

# ANALISIS KOMPARASI ALGORITMA C4.5, NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI KETEPATAN WAKTU LULUS MAHASISWA

Shakira Azzahra Hadi Putri<sup>a</sup>, Ekastini<sup>b\*</sup>, Juniardi Akhir Putra<sup>c</sup>

<sup>a,b,c</sup>*Fakultas Rekayasa Sistem, Universitas Teknologi Sumbawa*

*Corresponding Author:*

<sup>b</sup>*ekastini@uts.ac.id*

## ABSTRAK

Permasalahan kelulusan mahasiswa pada perguruan tinggi menjadi salah satu hal yang sangat penting dalam menunjukkan kualitas pembelajaran di perguruan tinggi terutama di kampus Universitas Teknologi Sumbawa (UTS). Tujuan dari penelitian ini adalah membandingkan tiga metode algoritma yaitu C4.5, Naive Bayes dan K-Nearest Neighbor (KNN) yang lebih baik dalam memprediksi ketepatan waktu lulus mahasiswa dengan menggunakan tools RapidMiner dengan metode Knowledge Discovery in Database (KDD). Dataset yang digunakan ketiga klasifikasi yaitu 330 data mahasiswa Informatika. Berdasarkan perbandingan dari ketiga algoritma tersebut dengan teknik splitting data, didapatkan bahwa algoritma C4.5 menghasilkan akurasi sebesar 73.49% dengan presisi sebesar 64.62% dan recall 41.89%. Algoritma Naive Bayes menghasilkan akurasi sebesar 72.79% dengan presisi 64.06% dan recall 38.11%. Sedangkan algoritma K-Nearest Neighbor (KNN) menghasilkan akurasi sebesar 76.08% dengan presisi 73.11% dan recall 41.92%. Dari hasil perbandingan ketiga algoritma tersebut, yang paling tepat dalam memprediksi ketepatan waktu lulus mahasiswa adalah algoritma K-Nearest Neighbor (KNN).

**Kata Kunci:** Data Mining, C4.5, Naive Bayes, K-Nearest Neighbor, RapidMiner

## ABSTRACT

The problem of student graduation in higher education is one of the most essential things in showing the quality of learning in higher education, especially on the Sumbawa University of Technology (UTS) campus. The purpose of this research is to compare three algorithm methods, namely C4.5, Naive Bayes, and K-Nearest Neighbor (KNN), which is better at predicting the timeliness of student graduation using RapidMiner tools with the Knowledge Discovery in Database (KDD) method. The dataset used by the three classifications is 330 Informatics student data. Based on the comparison of the three algorithms with data splitting techniques, it is found that the C4.5 algorithm produces an accuracy of 73.49% with a precision of 64.62% and a recall of 41.89%. The Naive Bayes algorithm produces an accuracy of 72.79% with a precision of 64.06% and a recall of 38.11%. Meanwhile, the K-Nearest Neighbor (KNN) algorithm produces an accuracy of 76.08% with a precision of 73.11% and a recall of 41.92%. From the comparison of the three algorithms, the most

appropriate for predicting the timeliness of student graduation is the K-nearest neighbor (KNN) algorithm.

**Keywords:** Data Mining, C4.5, Naive Bayes, K-Nearest Neighbour, RapidMiner

## PENDAHULUAN

Meningkatnya kualitas pendidikan dari tahun ke tahun adalah salah satu cara untuk mengukur keberhasilan pendidikan, terutama di perguruan tinggi. Tingkat kelulusan mahasiswa adalah salah satu indikatornya (Putri, et al., 2023). Tingkat kelulusan ini menunjukkan seberapa baik institusi pendidikan mendidik mahasiswa dan membimbing mereka menuju tingkat kelulusan yang memuaskan.

Universitas Teknologi Sumbawa adalah salah satu perguruan tinggi swasta yang berada di wilayah Nusa Tenggara Barat (NTB). Pada awal berdirinya Universitas Teknologi Sumbawa (UTS) pada tahun 2013 memiliki 4 program studi yaitu program studi metalurgi, teknik mesin, teknik informatika dan bioteknologi. Seiring berjalannya waktu, program pendidikan sarjana (S1) yang ada di UTS bertambah hingga sekitar 32 program studi dari 8 fakultas (UTS, 2024). Teknik Informatika yang kini sudah berganti nama menjadi Informatika merupakan salah satu program studi yang berada di fakultas Rekayasa Sistem. Berdasarkan data yang diperoleh melalui dari Sekretaris Program Studi Informatika UTS, tercatat ada sebanyak 1.178 mahasiswa informatika hingga pada tahun 2024.

Berdasarkan **Tabel 1** terlihat bahwa jumlah penerimaan mahasiswa program studi Informatika terus bertambah. Tetapi, terhitung tidak sampai setengah dari jumlah mahasiswa yang dapat lulus secara tepat waktu. Hal ini menyebabkan jumlah mahasiswa terus bertambah dan menghasilkan volume data mahasiswa semakin besar.

