

Model Algoritma K-Nearest Neighbor Dalam Menganalisis Kinerja Akademik Mahasiswa Program Studi Informatika Universitas Mulawarman

Masna Wati¹, Novianti Puspitasari², Adelowys Sinaga^{*3}

^{1,2,3}Jurusan Teknologi Informasi dan Komunikasi, Universitas Mulawarman, Samarinda
e-mail: ¹masnawati.ssi@gmail.com, ²miechan.novianti.com, ^{*3}adeloywsinaga11@gmail.com

Abstrak

Analisis kinerja akademik pada Program Studi Teknik Informatika Universitas Mulawarman sangat diperlukan untuk mengantisipasi terjadinya drop out (DO) mahasiswa. Penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) untuk menganalisis data akademik mahasiswa menggunakan 8 variabel terdiri dari Jenis Kelamin, Umur, IP Semester 1, IP Semester 2, IP Semester 3, IP Semester 4, IPK hingga Semester 4 dan SKS. Label waktu kelulusan terdiri dari Cepat, Tepat Waktu dan Terlambat. Data yang digunakan adalah data hasil akademik mahasiswa angkatan 2012, 2013 dan 2014 sebanyak 158. Hasil penelitian ini menunjukkan bahwa model algoritma KNN dalam menganalisis hasil akademik mahasiswa pada rasio split data yaitu 70:30 untuk data training dan data testing. Berdasarkan hasil pengujian, diketahui model terbaik memiliki tingkat akurasi sebesar 76,09%, specificity sebesar 78,87%, dan error rate sebesar 23,91%. IP Semester 2 menjadi faktor yang paling berpengaruh dalam label predikat kelulusan mahasiswa.

Kata kunci—Akademik Mahasiswa, Kelulusan, K-Nearest Neighbor

1. PENDAHULUAN

Pendidikan tinggi sebagai bagian dari sistem pendidikan nasional memiliki peran strategis dalam mencerdaskan kehidupan bangsa dan memajukan ilmu pengetahuan dan teknologi. Sistem Pendidikan Indonesia memperhatikan dan menerapkan nilai humaniora serta kebudayaan dan pemberdayaan bangsa Indonesia yang berkelanjutan [1], [2]. Perguruan tinggi merupakan sebuah institusi yang memiliki peran makro dalam pengembangan kualitas sumber daya manusia sebagai penyalur minat individu terhadap bidang tertentu sesuai minat dan bakat serta mengasah individu untuk menerapkan ilmu yang telah didapat demi stabilitas nasional [3].

Program Studi Teknik Informatika Universitas Mulawarman Samarinda merupakan salah satu penyelenggara pendidikan tinggi yang ada di Indonesia. Proses pendidikan di kampus ini juga turut serta dalam pengembangan pendidikan, penelitian dan sumber daya manusia yang bermartabat, inovatif di bidang informatika yang berwawasan wirausaha dan mampu bersaing di tingkat Nasional dan Internasional di pulau Kalimantan. Berhubungan dengan kualitas akademik, program studi berkewajiban untuk memantau perkembangan studi dari mahasiswa untuk mengantisipasi terjadinya *Drop Out* pada mahasiswa [4], [5]. Program studi perlu memberikan peringatan lebih dini kepada mahasiswa agar lulus tepat waktu dan tidak terjadi *Drop Out*. Oleh karena itu perlu dilakukan analisis untuk mengetahui kinerja akademik mahasiswa sehingga mahasiswa dapat diklasifikasikan berpeluang lulus cepat, tepat waktu atau terlambat [6]–[10].

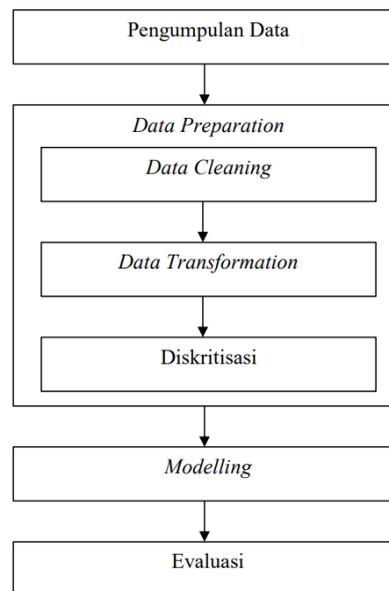
Teknik data mining dapat digunakan dalam menganalisis masalah ini dengan menerapkan *K-Nearest Neighbor* (KNN). Tujuan dari penelitian ini adalah untuk menerapkan metode KNN dalam memprediksi kelulusan mahasiswa agar dapat mengetahui cepat, tepat waktu dan

terlambatnya mahasiswa dalam menyelesaikan perkuliahan pada Program Studi Informatika Universitas Mulawarman.

