkk (2020) yang meneliti tentang kelulusan mahasiswa menggunakan metode naïve bayes classifier dan C4.5 menunjukkan nilai accuracy sebesar 78,46% menggunakan tool Rapidminer. Penelitian klasifikasi masyarakat miskin yang dilakukan oleh Anur menggunakan klasifikasi *naïve bayes* diperoleh akurasi sebesar 73%, yang termasuk kategori baik, serta nilai presisi sebesar 92% dan recall sebesar 86%, dapat disimpulkan bahwa sistem klasifikasi ini dapat dijadikan sebagai sumber informasi untuk mengambil keputusan. Dan penelitian oleh Saleh (2015) menggunakan klasifikasi naïve bayes dan metode Naïve Bayes ini berhasil memprediksi penggunaan listrik rumah tangga dengan tingkat keakuratan sebesar 73,333%.

Pada penelitian ini mengklasifikasi kelulusan mahasiswa berdasarkan data alumni mahasiswa memperoleh nilai akurasi *performance vector* sebesar 75% dan nilai presisi merupakan ukuran seberapa dekat hasil pengukuran atau perhitungan yang dilakukan berkali-kali satu sama lain, diperoleh nilai presisi sebesar 100%, sedangkan nilai recall adalah ukuran dari kemampuan sebuah model untuk menemukan semua instance positif yang ada dalam dataset, dan nilai recall sebesar 33,33%. Precision mengukur akurasi dari prediksi positif model, sedangkan recall mengukur kelengkapan dari prediksi positif tersebut. sehingga dikatakan algoritma naïve bayes metode klasifikasi data mining efektif digunakan untuk mengklasifikasi waktu kelulusan mahasiswa.

Simpulan (Penutup)

Berdasarkan hasil penelitian menggunakan data mining model klasifikasi dengan algoritma Naïve Bayes untuk memprediksi waktu kelulusan mahasiswa, diperoleh kesimpulan yaitu Data mining dengan model klasifikasi menggunakan algoritma Naïve Bayes terbukti efektif dalam mengklasifikasikan ketepatan waktu kelulusan mahasiswa. Dan Implementasi algoritma naïve bayes dalam mengklasifikasi ketepatan waktu kelulusan mahasiswa dengan persentase keakuratan sebesar

75%. Namun diketahui bahwa penelitian ini menggunakan dataset yang terbatas atau data dengan kualitas yang kurang baik, sehingga tidak mencakup semua faktor relevan yang mempengaruhi kelulusan mahasiswa sehingga diharapkan peneliti lain mempertimbangkan dengan menggunakan dataset yang lebih besar dan menambah variabel atau atribut dalam proses klasifikasi, serta menerapkan dua atau lebih metode, maka dapat melakukan perbandingan akurasi.

Daftar Pustaka

- Apridiansyah, Y., Veronika, N. D. M., & Putra, E. D. (2021). Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes. *JSAI: Journal Scientific and Applied Informatics*, 236–247. <http://dx.doi.org/10.36085/jsai.v4i2.1701>
- Bansal, A., Sharma, M., & Goel, S. (2017). Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining. *International Journal of Computer Applications*, 157(6), 35–40. <https://doi.org/10.5120/ijca2017912719>
- Budiman, I., & Ramadina, R. (2015). Penerapan Fungsi Data Mining Klasifikasi untuk Prediksi Masa Studi Mahasiswa Tepat Waktu pada Sistem Informasi Akademik Perguruan Tinggi. *Ijccs*, x, No.x(1), 1–5. <https://doi.org/10.5281/zenodo.3432716>
- Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika*, 8(1), 884–898. <https://dx.doi.org/10.26555/jifo.v8i1.a2086>
- Etriyanti, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, 13(1), 56–67. <https://doi.org/10.35671/telematika.v13i1.881>
- Fadillah, A. P. (2015). Penerapan Metode CRISP-DM untuk Prediksi Kelulusan Studi Mahasiswa Menempuh Mata Kuliah (Studi Kasus Universitas XYZ). *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(3), 260–270. <https://doi.org/10.28932/jutisi.v1i3.406>
- Hartati, S., & SAN, H. A. (2022). Algoritma Naive Bayes untuk Prediksi Kelulusan Mahasiswa. *Jurnal Cakrawala Informasi*, 2(2), 42–50. <https://doi.org/10.54066/jci.v2i2.234>
- Heryana, D. (2019). *Data Mining Untuk Memprediksi Kelulusan Mahasiswa Pendidikan Matematika Raden Intan Lampung Menggunakan Naive Bayes*. <http://repository.radenintan.ac.id/id/eprint/6430>
- Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *Ultimatika: Jurnal Ilmu Teknik Informatika*, 6(1), 16. <https://doi.org/10.31937/ti.v6i1.327>
- Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries. *Jurnal Mantik Penusa*, 1(2), 16–21. <https://api.semanticscholar.org/CorpusID:169472234>
- Maskoen, T. T., & Purnama, D. (2018). Area Under the Curve dan Akurasi Cystatin C untuk Diagnosis Acute Kidney Injury pada Pasien Politrauma. *Majalah Kedokteran Bandung*, 50(4), 259–264. <https://doi.org/10.15395/mkb.v50n4.1342>
- Putri, R. E., Suparti, & Rahmawati, R. (2014). Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbour Pada Analisis Data Status Kerja di Kab.Demak. *Jurnal Gaussian*, 3, 831–838. <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/8094%0Ahttps://ejournal3.undip.ac.id/index.php/gaussian/article/download/8094/7874>
- Rohmawan, E. P. (2018). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree dan Artificial Neural Network. *Jurnal Ilmiah MATRIK*, 3. <http://journal.binadarma.ac.id/index.php/jurnalmatrik/article/view/105>
- Saleh, A. (2019). Implementasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 1(2), 73–81. <https://doi.org/10.20895/inista.v1i2.73>
- Sumaraw, S. J. (2022). *Data Mining Mode Self-Organizing Maps (SOMs)*. CV. Bintang Semesta Media. http://repository.unima.ac.id:8080/bitstream/123456789/627/44/BUKU%20REFERENSI_DA

TA%20MINING%20MODEL_SYLVIA%20JANE%20A%20SUMARAUW%20%281%29.pdf

- Syukri Mustafa, M., Rizky Ramadhan, M., & Thenata, A. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Citec Journal*, 4(2), 151–162. <https://doi.org/10.24076/citec.2017v4i2.106>
- Widiastuti, N. A., Zyen, A. K., & Safik, N. (2023). Prediksi Penentuan Pemohon Kredit Sepeda Motor Menggunakan Algoritma Naive Bayes. *Jurnal Disprotek*, 10(2), 130–134. <https://doi.org/10.34001/jdpt.v10i2.4689>