

Penerapan *Machine Learning* Untuk Prediksi Masa Studi Mahasiswa di Perguruan Tinggi X

Isna Oktadiani^{1*}, Helmy Fitriawan², Muhammad Nurwahidin³, Herpratiwi⁴

^{1,3,4}Jurusan Ilmu Pendidikan Universitas Lampung, Bandar Lampung

²Jurusan Teknik Elektro Universitas Lampung, Bandar Lampung

Jl. Prof. Sumantri Brojonegoro No.1 Bandar Lampung 35145

¹isnaoktadiani@gmail.com

²helmy.fitriawan@eng.unila.ac.id

Intisari — Perguruan tinggi berperan dalam menghasilkan sumber daya yang berkualitas dari mahasiswa lulusannya, sehingga kualitas dan akreditasi perguruan tinggi menjadi hal yang perlu di perhatikan. Salah satu indikator pada akreditasi perguruan tinggi adalah kelulusan mahasiswa tepat waktu, sehingga kelulusan mahasiswa harus menjadi perhatian penting bagi perguruan tinggi. Berdasarkan hasil dokumentasi presentase kelulusan mahasiswa yang menyelesaikan studinya tepat waktu lebih rendah dari mahasiswa yang menyelesaikan studi tidak tepat waktu, oleh karena itu perlu adanya analisis masa studi mahasiswa untuk mengatasi masa studi yang lulus tidak tepat waktu menggunakan metode machine learning dengan algoritma Naïve Bayes Classifier untuk memprediksi masa studi mahasiswa. Metode penelitian menggunakan metode algoritma Naïve Bayes Classifier yang merupakan bagian dari Artificial Intelligence (AI), yang terdiri dari preprocessing, input, proses dan output. karena metode ini memiliki akurasi yang tinggi dan dapat bekerja lebih baik pada kasus di dunia nyata. Hasil prediksi ketepatan waktu masa studi mahasiswa pada perguruan tinggi X dengan 3553 data dengan metode algoritma Naïve Bayes Classifier, menggunakan tools WEKA berhasil memprediksi masa studi mahasiswa dengan data training 70% dan 30% sebagai data testing secara acak dengan sistem. Menggunakan 11 atribut, yaitu program studi, IPK, pekerjaan ibu, penghasilan ibu, periode masuk, pekerjaan ayah, penghasilan ayah, jalur masuk, jenis kelamin, dan sekolah asal, diperoleh presentase nilai presisi sebesar 54,545%, nilai recall sebesar 67,220%, dan tingkat akurasi mencapai 79,925% dikategorikan baik, dengan menggunakan perhitungan kurva ROC membentuk hampir mendekati (0,1) dengan nilai AUC sebesar 0,844 dikategorikan sangat baik. Berdasarkan hasil presentase tingkat akurasi, kurva ROC dan nilai AUC, sehingga Naïve Bayes Classifier dalam memprediksi kelulusan mahasiswa dengan kategori “Baik”.

Kata kunci — Masa studi, Naïve Bayes Classifier, Akurasi.

Abstract — Universities play a role in producing quality resources from their graduate students, so that the quality and accreditation of tertiary institutions are things that need attention. One indicator of higher education accreditation is student graduation on time, so student graduation must be an important concern for tertiary institutions. Based on the results of the documentation, the percentage of students graduating on time is lower than students who are not completing their studies on time, therefore it is necessary to analyze the student's study period to overcome the study period that graduates are not on time using the machine learning method with the Naïve Bayes Classifier algorithm to predict student study period. The research method uses the Naïve Bayes Classifier algorithm method which is part of Artificial Intelligence (AI), which consists of preprocessing, input, process and output. because this method has high accuracy and can work better in real-world cases. The results of predicting the timeliness of student study time at college X with 3553 data using the Naïve Bayes Classifier algorithm method, using WEKA tools succeeded in predicting student study time with 70% data training and 30% as random testing data with the system. Using 11 attributes, namely study program, GPA, mother's occupation, mother's income, entry period, father's occupation, father's income, route of entry, gender, and school of origin, obtained a percentage of precision value of 54.545%, recall value of 67.220%, and the accuracy level reaches 79.925% which is categorized as good, using the ROC curve calculation to form almost close to (0.1) with an AUC value of 0.844 which is categorized as very good. Based on the results of the percentage accuracy rate, ROC curve and AUC value, the Naïve Bayes Classifier predicts student graduation in the "Good" category.

Keywords— Period of study, Naïve Bayes Classifier, Accuracy.

I. PENDAHULUAN

Sistem pendidikan saat ini semakin maju dan berkembang, sehingga dibutuhkan dukungan Teknologi Informasi dan Komunikasi (TIK) dalam pelaksanaannya. Kemajuan teknologi telah mempengaruhi tatanan kehidupan dan tidak bisa dihindari karena hadirnya teknologi memberikan banyak manfaat dan kemudahan dalam pekerjaan serta pendidikan. Teknologi merupakan bagian dari Ilmu Pengetahuan dan Teknologi (IPTEK) memberikan banyak manfaat dan memudahkan pekerjaan, teknologi sebagai sistem yang mendukung dalam proses pendidikan (Mulyani & Haliza, 2021). Ketika melakukan sebuah penelitian, untuk melakukan analisis terhadap data yang sudah diperoleh harus dianalisis dan dihitung secara manual. Setelah adanya perkembangan IPTEK, semua tugas yang dikerjakan dengan manual dan membutuhkan waktu yang cukup lama, menjadi sesuatu yang mudah untuk dikerjakan, yaitu dengan menggunakan media teknologi, seperti dapat mengolah data dengan memanfaatkan berbagai program (Jamun, 2018).