2. METODE PENELITIAN

2.1 Tahapan Pelaksanaan Penelitian

Penelitian ini dilaksanakan dalam beberapa tahap kegiatan di mulai dari pengumpulan data hingga evaluasi. Seluruh kegiatan yang dilakukan dalam menganalisis hasil kegiatan akademik mahasiswa digambarkan dalam diagram tahapan pelaksanaan penelitian yang tersaji pada Gambar 1.



Gambar 1. Tahapan Pelaksanaan Penelitian

Pada proses pengolahan data awal atau *data preparation* terdiri dari *data cleaning*, *data transformation* dan diskritisasi. Pada tahap *data cleaning* dilakukan pembersihan pada *data raw* dari hasil pengumpulan data terhadap *record* yang memiliki *missing value* atau *record* yang duplikat. Pada tahap *data transformation* dilakukan transformasi terhadap nilai pada beberapa variabel sesuai dengan kebutuhan.

Pada tahap *modeling*, algoritma *K-Nearest Neighbor* digunakan untuk evaluasi kinerja akademik setiap mahasiswa terhadap dataset yang terbentuk pada tahap *data preparation*. Hasil evaluasi kinerja akademik berupa prediksi kelulusan mahasiswa dalam kategori/kelas lulus cepat, lulus tepat waktu atau lulus terlambat. Tahap evaluasi dilakukan untuk mengukur hasil prediksi kelulusan mahasiswa. Parameter evaluasi yang digunakan terdiri atas tingkat akurasi, presisi, *recall*, *specificity* dan *error rate* melalui *Confussion Matrix*.

2.2 Evaluasi Proses Pembelajaran

Penentuan prestasi untuk seorang mahasiswa terhadap penguasaan suatu mata kuliah ditentukan berdasarkan hasil kuis, ujian, praktikum, dan/atau tugas lain yang mencakup ranah kognitif, afektif, dan psikomotorik. Penilaian keberhasilan akademik mahasiswa didasarkan pada nilai bobot rata-rata atau Indeks Prestasi (IP). Indeks Prestasi dibedakan atas Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK). IPS dan IPK dihitung dari mata kuliah yang tertulis dalam kartu rencana studi, kecuali untuk mata kuliah yang termasuk dalam kelompok kegiatan akhir akademik yang belum memiliki nilai. Jumlah sks yang telah lulus dan IPK dipakai

sebagai dasar evaluasi keberhasilan studi seorang mahasiswa Program Vokasi (D3) dan Sarjana (S1). Evaluasi keberhasilan studi mahasiswa pada akhir tahun pertama, akhir tahun kedua, dan akhir tahun ketiga dilakukan oleh fakultas untuk menentukan seorang mahasiswa diperbolehkan melanjutkan studi bila memenuhi syarat [11].

2.3 Data Mining

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Istilah Data Mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Data Mining, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar [12], [13].

Proses data mining berdasarkan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) terdiri dari 6 fase yaitu [14]:

Business Understanding merupakan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan, serta kebutuhan dari perspektif bisnis. Kejadiannya antara lain menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

Data Understanding merupakan fase pengumpulan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai, mengidentifikasi masalah yang berkaitan dengan kualitas data dan mendeteksi subset yang menarik dari data untuk membuat hipotesa awal. *Data Preparation* merupakan fase yang padat karya. Aktivitas yang dilakukan antara lain memilih table dan field yang akan ditransformasikan ke dalam database baru untuk bahan data mining (set data mentah).

Modeling merupakan fase menentukan teknik data mining yang digunakan, menentukan tools data mining, teknik data mining, algoritma data mining dan menentukan parameter dengan nilai yang optimal. *Evaluation* merupakan fase interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama. *Deployment*, fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses data mining.

