

Penerapan Metode C4.5 dan K-Nearest Neighbor untuk Klasifikasi Kelulusan Mahasiswa Berdasarkan Data Akademik

Dina Amalia Putri^{1*}, Naza Sefti Prianita², Elkin Rilvani³

¹⁻³Teknik Informatika, Universitas Pelita Bangsa

Email : dinaamaliap33@gmail.com¹, nazasefti5@gmail.com², elkin.rilvani@gmail.com³

Korespondensi penulis : dinaamaliap33@gmail.com^{*}

Abstract. The issue of determining the number of students' graduation times is one of the important indicators in transmitting the quality and effectiveness of the higher education process in universities. The rate of on-time graduation not only impacts accredited institutions, but also becomes a concern for campus management in designing learning strategies and academic guidance. This study aims to apply and compare two classification algorithms in data mining, namely C4.5 and K-Nearest Neighbor (KNN), in predicting the accuracy of students' graduation times. Predictions are made based on academic attributes such as Grade Point Average (GPA), number of credits that have been achieved, and Semester Grade Point Average (IPS) as input variables. The method used in this study is Knowledge Discovery in Database (KDD) which includes data selection, preprocessing, transformation, data mining, and evaluation of results. The study was conducted using the RapidMiner tool, with a dataset of 279 Informatics Study Program students from the 2015 to 2019 intake. The data was classified into two categories: "graduated on time" and "not graduated on time". The test results showed that the KNN algorithm provided better performance compared to C4.5. KNN produced an accuracy of 76.08%, with a precision of 73.11% and a recall of 41.92%. Meanwhile, the C4.5 algorithm produced an accuracy of 73.49%, with a precision of 64.62% and a recall of 41.89%. This difference in accuracy indicates that KNN is more effective in capturing patterns in the data and providing more accurate predictions in this context. Thus, the KNN algorithm can be considered a more optimal method to assist universities in predicting potential student admissions in a timely manner, thus enabling early intervention for students at risk of late graduation. This research also contributes to the development of data mining-based academic decision support systems in higher education.

Keywords: C4.5, Data Mining, Graduation Prediction, K-Nearest Neighbor, Student Graduation.

Abstrak. Permasalahan ketepatan waktu kelulusan mahasiswa merupakan salah satu indikator penting dalam mengevaluasi kualitas dan efektivitas proses pendidikan tinggi di perguruan tinggi. Tingkat kelulusan tepat waktu tidak hanya berdampak pada akreditasi institusi, tetapi juga menjadi perhatian bagi manajemen kampus dalam merancang strategi pembelajaran dan bimbingan akademik. Penelitian ini bertujuan untuk menerapkan dan membandingkan dua algoritma klasifikasi dalam data mining, yaitu C4.5 dan K-Nearest Neighbor (KNN), dalam memprediksi ketepatan waktu kelulusan mahasiswa. Prediksi dilakukan berdasarkan atribut akademik seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, dan nilai Indeks Prestasi Semester (IPS) sebagai variabel input. Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD) yang mencakup tahap seleksi data, preprocessing, transformasi, data mining, dan evaluasi hasil. Penelitian dilakukan menggunakan tools RapidMiner, dengan dataset sebanyak 279 data mahasiswa Program Studi Informatika dari tahun angkatan 2015 hingga 2019. Data diklasifikasikan ke dalam dua kategori yaitu "lulus tepat waktu" dan "tidak lulus tepat waktu". Hasil pengujian menunjukkan bahwa algoritma KNN memberikan performa yang lebih baik dibandingkan dengan C4.5. KNN menghasilkan akurasi sebesar 76,08%, dengan precision 73,11% dan recall 41,92%. Sementara itu, algoritma C4.5 menghasilkan akurasi sebesar 73,49%, dengan precision 64,62% dan recall 41,89%. Selisih akurasi tersebut menunjukkan bahwa KNN lebih efektif dalam menangkap pola pada data dan memberikan prediksi yang lebih akurat dalam konteks ini. Dengan demikian, algoritma KNN dapat dipertimbangkan sebagai metode yang lebih optimal untuk membantu pihak kampus dalam memprediksi potensi kelulusan mahasiswa secara tepat waktu, sehingga memungkinkan adanya intervensi dini terhadap mahasiswa yang berisiko terlambat lulus. Penelitian ini juga memberikan kontribusi terhadap pengembangan sistem pendukung keputusan akademik berbasis data mining di lingkungan pendidikan tinggi.

