

KLASIFIKASI ALGORITMA *K-NEAREST NEIGHBOR*, *NAIVE BAYES*, *DECISION TREE* UNTUK PREDIKSI STATUS KELULUSAN MAHASISWA S1

¹⁾ Enggar Novianto, ²⁾ Arief Hermawan, ³⁾ Donny Avianto

^{1,2,3)} Magister Teknologi Informasi, Universitas Teknologi Yogyakarta

^{1,2,3)} Jl. Ring Road Utara, Sleman, D.I. Yogyakarta, Indonesia

E-mail : ¹enggarnoviyanto@gmail.com, ²ariefdb@uty.ac.id, ³donny@uty.ac.id

ABSTRAK

Mahasiswa merupakan faktor krusial yang harus diperhatikan dalam mengevaluasi program studi secara serius. Indikator keberhasilan program studi adalah lamanya dalam menyelesaikan studi. Masa studi adalah waktu dimana mahasiswa menyelesaikan studinya. Selain itu, waktu belajar mahasiswa mencerminkan tingkat kinerja belajar mahasiswa. Dalam perspektif yang lebih luas, rata-rata waktu studi mahasiswa mempengaruhi kualitas program studi dan karenanya waktu studi mahasiswa digunakan sebagai salah satu kriteria dalam menentukan penilaian oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN PT). Tujuan dari penelitian ini adalah untuk memahami seberapa baik kinerja *K-Nearest Neighbor*, *Naive Bayes*, *Decision Tree* untuk memprediksi Mahasiswa S1 Program Studi Ilmu Hukum Fakultas Hukum Universitas Sebelas Maret lulus tepat waktu dengan menggunakan aplikasi RapidMiner. Dari hasil proses pengujian dan prediksi dengan aplikasi RapidMiner menggunakan tiga metode yang telah dilakukan. Metode *K-Nearest Neighbor* (KNN) memperoleh hasil akurasi sebesar 96.67%, pada pengujian prediksi dengan metode *Naive Bayes* memperoleh nilai akurasi sebesar 77.33%, sedangkan metode *Decision Tree* memperoleh hasil akurasi sebesar 94.00%. Sehingga metode K-NN menjadi metode terbaik dalam klasifikasi komparasi dalam memprediksi kelulusan mahasiswa tepat waktu dengan prediksi nilai akurasi 96.67%.

Kata Kunci: K-Nearest Neighbor, Naive Bayes, Decision Tree, Klasifikasi, Mahasiswa

ABSTRACT

Students are a crucial factor that must be considered in seriously evaluating study programs. The indicator of the success of the study program is the length of time it takes to complete the study. The study period is the time when students complete their studies. In addition, student study time reflects the level of student learning performance. In a broader perspective, the average student study time affects the quality of study programs and therefore student study time is used as one of the criteria in determining the assessment by the National Accreditation Board for Higher Education (BAN PT). The purpose of this study was to understand how well the K-Nearest Neighbor, Naive Bayes, Decision Tree performed to predict undergraduate students of the Law Study Program, Faculty of Law, Sebelas Maret University, graduating on time using the RapidMiner application. From the results of the testing and prediction process with the RapidMiner application using the three methods that have been carried out. The K-Nearest Neighbor (KNN) method obtained an accuracy of 96.67%, in the prediction test using the Naive Bayes method it obtained an accuracy of 77.33%, while the Decision Tree method obtained an accuracy of 94.00%. So that the K-NN method is the best method in comparative classification in predicting student graduation on time with a predicted accuracy value of 96.67%.