Tahapan yang dilakukan dalam mengevaluasi hasil akademik mahasiswa menggunakan metode CRISP-DM yang terdiri dari 6 fase berikut:

1. *Business Understanding*

Pada penelitian ini dilakukan pemodelan untuk analisis kinerja akademik mahasiswa Program Studi Informatika Universitas Mulawarman. Tujuan dari pemodelan ini untuk memprediksi mahasiswa akan lulus cepat, tepat waktu atau lulus terlambat. Oleh karena itu, pemodelan yang dilakukan berupa klasifikasi menggunakan metode KNN dengan tiga kelas yaitu lulus cepat, lulus tepat waktu dan lulus terlambat. Dalam penelitian ini terdapat batasan masalah yaitu Data Training dan Data Testing menggunakan set data dan data record dari data diri serta nilai dan lama studi mahasiswa program studi informatika.

2. *Data Understanding*

Data dalam penelitian ini diperoleh dari bagian akademik fakultas di Universitas Mulawarman. Waktu kinerja akademik mahasiswa ini menggunakan data mahasiswa angkatan 2012 hingga angkatan 2014. Sumber data berasal dari buku yudisium Universitas Mulawarman mulai dari bulan Maret 2016 hingga bulan Maret 2018 dengan jumlah data sebanyak 158 mahasiswa.

3. *Data Preparation*

Pada proses pengolahan data awal atau *preparation* dilakukan berbagai proses agar mendapatkan dataset yang dapat diolah sesuai dengan tujuan penelitian.

4. *Modeling*

Algoritma yang digunakan adalah algoritma KNN untuk evaluasi kinerja akademik. Algoritma ini akan mengklasifikasikan mahasiswa ke dalam kelas lulus cepat, lulus tepat waktu atau lulus terlambat berdasarkan variabel yang digunakan dalam mengevaluasi hasil akademik mahasiswa.

5. *Evaluation*

Pada tahap evaluasi dilakukan pengukuran tingkat akurasi pemodelan KNN menggunakan *Confussion Matrix*.

6. *Deployment*

Setelah didapatkan hasil dari pemodelan algoritma KNN, hasil dapat dijadikan strategi baru yang dapat mengetahui bagaimana data mahasiswa yang diprediksi berpotensi lulus tidak sesuai target dan dapat mengantisipasi terjadinya *drop out*.

2.4 *K-Nearest Neighbor (KNN)*

Algoritma KNN merupakan salah metode klasifikasi dalam Data mining yang termasuk dalam algoritma supervised learning. Algoritma supervised learning merupakan algoritma yang bertujuan untuk menemukan pola pada data baru yang didasarkan pada analisis pola dalam Data Training. Komponen fundamental dalam proses klasifikasi yaitu, class untuk merupakan variabel dependen yang menunjukkan label dari suatu objek yang telah terklasifikasi. Predictors merupakan variabel independen yang menunjukkan karakter (atribut) dari data yang telah terklasifikasi. Training Dataset merupakan set data yang mengandung kedua komponen di atas dan digunakan sebagai model Training untuk mengenali class yang cocok berdasarkan predictors yang ada. Testing Dataset mengandung data baru yang akan diklasifikasikan berdasarkan Data Training dan nantinya dari akurasi dari klasifikasi tersebut akan dievaluasi [15], [16].

Algoritma KNN juga merupakan salah satu contoh dari instance-based learning, yang mana Data Training yang telah terklasifikasi digunakan sebagai pembanding dalam menemukan persamaan pada Data Testing dengan Data Training yang ada. Langkah pertama dalam metode KNN yaitu menentukan nilai k pada Data Training. Bila $k = 5$ maka 5 record dari Data Training tersebut diambil berdasarkan jarak yang paling dekat dari record baru. Kemudian record baru tersebut dibandingkan berdasarkan dari kelima record dari Data Training yang memiliki jarak yang paling mendekati dengan record baru tersebut. Model KNN jika digambarkan dalam bentuk diagram alur dijelaskan pada Gambar 2 [17], [18].