Kata kunci: C4.5, Data Mining, Kelulusan Mahasiswa, K-Nearest Neighbor, Prediksi Kelulusan.

1. LATAR BELAKANG

Kualitas pendidikan tinggi dapat diukur dari berbagai indikator, salah satunya adalah tingkat kelulusan mahasiswa tepat waktu. Dalam konteks ini, kemampuan untuk memprediksi kelulusan mahasiswa sangat penting untuk mendukung proses akademik dan kebijakan pendidikan. Universitas Teknologi Sumbawa mengalami peningkatan jumlah mahasiswa, namun hanya sebagian yang berhasil lulus tepat waktu, sehingga diperlukan strategi untuk mengatasi hal tersebut.

Data akademik yang tersedia dapat dimanfaatkan melalui penerapan teknik data mining untuk menghasilkan informasi prediktif. Dua algoritma yang umum digunakan dalam klasifikasi adalah C4.5 dan K-Nearest Neighbor (KNN). Keduanya memiliki karakteristik dan pendekatan yang berbeda dalam proses klasifikasi. Oleh karena itu, penelitian ini bertujuan untuk membandingkan performa kedua algoritma tersebut dalam memprediksi kelulusan mahasiswa.

Metode yang digunakan dalam penelitian ini adalah proses Knowledge Discovery in Database (KDD), yang terdiri dari beberapa tahapan, yaitu seleksi data, praproses data, transformasi, data mining, dan evaluasi hasil. Dataset yang digunakan dalam penelitian ini berasal dari program studi Informatika Universitas Teknologi Sumbawa dengan jumlah total 279 data mahasiswa dari angkatan 2015 hingga 2019. Atribut-atribut yang digunakan dalam proses klasifikasi meliputi Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, dan Indeks Prestasi Semester (IPS), yang dianggap sebagai indikator utama yang memengaruhi ketepatan waktu kelulusan mahasiswa.

Pengujian dilakukan dengan menggunakan perangkat lunak RapidMiner, dan hasil evaluasi performa algoritma dilakukan dengan menggunakan metrik akurasi, precision, dan recall. Hasil pengujian menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) memiliki tingkat akurasi tertinggi sebesar 76,08%, dengan nilai precision sebesar 73,11% dan recall 41,92%. Sementara itu, algoritma C4.5 memiliki akurasi sebesar 73,49%, precision 64,62%, dan recall 41,89%. Hal ini menunjukkan bahwa algoritma KNN lebih unggul dalam memprediksi kelulusan mahasiswa tepat waktu dibandingkan dengan C4.5 pada dataset yang digunakan.

Temuan ini memberikan implikasi penting bagi pihak pengelola perguruan tinggi, khususnya dalam mengembangkan sistem pendukung keputusan berbasis data akademik untuk memantau dan memfasilitasi proses belajar mahasiswa. Dengan mengadopsi algoritma klasifikasi yang paling optimal, institusi dapat secara proaktif mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu, sehingga dapat memberikan intervensi lebih awal seperti

pembimbingan akademik, konseling, atau program remedial. Ke depannya, integrasi data mining dalam sistem informasi akademik akan menjadi langkah strategis dalam meningkatkan kualitas pendidikan dan efisiensi manajemen perguruan tinggi.

2. KAJIAN TEORITIS

- a. Komparasi Algoritma: Komparasi algoritma adalah proses membandingkan dua atau lebih algoritma dalam menyelesaikan masalah untuk mengetahui kelebihan, kekurangan, serta efektivitas masing-masing.
- b. Algoritma C4.5: Merupakan pengembangan dari ID3 yang menggunakan gain ratio dalam pemilihan atribut pada pohon keputusan. C4.5 mampu menangani data numerik dan kategorikal serta dapat mengatasi missing value.
- c. Algoritma K-Nearest Neighbor: Merupakan metode klasifikasi berbasis instance yang menggunakan perhitungan jarak (biasanya Euclidean) untuk menentukan kelas berdasarkan data pelatihan terdekat.
- d. Kelulusan Mahasiswa: Kelulusan tepat waktu adalah mahasiswa yang menyelesaikan studinya dalam waktu ≤ 8 semester. Faktor akademik seperti IPS dan IPK sangat berpengaruh terhadap kelulusan.