IPTEK adalah awal dari kesuksesan bangsa, dan mendukung terciptanya teknologi-teknologi baru yang menandai adanya kemajuan zaman (Lestari, 2018). Hal tersebut dinyatakan oleh Prof. Agus Sartono, kesadaran akan pentingnya IPTEK yang telah disampaikan sejak 60 tahun yang lalu dalam pidato presiden Soekarno di Malang pada tahun 1958 bahwa “bangsa ini akan maju dan sejahtera jika pembangunannya dilandaskan pada ilmu pengetahuan dan teknologi” (Mulyani & Haliza, 2021). Dari pendapat ini, dengan teknologi dan pendidikan akan menempuh kemajuan dan perkembangan zaman.

Dunia pendidikan masih menghadapi tantangan yang cukup mendasar yaitu masalah mutu dan daya saing pendidikan, salah satunya pengelolaan pendidikan di perguruan tinggi. Perguruan tinggi mempunyai peranan yang sangat penting dalam meningkatkan kualitas pendidikan dan mampu melahirkan sumber daya yang unggul dan berkualitas. Kualitas perguruan tinggi salah satunya terlihat pada akreditasi

perguruan tinggi. Berdasarkan pedoman penilaian pemantauan dan evaluasi peringkat akreditasi, akreditasi perguruan tinggi – perguruan tinggi vokasi, perguruan tinggi negeri badan layanan umum, salah satu elemen penting pada borang akreditasi perguruan tinggi adalah data kelulusan, yaitu masa studi mahasiswa untuk setiap program dalam 3 tahun terakhir dan Persentase kelulusan tepat waktu untuk setiap program (Badan Akreditasi Nasional Perguruan Tinggi, 2021). Masa studi merupakan lamanya waktu yang dibutuhkan oleh mahasiswa dari masuk kuliah hingga lulus (Nalim et al., 2021).

Sedemikian pentingnya kualitas pendidikan, salah satunya adalah peningkatan kualitas lulusan. Pada institusi pendidikan di perguruan tinggi, bagian data kelulusan terdapat data nilai mahasiswa dan data jumlah kelulusan mahasiswa yang berukuran besar, yang dapat menghasilkan informasi yang berlimpah berupa jumlah kelulusan setiap tahunnya dan hasil akademik mahasiswa selama menempuh proses pembelajaran di perguruan tinggi.

Permasalahan utama dalam penelitian ini adalah perguruan tinggi di setiap tahun ajaran baru melaksanakan penerimaan mahasiswa baru, dengan jumlah pendaftar dan kuota daya tampung yang juga meningkat. Meskipun akhirnya mahasiswa akan lulus dan melaksanakan wisuda, namun terdapat mahasiswa yang lulus dengan masa studi tidak tepat waktu dan mahasiswa yang tidak aktif atau tidak melanjutkan studinya. Berdasarkan peraturan akademik perguruan tinggi X dijelaskan tentang penyelesaian kegiatan pendidikan di Universitas Lampung di bagian kesebelas tentang beban dan masa studi mahasiswa. Mahasiswa mengikuti kegiatan akademik untuk program sarjana dengan beban belajar minimal 144 sks batas maksimal 160 sks termasuk dengan tugas akhir, dengan masa studi 7 sampai 12 semester. Dalam kurikulum akademik yang dirancang diharapkan mahasiswa dapat menyelesaikan studinya selama 3,5 sampai 4 tahun. Mahasiswa yang lulus kuliah dalam waktu lebih dari 4 tahun bahkan sampai dengan 7 tahun, mahasiswa tersebut terlambat dalam menyelesaikan masa

studinya yang telah direncanakan. Menurut Nalim (Nalim et al., 2021), berdasarkan hasil penelitiannya, ada empat faktor yang secara signifikan mempengaruhi masa studi, yaitu jenis kelamin, jenis sekolah tinggi, IPK, dan jalur PMB. Perbedaan masa studi mahasiswa ditinjau dari jalur penerimaan mahasiswa baru terjadi dikarenakan masa studi berkaitan erat dengan prestasi belajar mahasiswa (IPK). Semakin baik prestasi belajar mahasiswa, maka masa studinya semakin cepat, begitu juga sebaliknya. Perbedaan rata-rata masa studi ini disebabkan oleh beberapa faktor, salah satunya yaitu prestasi mahasiswa sebelum masuk ke perguruan tinggi. Hal ini terlihat dari rata-rata masa studi yang tertinggi yaitu pada mahasiswa yang masuk perguruan tinggi melalui jalur SNMPTN. Jalur SNMPTN merupakan jalur masuk khusus yang diperuntukkan bagi siswa-siswa berprestasi di sekolah asalnya dan tidak memerlukan tes masuk perguruan tinggi. Hal ini tentu berkaitan erat dengan bagaimana pengelolaan waktu dan cara belajar mahasiswa ketika di bangku perkuliahan. Secara tidak langsung, pembiasaan di sekolah lama tentu mempengaruhi prestasi belajar dan tentunya masa studi mahasiswa di perguruan tinggi. Pada penelitian lainnya oleh Aurora (Putriku, 2018) bahwa pengaruh tingkat penghasilan orangtua terhadap