Gambar 2. Flowchart algoritma K-Nearest Neighbor

2.5 Confussion Matrix

Dalam algoritma klasifikasi penting untuk mengetahui apakah objek yang terklasifikasi tersebut bernilai benar atau salah. Metode *Confussion Matrix* adalah metode yang berbentuk tabel matriks digunakan untuk mengevaluasi adanya kesalahan dalam klasifikasi yang terjadi pada setiap objek. *Confussion Matrix* mengevaluasi seberapa akurat metode klasifikasi tersebut dalam memprediksikan *record* pada kelasnya [19], [20].

3. HASIL DAN PEMBAHASAN

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar.

Label kelas yang digunakan dalam penelitian ini adalah cepat, tepat dan terlambatnya mahasiswa dalam mencapai kelulusan. Atribut yang digunakan dalam menganalisis hasil kinerja akademik mahasiswa disajikan pada Tabel 1.

Tabel 1. Atribut Kinerja Akademik Mahasiswa

No	Atribut	Kode Atribut	Kategori	Skala Data
1	Jenis Kelamin	JK	Pria, Wanita	Nominal
2	Umur	U	Umur Genap Mahasiswa	Ordinal
3	IP Semester 1	IPS1		Ordinal
4	IP Semester 2	IPS2	1 (Indeks Prestasi ≤ 1.5)	Ordinal
5	IP Semester 3	IPS3	2 ($1.5 < \text{Indeks Prestasi} \leq 2.5$)	Ordinal
6	IP Semester 4	IPS4	3 ($2.5 < \text{Indeks Prestasi} \leq 3.5$)	Ordinal
7	IPK Hingga Semester 4	IPK	4 ($3.5 < \text{Indeks Prestasi} \leq 4.0$)	Ordinal
8	Jumlah SKS	SKS	SKS < 96 SKS ≥ 96	Ordinal

Penelitian ini menggunakan data akademik mahasiswa program studi Informatika Universitas Mulawarman angkatan 2012 hingga angkatan 2014 sebanyak 158 mahasiswa. Sebelum dilakukan pemodelan dalam mengklasifikasi kinerja akademik mahasiswa, data *raw* perlu dipastikan bersih dari gangguan sehingga dilakukan pembersihan data dan transformasi data. Setelah seluruh data diperiksa, tidak ditemukan data duplikat, *missing value* atau data yang tidak logis sehingga *record* dari variabel yang digunakan telah sesuai dengan data yang diperlukan.

Tahap transformasi data dilakukan dengan aturan-aturan sebagai berikut:

- a. Untuk setiap jenis kelamin laki-laki akan disimbolkan dengan “L” dan “P” untuk setiap jenis kelamin perempuan.
- b. Untuk setiap umur akan di kelompokkan berdasarkan umur mahasiswa saat mahasiswa tersebut terdaftar.
- c. Untuk setiap indek prestasi semester (IPS) di adaptasi dari predikat nilai bobot berdasarkan peraturan akademik Universitas Mulawarman tahun 2017 dengan range sebagai berikut:
 - 1) Untuk setiap Indeks prestasi semester mulai dari 0.0 sampai dengan 1.5 akan di bulatkan nilai bobotnya menjadi 1
 - 2) Untuk setiap Indeks prestasi semester mulai dari 1.6 sampai dengan 2.5 akan di bulatkan nilai bobotnya menjadi 2
 - 3) Untuk setiap Indeks prestasi semester mulai dari 2.6 sampai dengan 3.5 akan di bulatkan nilai bobotnya menjadi 3

- 4) Untuk setiap Indeks prestasi semester mulai dari 3.6 sampai dengan 4.0 akan di bulatkan nilai bobotnya menjadi 4
- d. Untuk lama masa studi yang ditempuh dengan range sebagai berikut:
 - 1) Lama masa studi diatas 4 tahun 5 bulan diberikan predikat “Terlambat”
 - 2) Lama masa studi dari 4 tahun 1 bulan sampai 4 tahun 5 bulan diberikan predikat “Tepat Waktu”
 - 3) Lama masa studi kurang dari 4 tahun diberikan predikat “Cepat”

Dari hasil tahap *data preparation* diperoleh dataset yang siap digunakan dalam proses *modeling* yang tersaji pada Tabel 2.