Keyword: K-Nearest Neighbor, Naive Bayes, Decision Tree, Classification, Student

PENDAHULUAN

Mahasiswa merupakan faktor krusial yang harus diperhatikan dalam mengevaluasi program studi secara serius. Indikator keberhasilan program studi adalah lamanya dalam menyelesaikan studi. Masa studi adalah waktu dimana mahasiswa menyelesaikan studinya. Selain itu, waktu belajar mahasiswa mencerminkan tingkat kinerja belajar mahasiswa. Dalam perspektif yang lebih luas, rata-rata waktu studi mahasiswa mempengaruhi kualitas program studi dan karenanya waktu studi mahasiswa digunakan

sebagai salah satu kriteria dalam menentukan penilaian oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN PT) [1]. Berdasarkan standar utama, mutu Perguruan Tinggi khususnya Program Studi di Indonesia. Salah satu poin akreditasi berkaitan dengan perguruan tinggi yaitu Mahasiswa yang lulus tepat waktu [2]. Perguruan Tinggi memerlukan ujian akreditasi yang membahas kelulusan mahasiswa. Ketepatan waktu lulus memiliki pengaruh penting terhadap evaluasi akreditasi perguruan tinggi dan program studi. Banyak perguruan tinggi dan program studi menekankan salah satu faktor yang

mempengaruhi ketepatan mahasiswa dalam menyelesaikan studinya. Waktu studi atau lama studi adalah waktu yang dibutuhkan mahasiswa untuk menyelesaikan masa studi pendidikannya [3]. Program studi melaksanakan fungsi Tridharma Perguruan Tinggi yaitu Pendidikan, Penelitian dan Pengabdian Kepada Masyarakat, serta mengelola Ilmu Pengetahuan dan Teknologi (IPTEK). Oleh karena itu, program studi harus mampu mengorganisasikan dirinya sendiri guna meningkatkan dan menjamin mutu program studi [4]. Sedangkan persoalan utama yang muncul adalah jika ada beberapa mahasiswa yang terlambat atau tidak lulus tepat waktu selama waktu yang sudah ditentukan, sehingga menjadi kendala bagi perguruan tinggi tersebut [5]. Program studi S1 Ilmu Hukum Fakultas Hukum Universitas Sebelas Maret merupakan salah satu program studi yang ada di lingkungan Universitas Sebelas Maret. Dengan memanfaatkan data mining pada data bidang pendidikan, sebuah institusi perguruan tinggi bisa memperoleh suatu informasi yang berguna untuk melakukan perbaikan dan meningkatkan kualitas pada program studi tersebut. Data-data bidang pendidikan pada umumnya berupa data profil mahasiswa, kode mata kuliah, nama mata kuliah, kartu rencana studi (KRS), kartu hasil studi (KHS) yang didapatkan dari Sistem Informasi Akademik (SIKAD). Data yang didapatkan dari SIKAD digunakan untuk memproses data mining dengan model komparasi *K-Nearest Neighbor*, *Naive Bayes*, dan *Decision Tree*. Dari komparasi metode ini digunakan untuk memprediksi kelulusan mahasiswa program studi S1 Ilmu Hukum Fakultas Hukum Universitas Sebelas Maret. Penelitian yang menargetkan waktu kelulusan mahasiswa juga telah banyak dilakukan, diantaranya penelitian yang dilakukan oleh [1] dengan menggunakan metode algoritma *Naive Bayes Classifier* dan *C4.5* dengan memprediksi masa studi tepat waktu pada STMIK Bina Nusantara Lubuklinggau. Dari 162 data mahasiswa yang terdapat pada program studi sistem informasi selama tahun 2013 dan 2014, terdapat 117 dari mereka menyelesaikan masa studinya tepat waktu, sementara 45 lainnya

terlambat dalam masa studi. Kesimpulan dari penelitian ini adalah berdasarkan data set yang diimplementasikan menggunakan metode *Naive Bayes Classifier* dan prediksi status kelulusan mahasiswa di STMIK Bina Nusantara Jaya Lubuklinggau dengan tingkat akurasi masing-masing 78,46% dan 79,08%, sehingga algoritma *C4.5* dapat digunakan untuk memprediksi status kelulusan mahasiswa dengan tingkat akurasi sebesar 79,08%.