prestasi belajar mahasiswa, secara partial tingkat penghasilan orangtua memiliki pengaruh positif terhadap prestasi belajar mahasiswa. Semakin tinggi tingkat penghasilan orangtua akan lebih mudah dalam membiayai dan mencukupi segala kebutuhan anak terlebih dalam melaksanakan pendidikan anak, sehingga hal ini mempengaruhi prestasi belajar mahasiswa. Namun tidak disangka bahwa sebagian anak lainnya yang serba kekurangan menjadikan kondisi yang serba kekurangan akibat kondisi orangtua ini sebagai batu loncatan untuk lebih giat belajar dan mendapatkan prestasi belajar yang baik, memiliki motivasi yang tinggi untuk tetap melanjutkan pendidikan hingga ke perguruan tinggi dan memiliki cita-cita besar untuk memerangi kemiskinan dengan harapan kelak mampu mengubah perekonomian mereka kelak. Berdasarkan hasil penelitian sebelumnya faktor yang mempengaruhi masa studi mahasiswa, diantaranya jenis kelamin, jenis sekolah tinggi, IPK, dan jalur PMB, dan penghasilan orangtua, maka peneliti tertarik untuk melakukan penelitian dengan variabel tersebut pada masa studi mahasiswa di perguruan tinggi yang berbeda dari penelitian sebelumnya. Berikut data Masa Studi mahasiswa angkatan 2017 di perguruan tinggi X:

Tabel 1. Data Masa Studi Mahasiswa Angkatan Masuk Tahun 2017 perguruan tinggi X

Tahun Lulus & semester	Semester	Jumlah Mahasiswa		
		Lulus	Tidak Aktif/DO	Aktif
2017/2018 ganjil	1			
2017/2018 genap	2			
2018/2019 ganjil	3		1	
2018/2019 genap	4		0	
2019/2020 ganjil	5		64	
2019/2020 genap	6		1	
2020/2021 ganjil	7	0	21	
2020/2021 genap	8	125	16	
2021/2022 ganjil	9	236	3	
2021/2022 genap	10	196	0	
2022/2023 ganjil	11	59	117	263
Total		616	223	263

Sumber: Data Akademik perguruan tinggi X.

Berdasarkan data tersebut diketahui bahwa dari 1102 mahasiswa angkatan 2017, terdapat mahasiswa yang lulus tepat waktu dan telah menempuh perkuliahan selama 7 dan 8

Semester sebanyak 125 mahasiswa atau 11,34%, sedangkan mahasiswa lulus yang menempuh perkuliahan selama > 8 Semester sebanyak 491 mahasiswa atau 44,56%,

terdapat mahasiswa yang status studinya tidak aktif (tidak menyelesaikan studinya) sebanyak 223 mahasiswa atau 20,24%, dan sebanyak 263 mahasiswa atau 23,87% yang belum menyelesaikan studinya hingga tahun akademik 2022/2023 ganjil. Mahasiswa yang lulus > 8 semester lebih banyak dari mahasiswa yang lulus ≤ 8 semester (mahasiswa yang lulus tepat waktu). Melihat data masa studi mahasiswa di atas, setelah dilakukan observasi kepada wakil dekan bidang akademik fakultas keguruan dan ilmu pendidikan, bahwa proses bagi peserta yang putus studi atau DO terdeteksi melalui sistem dari akademik universitas dan fakultas, dan dari fakultas memberikan surat peringatan kepada mahasiswa tersebut melalui masing-masing prodi. Adapun tentang mahasiswa yang lulus tidak tepat waktu akan diberi peringatan dari masing-masing prodi, dan saat ini kebijakan yang diambil berupa peringatan kepada mahasiswa yang bersangkutan, dan belum adanya pengembangan kebijakan lebih dalam hal menelusuri faktor-faktor yang mempengaruhi masa studi mahasiswa, belum ada alat untuk memprediksi masa studi mahasiswa, serta belum adanya analisis menggunakan algoritma komputasi dalam menganalisis dan memprediksi masa studi mahasiswa.

Berdasarkan permasalahan pada masa studi tersebut, diperlukan sebuah sistem untuk prediksi kelulusan mahasiswa tepat waktu dengan sebuah analisis komputasi yang mampu memberikan gambaran dan peluang kelulusan mahasiswa. Adanya informasi mengenai lama studi mahasiswa tentu akan menjadi pendukung suatu pengambilan keputusan yang tepat bagi manajemen Perguruan Tinggi dalam mengambil langkah berikutnya. Dengan demikian, dapat menjadi dasar dalam pengambilan kebijakan dan manajemen pendidikan.

Machine Learning adalah suatu area dalam *Artificial Intelligence* (AI) atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. *Artificial Intelligent* (AI) atau Kecerdasan Buatan adalah teknik yang menjadikan komputer dapat berpikir secerdas atau melampaui kecerdasan manusia.

Tujuannya agar komputer memiliki kemampuan berperilaku, berpikir, dan mengambil keputusan layaknya manusia (Cholissodin et al., 2020). Arthur Samuel, seorang pionir dalam pengembangan permainan komputer dan kecerdasan buatanlah yang pertama kali mengeluarkan istilah "*Machine Learning*" ke publik pada tahun 1959. Perkembangan pembelajaran mesin tumbuh berkat berkembangnya bidang kecerdasan buatan atau *Artificial Intelligence* (AI). Banyak peneliti di bidang AI tertarik untuk memiliki mesin yang dapat belajar dari data (Ibnu, 2021).