Tabel 2. Hasil Kinerja Akademik Mahasiswa Informatika Angkatan 2012-2014

No	Mahasiswa	JK	UMUR	IPS1	IPS2	IPS3	IPS4	SKS	Predikat
1	Mhs1	1	18	4	4	4	4	144	Cepat
2	Mhs2	1	15	3	3	4	4	146	Terlambat
3	Mhs3	2	17	3	4	4	4	148	Cepat
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
158	Mhs158	2	17	4	4	4	4	146	Cepat

Berdasarkan Tabel 2, dapat diketahui variabel mana yang paling berpengaruh terhadap sebuah model data. Hal ini dapat dilihat pada nilai koefisien dari masing-masing variabel yang diperoleh dari hasil pengolahan data menggunakan Aplikasi *IBM SPSS Statistic 23*. Nilai koefisien setiap variabel yang telah diperoleh ditunjukkan pada Gambar 3.

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.429	3.338		-.428	.669
	JENIS KELAMIN	-.151	.087	-.128	-1.726	.086
	UMUR SAAT MENDAFTAR	-.030	.052	-.041	-.567	.572
	IPS 1	-.136	.114	-.126	-1.199	.232
	IPS 2	-.211	.121	-.182	-1.742	.084
	IPS 3	-.059	.097	-.053	-.613	.541
	IPS 4	-.035	.096	-.028	-.358	.721
	IPK HINGGA SMT 4	-.162	.145	-.138	-1.117	.266
	JUMLAH SKS	.040	.021	.138	1.866	.064

Gambar 3. Nilai Koefisien Variabel

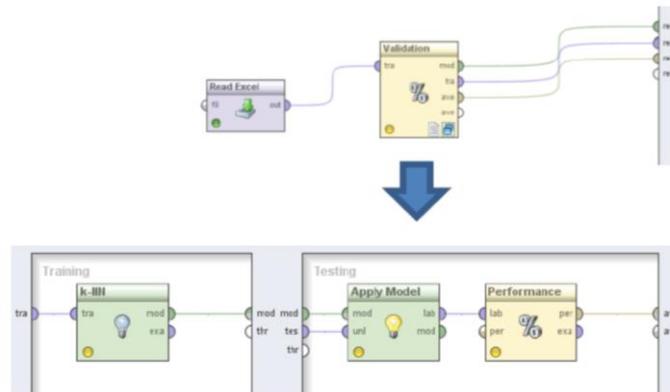
Dari Gambar 3 diketahui bahwa nilai koefisien masing-masing variabel yaitu, Jenis Kelamin = -0,151, umur = -0,030, ips1 = -0,136, ips2 = -0,211, ips3 = -0,059, ips4 = -0,035, ipk4 = -0,162, dan jumlah sks = 0,40. Variabel yang paling berpengaruh adalah variabel yang memiliki nilai koefisien yang paling menjauhi nilai Nol. Oleh karena itu, variabel yang paling berpengaruh dalam penelitian ini adalah Indeks Prestasi Semester 2 dengan nilai koefisien = -0,211

Pada split data dengan *ratio* 10% hingga 90% sebagai *Data Training* untuk diolah dalam pemodelan. *Data Training* diperoleh dengan mengambil data sesuai urutan dimulai dari *record* pertama (*Linear Sampling*) sehingga tidak perlu mengacak urutan datanya kembali. *Data Testing* dimulai dari nomor urut selanjutnya hingga ke data paling bawah.

Metode KNN diterapkan pada tahap *modeling*. Pada tahap ini, dataset pada Tabel 2 yang telah dibagi menjadi Data Training dan Data Testing dalam beberapa skenario *ratio* digunakan untuk menganalisis model klasifikasi KNN terhadap hasil kinerja mahasiswa. Langkah-langkah penerapan metode KNN dalam menganalisis hasil kinerja akademik mahasiswa sebagai berikut:

1. Menentukan nilai K. Dalam penelitian ini digunakan nilai K=5 dan Euclidean Distance terdekat dari Data Training.
2. Menghitung jarak Euclidean Distance untuk mendapatkan nilai K.
3. Memilih nilai K dengan jarak terdekat dari Data Training ke Data Testing
4. Mengurutkan nilai K dari yang terendah
5. Menentukan label berdasarkan mayoritas yang telah didapat.