**Tabel 1. Mahasiswa Informatika UTS Tahun 2015-2019**

Angkatan	Mahasiswa	Lulus Tepat Waktu	Tidak Tepat Waktu	Dikeluarkan	Mengundurkan Diri
2015	117	64	6	25	6
2016	101	31	31	24	7
2017	133	48	17	20	10
2018	144	39	37	25	34
2019	136	39	18	15	25
<b>Jumlah</b>	<b>631</b>	<b>221</b>	<b>109</b>	<b>109</b>	<b>82</b>

Salah satu cara untuk mengatasi adanya ketidaktepatan waktu lulus mahasiswa adalah dengan memanfaatkan algoritma machine learning yang dapat memproses data dan membuat prediksi berdasarkan pola yang terlihat. Tiga algoritma yang paling umum digunakan adalah C4.5, Naive Bayes dan K-Nearest Neighbor (KNN). Ketiga algoritma tersebut memiliki kelebihan dan kekurangannya sendiri sehingga perlu dilakukan analisis komparasi untuk mengetahui mana yang paling efektif dalam memprediksi ketepatan waktu lulus mahasiswa.

Pada penelitian yang dilakukan oleh (Putri, et al., 2023) telah dilakukan perbandingan komparasi antara algoritma *KNN*, *Naive Bayes* dan *SVM*. Hasil akhir dari penelitian yang dilakukan didapatkan bahwa algoritma *KNN* adalah algoritma yang paling cocok dalam melakukan prediksi kelulusan tingkat akhir. Sedangkan pada penelitian lainnya yang

dilakukan oleh (Rahmayanti, Ili, & Suratno, 2022) dilakukan perbandingan antara algoritma *C4.5* dan *Naive Bayes*. Hasil akhir didapatkan bahwa algoritma *C4.5* merupakan algoritma yang paling cocok dalam melakukan prediksi ketepatan waktu lulus mahasiswa.

Penelitian sebelumnya telah menunjukkan bahwa algoritma *C4.5*, *Naive Bayes* dan *KNN* dapat digunakan dalam memprediksi ketepatan waktu lulus mahasiswa. Namun, belum ada studi yang secara khusus membandingkan ketiga algoritma tersebut dalam memprediksi ketepatan waktu lulus mahasiswa. Oleh karena itu, penelitian ini bertujuan untuk melakukan perbandingan antara algoritma *C4.5*, *Naive Bayes* dan *KNN* untuk memprediksi ketepatan waktu lulus mahasiswa. Hasil penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang kelebihan dan kekurangan masing-masing algoritma dalam memprediksi ketepatan waktu lulus mahasiswa serta dapat digunakan sebagai acuan untuk menentukan langkah dan kebijakan bagi mahasiswa untuk yang menargetkan lulus tepat waktu.

Hasil penelitian ini berkontribusi terhadap pengetahuan *Data Mining* maupun bagi pihak-pihak yang terkait dengan kelulusan mahasiswa. Bagi *Data Mining*, hasil penelitian ini berperan dalam memperkaya ilmu dan pemahaman terkait algoritma klasifikasi yang cocok digunakan dalam memprediksi ketepatan waktu lulus mahasiswa. Bagi pihak lain, terutama universitas, hasil penelitian ini dapat berguna sebagai informasi sehingga didapatkannya sebuah strategi atau solusi dalam mengatasi faktor-faktor yang menyebabkan ketidaktepatan waktu lulus mahasiswa lainnya.

## LITERATUR REVIEW

Dalam penelitian ini, penulis menggunakan beberapa tinjauan pustaka sebagai dasar teori. Berikut beberapa dasar teori yang digunakan.

### 1. Komparasi Algoritma

Komparasi menurut Wikipedia diartikan sebagai perbandingan. Dengan kata lain, komparasi adalah proses memeriksa dua atau lebih objek untuk menentukan relevansinya, perbedaan, dan tingkat perbedaan masing-masing (Comparison, 2023). Berdasarkan pengertian tersebut, komparasi algoritma diartikan sebagai proses membandingkan berbagai algoritma dalam menyelesaikan suatu masalah dengan cara mempelajari persamaan, perbedaan, dan keunggulan di antara mereka.

Komparasi algoritma bertujuan untuk meningkatkan pemahaman tentang topik yang dibandingkan atau untuk membuat keputusan yang lebih baik berdasarkan hasil analisis. Komparasi algoritma digunakan untuk mendapatkan pemahaman yang lebih baik tentang algoritma yang dibandingkan, menemukan peluang dan ancaman, menemukan kekuatan dan kelemahan, dan membuat keputusan yang lebih baik (Gunawan, Zarlis, & Roslina, 2021).

### 2. Algoritma Klasifikasi

#### a. Algoritma C4.5

Algoritma decision tree menggunakan klasifikasi data mining untuk mengonstruksi pohon keputusan dalam pelatihan data dengan rekaman dalam basis data (Suntoro, 2019). Algoritma ini dibagi menjadi beberapa jenis yaitu ID3, C4.5, C5.0, CHAID, dan MARS. Menurut (Suwitno, 2017) algoritma

C4.5 dikembangkan oleh J. Ross Quinlan dan diberi nama C4.5 karena merupakan turunan dari algoritma ID3. Algoritma ini mampu mengolah data numerik dan kategorik serta mempertahankan akurasi model meskipun terjadi perubahan pada data.