Algoritma Naïve Bayes Classifier merupakan sebuah metode klasifikasi menggunakan metode *Probabilitas* dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma Naïve Bayes Classifier memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari Naïve Bayes Classifier ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi / kejadian (Mochamad Farid et al., 2019).

Naïve Bayes Classifier adalah suatu metode yang digunakan untuk dapat memperkirakan atau memprediksi suatu class dari suatu objek yang kelasnya tidak diketahui dari masing-masing kelompok atribut yang ada, dan menentukan class mana yang paling optimal berdasarkan pengaruh yang didapat dari hasil pengamatan. Klasifikasi– klasifikasi Bayes adalah klasifikasi statistik yang dapat digunakan untuk memprediksi *Probabilitas* keanggotaan suatu *class* (Ling et al., 2014).

Naïve Bayes Classifier didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, *Probabilitas* mengamati secara bersama adalah produk dari *Probabilitas* individu. Keuntungan penggunaan Naïve Bayes Classifier adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Saleh, 2015).

Berdasarkan teorema bayes yang diutarakan oleh Thomas Bayes berpedoman pada konsep

Probabilitas bersyarat, yang menjelaskan tentang peluang sebuah kejadian berdasarkan pengetahuan awal (prior) dari kondisi yang dari kondisi yang berhubungan dengan kejadian tersebut atau dikenal sebagai teorema yang melakukan prediksi *Probabilitas* di masa depan dengan menggunakan dasar dari pengalaman yang ada di masa sebelumnya (Sari, 2016)

Prediksi Bayes didasarkan pada formula teorema Bayes berikut dasar dari algoritma *Naïve Bayes Classifier*:

$$P(C|F) = \frac{P(F|C).P(C)}{P(F)} \quad (1)$$

dimana:

$P(C)$: *Probabilitas* dari kelas C (prior probability)

$P(F)$: *Probabilitas* dari F (evidence)

$P(C|F)$: *Probabilitas* kelas C berdasarkan kondisi F (posterior probability)

$P(F|C)$: *Probabilitas* F berdasarkan kondisi kelas

Naïve Bayes Classifier dalam proses klasifikasi data memerlukan banyak petunjuk berupa atribut untuk mendapatkan kesimpulan berupa label kelas yang sesuai untuk sampel data. semakin banyak karakteristik atau atribut dari data maka semakin rumit faktor-faktor syarat yang mempengaruhi nilai *Probabilitas*, dengan adanya asumsi independent pada *Naïve Bayes* membuat masing-masing atribut (F_1, F_2, \dots, F_n) saling bebas antara satu dengan yang lain, dengan adanya asumsi independensi yang kuat membuat syarat peluang menjadi lebih sederhana seperti persamaan berikut (Dahri et al., 2016):

$$P(C|F_1, \dots, F_n) = P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots \\ = P(C) \prod_{i=1}^n P(F_i|C) \quad (2)$$

Keuntungan menggunakan metode *naïve bayes* adalah metode ini hanya memerlukan jumlah data pelatihan (*training data*) yang kecil untuk menentukan parameter yang diperlukan dalam proses pengklasifikasian atau biasa disebut *Naïve Bayes Classifier*. Kelebihan metode *Naïve Bayes Classifier* lainnya yaitu implementasi yang sederhana, bekerja dengan cepat, memberikan hasil yang baik, serta memiliki tingkat akurasi yang

tinggi dengan perhitungan yang sederhana (Heksaputra et al., 2013).

Metode ini menunjukkan kecepatan dan akurasi yang tinggi saat digunakan pada database yang besar (Handayani & Pribadi, 2015).

II. METODE PENELITIAN

Pada penelitian ini akan dilakukan pengembangan model analisis menggunakan *machine learning* yang merupakan salah satu bagian dari kecerdasan buatan (*Artificial Intelligence*). *Machine learning* merupakan sub dari bidang keilmuan kecerdasan buatan (*Artificial intelligence*) yang banyak diteliti dan digunakan untuk memecahkan berbagai masalah. Menggunakan algoritma *machine learning* untuk memecahkan masalah sesuai dengan kebutuhan pada bidang masing-masing (Roihan et al., 2020). Algoritma *machine learning* pun terus dikembangkan oleh beberapa penelitian. Tren terkini menunjukkan bahwa algoritma *machine learning* banyak dikembangkan dalam berbagai bidang salah satunya bidang pendidikan seperti (Adnyana, 2019); (Khasanah et al., 2022); (Pratama et al., 2020); (Martanto et al., 2019).

Salah satu algoritma dalam *machine learning* adalah *Naïve Bayes Classifier* yang merupakan metode pengklasifikasian statistik yang dapat digunakan untuk memprediksi *Probabilitas* keanggotaan suatu kelas (Damayanti, 2019).

Penelitian terdahulu menurut (Widaningsih, 2019), Algoritma *Naïve Bayes Classifier* yang paling baik untuk memprediksi tingkat kelulusan, karena memiliki nilai akurasi tertinggi dan error terkecil dibandingkan dengan algoritma lainnya.

Solusi berdasarkan permasalahan adalah dengan mengidentifikasi akurasi tepat waktu atau tidak tingkat kelulusan mahasiswa dengan model *Naïve Bayes Classifier*. Model yang diambil dalam penelitian ini adalah *Naïve Bayes Classifier*, kelebihan *Naïve Bayes Classifier* dibandingkan algoritma lain adalah pada kemampuannya mengklasifikasi dokumen dengan kesederhanaan dan kecepatan komputasinya namun memiliki

komputasi tinggi, metode Naïve Bayes Classifier juga memiliki kinerja yang baik terhadap pengklasifikasian data dokumen yang mengandung angka maupun teks (Nathaniel Chandra et al., 2019).