3. METODE PENELITIAN

Pemahaman Domain dan Tujuan KDD

Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna bagi prodika di manfaatkan dengan maksimal. Salah satu cara untuk memanfaatkan data tentang mahasiswa yang lulus ini adalah dengan mengolahnya menggunakan *data mining*. Dengan proses *data mining* ini dapat digunakan untuk menghasilkan suatu informasi seperti tepat atau tidaknya kelulusan mahasiswa.

Untuk mencapai hal tersebut, akan dilakukan upaya pengumpulan data mengenai mahasiswa yang berhasil menyelesaikan program Studi Informatika menggunakan pendekatan model *knowledge discovery in databases* (KDD). Tujuan yang diinginkan dari proses KDD ini adalah untuk memperoleh informasi tentang status kelulusan mahasiswa berdasarkan kategori jenis kelamin, asal daerah, IPK, serta nilai TOEFL-nya.

Pemilihan dan Penambahan Data

Data yang akan digunakan dalam proses KDD adalah data tentang mahasiswa yang lulus Prodi Informatika dari Januari 2015 sampai Agustus 2019. Data tentang mahasiswa yang lulus tersebut di ambil dari aplikasi repositori lulusan. Tidak semua data tentang mahasiswa yang

lulus akan digunakan. Data yang akan diambil adalah nama mahasiswa, jenis kelamin, IPK, nilai TOEFL, dan lama studi. Jenis kelamin, asal daerah, IPK, dan nilai TOEFL akan digunakan sebagai atribut dan lama studi akan digunakan sebagai kelas.

Dalam pembangunan aplikasi prediksi kelulusan mahasiswa, IPK menggambarkan performa akademik mahasiswa. Nilai TOEFL menggambarkan pemahaman mahasiswa dalam memahami literatur pembelajaran yang menggunakan bahasa Inggris. Asal daerah menggambarkan pengaruh faktor keluarga dan perbedaan kultur terhadap performa akademik mahasiswa. Keluarga dan kemandirian belajar merupakan faktor yang menentukan prestasi mahasiswa. Selain itu fenomena *culture shock* tersebut dapat menimbulkan efek stres yang dapat mempengaruhi prestasi mahasiswa. Sedangkan jenis kelamin menggambarkan gender terhadap performa akademik mahasiswa. Dalam sebuah penelitian yang meneliti pengaruh gender dan motivasi belajar terhadap prestasi siswa, perempuan dinilai lebih berprestasi.

Tabel 1. Tabel Transformasi Data

| No | Atribut | Nilai Atribut | Keterangan |
|----|-------------|----------------|---|
| 1 | Asal daerah | Jateng | Provinsi bernilai Jawa Tengah |
| | | Luar Jateng | Provinsi selain Jawa Tengah yang masih berada di Pulau Jawa |
| 2 | IPK | Luar Jawa | Provinsi di luar Pulau Jawa |
| | | Memuaskan | IPK kurang dari 2,75 |
| | | Sangat | IPK di antara 3,50 |
| | | Memuaskan | sampai dengan 3,50 |
| 3 | TOEFL | Dengan Pujian | IPK lebih dari 3, |
| | | Dasar | Nilai TOEFL kurang dari 420 |
| | | Menengah | Nilai TOEFL di antara |
| | | Bawah | 421 sampai 480 |
| 4 | Lama studi | Menengah Atas | Nilai TOEFL di antara |
| | | Mahir | 481 sampai 520 |
| | | <5 tahun | Nilai TOEFL lebih dari 5 tahun |
| | | ≥ 5 tahun | Lama studi kurang dari 5 tahun |
| | | | Lama studi lebih dari sama Dengan 5 tahun |

Dari pada laki-laki. Hal ini dikarenakan perempuan lebih tekun dan rajin dari pada laki-laki. Jumlah data yang digunakan adalah 382 data dengan 212 data untuk kelas <5 tahun dan 170 data untuk kelas ≥ 5 tahun.