Dengan menggunakan bantuan software Rapid Miner, desain pemodelan algoritma K-NN dapat dilihat pada Gambar 4.



Gambar 4. Ilustrasi Model Split Validation

Dari Gambar 4 diketahui beberapa fungsi yaitu:

- a. *Apply Model* : Untuk menerapkan model yang di peroleh dari Data Training ke data set yang tidak berlabel. Output-nya untuk melihat tampilan hasil yang menunjukkan rumus regresi linier dengan koefisien yang diperoleh dari proses Data Training.
- b. *Performance* : Daftar nilai kriteria kinerja. output-nya berupa kriteria kinerja yang telah dihitung oleh operator performance (perhitungan kinerja vector)
- c. *Validation* : Untuk melakukan validasi sederhana yaitu secara acak membagi example set menjadi satu Data Training dan mengetest Data Testing serta mengevaluasi model. dalam hal ini menggunakan Split Validation, Split 40 Validation adalah metode untuk memperkirakan seberapa akurat suatu model yang digunakan.
- d. *Read excel* : Operator ini untuk memuat atau membaca data dari ms. excel yang berbentuk table.
- e. *K-NN* : Operator ini menghasilkan model K-Nearest Neighbor yang digunakan untuk klasifikasi atau regresi.

Tahap evaluasi menggunakan bantuan *Confussion Matrix* dengan label cepat, tepat waktu dan terlambat untuk menguji model klasifikasi yang dibentuk oleh metode KNN. Hasil dari evaluasi model algoritma KNN dalam menganalisis kinerja akademik mahasiswa yang telah dilakukan dapat dilihat pada Tabel 3.

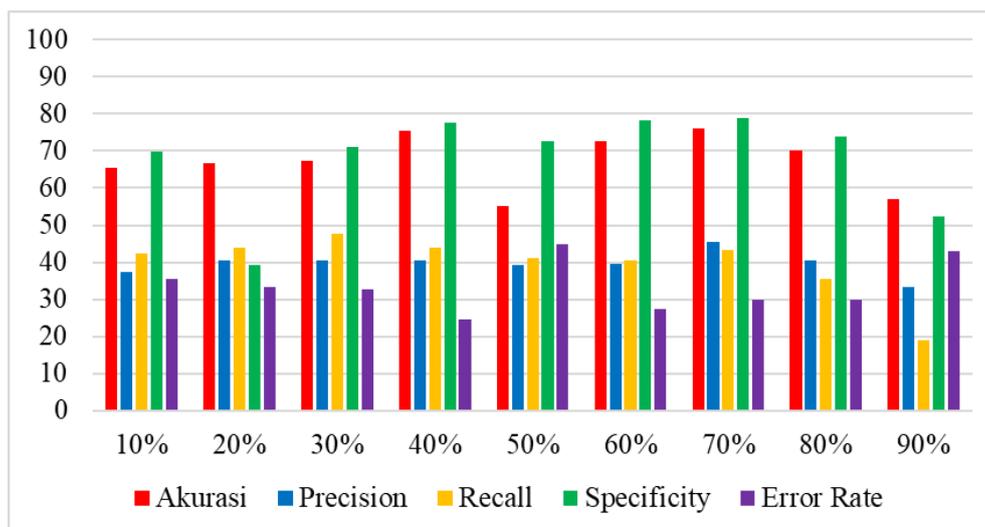
Tabel 3. Evaluasi Pemodelan Algoritma KNN

Total Data	Evaluasi Algoritma							
	Rasio (%)	Jumlah Data Training	Jumlah Data Testing	Akurasi (%)	Precision (%)	Recall (%)	Specificity (%)	Error Rate (%)
158 Data	10%	16	142	65,49%	37,51%	42,47%	69,94%	34,51%
	20%	32	126	66,70%	40,42%	43,98%	39,33%	33,30%
	30%	48	110	67,30%	40,42%	47,62%	70,93%	32,70%
	40%	64	94	75,53%	40,42%	43,98%	77,60%	24,47%
	50%	80	78	55,26%	39,31%	41,13%	72,64%	44,74%