Untuk memilih atribut menjadi node, algoritma C4.5 menggunakan perhitungan entropy, information gain, split info dan gain ratio (Suntoro, 2019).

$$SplitInfo_A(D) = - \sum_{j=1}^p \frac{|D_j|}{|D|} \times \log_2 \left( \frac{|D_j|}{|D|} \right) \quad (1)$$

$$GainRatio = \frac{Gain(S,A)}{SplitInfo_A(D)} \quad (2)$$

Algoritma C4.5 bekerja dengan membangun pohon keputusan untuk merepresentasikan aturan untuk mengklasifikasi data baru (Anisa & Mesran, 2018). Cara kerja algoritma C4.5 sebagai berikut:

- 1) Persiapan data
- 2) Pemilihan atribut akar
- 3) Pembentukan cabang pohon
- 4) Pemangkasan pohon (opsional)
- 5) Prediksi

b. Algoritma Naive Bayes

Pendekatan Naive Bayesian pertama kali diterapkan oleh Pendeta Thomas Bayes (1702-1761) dalam esainya (1763) *Essay Towards a Problem in the Doctrine of Chances* (Suwitno, 2017). Menurut Thomas, Naive Bayes merupakan prediksi kemungkinan masa depan berdasarkan pengalaman masa lalu. Hal ini didasarkan pada asumsi penyederhanaan bahwa nilai atribut tidak bergantung satu sama lain secara kondisional, mengingat nilai awalnya. Nilai keluaran, atau probabilitas yang dipertimbangkan bersama-sama, adalah hasil dari probabilitas individu.

Teorema Bayesian menghitung nilai posterior probability  $P(H|X)$  menggunakan probabilitas  $P(H)$ ,  $P(X)$ , dan  $P(X|H)$ , di mana nilai  $X$  adalah data testing yang kelasnya belum diketahui. Nilai  $H$  adalah hipotesis data  $X$  yang merupakan suatu kelas yang lebih spesifik. Nilai  $P(X|H)$  disebut likelihood yang merupakan probabilitas hipotesis  $X$  berdasarkan kondisi  $H$ . Nilai  $P(H)$  disebut prior probability merupakan probabilitas hipotesis  $H$ . Sedangkan nilai  $P(X)$  disebut predictor prior probability merupakan probabilitas  $X$ .

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (3)$$

Langkah-langkah algoritma *Naive Bayes* adalah sebagai berikut:

1. Mempersiapkan data.
2. Menentukan kelas data *training*.
3. Menghitung jumlah data yang serupa dengan kelas yang sama.
4. Kalikan semua hasil sesuai dengan data *testing* untuk menentukan kelas baru.

5. membandingkan hasil per kelas, kelas baru merupakan kelas dengan nilai yang paling tinggi.

c. Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor (KNN) merupakan metode pencarian menghitung jarak antara kasus baru dan kasus lama. Metode ini dikenal sebagai algoritma pembelajar yang malas (*lazy learners*) karena hanya melihat jarak dengan tetangga (*neighbor*). Menurut (Harrington, 2012) algoritma KNN sering digunakan peneliti karena memiliki akurasi tinggi, insentive terhadap outlier, dan tidak ada asumsi terhadap data. Namun kelemahannya yaitu perlu menentukan nilai *k* optimal, komputasi mahal dan membutuhkan banyak memori.

Rumus *Eclidean Distance*:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

di mana,

- $d(x, y)$  adalah jarak antara data  $x$  ke data  $y$
- $x_i$  adalah data *testing* ke- $i$
- $y_i$  adalah data *training* ke- $i$

Langkah-langkah algoritma *KNN* adalah:

1. Mempersiapkan data.
2. Menentukan nilai  $k$  secara manual.
3. Menghitung jarak antara data *training* dan data *testing* dengan metode *Euclidean Distance*.
4. Mengurutkan data *training* dari jarak terkecil.
5. menentukan kelas dengan kelas yang dipilih adalah kelas dengan jumlah nilai  $k$  terbanyak pada data *testing*.

