

PREDIKSI KELULUSAN TEPAT WAKTU BERDASARKAN RIWAYAT AKADEMIK MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*

Imam Riadi¹, Rusydi Umar², Rio Anggara^{*3}

^{1,2,3}Universitas Ahmad Dahlan, Yogyakarta

Email: ¹imam.riadi@is.uad.ac.id, ²rusydi_umar@rocketmail.com, ³anggarariocopasga@gmail.com

*Penulis Korespondensi

(Naskah masuk: 01 Juni 2024, diterima untuk diterbitkan: 02 April 2024)

Abstrak

Lulus tepat waktu adalah pencapaian paripurna ingin dicapai oleh semua mahasiswa atau calon mahasiswa. Kelulusan memiliki 2 klasifikasi seperti lulus tidak tepat waktu dan lulus tepat waktu. Kelulusan menjadi suatu penilaian akreditasi universitas dan penilaian oleh masyarakat secara luas. Perguruan tinggi meluluskan mahasiswa-mahasiswa dengan beberapa kriteria standar yang harus dimiliki. Diharapkan mahasiswa lulus memenuhi syarat standar kelulusan dalam waktu maksimal 4 tahun masa studi. Evaluasi dan pemantauan kelulusan sangat penting dilakukan, salah satunya dengan mempelajari data *history* mahasiswa yang telah lulus sebagai upaya mahasiswa lulus tidak melebihi waktu standar yang telah ditetapkan. Prediksi kelulusan yang dilakukan pada riset menggunakan kaidah klasifikasi *K-Nearest Neighbor* dengan objek penelitian yaitu mahasiswa. Atribut yang dipakai dalam penelitian metode klasifikasi yaitu nama, asal SMA/SMK, wilayah asal SMA/SMK, nilai rata-rata matematika dan lama studi. Tahapan penelitian diawali dengan pengumpulan data, pemilihan atribut, pembersihan data, transformasi data, pemilihan data testing dan data training. Pengujian akurasi yang didapatkan pada penelitian metode klasifikasi dengan klaster data $k=1$, $k=2$, $k=3$, $k=4$, $k=5$, $k=6$ dan $k=7$ menghasilkan klaster dengan nilai $k=3$ paling tinggi. Hasil pengujian akurasi prediksi penelitian menggunakan *confusion matrix* menghasilkan akurasi paling besar sesuai target yaitu mencapai 78% menggunakan objek penelitian sebanyak 93 data mahasiswa terdiri dari 78 training dan 12 data testing. Hasil pengujian *point* $k=1$ sampai *point* $k=7$, $k=3$ merupakan nilai akurasi prediksi yang paling tinggi sehingga hasil penelitian menjadi sumber pengetahuan untuk fakultas dalam prediksi kelulusan mahasiswa.

Kata Kunci : Prediksi Kelulusan, *Data Mining*, Klasifikasi, *K-Nearest Neighbor*, *Confusion Matrix*.

PREDICTED TIMELY GRADUATION BASED ON ACADEMIC HISTORY BEFORE COLLEGE WITH *K-NEAREST NEIGHBOR* METHOD

Abstract

*Graduating on time is a plenary achievement to be achieved by all students or prospective students. Graduation has 2 classifications such as graduating not on time and graduating on time. Graduation becomes an assessment of university accreditation and an assessment by the wider community. Universities graduate students with several standard criteria that must be possessed. It is expected that graduating students meet the graduation standard requirements within a maximum of 4 years of study period. Evaluation and monitoring of graduation is very important to do, one of which is by studying the history data of students who have graduated as an effort for students to graduate not to exceed the standard time that has been set. The graduation predictions carried out in research use the *K-Nearest Neighbor* classification rules with the research object being students. The attributes used in the research classification method are name, high school/vocational high school origin, high school/vocational high school origin, average grade in mathematics and length of study. The research phase begins with data collection, attribute selection, data cleaning, data transformation, selection of testing data and training data. The accuracy test obtained in the classification method research with data clusters $k = 1$, $k = 2$, $k = 3$, $k = 4$, $k = 5$, $k = 6$ and $k = 7$ produces a cluster with the highest $k = 3$ value. The results of testing the accuracy of research predictions using the *confusion matrix* produced the greatest accuracy according to the target, reaching 78% using a research object of 93 student data consisting of 78 training and 12 testing data. The test results *point* $k=1$ to *point* $k=7$, $k=3$ is the highest prediction accuracy value so that the research results become a source of knowledge for the faculty in predicting student graduation.*

Keywords: Graduation Prediction, *Data Mining*, Classification, *K-Nearest Neighbor*, *Confusion Matrix*.