Total Data	Evaluasi Algoritma							
	Rasio (%)	Jumlah Data Training	Jumlah Data Testing	Akurasi (%)	Precision (%)	Recall (%)	Specificity (%)	Error Rate (%)
	60%	96	62	72,58%	39,48%	40,55%	78,08%	27.42%
	*70%	112	46	76,09%	45,57%	43,30%	78,87%	23,91%
	80%	128	30	70,00%	40,47%	35,45%	73,80%	30,00%
	90%	144	14	57,14%	33,33%	19,04%	52,38%	42.86%

Berdasarkan Tabel 3, hasil akurasi, *precision*, *recall*, *specificity* dan *error rate* dari algoritma *K-Nearest Neighbor* untuk menganalisa kinerja akademik mahasiswa berdasarkan masing-masing pembagian *Data Training* yang dimulai dari 10% hingga 90% didapat akurasi dan *precision* tertinggi pada rasio 70%, dimana akurasi menunjukkan angka 76,09%, *precision* 45,57%, *specificity* 78,87% dan *error rate* terendah dengan 23,91% sedangkan *recall* tertinggi didapat pada rasio 30% dengan jumlah 47,62%. Hal ini menandakan bahwa model algoritma KNN dengan hasil terbaik dalam menganalisa kinerja akademik mahasiswa adalah pada rasio 70%.

Hasil evaluasi pemodelan algoritma KNN pada Tabel 3 dapat disajikan dalam bentuk diagram grafik agar mudah dalam melihat tinggi rendahnya persentase yang didapat. Diagram grafik tersebut dapat dilihat pada Gambar 5.



Gambar 5. Diagram Hasil *Confusion Matrix* Metode KNN

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

1. Model klasifikasi terbaik dalam menganalisis kinerja akademik mahasiswa untuk mengklasifikasikan mahasiswa akan lulus Cepat, Tepat Waktu atau Terlambat menggunakan Algoritma KNN dengan Euclidian Distance, nilai K=5 dan split data pada ratio 70% Data Training.
2. Hasil evaluasi model KNN terbaik yaitu pada split 70:30 dengan 112 Data Training dan 46 Data Testing diperoleh nilai akurasi sebesar 76,09%, *precision* sebesar 45,57%, *recall* sebesar 43,30%, *specificity* sebesar 78,87% dan *error rate* sebesar 23,91%.

3. Berdasarkan hasil evaluasi koefisien variabel diketahui bahwa IP Semester 2 menjadi faktor yang paling berpengaruh dalam kinerja akademik mahasiswa.

5. SARAN

Dari hasil analisis yang dilakukan maka perlu diteliti lebih lanjut tentang variabel atau atribut lain yang memiliki pengaruh dalam waktu kelulusan mahasiswa Program Studi Informatika Universitas Mulawarman.