3. Kelulusan Mahasiswa

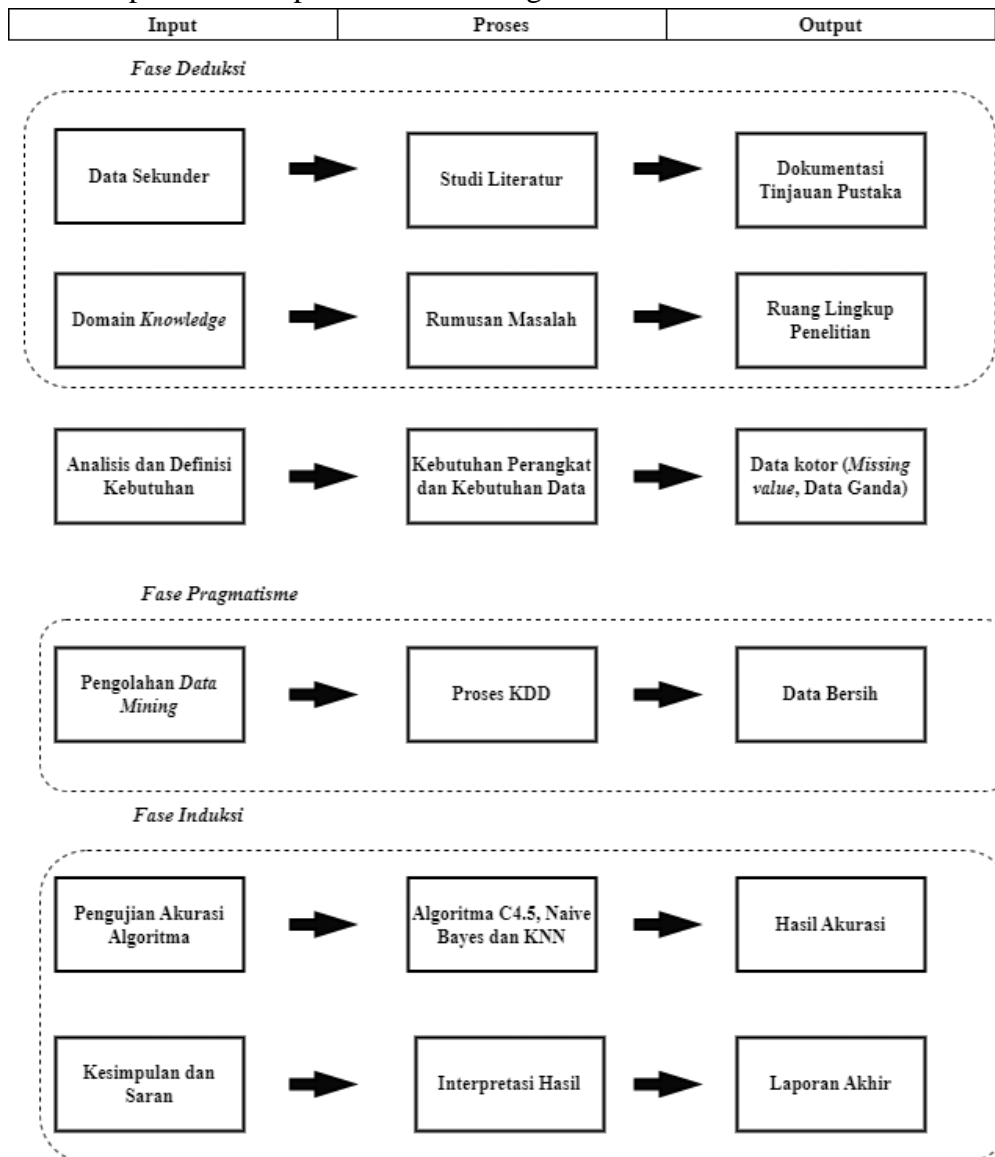
Dibandingkan dengan kelompok usia lainnya atau kelompok usia lainnya, mahasiswa adalah kelompok masyarakat dengan intelektualitas yang lebih luas. Intelektual ini membantu siswa bersaing di dunia kerja karena memungkinkan mereka untuk menangani dan menyelesaikan masalah sehari-hari. (Cahyani & Setyawati, 2016).

Kelulusan mahasiswa diatur dalam peraturan hukum dan peraturan menteri, yaitu Permendikbudristek No. 53 Tahun 2023. Kelulusan mahasiswa merupakan hal penting yang harus diperhatikan, karena menjadi salah satu faktor yang memengaruhi penilaian pemerintah serta status akreditasi sebuah program studi. Salah satu parameter keberhasilan mahasiswa dalam menyelesaikan masa studinya dan memperoleh gelar sarjana adalah lulus tepat waktu. Mahasiswa dikatakan lulus tepat waktu selama kurang dari atau sama dengan 8 semester atau 4 tahun, sedangkan jika lebih dari 4 tahun maka dikatakan lulus tidak tepat waktu (Amelia, Lumenta, & Jacobus, 2017).

## METODE

Pada dasarnya, metode penelitian merupakan teknik ilmiah untuk mengumpulkan data dengan tujuan menemukan jawaban atas suatu pertanyaan atau masalah dengan sabar, cermat, terencana, dan sistematis (Sugiyono, 2019). Dengan kata lain, metode penelitian adalah pendekatan ilmiah yang bertujuan untuk mendapatkan fakta serta menguji kebenarannya. Tahapan penelitian digambarkan sebagai berikut:

Dengan merujuk pada urutan langkah-langkah yang ada pada gambar, prosedur yang dilakukan dalam penelitian dapat diuraikan sebagai berikut:



**Gambar 1. Metode Penelitian**

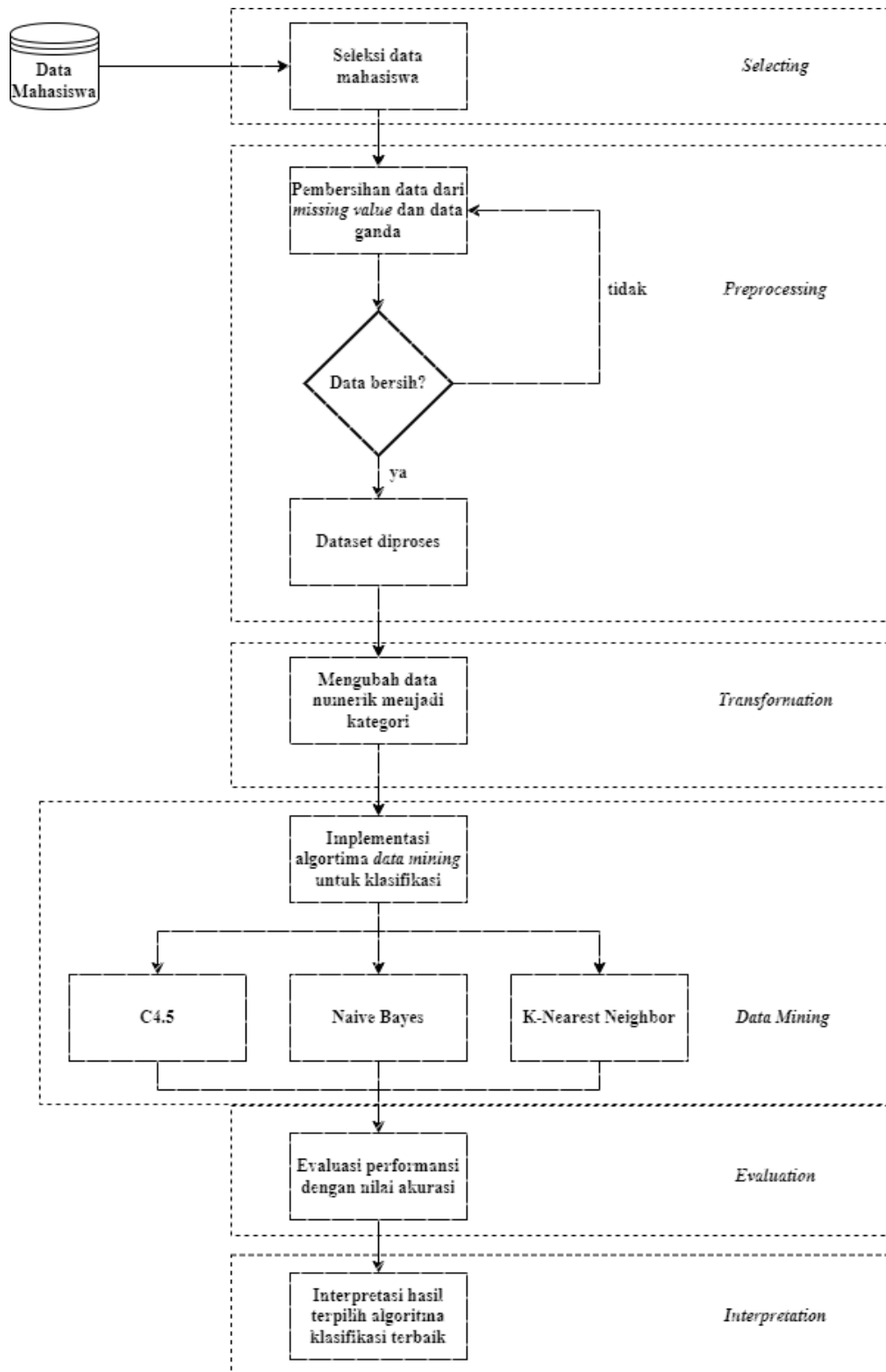
1. Data sekunder digunakan untuk mengidentifikasi masalah dan menentukan kriteria evaluasi data. Membaca artikel ilmiah, buku, jurnal dan referensi lain yang relevan dengan topik penelitian merupakan bagian yang sangat penting dari tinjauan pustaka.
2. *Knowledge* adalah informasi yang sudah disintesis dan dirangkum dalam organisasi untuk meningkatkan pemahaman dan pengetahuan. Masalah yang sudah dirumuskan akan menghasilkan *output* berupa ruang lingkup penelitian, yaitu suatu metode pembatasan permasalahan dalam ilmu yang akan dikaji dalam kajian ilmiah.

3. Analisis dan penentuan kebutuhan melibatkan identifikasi semua kebutuhan sistem melalui analisis kebutuhan perangkat dan kebutuhan data.
4. Analisis data melibatkan proses *Knowledge Discovery in Database* (KDD). Dalam proses ini, akan dilakukan pembersihan data dari *missing value*, data kosong dan data duplikat/ganda dengan menggunakan metode KDD. Tujuannya adalah untuk menghasilkan data baru yang bersih dan sesuai dengan kebutuhan penelitian.
5. Pengujian akurasi algoritma  
Pada tahap ini dilakukan pengujian akurasi terhadap data oleh masing-masing algoritma untuk kemudian didapatkan akurasi perbandingannya.
6. Kesimpulan dan saran  
Pada tahap ini, hasil penelitian akan disajikan. Selanjutnya, di bagian saran, peneliti akan membuat saran untuk digunakan sebagai referensi untuk penelitian serupa di masa mendatang.

Metode analisis data yang digunakan pada penelitian ini memanfaatkan pengolahan data mining mengikuti proses Knowledge Discovery in Database (KDD) untuk menghasilkan informasi sesuai dengan alur yang ditentukan, meliputi tahapan sebagai berikut:

ti tahapan sebagai berikut:

1. Pada tahap *selection*, dilakukan pemilihan data lulusan yang terdiri dari variabel-variabel prediktor dan satu variabel target. Variabel target adalah klasifikasi kelulusan tepat waktu, yaitu lulus dalam waktu 4 tahun atau kurang dengan IPK minimal 3,00. Sedangkan variabel prediktor adalah jenis kelamin, indeks prestasi semester 4-6, dan total SKS.
2. Pada tahap *preprocessing*, dilakukan pembersihan data untuk mendapatkan data yang relevan. Kegiatan ini meliputi memeriksa, memperbaiki, dan membuang data yang tidak relevan seperti *missing value*, data kosong dan data ganda.
3. Pada tahap *transformation*, dilakukan perubahan skala pengukuran data asli menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi analisis yang sesuai untuk diproses.
4. Pada tahap data mining, dipilih algoritma data mining yang sesuai. Untuk fungsi klasifikasi, digunakan algoritma *C4.5*, *Naive Bayes*, dan *KNN*. Pada tahap ini, data akan dibagi (*splitting data*) menjadi dua bagian, yaitu data pelatihan (*training*) dan data pengujian (*testing*). Rasio pembagian data yang digunakan adalah 90:10, 80:20, 70:30, 60:40 dan 50:50.
5. Pada tahap *evaluation*, dilakukan evaluasi terhadap hasil prediksi yang dihasilkan oleh ketiga algoritma. Algoritma yang menghasilkan nilai akurasi paling tinggi akan dipilih.



Gambar 2. Analisa Proses KDD

## HASIL

### 1. Pengumpulan Data

Dalam penyusunan skripsi ini, teknik pengumpulan data yang digunakan adalah dokumentasi. Dalam proses ini, dilakukan pengumpulan data dan informasi mahasiswa



Informatika dari Sekretaris Program Studi Informatika, khususnya informasi data mahasiswa Informatika tahun 2015 hingga 2019.

2. Proses Analisa Data

a) Data Sekunder

Pada proses ini data didapatkan secara langsung dari Sekretaris Program Studi Informatika di Universitas Teknologi Sumbawa. Data yang dikumpulkan merupakan data akademik mahasiswa selama periode 2015-2019 dengan total 631 data. Dari data tersebut, terdiri dari 221 data mahasiswa yang lulus tepat waktu dan 109 data mahasiswa yang tidak lulus tepat waktu. Data-data ini selanjutnya akan digunakan sebagai data pelatihan sekaligus data pengujian, dengan menerapkan pemisahan data untuk menguji akurasi.

3. *Data Selection*

Ditahap awal, sumber data yang sudah didapat akan dipisah antara data yang akan digunakan dan data yang tidak digunakan. Data yang digunakan adalah data yang sesuai dengan variabel perhitungan yang diperlukan untuk dilakukan analisis.

**Tabel 2. Data Awal Mahasiswa Informatika Tahun 2015-2019**

NIM	Tanggal Masuk	Jenis Kelamin	IPS 4	IPS 5	IPS 6	IPK	SKS	Status
15001	2015	L	3.87	3.82	3.28	3.58	135	Tidak Tepat Waktu
15005	2015	L	3.50	3.31	3.24	3.16	132	Tidak Tepat Waktu
15010	2015	L	3.85	3.58	3.30	3.35	132	Tidak Tepat Waktu
...	...	...	...	...	...	...	...	...
19076	2019	L	3.50	2.93	3.74	3.26	130	Tepat Waktu

4. *Data Preprocessing*

Pada tahap ini dilakukan proses pembuatan dataset dengan dilakukan pembersihan data dari missing value, ganda atau kosong karena keberadaannya dapat mempengaruhi tingkat akurasi dari hasil data mining nantinya. Tahap processing ini dilakukan secara manual menggunakan excel dari data yang telah diperoleh. Tahap final dari dilakukannya tahap preprocessing didapatkan data bersih bebas dari missing value, ganda dan kosong berjumlah 279 data.

5. *Data Transformation*

Pada proses ini, data akan diubah kedalam bentuk data yang sesuai dengan proses pengolahan data yang dapat dianalisis. Pada tahap ini dilakukan penyederhanaan atribut dan menghasilkan *dataset final*. Jumlah atribut yang digunakan sesuai dengan kebutuhan penelitian. Berikut penentuan atribut serta pembagian skala atribut:

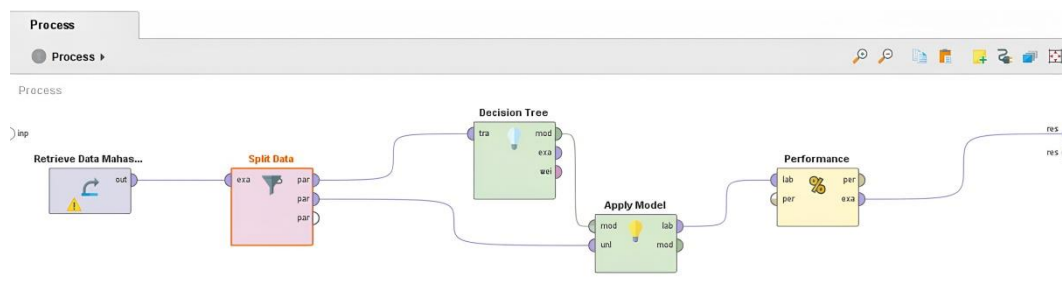
**Tabel 3. Data Mahasiswa Setelah Transformasi**

NIM	Angkatan	Jenis Kelamin	IPS 4	IPS 5	IPS 6	IPK	SKS	Status
15001	2015	L	Besar	besar	sedang	besar	besar	Tidak Tepat Waktu

NIM	Angkatan	Jenis Kelamin	IPS 4	IPS 5	IPS 6	IPK	SKS	Status
15005	2015	L	Besar	sedang	sedang	sedang	besar	Tidak Tepat Waktu
15006	2015	L	Besar	besar	sedang	sedang	besar	Tidak Tepat Waktu
15008	2015	L	Sedang	sedang	sedang	sedang	besar	Tidak Tepat Waktu

### 6. Data Mining

Dalam proses ini dilakukan penerapan teknik *data mining* untuk menemukan pola atau model dari data, menggunakan hasil transformasi data yang telah dilakukan sebelumnya. Proses penerapan *data mining* dilakukan dengan bantuan perangkat lunak *RapidMiner*, dengan menggunakan operator-operator yang tersedia pada *RapidMiner* untuk mengolah data.