Penelitian ini menggunakan pendekatan Knowledge Discovery in Database (KDD) yang terdiri dari lima tahapan utama: Selection, Preprocessing, Transformation, Data Mining, dan Evaluation. Seluruh proses dilakukan untuk menghasilkan model klasifikasi yang dapat memprediksi kelulusan mahasiswa berdasarkan data akademik.

Selection

Pada tahap ini, dilakukan pemilihan atribut yang relevan untuk klasifikasi kelulusan mahasiswa. Atribut yang digunakan meliputi:

- Jenis Kelamin
- IPS Semester 4, 5, dan 6
- IPK Akhir
- Jumlah SKS
- Status Kelulusan (Tepat Waktu atau Tidak Tepat Waktu) Data diperoleh dari rekap akademik mahasiswa Program Studi Informatika angkatan 2015–2019 sebanyak 279 data.

Preprocessing

Tahapan ini bertujuan membersihkan data dari kesalahan atau ketidaksesuaian. Proses meliputi:

- Menghapus data kosong (*missing value*)
- Menghilangkan data duplikat
- Validasi format data Pembersihan dilakukan secara manual menggunakan Microsoft Excel agar data siap diproses lebih lanjut.

Transformation

Data yang telah dibersihkan kemudian ditransformasikan agar sesuai dengan kebutuhan algoritma klasifikasi. Proses ini melibatkan:

- Normalisasi nilai numerik menjadi kategori (misal: rendah, sedang, tinggi)
- Pengkodean variabel kategorik
- Penyederhanaan atribut untuk mengurangi kompleksitas model Transformasi dilakukan agar data lebih mudah dianalisis oleh algoritma C4.5 dan KNN.

Data Mining

Untuk mencapai tujuan dalam proses KDD, metode pohon keputusan akan diterapkan sebagai teknik penambangan data, dengan penggunaan algoritma C4.5 dalam penyusunan pohon keputusan. Prosedur pemangkasan pada pohon keputusan akan menggunakan metode *error-based pruning*. Disamping itu, dengan adanya ketidakseimbangan data, data yang

dipakai dalam pembuatan pohon keputusan akan ditangani menggunakan algoritma *random over sampling* (ROS).

Tahap *data mining* diawali dengan memisahkan informasi mengenai mahasiswa yang telah lulus ke dalam data latih dan data uji. Data latih akan dimanfaatkan dalam pembentukan pohon keputusan, sedangkan data uji akan berfungsi untuk menilai kinerja pohon keputusan yang telah di rekayasa. Sebelum digunakan, data latih yang akan dipakai dalam proses pembuatan pohon keputusan perlu diseimbangkan agar tidak terdapat bias terhadap kelas mayoritas di dalam pohon yang dihasilkan. Proses untuk mencapai keseimbangan data latihan dilakukan melalui pendekatan *random over sampling* (ROS). Proses ROS dimulai dengan menghitung perbedaan antara jumlah kelas mayoritas dan kelas minoritas. Selanjutnya, satu data acak dari kelas minoritas dipilih. Data tersebut kemudian ditambahkan ke kelas minoritas. Proses penambahan data ini akan diulang hingga jumlah data di kelas mayoritas dan kelas minoritas menjadi setara. Metode ROS diilustrasikan dalam bentuk flowchart.

Setelah dataset pelatihan seimbang, dataset tersebut siap untuk digunakan dalam proses pembuatan pohon keputusan dengan algoritma C4.5. Algoritma C4.5 merupakan sebuah kemajuan dari algoritma ID3 yang ditemukan oleh J.Ross Quinlan pada tahun 1993. Detail dari algoritma C4.5 adalah sebagai berikut:

- Menghitung jumlah kasus total, jumlah kasus dengan keputusan <5 tahun, kasus dengan keputusan ≥ 5 tahun, dan *entropy* dari semua kasus dan kasus yang dibagi berdasarkan nilai atribut.

$$\text{Entropy}(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

S = kumpulan data

K = banyaknya kelas dalam S

P_j = probabilitas kelas C_j

Jika kasus total hanya memiliki satu kelas (<5 tahun atau ≥ 5 tahun), maka jadikan node sebagai daun dengan nilai yang mayoritas.