Pada penelitian ini akan digunakan data akademik mahasiswa perguruan tinggi X. Data tersebut terdiri dari banyak fitur. Adapun class yang digunakan yaitu lulus tepat waktu dan tidak tepat waktu. Oleh karena itu, peneliti melakukan suatu penelitian penerapan kecerdasan buatan (*Artificial Intelligence*) dalam pendidikan, pemanfaatan teknologi pada pendidikan dalam mengidentifikasi masa studi mahasiswa menggunakan metode *Naïve Bayes Classifier*.

Objek penelitian ini pada mahasiswa di perguruan tinggi X diharapkan adanya rekomendasi dan tindak lanjut terhadap proses pembelajaran dapat menjadi lebih baik lagi sehingga meminimalisir mahasiswa mutasi, lulus tidak tepat waktu, dan mahasiswa drop out. Adapun Subjek pada penelitian ini merupakan mahasiswa perguruan tinggi X.

Pengumpulan data dilakukan dalam berbagai *setting*, berbagai sumber, dan berbagai cara (Sugiyono, 2015). Pengumpulan data yang dilakukan penulis dalam penelitian ini adalah dokumentasi. Kegiatan pengumpulan data pada penelitian ini adalah melakukan kajian terhadap beberapa teori yang berasal dari jurnal-jurnal maupun penelitian yang relevan dihubungkan dengan potensi dan identifikasi masalah yang dapat menjadi penguat dan solusi untuk memprediksi masa studi mahasiswa. Pada penelitian ini dilakukan pengumpulan data dan informasi yang akurat, sehingga dapat mendukung proses penelitian. Berikut adalah teknik pengumpulan data tersebut:

A. Studi literatur

Mempelajari dan memahami teori-teori yang berhubungan dengan penelitian baik secara langsung maupun tidak langsung, sumbernya adalah buku, jurnal, karya ilmiah, paper dan sumber ilmiah lainnya.

B. Dokumentasi

Metode dokumentasi adalah mencari data mengenai hal-hal yang berupa catatan transkrip, buku, surat, dokumentasi, dan sebagainya. Adapun variabel atribut yang akan digunakan pada penelitian ini adalah sebagai berikut:

Tabel 1. Variabel Penelitian

Variabel	Kategori		
Variabel Depende n	Periode Masuk (Tahun Angkatan)	2014	
		2015	
		2016	
		2017	
	IPK	2,10-2,75	
		2,76-3,00	
		3,01-3,50	
		3,51-4,00	
	Jenis Kelamin	Laki-laki	
		Perempuan	
	Jenis Sekolah Asal	SMAN	
		SMAS	
		MAN	
		MAS	
		SMKN	
	Jalur Masuk Kampus	SNMPTN	
		SBMPTN	
		Mandiri	
		Paralel	
		Pmpap	
		Prestasi Khusus	
		Afirmasi	
		Kerjasama	
		Program Studi	Pend. Matematika
			Pend. Fisika
	Pend. Kimia		
	Pend. Biologi		
PPKN			
Pend. Sejarah			
Pend. Geografi			
Pend. Ekonomi			
Pendidikan Jasmani			
Bimb. Konseling			
Pend. Bahasa Indonesia			
Pend. B. Inggris			
PGSD			
Pendidikan Tari			
Pendidikan Anak Usia Dini			
Bahasa Perancis			
Pekerjaan Ayah	BUMN		
	Buruh		
	Dosen		

Variabel	Kategori		
	Guru		
	Nelayan		
	Pedagang		
	Pensiunan		
	Petani		
	PNS		
	Polri		
	Supir		
	Wirastawa		
	Tidak Bekerja		
	Pekerjaan Ibu	BUMN	
		Buruh Harian Lepas	
		Guru	
		Ibu Rumah Tangga	
		Pedagang	
		Pensiunan	
		Petani	
		PNS	
		TNI	
		Wiraswasta	
		Ibu Rumah Tangga	
		Tidak bekerja	
	Penghasilan Ayah	>5.000.000	
		4.000.001 – 5.000.000	
		3.000.001 – 4.000.000	
		2.000.001 – 3.000.000	
		1.000.001 – 2.000.000	
		1 – 1.000.000	
		0	
		Penghasilan Ibu	>5.000.000
			4.000.001 – 5.000.000
			3.000.001 – 4.000.000
	2.000.001 – 3.000.000		
1.000.001 – 2.000.000			
1 – 1.000.000			
0			
Variabel Independen	Status (Ketepatan waktu kelulusan)		
	Tepat Waktu Tidak Tepat Waktu		