1. PENDAHULUAN

Pendidikan yang berkualitas pada suatu negara berawal dari kualitas pendidikan masing-masing daerah. Kualitas pendidikan masing-masing daerah di negara Indonesia belum cukup merata dan belum maksimal (Gibran Satya Nugraha, Hairani, 2018). Kualitas pendidikan menjadi kurang bagus bisa terjadi karena terletak di daerah yang jaraknya jauh dari pusat kota (Khasanah, Salim, Afni, Komarudin, & Maulana, 2022). Hal ini selaras dengan perhatian dari pemerintah daerah yang kurang perhatian pada daerah-daerah yang jauh dari pandangan kota. Kualitas Pendidikan ini bukan distribusi nasional dan subsidi Pendidikan oleh pemerintah pusat, namun menjadi tanggung jawab dan anggaran dari pemerintah daerah (Ramadhayanti, 2018). Sehingga menimbulkan tidak meratanya tingkat baik buruknya pendidikan satu wilayah dengan wilayah lain yang berdekatan dengan kota di setiap daerah (Nurhanifa, 2020). Sebagai bahan penelitian yang salah satu variabel menggunakan Pendidikan tingkat atas, dan untuk mengatasi kualitas pendidikan yang tidak sama. Dilakukan proses klasterisasi kualitas pendidikan di setiap provinsi di Indonesia untuk menunjukkan daerah mana saja yang memiliki kualitas pendidikan tinggi (Gede Aditra Pradnyana, 2018). Hal ini dilakukan untuk kualitas pendidikan rendah tidak tergabung dengan kualitas pendidikan yang tinggi. Klasterisasi menjadi upaya dalam keadilan pengolahan data penelitian antara daerah yang tinggi kualitas pendidikan dan daerah yang rendah kualitas Pendidikan.

Universitas adalah lembaga pendidikan formal yang mencetak mahasiswa-mahasiswa yang kreatif, mandiri dan inovatif sehingga dapat berguna bagi masyarakat, agama dan negara (Pamulang, Surya, No, Barat, & Selatan, 2021). Biasanya perguruan tinggi memiliki ribuan mahasiswa yang menempuh pendidikan. Mahasiswa atau akademisi ini tentunya dapat mengembangkan, menerapkan dan menemukan pengetahuan baru (Heryana et al., 2020). Mahasiswa memiliki kemampuan atau skill sesuai sektor yang diminati dalam penerapan kedalam masyarakat setelah lulus (Ahmad R. Pratama, 2022). Sehingga kelulusan sangat diharapkan oleh mahasiswa dan masyarakat supaya kemampuan yang dimiliki bisa berkembang, berguna dan merubah kehidupan sekitar dan masyarakat (Novendra Adisaputra Sinaga, B. Herawan Hayadi, 2022). Tingginya kepercayaan masyarakat untuk mensekolahkan anak-anak ke perguruan tinggi tidak imbang dengan total banyaknya berdiri perguruan tinggi. Sehingga perguruan tinggi berlomba dalam mencetak lulusan yang berkualitas supaya mendapat eksistensi dimata masyarakat, dan lulusan yang berkualitas tentunya menjadi nilai lebih dalam persaingan perguruan tinggi. Tingginya persaingan perguruan tinggi sangat mempengaruhi sedikit dan banyaknya dalam menampung mahasiswa (Wiliani, 2020). Perguruan tinggi yang tidak mendapatkan kepercayaan di

masyarakat hanya akan memiliki peserta didik sedikit, sedangkan perguruan tinggi yang eksis dimasyarakat akan mendapat peserta didik yang tidak sedikit.

Setiap status mahasiswa harus melewati masa sekolah menengah atas atau sederajat. Karena sekolah menengah atas merupakan syarat yang wajib dimiliki sebagai kelengkapan administrasi untuk mendaftar perguruan tinggi. Diketahui bahwa tidak semua SMA atau SMK sederajat tidak memiliki kualitas pendidikan yang sama (Riyan Latifahul Hasanah, 2019). Dengan adanya data mahasiswa dan akademik sebelum kuliah dapat menjadi acuan pengelompokan kualitas pendidikan (Sumarlin, 2018). Proses pengolahan dan pengelompokan data, diharapkan menjadi sebuah pengetahuan atau informasi setiap mahasiswa untuk lulus kuliah tepat waktu (Wisdayani, Nur, & Wasono, 2019). Kelulusan mahasiswa terbagi menjadi 2 klasifikasi yaitu kluster lulus tepat waktu dan kluster lulus tidak tepat waktu (Naufal, 2017). Kuliah lulus tepat waktu setiap mahasiswa menjadi tugas perguruan tinggi. Apabila banyak mahasiswa yang tidak lulus tepat waktu dapat menjadi ancaman bagi perguruan tinggi, ancaman bisa berasal dari wali maupun pemerintah. Dapat dipastikan perguruan tinggi tersebut tidak akan bertahan lama beroperasi.

Kelulusan mahasiswa telah diteliti pada tahun 2019 dengan metode yang sama yaitu *K-Nearest Neighbor* menggunakan objek mahasiswa namun dengan data yang berbeda. Prediksi kelulusan pada penelitian tahun ini menghasilkan 5 tingkatan akurasi yang di tampilkan ke dalam *ROC Curve*. Menghasilkan validasi dengan akurasi 0,90-1,00 sampai akurasi 0,50-0,60 (Muliono, 2019).

Banyak metode penelitian dalam prediksi kelulusan mahasiswa seperti penelitian tahun 2020 perbandingan 3 metode sekaligus yaitu *k-Nearest Neighbor*, *Naïve Bayes* dan *Decision Tree* (Hozairi, 2021). Metode *K-Nearest Neighbor* sangat bisa diprioritaskan dalam prediksi atau klasifikasi seperti hasil penelitian tahun 2020 kriteria skor kualifikasi Ujian Nasional (UN) tingkat Sekolah Menengah Pertama (SMP) dengan variabel yang digunakan dalam prediksi dan klasifikasi ini adalah nilai *tryout* sedangkan atribut siswa yang digunakan adalah NIS, Nama, IPA1 dan IPA2. Menggunakan nilai ujian siswa selama 3 tahun terakhir ada 744 data. Setelah data diolah, menghasilkan akurasi rata-rata sebesar 88,42%. Dengan menggunakan *k-value* yaitu 7, akurasi *mean* 96,26%, akurasi 96,17%, dan *recall* 97,32% (Sri Mulyati, Syepri Maulana Husein, 2020).