Penelitian lainnya yang dilakukan oleh [6] berdasarkan hasil pengujian algoritma *C4.5* untuk memprediksi kelulusan mahasiswa program studi Informatika dengan atribut yang digunakan adalah IPK, TOEFL, asal daerah dan jenis kelamin. Algoritma *C4.5* dapat digunakan pada situasi saat ini untuk mendapatkan prediksi keberhasilan dengan rata-rata presisi 63,93%, *recall* 60,73%, dan akurasi 60,52%. Sebaliknya, hasil metode *error based pruning* dengan tingkat kepercayaan 0,4 menghasilkan presisi 70,70%, *recall* 50,65%, dan akurasi 61,57%. Pengujian yang dilakukan dengan tingkat kepercayaan 0,25 menghasilkan presisi 73,77%, *recall* 48,84%, dan akurasi 62,44%. Dari hasil tersebut, dapat disimpulkan bahwa menggunakan tingkat kepercayaan 0,25 meningkatkan *output* lebih baik daripada menggunakan tingkat kepercayaan 0,4.

Pada penelitian yang dilakukan oleh [7] dalam penelitian ini digunakan data dari sekitar 141 mahasiswa Fakultas Ilmu Komputer angkatan 2015 dan 2016. Atribut yang digunakan untuk menentukan tingkat keberhasilan seorang mahasiswa adalah informasi dari catatan akademiknya seperti IPS 1-4, jenis kelaminnya, SKS yang telah diselesaikan, dan beberapa informasi lainnya seperti penghasilan bulanan dan keterlibatannya dalam kegiatan organisasi. Algoritma ini berperan dalam menentukan kedekatan kasus baru dengan kasus lama, populasi terbesar berada pada interval K dengan nilai terdekat yang diprediksi siswa akan selesai tepat waktu atau tidak. Hasil penelitian ini menunjukkan hasil prediksi dengan pengujian nilai $K = 5$ sampai $K = 11$. Jumlah data pengujian. Hasil pengujian terbaik dengan nilai $K = 11$ adalah akurasi 76%, akurasi 70%, mengingat waktu

klasifikasi 77% dan tidak tepat waktu.

Penelitian lainnya yang dilakukan oleh [8] tujuan dari penelitian ini adalah untuk mengidentifikasi studi pustaka yang digunakan dan menunjukkan bahwa metode data mining *Naïve Bayes* dapat memprediksi seberapa cepat mahasiswa akan menyelesaikan studi. Mengenai tingkat akurasi dari ketiga literatur tersebut memberikan akurasi lebih dari 90% meskipun menggunakan atribut dan aplikasi data yang berbeda. Salah satu atribut yang ditemukan di seluruh literatur yang dapat membuat prediksi adalah IPK (Indeks Prestasi Kumulatif).

Penelitian lainnya [9] metode yang digunakan dalam penelitian ini adalah algoritma C4.5 dengan variable satuan kredit semester (SKS) dari semester 2 sampai dengan semester 4 dan Indeks Prestasi Kumulatif (IPK). Sementara dari Semester 2 sampai Semester 4. Pembuktian konsep dilakukan untuk aplikasi. dibangun dengan skor kelayakan sebesar 67,27% yang berarti aplikasi yang dibangun layak dan juga diuji akurasinya menggunakan rumus matriks konfusi dengan akurasi sebesar 74,05%.