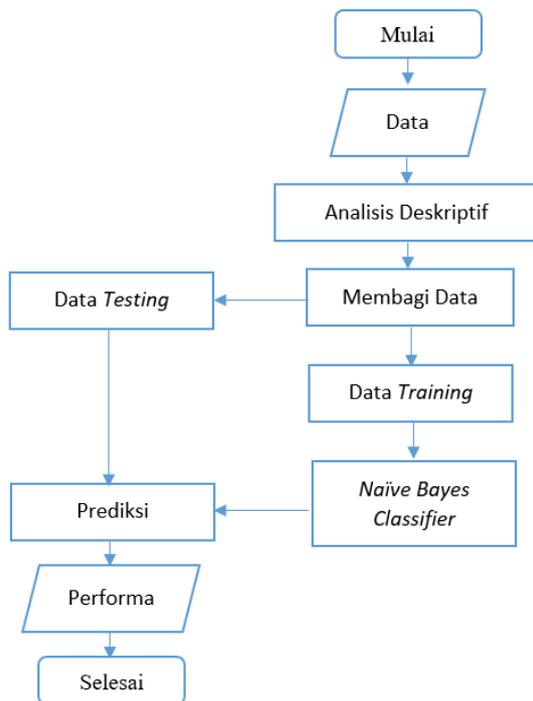
Metode deskriptif diterapkan dalam mendeskripsikan proses penyusunan untuk memprediksi kelulusan mahasiswa tepat waktu. Adapun teknik analisis metode *Machine learning* dengan algoritma *Naïve*

Bayes Classifier. Teknik analisis dan prediksi data yang digunakan pada penelitian adalah sebagai berikut:

1. Menganalisis data secara deskriptif untuk mengetahui gambaran umum mengenai data Mahasiswa perguruan tinggi X serta menentukan variabel yang diperlukan dalam proses klasifikasi.
2. Membagi data menjadi data training dan data testing.
3. Pembuatan model klasifikasi dari data training dengan menggunakan algoritma *Naïve Bayes Classifier*, dengan tahapan-tahapan sebagai berikut:
 - Menghitung *Probabilitas* prior pada data training, yakni *Probabilitas* mahasiswa lulus tepat waktu dan tidak tepat waktu.
 - Menghitung *Probabilitas* setiap variabel pada data training terhadap masing-masing kelas.
 - Menggunakan data testing yang berdasarkan *Probabilitas* data training, hitung perkalian *Probabilitas* prior dengan atribut pada masing-masing kelas.
 - Mencari nilai *Probabilitas* maksimal pada masing-masing kelas.
 - Menghitung ketepatan klasifikasi dengan cara menghitung jumlah *error* (APER), *akurasi*, *sensitivity* dan *specificity* dalam proses pengklasifikasian.

Berdasarkan langkah-langkah yang dijelaskan diatas, penelitian ini bertujuan untuk memprediksi masa studi mahasiswa dengan menggunakan machine learning algoritma *Naïve Bayes*. Langkah pertama yang dilakukan adalah data pre-processing merupakan tahap persiapan data.

Tahapan diatas dapat dilihat pada Gambar 1 berikut:



Gbr.1 Diagram Alir Tahapan Analisis dan Prediksi

III. HASIL DAN PEMBAHASAN

A. Preprocessing Data

Tahap preprocessing merupakan tahap persiapan data. Pada tahap ini, data yang digunakan merupakan gabungan dari data akademik, keuangan, dan siakadu. Ketelitian pada penyusunan data sangat diperlukan karena jumlah data yang banyak. Setelah data tersusun, kemudian dilakukan proses pembersihan dataset (*Cleaning dataset*). Proses pembersihan data dilakukan dengan memeriksa data yang tidak konsisten, memperbaiki data yang *missing values*, dan *redundant data*.

Dataset hasil proses pembersihan akan digunakan sebagai masukan dalam proses imputasi *missing values*. Berdasarkan data yang diperoleh, fitur yang akan digunakan sebanyak 12 fitur yang terdiri dari, periode masuk, jenis kelamin, jalur masuk, sekolah asal, pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, kode program studi, ipk lulus, dan status kelulusan tepat waktu atau tidak tepat waktu.

Selanjutnya data diinput menggunakan tools WEKA. Berdasarkan hasil prediksi dari

Metode Machine Learning menggunakan Algoritma Naïve Bayes Classifier digunakan rumus akurasi, recall dan presisi. Setelah dilakukan pengolahan data diperoleh akurasi pada model tersebut. Akurasi pada model dihitung dengan menggunakan *confusion matrix*. Berikut ini *confusion matrix* dengan metode Naïve Bayes Classifier dapat dilihat pada Gambar 2 berikut.

```

=== Confusion Matrix ===
      a  b  <-- classified as
162 135 |   a = 1
 79 690 |   b = 2
  
```

Gbr.2 Confusion Matrix

Gambar 2 merupakan *confusion matrix* dengan metode Naïve Bayes Classifier yang diperoleh melalui tools WEKA. Adapun *confusion matrix* dengan metode Naïve Bayes Classifier tersebut dalam bentuk Tabel dapat dilihat pada Tabel 1.2 Berikut.

Tabel 2. Tabel Confusion Matrix

Correct Classification	Classified as	
	<i>Predicted "+"</i>	<i>Predicted "-"</i>
<i>Actual "+"</i>	162	135
<i>Actual "-"</i>	79	690

Pada Tabel 2 dapat dilihat *confusion matrix* sebanyak 162 record data *True Positive (TP)* di prediksi tepat sebagai lulus tepat waktu dan sebanyak 135 record *False Positive (FP)* di prediksi tidak tepat sebagai lulus tidak tepat waktu, kemudian sebanyak 690 record data *True Negative (TN)* di prediksi tepat sebagai lulus tidak tepat waktu dan sebanyak 79 record *False Negative (FN)* di prediksi tidak tepat sebagai lulus tidak tepat waktu.