Berdasarkan hasil studi tahun 2019, dihitung nilai akurasi menggunakan *Confusion Matrix* yang menyimpulkan bahwa kaidah *k-nearest neighbor* (k-NN) menggunakan dasar PSO mengungguli algoritma *Decision Tree* dalam mengklasifikasikan pekerjaan lulusan. (C4.5) dengan nilai akurasi 76,69%. Peningkatan nilai akurasi kaidah k-NN (= 6.88%) meningkat dibandingkan dengan kaidah

Decision Tree (= 5.80%). Nilai akurasi sebelum dan sesudah penambahan fitur PSO lebih tinggi pada algoritma k-NN. Atribut Indeks Prestasi Semester (IPS) 6 memiliki bobot tertinggi baik untuk tahapan kaidah *Decision Tree* (C4.5) maupun Algoritma k-NN (Moh. Zainuddin, 2019).

Penelitian Universitas Amikom (2018) menggunakan rumus kaidah *k-nearest neighbour* untuk menguji kuantitas dari klasifikasi dengan k=1 sampai k=20 dan *k-fold cross validation* 2 sampai 10. Berdasarkan percobaan presisi, akurasi dan *recall* metode k-NN, pengujian dengan *k-value* hingga 1 sampai 20 dan pengujian dengan *k-fold* hingga 2-10 dengan nilai tertinggi presisi, dan *recall*. Pada percobaan *k-fold=5* dengan k=14 dan k=18, maka dilakukan pengamatan tambahan antara k = 14 dan k = 18 yang seringkali memiliki nilai terbesar diantara kedua nilai k tersebut, yaitu k = 14. Dengan demikian, telah terlihat k metode k-NN paling tinggi untuk memprediksi kelulusan mahasiswa dengan rata-rata nilai semester adalah k = 14 dan *k-fold crossvalidation* dengan presisi 5 = 98,6%, presisi = 99,53, *recall* = 97,6 % (Susanto, Kusri, & Fatta, 2018).

Penelitian tahun sebelumnya memiliki maksud menerapkan kaidah k-NN dengan *Euclidean* dan k-NN (Wiyana Yusuf, Rahma Witri, 2022) dengan *Manhattan* untuk mengklasifikasi akurasi kelulusan. Kelulusan mahasiswa ditentukan oleh variabel jenis kelamin, jurusan, jumlah SKS semester I, jumlah SKS semester II, jumlah SKS semester III, IP semester I, IP semester II, IP semester semester ketiga, dan usia. Variabel-variabel tersebut menentukan ketepatan kelulusan mahasiswa, tepat waktu maupun tidak tepat waktu. Penerapan kaidah k-NN dilakukan dengan menggunakan perangkat lunak *Rapidminer*. Hasil diperoleh setelah dilakukan pengujian terhadap 380 data latih dan 163 data uji. Sistem akurasi tertinggi dicapai dengan nilai K=7 hasil perhitungan 85.28% (Hidayati & Hermawan, 2021).

2. METODE PENELITIAN

2.1 Dataset Penelitian

Pada riset yang dilakukan menggunakan dataset sumber dari Fakultas Teknologi Industri (FTI) UAD, yaitu mahasiswa angkatan 2014-2015. Data penelitian terdiri dari gabungan antara data sebelum kuliah dan data pada masa kuliah. Data sebelum kuliah memuat atribut mulai dari asal sma atau smk mahasiswa, asal daerah sma atau smk, dan nilai matematika raport. Data sebelum kuliah dirasa sangat mempengaruhi prediksi kelulusan setiap mahasiswa, karena kualitas dan lingkungan pendidikan yang baik akan mencetak mahasiswa rajin dan berprestasi.

Nilai matematika raport dikumpulkan dari data Penerimaan Mahasiswa Didik (PMDK) (Rathore & Arjaria, 2019) saat mahasiswa mendaftar ke UAD. Data penelitian terdiri dari sebelum kuliah dan data pada masa perkuliahan yang digunakan dalam

penelitian (Rustam & Annur, 2019). Data berupa atribut NIM mahasiswa, nama mahasiswa, prodi mahasiswa, nilai IPK, nilai TOEFL dan lama studi.

Total data penelitian merupakan gabungan dari data sebelum kuliah dan data kuliah. Atribut data penelitian meliputi NIM mahasiswa, nama mahasiswa, prodi mahasiswa, nama SLTA, Provinsi SLTA, nilai rata matematika, nilai IPK, nilai TOEFL dan lama studi. *Dataset* dari atribut diatas dapat dilihat pada Tabel 1 dan Tabel 2 *Dataset* Penelitian.