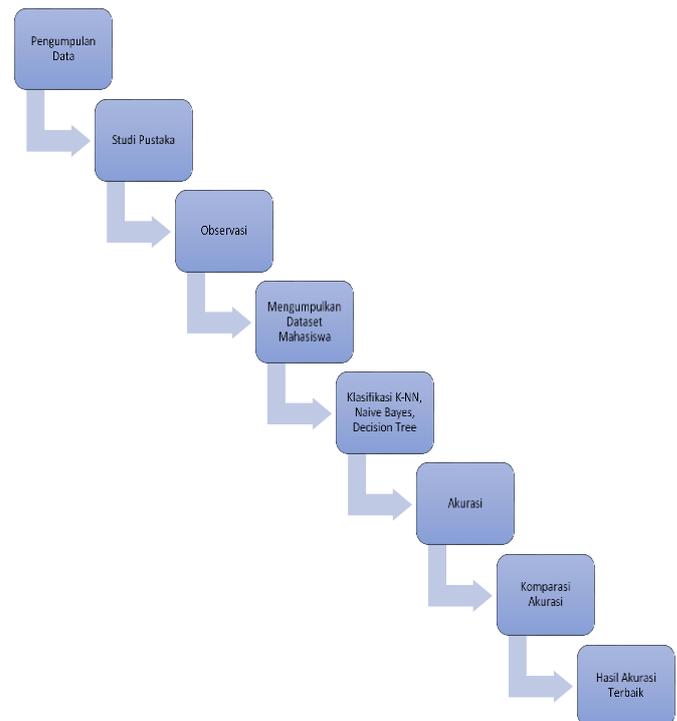
Berdasarkan informasi di atas dan penelitian sebelumnya, tujuan dari penelitian ini adalah untuk memahami seberapa baik kinerja *K-Nearest Neighbor*, *Naive Bayes*, *Decision Tree* untuk memprediksi Mahasiswa S1 Program Studi Ilmu Hukum Fakultas Hukum Universitas Sebelas Maret lulus tepat waktu dengan menggunakan aplikasi RapidMiner.

METODE

Teknik Pengumpulan Data

Metodologi pengumpulan data yang digunakan dalam penelitian ini adalah studi literatur dan observasi. Kajian dilakukan dengan mencari referensi di berbagai jurnal atau makalah, sedangkan observasi dilakukan dengan wawancara dengan staff akademik.

Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Dataset Mahasiswa

Penelitian ini menggunakan data dari Bagian Akademik Fakultas Hukum Universitas Sebelas Maret. Dataset yang digunakan yaitu data mahasiswa program studi S1 Ilmu Hukum tahun kelulusan 2018 dan 2019 yang berjumlah 300 dataset. Atribut yang digunakan yaitu NIM, Jenis Kelamin, IPS1, IPS2, IPS3, IPS4, IPS5, IPS6, IPS7, IPK, dan Status sebagai atribut hasil. Dataset yang digunakan dengan format xlsx Microsoft Excel. Adapun contoh dataset yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Dataset Mahasiswa

No	NIM	Jenis Kelamin	IPS1	IPS2	IPS3	IPS4	IPS5	IPS6	IPS7	IPK	Status
1	E0011145	Laki-laki	2.76	2.95	3.20	3.17	2.98	3.04	3.03	3.07	Terlambat
2	E0012133	Perempuan	3.00	3.30	3.14	3.14	2.84	3.13	3.25	3.17	Terlambat
3	E0012245	Perempuan	3.50	3.30	3.70	3.29	3.53	3.72	3.73	3.54	Terlambat

4	E0012358	Laki-laki	3.17	3.41	3.61	3.36	3.48	3.63	3.46	3.41	Terlambat
5	E0013094	Perempuan	2.90	2.89	3.30	2.85	2.98	3.00	3.08	3.09	Terlambat
299	E0015442	Perempuan	3.43	3.56	3.55	3.67	3.35	3.66	3.60	3.59	Tepat
300	E0015443	Perempuan	2.57	3.02	3.13	3.25	3.29	3.58	1.91	3.53	Tepat

K-Nearest Neighbor

K-NN juga dikenal sebagai *K-Nearest Neighbor* adalah algoritma yang mendeteksi dengan membandingkan beberapa kondisi terkini dengan kondisi lama. Memahami hubungan antara data yang harus dievaluasi dan data yang sudah dikenal sebagai K-NN. Pengujian data dan pelatihan data nilai jarak dilakukan secara berkelanjutan dengan beberapa tingkatan [10]. Agar dapat menyediakan data dengan gaya kuantitatif sebagai dasar pengambilan keputusan, ambang batas K-NN akan diterapkan pada sistem yang dibangun dengan arsitektur berbasis model. Tujuan perancangan *modelbase* adalah agar data yang dimasukkan pengguna dapat diproses dan menghasilkan kesimpulan yang akurat [11].