Selanjutnya mengukur kemampuan model klasifikasi yang dibangun, yakni *precision*, *recall* dan akurasi. Nilai *precision* adalah nilai sensitifitas atau nilai ketepatan model antara informasi yang diberikan oleh sistem untuk menunjukkan secara benar data kelas negatif atau kelas positif.

Sedangkan nilai *recall* adalah nilai yang menunjukkan tingkat keberhasilan atau spesifisitas untuk mengetahui kembali sebuah informasi secara benar tentang data yang

kelas negatif atau pun konten teks positif. Sedangkan akurasi adalah nilai yang menunjukkan tingkat kedekatan antara nilai prediksi sistem dengan nilai prediksi manusia. Perhitungan nilai akurasi, presisi dan recall adalah sebagai berikut.

1) Akurasi

Perbandingan jumlah item yang di prediksi benar dengan total seluruh prediksi yang dilakukan.

$$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{162 + 690}{162 + 690 + 79 + 135} \\ &= 79,925 \end{aligned}$$

2) Recall atau Sensitivity atau TPR (True Positive Rate)

Perbandingan jumlah item yang relevan diidentifikasi benar dengan seluruh item yang benar.

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ &= \frac{162}{162 + 79} \\ &= 67,220 \end{aligned}$$

3) Presisi

Perbandingan jumlah item yang di identifikasi sebagai positif secara benar terhadap jumlah item yang di identifikasi positif.

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ &= \frac{162}{162 + 135} \\ &= 54,545 \end{aligned}$$

Hasil Ketepatan model Naïve Bayes Classifier dapat dilihat pada Tabel 4.15 sebagai berikut.

Tabel 3. Performa prediksi algoritma Naïve Bayes Classifier

Akurasi	Presisi	Recall
79,925	54,545	67,220

Selanjutnya, perhitungan *Missclassification (Error) Rate*, merupakan presentase jumlah record data yang

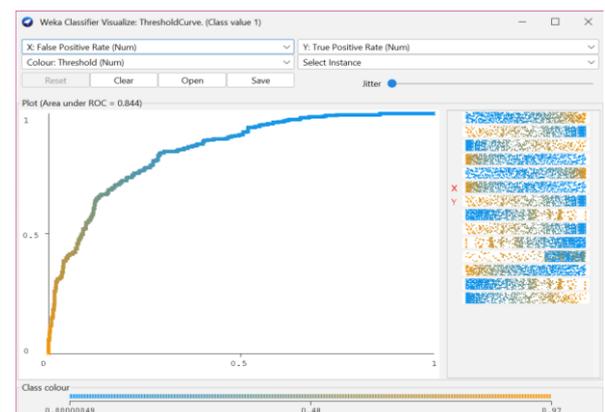
dipredikdi secara salah oleh algoritma yang digunakan, perhitungannya sebagai berikut.

Missclassification (Error) Rate

$$\begin{aligned} &= \frac{FP + FN}{\text{Total Data}} \\ &= \frac{135 + 79}{1066} \\ &= 20,075\% \end{aligned}$$

Berdasarkan hasil perhitungan diatas, diperoleh presentase keakuratan menunjukkan keefektifan dataset kelulusan mahasiswa yang diterapkan ke dalam metode Naïve Bayes Classifier, pada nilai presisi sebesar 54,545%, hal ini menunjukkan kemampuan metode Naïve Bayes Classifier dalam mencocokkan antara bagian data yang diambil dengan informasi yang dibutuhkan. Selanjutnya nilai recall diperoleh sebesar 67,220%, yang menunjukkan keberhasilan metode Naïve Bayes Classifier dalam menemukan kembali suatu informasi pada prediksi ketepatan kelulusan mahasiswa. Dan nilai akurasi mencapai 79,925%, menunjukkan keakuratan metode Naïve Bayes Classifier dalam memprediksi kelulusan mahasiswa perguruan tinggi X dengan tingkat akurasi “Baik” dengan presentase eror sebesar 20,075%.

Hasil pengujian metode Naïve Bayes Classifier dengan kurva ROC dari tools WEKA dapat dilihat pada Gambar 4.21 dan sebaran data dapat dilihat pada Gambar 1.3 berikut.



Gbr. 3 Kurva ROC metode Naïve Bayes Classifier



Gbr.4 Sebaran data set berdasarkan kurva ROC

Pada Gambar 3 dan 4 terlihat bahwa kurva membentuk hampir mendekati (0,1) dengan nilai AUC sebesar 0,844, sehingga dapat dikatakan Naïve Bayes Classifier dalam memprediksi kelulusan mahasiswa dengan kategori “Sangat Baik”.

IV. KESIMPULAN

Berdasarkan hasil prediksi masa studi mahasiswa dengan metode machine learning menggunakan Algoritma Naïve Bayes Classifier pada data masa studi mahasiswa diperoleh nilai presisi sebesar 54,545%, hal ini menunjukkan kemampuan metode Naïve Bayes Classifier dalam mencocokkan antara bagian data yang diambil dengan informasi yang dibutuhkan. Selanjutnya nilai recall diperoleh sebesar 67,220%, yang menunjukkan keberhasilan metode Naïve Bayes Classifier dalam menemukan kembali suatu informasi pada prediksi ketepatan kelulusan mahasiswa. Dan nilai akurasi mencapai 79,925%, menunjukkan keakuratan metode machine learning menggunakan algoritma Naïve Bayes Classifier dalam memprediksi kelulusan mahasiswa perguruan tinggi X dengan tingkat akurasi “Baik”.