Tabel 1. *Dataset* Penelitian

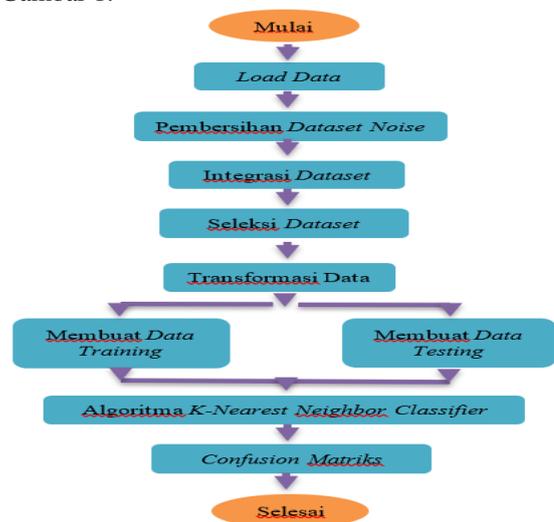
No	NIM	Nama	Nama SLTA	Provinsi
1	1400018012	Mackands Leonardo Oktano	SMA Gajah Mada, Yogyakarta	DI Yogyakarta
2	1400018016	Muhammad Rasyid Ridho	MA Negeri 3, Yogyakarta	DI Yogyakarta
3	1400018017	Dedy Saputra	SMA Negeri 1, Babakan	Jawa Barat
...
92	1500018310	Faiz Isnan Abdurrahman	MA KH Mustofa, Sukamanah	Jawa Barat

Tabel 2. *Dataset* Penelitian

No	NIM	Nama	Rata matematika	Lama Studi
1	1400018012	Mackands Leonardo Oktano	83	3 Th,11 Bln,30 Hr
2	1400018016	Muhammad Rasyid Ridho	78	3 Th,11 Bln,30 Hr
r3	1400018017	Dedy Saputra	89	4 Th,1 Bln,23 Hr
...
92	1500018310	Faiz Isnan Abdurrahman	82	3 Th,11 Bln,14 Hr

2.2 Tahapan Penelitian

Pada point ini, penjabaran tahapan penelitian dalam sistem prediksi kelulusan ditampilkan pada Gambar 1.



Gambar 1. Tahapan sistem prediksi

Terlihat Gambar 1 merupakan tahapan alur sistem prediksi kelulusan dimulai dari *load data*, yang merupakan pembukaan dataset yang telah dimasukkan kedalam sistem. Pada pembersihan *dataset noise* merupakan tahapan penghapusan data mahasiswa apabila data mahasiswa tidak lengkap pada atribut yang digunakan (M. Reza Noviansyah,

Tedy Rismawan, 2018). Integrasi *dataset* merupakan langkah penelitian yang menggabungkan data mahasiswa dari beberapa *file* menjadi satu *file* seperti data riwayat sebelum kuliah dan data setelah kuliah sesuai nama mahasiswa (Hermawan, 2018). Seleksi *dataset* merupakan tahapan seleksi atribut yang akan digunakan dalam penelitian, dari banyaknya atribut yang terkumpul dalam satu file hanya akan diambil atribut NIM, nama, nama SLTA, Provinsi SLTA, nilai rata-rata matematika dan lama studi (Kholil, 2018).

Transformasi data mahasiswa merupakan tahapan merubah data mahasiswa dari awalnya berbentuk string menjadi numerik (Allan, 2020). Perubahan data ini dilakukan pada atribut nama SLTA, Provinsi SLTA dan atribut lama studi. Pada atribut nama SLTA masing-masing mahasiswa akan dikategorikan menjadi 3 kolom terdiri dari kategori MA (Madrasah Aliyah), SMK (Sekolah Menengah Kejuruan) dan SMA (Sekolah Menengah Atas). Begitu juga dengan atribut Provinsi asal SLTA masing-masing mahasiswa akan dikategorikan menjadi 3 kolom wilayah terdiri dari wilayah I, wilayah II dan wilayah III. Pembagian wilayah dalam riset yang dilakukan berpacu pada penelitian ditahun 2018 (Gibran Satya Nugraha, Hairani, 2018) berdasarkan perhitungan kualitas pendidikan di Indonesia seperti Tabel 3. Kualitas Pendidikan

Tabel 3. Kualitas Pendidikan

Kelompok	Provinsi
Wilayah 1	Maluku Utara, Kalimantan Tengah.
Wilayah 2	Kalimantan Barat, Kep. Bangka, Aceh, NTB, Sulawesi Barat, Sulawesi Selatan, Bali, Jawa Tengah, Jawa Timur, Sulawesi Utara, Riau, Sumatera Barat, Lampung, NTT, Sulawesi Tengah, Banten, DI Yogyakarta, Kalimantan Selatan, Bengkulu, Sumatera Selatan, Gorontalo, Sulawesi Tengah, Jakarta, Jambi, Jawa Barat, dan Sumatera Utara.
Wilayah 3	Papua Barat, Maluku, Kalimantan Timur dan Papua.

Pada atribut lama studi data penelitian diubah kedalam 2 kategori, yaitu kategori lulus tepat waktu dan kategori lulus tidak tepat. Dikatakan lulus tepat waktu apabila kurang dari 1.440 hari masa studi atau tidak lebih dari 4 tahun, sebaliknya lulus tidak tepat waktu apabila lama studi lebih dari 4 tahun atau lebih dari 1.440 hari. Kategori penelitian ini sesuai dengan standar nasional perguruan tinggi dari Peraturan Menteri Pendidikan dan Budaya (PERMENDIKBUD).