- Menghitung *information gain* untuk setiap atribut.

$$(2) \quad \text{gain}(A) = \text{entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{entropy}(S_i)$$

S = kumpulan data

A = atribut

A_i = nilai atribut ke- i

$|S_i|$ = jumlah data untuk A_i

$|S|$ = jumlah data dalam S

k = jumlah nilai atribut

c) Menghitung *split info* untuk setiap atribut.

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{s_i}{S} \log_2 \frac{s_i}{S} \quad (3)$$

\hookrightarrow = kumpulan data

A = atribut

A_i = nilai atribut ke- i

n = jumlah nilai atribut

S_i = jumlah data untuk A_i

d) Menghitung *gain ratio* untuk setiap atribut

$$GainRatio(A) = \frac{gain(A)}{SplitInfo(S, A)} \quad (4)$$

A = atribut

S = kumpulan data

e) Memilih atribut dengan nilai *gain ratio* sebagai *node*.

f) Membagi data berdasarkan nilai atribut dari atribut terpilih. Kemudian menggunakannya untuk melakukan langkah selanjutnya.

g) Ulangi langkah 1 sampai 6 hingga seluruh atribut digunakan atau memenuhi suatu kondisi berhenti.

Pada tahap ini, dua algoritma klasifikasi diterapkan:

- **C4.5:** Digunakan untuk membentuk model pohon keputusan berdasarkan atribut yang dipilih. C4.5 menghitung Gain Ratio untuk menentukan atribut utama dalam klasifikasi.
- **K-Nearest Neighbor (KNN):** Mengklasifikasikan data berdasarkan kedekatan jarak (menggunakan Euclidean Distance). Nilai k yang digunakan adalah $k = 3$, ditentukan berdasarkan eksperimen awal. Proses klasifikasi dilakukan menggunakan perangkat lunak RapidMiner. Dataset dibagi menjadi data latih (training set) dan data uji (testing set) dengan rasio 80:20.

Evaluation

Setelah model terbentuk, evaluasi dilakukan dengan mengukur kinerja masing-masing algoritma. Parameter evaluasi meliputi:

- Akurasi: Proporsi prediksi yang benar dari seluruh data uji
- Precision: Ketepatan prediksi kelas positif
- Recall: Kemampuan model dalam mendeteksi kelas positif secara lengkap. Hasil evaluasi digunakan untuk menentukan algoritma yang paling efektif dalam klasifikasi kelulusan mahasiswa.

4. HASIL DAN PEMBAHASAN

Hasil Pengujian:

Pada bagian ini disajikan hasil penelitian yang berupa implementasi aplikasi prediksi kelulusan serta interpretasi dan evaluasi

a. Implementasi

Implementasi dari fungsional membuat dan menampilkan pohon keputusan beserta pengaturan nilai confidence factor dapat dilihat pada tabel 2. Implementasi dari fungsional melakukan pengukuran kinerja dan menampilkan hasilnya disajikan pada tabel 3. Yang terakhir adalah fitur mengidentifikasi kelulusan mahasiswa berdasarkan kriteria yang ditentukan dapat dilihat tampilan implementasinya pada tabel 4, yang hasilnya disajikan pada tabel 5.

Tabel 2. Detail Pengukuran Kinerja

| Pengukuran Ke- | Data Latih | Data Uji | <i>prunning</i> | Nilai confidence |
|-----------------------|-------------------|-----------------|-----------------|-------------------------|
| 1 | 304 | 115 | Tidak | - |
| 2 | 304 | 115 | Ya | 0,25 |
| 3 | 304 | 115 | Ya | 0,4 |
| 4 | 308 | 115 | Tidak | - |
| 5 | 308 | 115 | Ya | 0,25 |
| 6 | 308 | 115 | Ya | 0,4 |
| 7 | 302 | 115 | Tidak | - |
| 8 | 302 | 115 | Ya | 0,25 |
| 9 | 302 | 115 | Ya | 0,4 |
| 10 | 298 | 115 | Tidak | - |
| 11 | 298 | 115 | Ya | 0,35 |
| 12 | 298 | 115 | Ya | 0,4 |
| 13 | 290 | 115 | Tidak | - |
| 14 | 290 | 115 | Ya | 0,25 |
| 15 | 290 | 115 | Ya | 0,4 |