UCAPAN TERIMA KASIH

Terima kasih kepada semua pihak yang telah memberikan kontribusi dalam saran dan motivasi pada penulisan penelitian ini.

REFERENSI

- [1] Adnyana, I. M. B. (2019). Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa. *Jurnal Sistem Dan Informatika*, 13(2), 72–76.
- [2] Badan Akreditasi Nasional Perguruan Tinggi, (BAN PT). (2021). *Pedoman Penilaian Pemantauan dan Evaluasi Peringkat Akreditasi Perguruan Tinggi Akademik Perguruan Tinggi Negeri – Badan Layanan Umum* (Versi upda). <https://www.banpt.or.id>
- [3] Cholissodin, I., Sutrisno, Soebroto, A. A., Hasanah, U., & Febiola, Y. I. (2020). *Buku Ajar AI , Machine Learning & Deep Learning (Teori & Implementasi)*. Fakultas Ilmu Komputer Universitas Brawijaya.
- [4] Dahri, D., Agus, F., & Khairina, D. M. (2016). Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, 11(2), 29. <https://doi.org/10.30872/jim.v11i2.211>
- [5] Damayanti, A. (2019). Algoritma Naïve Bayes Untuk Prediksi Jumlah Pendaftar Ulang Pada Penerimaan Mahasiswa Baru. *Jurnal Multi Media Dan IT (JOMMIT)*, 3(2), 2–6. <https://doi.org/10.46961/jommit.v3i2.338>
- [6] Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, 7(1), 19–24.
- [7] Heksaputra, D., Naimah, Z., Azani, Y., & Iswari, L. (2013). Penentuan Pengaruh Iklim Terhadap Pertumbuhan Tanaman dengan Naïve Bayes. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 34–39.
- [8] Ibnu, D. I. (2021). Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python. In *UR Press*. <https://doi.org/10.5281/zenodo.5113507>
- [9] Jamun, Y. M. (2018). Dampak Teknologi Terhadap Pendidikan - Pdf. *Jurnal Pendidikan Dan Kebudayaan Missio*, 10(1), 48–52. <http://jurnal.unikastpaulus.ac.id/index.php/jpkm/article/view/54>
- [10] Khasanah, N., Salim, A., Afni, N., Komarudin, R., & Maulana, Y. I. (2022). Prediksi Kelulusan Mahasiswa dengan Metode Naive Bayes. *Technologia*, 13(3), 207–214. <https://doi.org/10.31602/tji.v13i3.7312>

- [11] Lestari, S. (2018). Peran Teknologi dalam Pendidikan di Era Globalisasi. *Edureligia: Jurnal Pendidikan Agama Islam*, 2(2), 94–100.
<https://doi.org/10.33650/edureligia.v2i2.459>
- [12] Ling, J., N. Kencana, I. P. E., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92–99.
<https://doi.org/10.24843/mtk.2014.v03.i03.p070>
- [13] Martanto, M., Ali, I., & Mulyawan, M. (2019). Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Machine Learning dengan Teknik Deep Learning. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, 4(2–2), 191–194.
<https://doi.org/10.30591/jpit.v4i2-2.1877>
- [14] Mochamad Farid, R., Jatnika, H., & Valentino, B. (2019). Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS). *Jurnal Pengkajian Dan Penerapan Teknik Informatika*, 12(2), 131–144.
<https://doi.org/10.33322/petir.v12i2.471>
- [15] Mulyani, F., & Haliza, N. (2021). Analisis Perkembangan Ilmu Pengetahuan dan Teknologi (IPTEK) dalam Pendidikan. *Jurnal Pendidikan Dan Konseling (JPdK)*, 3(1), 101–109.
- [16] Nalim, N., Dewi, H. L., & Safii, M. A. (2021). Analisis Faktor-Faktor yang Mempengaruhi Keberhasilan Studi Mahasiswa di PTKIN Provinsi Jawa Tengah. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, 7(4), 1003.
<https://doi.org/10.33394/jk.v7i4.3430>
- [17] Nathaniel Chandra, D., Indrawan, G., & Nyoman Sukajaya. (2019). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram. *Jurnal Ilmu Komputer Indonesia (JIKI)*, 4(2), 11–19.
- [18] Pratama, K. A., Pradnyana, G. A., & Arthana, I. K. R. (2020). Pengembangan Sistem Cerdas Untuk Prediksi Daftar Kembali Mahasiswa Baru Dengan Metode Naive Bayes (Studi Kasus: Universitas Pendidikan Ganesha). *Science and Information Technology (SINTECH) Journal*, 3(1), 22–34.
<https://doi.org/10.31598/sintechjournal.v3i1.523>
- [19] Putriku, A. E. (2018). *Prestasi Mahasiswa Jurusan Manajemen Fakultas Ekonomi Stambuk 2014 Universitas Hkbp Nommensen*. 7(1), 50–58.
- [20] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
<https://doi.org/10.31294/ijcit.v5i1.7951>
- [21] Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 1(3), 73–81.
<https://doi.org/10.20895/inista.v1i2.73>
- [22] Sari, C. R. (2016). Teknik Data Mining Menggunakan Classification Dalam Sistem Penunjang Keputusan Peminatan SMA Negeri 1 Polewali. *IJNS – Indonesian Journal on Networking and Security*, 5(1), 48–54.
- [23] Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25.
<https://doi.org/10.36787/jti.v13i1.78>