2.3 Metode Penelitian

K-Nearest Neighbors (k-NN) merupakan salah satu algoritma klasifikasi dalam data mining termasuk dalam *supervised learning* dalam memperhitungkan jarak terdekat satu data baru dengan beberapa kelompok yang sudah berlabel sehingga dapat disimpulkan label data baru. Pencarian jarak terdekat dengan algoritma k-NN (Anton Yudhana, Sunardi,

2020) salah satu cara dengan persamaan (1) yaitu perhitungan rumus jarak *euclidean distance*. Pada persamaan (1), x_i merupakan total sampel data dan y_i merupakan total data uji. Sedangkan i merupakan atribut data sedangkan d merupakan jarak.

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{1}$$

Tahapan perhitungan metode *k-nearest neighbors* pada umumnya dimulai dari menentukan nilai K, lalu dengan nilai K tersebut akan dihitung jarak terdekat dengan data yang sudah memiliki label (Rachmat et al., 2020). Tahapan dapat terlihat pada Gambar 2.



Gambar 2. Perhitungan metode *k-nearest neighbors*

2.4 Pengujian Akurasi

Pengujian akurasi salah satunya menggunakan *confusion matriks*, yaitu dengan menghitung *recall*, *precision* dan *accuracy* (Hakim, Rizal, & Ratnasari, 2019). Bukan hanya untuk menghitung akurasi, *confusion matriks* juga memberi pengetahuan penting dalam mengevaluasi kekuatan dan kelemahan model klasifikasi dalam memprediksi data. Penggunaan *confusion matriks* pada penelitian ini fokus pada akurasi hasil prediksi. *Accuracy* yaitu nilai tingkat kedekatan prediksi dan nilai *actual* ($\frac{TP+TN}{TOTAL}$), nilai *precision* seperti keakuratan data yang diminta dengan hasil prediksi oleh *system* ($\frac{TP}{TP+FP}$), dan nilai *recall* adalah keberhasilan *system* dalam menemukan kembali sebuah informasi ($\frac{TP}{TP+FN}$) (Wiyana Yusuf, Rahma Witri, 2022). Terlihat Tabel 2 merupakan rumus *confusion matriks*.

Tabel 4. *Confusion Matriks*

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Rumus akurasi *confusion matriks* berbentuk matriks 2x2 terdiri dari empat bagian yaitu *true positive*, *true negative*, *false positive* dan *false negative*. *True positive (TP)* merupakan total data yang bernilai positif dan diprediksi benar sebagai positif, sedangkan *false positive (FP)* adalah total data yang bernilai Negatif tetapi diprediksi sebagai positif. *True negative (TN)* merupakan total data yang bernilai negatif dan diprediksi benar sebagai negatif, sedangkan *false negative (FN)* adalah total data yang bernilai positif dan diprediksi benar sebagai negatif.

3. HASIL DAN PEMBAHASAN

Pembahasan point penelitian prediksi kelulusan mahasiswa menggunakan perangkat *Computer Acertype Aspire ES 14* dengan model *ES1-432-C6AH Intel Celeron N3350, system operasi Windows 10 Pro 32*, dengan bahasa *python 3.0*.

Data penelitian terdiri dari beberapa file yang digabungkan menjadi satu melalui proses integrasi data. Integrasi data merupakan penggabungan atau menyatukan beberapa file dari berbagai sumber yang berbeda menjadi sebuah file. Penggabungan file berdasarkan variabel nama mahasiswa. Hasil dari penggabungan menghasilkan banyak atribut. Tahapan berikutnya penghapusan atribut-atribut yang tidak digunakan sehingga menghasilkan atribut NIM mahasiswa, nama mahasiswa, prodi mahasiswa, nama SLTA, provinsi SLTA, nilai rata matematika dan lama studi.

Transformasi akan dilakukan pada atribut-atribut seperti atribut nama SLTA, Provinsi SLTA dan lama studi. Pada atribut SLTA ditransformasikan kedalam kolom SMA, MA dan SMK. Selanjutnya atribut Provinsi SLTA ditransformasikan kedalam kolom wilayah I, wilayah II dan wilayah III. Tahapan terakhir transformasi, dikerjakan pada atribut lama studi. Menggunakan kondisi apabila melewati semester 8 maka termasuk kategori lulus tidak tepat waktu dan sebaliknya apabila lulus 8 semester maka mahasiswa lulus tepat waktu. Hasil transformasi dari atribut nama SLTA, provinsi SLTA dan lama studi terlihat pada Tabel 5.