**Gambar 3. Model Klasifikasi Algoritma**

Gambar di atas merupakan gambar klasifikasi algoritma C4.5. model serupa diterapkan pada algoritma Naive Bayes dan KNN untuk mendapatkan hasil pengujian akurasi.

### 7. Evaluation dan Interpretation

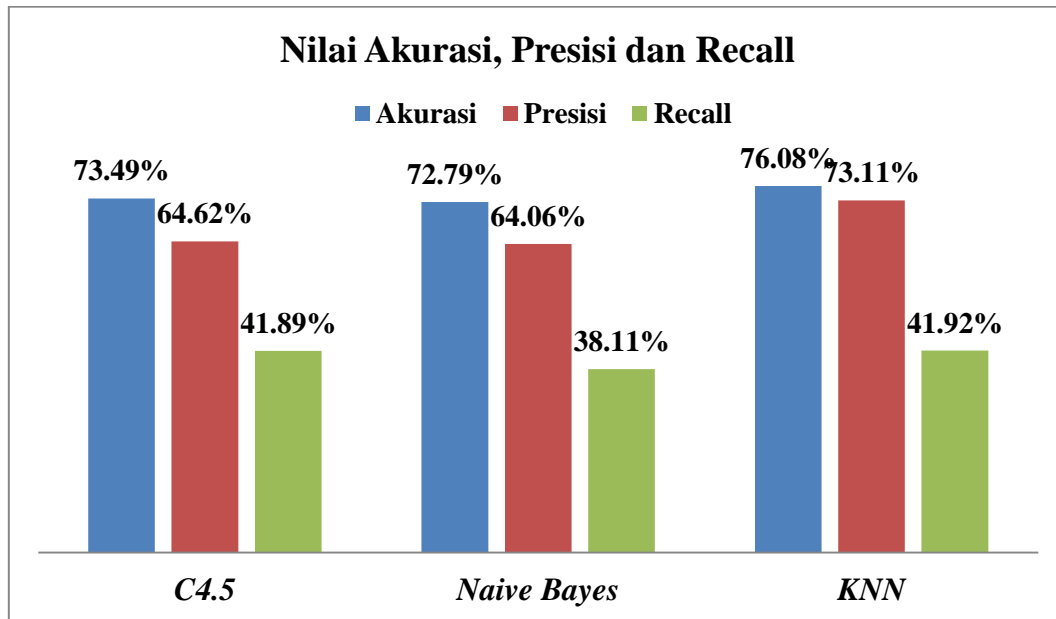
Berdasarkan pengujian akurasi *data mining* yang telah dilakukan maka beberapa metode yakni *C4.5*, *Naive Bayes* dan *KNN* layak digunakan sebagai model prediksi ketepatan waktu lulus mahasiswa informatika UTS, kelayakan metode yang dapat didukung dengan tingkat akurasi dari ketiga metode yang digunakan dalam penelitian ini ditunjukkan pada tabel dibawah ini.

**Tabel 4. Hasil Komparasi Ketiga Algoritma**

	C4.5	Naive Bayes	KNN
Accuracy	73.49	72.79	76.08
Precision	64.62	64.06	73.11
Recall	41.89	38.11	41.92

**Tabel 4** membandingkan akurasi, presisi dan recall dari setiap algoritma. Terlihat bahwa nilai akurasi algoritma *KNN* merupakan akurasi yang paling tinggi jika dibandingkan dengan algoritma lainnya. Nilai presisi dan recall yang paling tinggi juga dihasilkan oleh

algoritma *KNN*. Grafik nilai akurasi, presisi dan recall masing-masing algoritma terlihat pada **Gambar 4** berikut.



**Gambar 4. Nilai Akurasi, Presisi dan Recall Masing-Masing Algoritma**

Berdasarkan **Gambar 4** menunjukkan hasil perbandingan performa ketiga algoritma yaitu C4.5, Naive Bayes dan KNN berdasarkan tiga matriks evaluasi yaitu akurasi, presisi dan recall. Algoritma C4.5 menghasilkan akurasi sebesar 73.49% dengan presisi 64.62 dan recall 41.89%. Algoritma Naive Bayes menghasilkan akurasi sebesar 72.79 dengan presisi 64.06% dan recall 38.11%. Sedangkan algoritma KNN menghasilkan akurasi sebesar 76.08% dengan presisi 73.11% dan recall 41.92%.