Tabel 3. Confusion Matrix pengukuran ke-1

| | Prediksi <5 tahun | Prediksi ≥5 tahun |
|-----------------|---------------------|---------------------|
| Aktual <5 tahun | True positives: 38 | False negatives: 22 |
| Aktual ≥5 tahun | False positives: 23 | True negatives: 32 |

Tabel 4. Hasil Pengukuran Kinerja

| Pengukuran Ke- | Precision | Recall | Akurasi |
|----------------|-----------|--------|---------|
| 1 | 62.3% | 63.33% | 60.86% |
| 2 | 69.7% | 38.33% | 59.13% |
| 3 | 69.7% | 38.33% | 59.13% |
| 4 | 64% | 55.17% | 61.74% |
| 5 | 66.07% | 63.79% | 65.22% |
| 6 | 64.58% | 53.45% | 61.74% |
| 7 | 60.66% | 60.66% | 58.26% |
| 8 | 69.23% | 59.02% | 64.35% |
| 9 | 69.23% | 59.02% | 64.35% |
| 10 | 65.52% | 60.32% | 60.87% |
| 11 | 83.87% | 41.27% | 63.48% |
| 12 | 83.87% | 41.27% | 63.48% |
| 13 | 67.19% | 64.18% | 60.87% |
| 14 | 80% | 41.79% | 60% |
| 15 | 66.13% | 61.19% | 59.13% |

Tabel 5. Hasil Rata-rata Pengukuran Kinerja

| Pruning | Nilai confidence | precision | Recall | Akurasi |
|---------|------------------|-----------|--------|---------------|
| Tidak | - | 63.93% | 60.73% | 60.52% |
| Ya | 0,25 | 73.77% | 48.84% | 62.44% |
| Ya | 0,4 | 70.70% | 50.56% | 61.57% |

- C4.5: Akurasi 73,49%, Precision 64,62%, Recall 41,89%
- KNN: Akurasi 76,08%, Precision 73,11%, Recall 41,92%

b. Interpretasi

Pola yang dihasilkan dari proses *data mining* dapat ditampilkan dalam bentuk pohon keputusan dan *rules*. Berikut adalah contoh pohon keputusan yang dihasilkan:

```
IPK = Sangat Memuaskan = ≥5 Tahun
IPK = Dengan Pujian
| TOEFL = Dasar = <5 Tahun
| TOEFL = Menengah Bawah = <5 Tahun
| TOEFL = Menengah Atas = <5 Tahun
| TOEFL = Mahir = ≥5 Tahun
IPK = Memuaskan = ≥5 Tahun
```

Dari pohon Keputusan tersebut dapat diambil *rules* sebagai berikut:

- IF IPK sangat memuaskan THEN ≥ 5 tahun
- IF IPK dengan pujian AND TOEFL dasar THEN <5 tahun
- IF IPK dengan pujian AND TOEFL menengah bawah THEN <5 tahun
- IF IPK dengan pujian AND TOEFL menengah atas THEN <5 tahun
- IF IPK dengan pujian AND TOEFL mahir AND asal daerah Jateng AND jenis kelamin Perempuan THEN <5 tahun
- IF IPK dengan pujian AND TOEFL mahir THEN ≥ 5 tahun
- IF IPK memuaskan THEN ≥ 5 tahun

Pembahasan:

KNN menunjukkan performa yang lebih baik dalam hal akurasi dan precision dibandingkan C4.5. Hal ini menunjukkan bahwa KNN lebih sesuai digunakan untuk data dengan struktur yang tidak terlalu kompleks dan pola yang dapat dikenali melalui kedekatan data.

Visualisasi

Model pohon keputusan C4.5 menunjukkan atribut utama yang mempengaruhi kelulusan adalah IPK dan IPS semester akhir. Untuk KNN, hasil prediksi sangat tergantung pada pemilihan nilai k.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa metode KNN memiliki akurasi yang lebih tinggi dibandingkan metode C4.5 dalam memprediksi kelulusan mahasiswa berdasarkan data akademik. Algoritma KNN sangat efektif karena mempertimbangkan kedekatan nilai antar data. Namun, C4.5 juga tetap relevan karena memberikan interpretasi model melalui pohon keputusan.

Berdasarkan kesimpulan dari peneliti yang telah dijelaskan diatas, maka muncul saran untuk

penelitian ini ke depannya yaitu:

- Penambahan jumlah data dan atribut dapat meningkatkan akurasi model.