Tabel 5. Hasil *preprocessing system dataset*

No	SMA	MA	SMK	Wil. I	Wil. II	Wil. III	Rate Mate	Status Kelulusan
1	1	0	0	0	1	0	98	Tepat Waktu
2	1	0	0	0	1	0	98	Tepat Waktu
3	0	1	0	0	1	0	83	Tidak Tepat Waktu
4	1	0	0	0	1	0	86	Tidak Tepat Waktu
5	1	0	0	0	1	0	89	Tidak Tepat Waktu
6	1	0	0	0	1	0	77	Tidak Tepat Waktu
7	1	0	0	0	1	0	88	Tidak Tepat Waktu
8	1	0	0	0	1	0	81	Tidak Tepat Waktu
9	0	1	0	0	1	0	88	Tidak Tepat Waktu
10	1	0	0	1	1	0	88	Tidak Tepat Waktu
11	1	0	0	0	1	0	79	Tidak Tepat Waktu

Tahapan transformasi ini menggunakan tahapan tipe *one hot encoding*. *One hot encoding* yaitu tahapan mengubah kategori menjadi vector biner dengan nilai 1 sesuai kategori dan nilai 0 untuk kategori lainnya. Penerapan ini seperti terlihat pada Gambar 3 diatas. Penelitian yang dilakukan juga menggunakan bantuan sistem penelitian. Proses penerapan sistem penelitian menggunakan

perhitungan kaidah *k-Nearest Neighbor*. Penerapan perhitungan menggunakan pendekatan jarak *Euclidian Distance* kedekatan antara satu data mahasiswa dengan data mahasiswa lainnya yang telah memiliki label lulus tepat waktu dan lulus tidak tepat waktu. Perhitungan kedekatan jarak dilakukan dari masing-masing atribut pada data mahasiswa, sehingga menghasilkan klasifikasi. Prediksi kelulusan berdasarkan kedekatan data mayoritas yang memiliki label lulus tepat waktu atau lulus tidak tepat waktu sehingga akan direkomendasikan sesuai label mayoritas terdekat.

Prediksi kelulusan oleh sistem dengan objek penelitian mahasiswa tehnik informatika FTI UAD sebanyak 93 data terdiri dari 78 training dan 12 data testing. Sistem prediksi diuji dengan data mahasiswa yang belum lulus dengan detail data nama Isnan Arif Cahyadi, asal sekolah sman 3 karawang, asal provinsi SLTA jawa barat dan nilai rata matematika 85. Terlihat pada Gambar 3 data uji prediksi.

Data Mahasiswa

	NIM	NAMA	ASAL SEKOLAH	PROVINSI	RATA MATE	Wils
1	1700018101	ISNAN ARIF CAHYADI	SMAN 3 Karawang	Jawa Barat	85	

Gambar 3. Data uji prediksi

Confusion Matrix

	0	1
0	8	2
1	0	2

	TEPAT	TIDAK TEPAT
TEPAT	8	2
TIDAK TEPAT	0	2

precision recall f1-score support

TP	1.00	0.80	0.89	10
TTP	0.50	1.00	0.67	2

accuracy 0.75
macro avg 0.75 0.90 0.78 12
weighted avg 0.92 0.83 0.85 12

Choose a Excel file

Drag and drop file here
Limit 200MB per file • CSV, XLSX

Data Untuk Prediksi PAPER 2.xlsx 0.6KB

Data Mahasiswa

	PROVINSI	RATA MATE	Wilayah I	Wilayah II	Wilayah III	SMA	SMK	MA
1	Jawa Barat	85	0	1	0	1	0	0

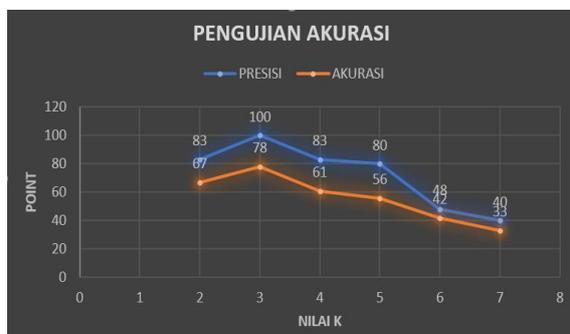
	RATA MATE	Wilayah I	Wilayah II	Wilayah III	SMA	SMK	MA	0_x	1	0_y
1	85	0	1	0	1	0	0	1	0	TEPAT

Gambar 4. Hasil prediksi sistem

Pada Gambar 4 terlihat hasil prediksi dari data mahasiswa dengan penerapan algoritma *k-Nearest Neighbor* mahasiswa tersebut direkomendasikan lulus tepat waktu, berdasarkan pendekatan *dataset training* dan *data testing*. Dimana kedekatan mayoritas *dataset training* dan *data testing* memiliki label lulus tepat waktu. Hasil rekomendasi prediksi diuji akurasi menggunakan pengujian *confusion matriks*. Pengujian akurasi bertujuan untuk melihat

seberapa besar nilai akurasi hasil rekomendasi dengan algoritma k-NN.

Pengujian akurasi dari penelitian menggunakan *Python*, dengan nilai uji k-2 sampai k-7. Nilai uji k-1 sampai k-5 untuk mencari nilai akurasi yang tinggi. Perhitungan akurasi dengan perhitungan *Confusion Matrix* terlihat Gambar 5 dibawah ini.



Gambar 5. Pengujian akurasi

Perhitungan pengujian akurasi dengan nilai uji k-2 sampai k-7 dengan sistem membuktikan akurasi tertinggi nilai k sama dengan 3. Dengan nilai k-3 sistem prediksi kelulusan memiliki nilai akurasi sebesar 0.78 atau 78%. Hasil pengujian akurasi k-2 memiliki nilai sedikit lebih rendah dengan nilai akurasi 67% dan nilai presisi 83%. Pada k sama dengan 4 pengujian akurasi juga memiliki nilai lebih rendah dengan nilai akurasi 61% dan presisi 83%. Begitu juga dengan k-5, k-6 dan k-7. Nilai akurasi dan presisi semakin rendah. Hasil perhitungan tersebut berdasarkan 93 *dataset* terdiri dari 78 *data training* dan 12 *data testing*.