#### 8. *Knowledge*

Berdasarkan hasil evaluasi yang telah dilakukan sebelumnya dan visualisasi yang terdapat pada **Gambar 4** dapat ditarik kesimpulan bahwa dari ketiga algoritma klasifikasi yang digunakan dalam mengolah data, algoritma KNN lah yang memiliki tingkat akurasi paling tinggi yakni sebesar 76.08% dengan presisi 73.11% dan recall 41.92%. Dari tabel hasil komparasi maupun visualisasi terlihat jelas bahwa algoritma KNN lebih unggul dalam akurasi, presisi dan recall jika dibandingkan dengan dua algoritma lainnya. Dengan demikian, dapat disimpulkan bahwa algoritma KNN merupakan algoritma yang paling tepat dalam hal memprediksi ketepatan waktu lulus mahasiswa.

## SIMPULAN

Setelah melalui langkah-langkah penelitian dalam Analisis Komparasi algoritma C4.5, Naive Bayes dan K-Nearest Neighbor Dalam Memprediksi Ketepatan Waktu Lulus Mahasiswa Informatika UTS dapat ditarik kesimpulan bahwa hasil komparasi algoritma dihasilkan bahwa tingkat akurasi algoritma C4.5 sebesar 73.49% presisi 64.62% dan recall 41.89%. Algoritma Naive Bayes menghasilkan akurasi sebesar 72.79% presisi 64.06% dan recall 38.11%. Algoritma KNN menghasilkan akurasi sebesar 76.08% presisi 73.11% dan recall 41.92%. Dari komparasi ketiga algoritma, diketahui bahwa algoritma KNN memiliki tingkat akurasi paling tinggi yaitu sebesar 76.08% dan tepat untuk memprediksi ketepatan waktu lulus mahasiswa. Dari penelitian yang telah dilakukan, terdapat beberapa hal yang dapat

mempengaruhi ketepatan waktu lulus mahasiswa diantaranya jumlah SKS, nilai IP Semester dan IPK. Setelah dilakukan penelitian dan hasil analisis ketepatan waktu lulus mahasiswa, penelitian ini dapat dijadikan acuan sebagai pengambilan keputusan untuk langkah berikutnya dalam mengatasi adanya ketidaktepatan waktu lulus mahasiswa. CONCLUSION

## SARAN

Berdasarkan kesimpulan dari peneliti yang telah dijelaskan diatas, maka muncul saran untuk penelitian ini ke depannya yaitu:

1. Dapat menambahkan jumlah data mahasiswa serta atribut yang digunakan sehingga analisis komparasi algoritma dapat dilakukan dalam skala besar dengan keakuratan yang tinggi.
2. Diharapkan menggunakan algoritma *data mining* lainnya sehingga nantinya dapat dijadikan sebagai bahan pertimbangan hasil akurasi algoritma mana yang jauh lebih baik.
3. Diharapkan menggunakan *tools* pendukung berbeda selain *RapidMiner*.
4. Diharapkan adanya strategi atau solusi yang efektif dalam mengatasi peningkatan lulus tepat waktu dengan memperhatikan beberapa faktor seperti IP Semester dan IPK yang rendah.
5. Dapat dilakukan pengembangan penelitian berupa Sistem Pengambilan Keputusan (SPK) guna untuk mengatasi beberapa faktor yang menyebabkan adanya ketidaktepatan waktu lulus mahasiswa.

## DAFTAR PUSTAKA

- Amelia, M. W., Lumenta, A. S., & Jacobus, A. (2017). Prediksi Masa Studi Mahasiswa dengan Menggunakan Algoritma Naive Bayes. *E-Journal Teknik Informatika*, 8301-8364.
- Cahyani, H., & Setyawati, R. W. (2016). Pentingnya Peningkatan Kemampuan Pemecahan Masalah melalui PBL untuk Mempersiapkan Generasi Unggul Menghadapi MEA. *Seminar Nasional Matematika X Universitas Negeri Semarang 2016*, 151-152.
- Comparison. (2023, Agustus 16). Diambil kembali dari Wikipedia: <https://en.wikipedia.org/>
- Gunawan, M., Zarlis, M., & Roslina. (2021). Analisis Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Media Informatika Budidarma*, 513-523.
- Harrington, P. (2012). *Machine Learning in Action*. Shelter Island, New York: Manning Publications Co.
- Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. Canada: A John Wiley & Sons, Inc.
- Karamouzis, S., & Andreas, V. (2009). *Sensitivity Analysis of Neural Network Parameters for Identifying the Factors for College Student Success*. Los Angeles: WIR World Congress on Computer Science and Information Engineering.
- Putri, A., Hardiana, C. S., Elma, N., Siregar, F. T., Rahmadden, & Wahyuni, R. (2023). Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan

- Mahasiswa Tingkat Akhir. *MALCOM: Indonesia Journal of Machine Learning and Computer Science*, 20-26.
- Rahmayanti, A., Ili, R. U., & Suratno, S. (2022). Perbandingan Metode Algoritma C4.5 dan Naive Bayes untuk Memprediksi Kelulusan Mahasiswa. *Walisongo Journal of Information Technology*, 11-22.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Bandung: ALFABETA.
- Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: PT Elex Media Komputindo.
- Suwitno. (2017). Prototype for Predicting Graduation on Time at Buddhi Dharma University Using a Comparison of the C4.5 and Naive Bayes Algorithms. *JURNAL TECH-E*, 29-36.
- UTS. (2024). *Akademik*. Diambil kembali dari Universitas Teknologi Sumbawa: <https://uts.ac.id/>