DAFTAR PUSTAKA

- [1] E. Simatupang and I. Yuhertiana, "Merdeka belajar kampus merdeka terhadap perubahan paradigma pembelajaran pada pendidikan tinggi: Sebuah tinjauan literatur," *Jurnal Bisnis, Manajemen, Dan Ekonomi*, vol. 2, no. 2, pp. 30–38, 2021.
- [2] A. Widiensyah, "Peranan sumber daya pendidikan sebagai faktor penentu dalam manajemen sistem pendidikan," *Cakrawala-Jurnal Humaniora*, vol. 18, no. 2, pp. 229–234, 2018.
- [3] B. Lian, "Tanggung jawab Tridharma perguruan tinggi menjawab kebutuhan masyarakat," in *Prosiding Seminar Nasional Program Pascasarjana Universitas PGRI Palembang*, 2019, pp. 100–106.
- [4] N. Nurjayadi, K. Andesa, T. Nasution, and H. Herwin, "Peringatan Dini Masa Studi Dan Ipk Mahasiswa Berbasis Framework for Application of System Thinking," *JURNAL TEKNOLOGI DAN OPEN SOURCE*, vol. 3, no. 2, pp. 261–271, 2020.
- [5] J. O. Isakh, "Sistem Informasi Pemantauan Perkembangan Mahasiswa yang Memiliki Kendala dalam Proses Belajar Mengajar," *Jurnal STRATEGI-Jurnal Maranatha*, vol. 2, no. 2, pp. 581–591, 2020.
- [6] M. Wati, Haeruddin, and W. Indrawan, "Predicting degree-completion time with data mining," *Proceeding - 2017 3rd International Conference on Science in Information Technology: Theory and Application of IT for Education, Industry and Society in Big Data Era, ICSITech 2017*, vol. 2018-Janua, pp. 732–736, 2018, doi: 10.1109/ICSITech.2017.8257209.
- [7] M. Wati, H. S. Pakpahan, and N. Novirasari, "Comparative Analysis of Multi-Criteria Decision Making for Student Degree Completion Time based on Entropy Weighted," *Proceedings of ICAITI 2018 - 1st International Conference on Applied Information Technology and Innovation: Toward A New Paradigm for the Design of Assistive Technology in Smart Home Care*, pp. 56–61, 2019, doi: 10.1109/ICAITI.2018.8686746.
- [8] M. Wati, N. Novirasari, and H. S. Pakpahan, "Evaluation of scholarly performance student using multi-criteria decision-making with objective weight," *International Electronics Symposium on Knowledge Creation and Intelligent Computing, IES-KCIC 2018 - Proceedings*, pp. 56–61, 2019, doi: 10.1109/KCIC.2018.8628490.
- [9] M. Wati, N. Novirasari, E. Budiman, and Haeruddin, "Multi-Criteria Decision-Making for Evaluation of Student Academic Performance Based on Objective Weights," in *the Third International Conference on Informatics and Computer*, 2019, no. 11, pp. 1–5. doi: 10.1109/iac.2018.8780421.
- [10] M. Wati, W. Indrawan, J. A. Widians, and N. Puspitasari, "Data Mining For Predicting Students ' Learning Result," in *2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT)*, 2017, pp. 1-4. doi: 10.1109/CAIPT.2017.8320666.
- [11] Senat Universitas Mulawarman, *Peraturan Akademik Universitas Mulawarman*. 2020.
- [12] D. Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Indonesia: Deepublisher, 2014.

-
- [13] T. R. Shultz and S. E. Fahlman, *Encyclopedia of Machine Learning and Data Mining*. 2017. doi: 10.1007/978-1-4899-7687-1.
- [14] M. Tiwari, R. Dixit, and A. Kesharwani, *Data Mining Principles, Process Model and Applications*. India: Educreation Publishing, 2111.
- [15] J. Angelina Widians, H. Santoso Pakpahan, E. Budiman, and M. Soleha, “Klasifikasi Jenis Bawang Menggunakan Metode K-Nearest Neighbor Berdasarkan Ekstraksi Fitur Bentuk dan Tekstur,” *JURTI*, vol. 3, no. 2, 2019.
- [16] R. R. Waliyansyah and C. Fitriyah, “Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN),” *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 2, pp. 157–163, Aug. 2019, Accessed: Jan. 20, 2022. [Online]. Available: <https://jurnal.untan.ac.id/index.php/jepin/article/view/32473>
- [17] A. A. WPR, F. Rozi, and F. Sukmana, “Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor,” *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 1, pp. 155–160, 2021.
- [18] A. R. D. Nugraha, K. Auliasari, and Y. A. Pranoto, “Implementasi Metode K-Nearest Neighbor (KNN) Untuk Seleksi Calon Karyawan Baru (Studi Kasus: BFI Finance Surabaya),” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 2, pp. 14–20, 2020.
- [19] E. Budiman, Havaluddin, N. Dengan, A. H. Kridalaksana, M. Wati, and Purnawansyah, *Performance of Decision Tree C4.5 Algorithm in Student Academic Evaluation*, vol. 488. 2018. doi: 10.1007/978-981-10-8276-4_36.
- [20] M. Wati, H. S. Pakpahan, A. Prafanto, F. Akbar, Havaluddin, and A. W. D. Boernama, “Application of C4.5 Classification Algorithm for Chronic Kidney Disease Diagnosis,” in *ICEEIE 2019 - International Conference on Electrical, Electronics and Information Engineering: Emerging Innovative Technology for Sustainable Future*, 2019. doi: 10.1109/ICEEIE47180.2019.8981458.
-