4. KESIMPULAN DAN SARAN

Hasil penelitian sebagai pengembangan ilmu pengetahuan dan membuktikan apakah algoritma *K-Nearest Neighbor* bisa digunakan sebagai metode klasifikasi kelulusan mahasiswa tepat waktu dan kelulusan mahasiswa tidak tepat waktu. Setelah dilakukan tahapan-tahapan riset, dapat ditarik kesimpulan bahwa klasifikasi kelulusan mahasiswa melalui penerapan algoritma *K-Nearest Neighbor* membuktikan bisa digunakan untuk klasifikasi kelulusan mahasiswa tepat waktu dan lulus tidak tepat waktu. Proses klasifikasi kelulusan mahasiswa menjadi semakin akurat apabila kapasitas *dataset* penelitian yang digunakan dengan kuantitas banyak dan atribut yang dipakai lebih *variatif*. Klasifikasi pada penelitian ini menggunakan *dataset* mahasiswa FTI UAD sebanyak 93 data mahasiswa terdiri 78 *data training* dan sebanyak 12 mahasiswa *data testing*. Kaidah penelitian dengan *K-Nearest Neighbor* untuk mengukur akurasi penelitian dengan metode *Confusion Matrix*. Penelitian menggunakan variabel nama, asal SMA/SMK, wilayah asal SMA/SMK, dan nilai rata-rata matematika. Pengujian klasifikasi diuji dengan *cluster data* mulai dari k=2, k=3, k=4, k=5, k=6 dan k=7. Dari pengujian *cluster* menunjukkan *cluster* dengan nilai k=3 adalah nilai paling tinggi dari

nilai yang lainnya, dengan hasil pengujian akurasi 78% dan presisi 100%. Sistem prediksi yang dibangun dari penelitian ini menjadi pemicu sistem prediksi kelulusan lainnya dan dengan metode lebih *variatif* lainnya. Sehingga ilmu pengetahuan dapat berkembang dan sistem prediksi memiliki berbagai macam metode.

DAFTAR PUSTAKA

- AHMAD R. PRATAMA, R. R. A. 2022. 2022. Model Klasifikasi Calon Mahasiswa Baru Untuk Sistem Rekomendasi Program Studi Sarjana Berbasis *Machine Learning*. 9(4). <https://doi.org/10.25126/jtiik.202294311>
- ALLAN, S. 2020. *Migration And Transformation: A Sociomaterial Analysis Of Practitioners' Experiences With Online Exams*. *Research In Learning Technology*. <https://doi.org/10.25304/rlt.v28.2279>
- ANTON YUDHANA, SUNARDI, dan A. J. S. H. 2020. Algoritma K-NN dengan *Euclidean Distance* untuk Prediksi Hasil Penggajian Kayu Sengon. *Undip E-Journal System*, (4).
- GEDE ADITRA PRADNYANA, A. A. J. P. 2018. Sistem Pembagian Kelas Kuliah Mahasiswa Dengan Metode *K-Means* dan *K-Nearest Neighbors* Untuk Meningkatkan Kualitas Pembelajaran. 59–68.
- GIBRAN SATYA NUGRAHA, HAIRANI, R. F. P. A. 2018. Aplikasi Pemetaan Kualitas Pendidikan Di Indonesia Menggunakan Metode *K-Means*. 17(2).
- HAKIM, L. A. R., RIZAL, A. A., & RATNASARI, D. 2019. Aplikasi Prediksi Kelulusan Mahasiswa Berbasis *K-Nearest Neighbor* (K-NN). *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, 1(1), 30–36. <https://doi.org/10.35746/jtim.v1i1.11>
- HERMAWAN, H. 2018. Analisis Pengaruh Bauran Pemasaran Terhadap Keputusan, Kepuasan Dan Loyalitas Konsumen Dalam Pembelian Roti Ceria Di Jember. *Jurnal Manajemen Dan Bisnis Indonesia*.
- HERYANA, N., KARAWANG, U. S., JUARDI, D., KARAWANG, U. S., SOLEHUDIN, A., KARAWANG, U. S., & MINING, D. 2020. Kajian Model Prediksi Ketepatan Kelulusan Mahasiswa : Sebuah Literatur Review Kajian Model Prediksi Ketepatan Kelulusan Mahasiswa : Sebuah *Literatur Review*. (September). <https://doi.org/10.35706/syji.v9i1.3388>
- HIDAYATI, N., & HERMAWAN, A. 2021. *K-Nearest Neighbor* (k-NN) *algorithm with Euclidean and Manhattan in classification of student graduation*. 2(2), 86–91.
- HOZAIRI, A. S. A. 2021. Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa

- Dengan Model *K-Nearest Neighbor*, *Decision Tree* Serta *Naive Bayes Orange Data Mining Implementation For Student Graduation Classification Using K-Nearest Neighbor*. *Jurnal Ilmiah NERO*, 6(2), 133–144.
- KHASANAH, N., SALIM, A., AFNI, N., KOMARUDIN, R., & MAULANA, Y. I. 2022. Prediksi Kelulusan Mahasiswa Dengan Metode *Naive Bayes*. 13(3), 207–214.
- KHOLIL, M. 2018. Penerapan Metode *K Nearest Neighbord* Dalam Proses Seleksi Penerima Beasiswa.
- M. REZA NOVIANSYAH, TEDY RISMAWAN, D. M. M. 2018. Penerapan Data Mining Menggunakan Metode *K-Nearest Neighbor* Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data AWS (*Automatic Weather Station*) (Studi Kasus: Kabupaten Kubu Raya). *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48–56.
- MOH. ZAINUDDIN, A. N. 2019. Studi Komparasi Algoritma *Decesion Tree* (C4.5) Dengan Algoritma K-NN Dalam Memprediksi Kelulusan Tepat Waktu Mahasiswa. 10.
- MULIONO, R. 2019. Analisis Algoritma K-Nearest Neighbors dalam Prediksi Waktu Kelulusan Mahasiswa. 2(November), 12–16.
- NAUFAL, M. A. 2017. Implementasi Metode Klasifikasi *K-Nearest Neighbor* (K-NN) Untuk Pengenalan Pola Batik Motif Lampung.
- NURHANIFA, T. S. 2020. Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung dan Pinjam Buku di Perpustakaan Menggunakan Metode C4.5. *Jurnal Sarjana Teknik Informatika*, 8(2), 24–33.
- NOVENDRA ADISAPUTRA SINAGA, B. HERAWAN HAYADI, Z. S. 2022. Perbandingan Akurasi Algoritma *Naive Bayes*, K-NN Dan SVM Dalam Memprediksi Penerimaan Pegawai. 5, 27–34. <https://doi.org/10.37600/tekinkom.v5i1.446>
- PAMULANG, U., SURYA, J., NO, K., BARAT, P., & SELATAN, T. 2021. Prediksi Kelulusan Tepat Waktu Menggunakan Algoritma SVM dan *K-Nearest Neighbour* Berbasis *Particle Swarm*. XVI(01), 55–65.
- RACHMAT, R., CHRISNANTO, Y. H., UMBARA, F. R., INFORMATIKA, S., JENDERAL, U., & YANI, A. 2020. Sistem Prediksi Mutu Air Di Perusahaan Daerah Air Minum Tirta Raharja Menggunakan K – *Nearest Neighbors* (K – NN). 189–193.
- RAMADHAYANTI, A. 2018. Analisis Strategi Belajar Dengan Metode Bimbel Online Terhadap Kemampuan Pemahaman Kosakata Bahasa Inggris dan *Pronunciation* (Pengucapan/pelafalan) Berbahasa Remaja Saat Ini. *KREDO : Jurnal Ilmiah Bahasa Dan Sastra*. <https://doi.org/10.24176/kredo.v2i1.2580>
- RATHORE, A. S., & ARJARIA, S. K. 2019. *Intelligent Tutoring System*. In *Utilizing Educational Data Mining Techniques For Improved Learning: Emerging Research And Opportunities*. <https://doi.org/10.4018/978-1-7998-0010-1.ch006>
- RIYAN LATIFAHUL HASANAH, F. F. W. & D. R. 2019. TAM Analysis on the Factors Affecting Admission of Students for *Ruangguru Application*. *Jurnal Sistem Informasi*.
- RUSTAM, S., & ANNUR, H. 2019. Akademik Data Mining (Adm) K-Means Dan K-Means k-NN Untuk Mengelompokan Kelas Mata Kuliah Kosentrasi. 11(28), 260–268.
- SRI MULYATI, SYEPRY MAULANA HUSEIN, R. 2020. Rancang Bangun Aplikasi Data Mining Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma (k-NN) *K-Nearest Neighbor* Dengan Metode *Euclidean Distance* Pada Smpn 2 Pagedangan. *Jurnal Teknik Informatika (JIKA) Universitas Muhammadiyah Tangerang*, 65–73.
- SUSANTO, E. S., KUSRINI, & FATTA, H. AI. 2018. Informatika Universitas Amikom Yogyakarta Menggunakan Metode *K-Nearest Neighbor*. *Jurnal Teknologi Informasi*, XIII, 67–72. <http://jti.respati.ac.id/index.php/jurnaljti/article/view/260/239>.
- SUMARLIN, D. A. 2018. Implementasi *K-Nearest Neighbord* Pada *Rapidminer* Untuk Prediksi Kelulusan Mahasiswa. HOAQ: JURNAL TEKNOLOGI INFORMASI, 10, 35–41.
- WILIANI, A. B. I. E. P. P. A. P. N. 2020. Komparasi Algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbor* Untuk Memprediksi Mahasiswa Lulus. 5(2), 265–270
- WISDAYANI, D. S., NUR, I. M., & WASONO, R. 2019. Penerapan Algoritma *K-Nearest Neighbor* dalam Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Jawa Tengah. *Prosiding Mahasiswa Seminar Nasional Unimus*, 2, 373–380.
- WIYANA YUSUF, RAHMA WITRI, C. J. 2022. Model Prediksi Penjualan Jenis Produk Tekstil Menggunakan Algoritma *K-Nearest Neighbor* (K-NN). (*Indonesian Journal on Computer and Information Technology*), 7(1), 1–6.

Halaman ini sengaja dikosongkan