- Diperlukan eksperimen dengan algoritma lain untuk memperluas wawasan komparasi.
- Pengembangan sistem berbasis web yang mengintegrasikan hasil prediksi kelulusan secara real-time dapat membantu pihak akademik.

DAFTAR REFERENSI

- Amri, Z., Kusrini, & Kusnawi. (2023). Prediksi tingkat kelulusan mahasiswa menggunakan algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM. *Edumatic: Jurnal Pendidikan Informatika*, 7(2), 187–196. <https://doi.org/10.29408/edumatic.v7i2.18620>
- Anggresta, V. (2015). Analisis faktor-faktor yang memengaruhi belajar mahasiswa Fakultas Ekonomi Universitas Negeri Padang. *Journal of Economic and Economic Education*, 4, 19-29. <https://doi.org/10.22202/economica.2015.v4.i1.325>
- Azizah Adha, D. A. R., Allanda, A. N., & Fatmasari, D. A. (2023). Performansi algoritma C4.5 untuk prediksi kelulusan mahasiswa. *Jurnal Cakrawala Informasi*, 3(2). <https://doi.org/10.54066/jci.v3i2.339>
- Devinta, M. (2015). Fenomena culture shock (gegar budaya) pada mahasiswa perantauan di Yogyakarta. *Jurnal Pendidikan Sosiologi*.
- Fitriyanti, V., & Hermawan, D. (2021). Klasifikasi predikat kelulusan mahasiswa menggunakan algoritma C4.5. *Jurnal Saintekom*, 11(1), 45-52.
- Hasibuan, T., & Mahdiana, D. (2023). Prediksi kelulusan mahasiswa tepat waktu menggunakan algoritma C4.5 pada UIN Syarif Hidayatullah Jakarta. *SKANIKA*, 6(1), 61–74. <https://doi.org/10.36080/skanika.v6i1.2976>
- Kusnia, Y. (2017). Pengaruh karakteristik gender dan motivasi belajar terhadap prestasi belajar matematika siswa Kelas X IPA 1 di MAN 2 Semarang. (Tesis).
- Purwanto, E., Kusrini, & Sudarmawan. (2024). Prediksi kelulusan tepat waktu menggunakan metode C4.5 dan K NN. *Techno Jurnal Informatika*, 15(2), 78-86.
- Putri, S. A. H., Ekastini, & Akhir Putra, J. (2024). Analisis komparasi algoritma C4.5, Naive Bayes dan K-Nearest Neighbor untuk memprediksi ketepatan waktu lulus mahasiswa. *Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP)*, 7(2), 172–184. <https://doi.org/10.34012/jutikomp.v7i2.5575>
- Quinlan, J. R. (1993). C4.5: Programs for machine learning. San Mateo, CA: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-08-050058-4.50004-8>
- Rahmayanti, A., Rusdiana, L., & Suratno, S. (2022). Perbandingan metode algoritma C4.5 dan Naive Bayes untuk memprediksi kelulusan mahasiswa. *Walisoongo Journal of Information Technology*, 4(1), 193–199. <https://doi.org/10.21580/wjat.2022.4.1.9654>
- Saifudin, A., & Wahono, R. S. (2015). Pendekatan level data untuk menangani ketidakseimbangan kelas pada prediksi cacat software. *Journal of Software Engineering*, 1, 76-85.
- United authors (Abu Tholib et al.). (2023). Comparison of C4.5 and Naive Bayes for predicting student graduation using machine learning algorithms. *International Journal of Engineering and Computer Science Applications*, 2(2), 71–78. <https://doi.org/10.30812/ijecsa.v2i2.3364>

Wati, E. F., & Rudianto, B. (2023). Penerapan algoritma KNN, Naive Bayes dan C4.5 dalam memprediksi kelulusan mahasiswa. *FORMAT: Jurnal Ilmiah Komputer*, 13(2), 90-97. <https://doi.org/10.22441/format.2022.v11.i2.009>

Widaningsih, S. (2019). Perbandingan metode data mining C4.5, Naive Bayes, KNN dan SVM untuk prediksi kelulusan mahasiswa. *Jurnal Tekno*, 8(1), 12-20.

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>