Naive Bayes

Secara statistik, metode yang dikenal sebagai *Naive Bayes* dapat digunakan untuk memperkirakan kemungkinan terjadinya suatu kelas. *Naive Bayes* sering digunakan untuk mengkategorikan data kumpulan yang diperoleh melalui pelatihan intensif [10]. Selain itu, untuk klasifikasi probabilistik yang kuat yang mengurangi probabilitas total dengan menggabungkan frekuensi dan menggabungkan nilai numerik dari kumpulan data yang tersedia [12]. *Naive Bayes* adalah teknik yang menggunakan probabilitas sederhana yang didasarkan pada teori inferensi Bayesian dengan asumsi independensi yang kuat. Manfaat lain dari *Naive Bayes* adalah dapat digunakan untuk menganalisis variable dalam situasi yang paling bermasalah [13]. *Nave Bayes* adalah

statistik klasifikasi yang dapat digunakan untuk memperkirakan kemungkinan setiap kelas tertentu akan memiliki anggota. Saat digunakan untuk database dengan jumlah data yang besar. *Bayesian Classification* memiliki kecepatan dan akurasi yang kuat [14]. Menurut [10] Metode *Nave Bayes* yang tergolong teknik statistik, dapat memprediksi kemungkinan terjadinya kelas kehancuran. *Nave Bayes* sering digunakan untuk mengkategorikan data himpunan dari pelatihan intensif. Terorema Bayes sebagai berikut:

$$P(C | A) = \frac{P(A|C)P(C)}{P(A)}$$

Decision Tree

Merupakan satu-satunya metode paling populer yang digunakan dalam klasifikasi. Mengungkapkan pohon keputusan karena model yang ditemukan menggunakan data yang disajikan sebagai pohon. Untuk mengkategorikan sekumpulan data tertentu, setiap atribut harus diterapkan melalui rentang node yang ada di *node* keputusan, dan setelah atribut mencapai *node* daun, data akan diklasifikasikan sesuai dengan kelas yang ada [15]. Menurut [15] pada pohon keputusan ada tiga jenis *node* yang berbeda yaitu:

1. ***Root Node***

adalah node yang secara fisik terletak di akar pohon. Tidak ada *input* untuk *Root Node*, menandakan bahwa tidak ada yang akan memasuki *node* ini.

2. ***Internal Node***

Internal Node adalah simpul percabangan pada *node* ini hanya ada

satu input dan mungkin memiliki satu atau lebih *output*.

3. *Leaf Node*

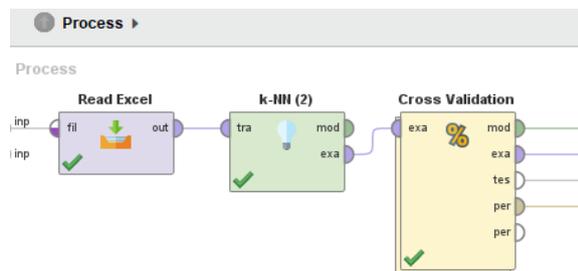
Node terakhir pohon keputusan dikenal sebagai *node* daun atau *node* terminal.

Node saat ini hanya memiliki satu input dan tidak ada *output*.

HASIL

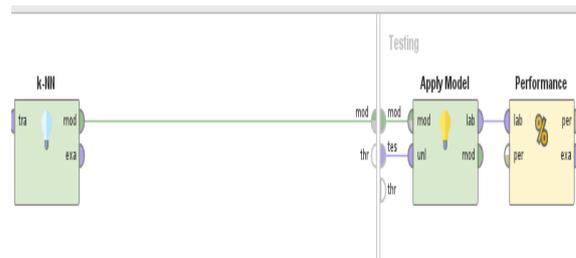
Model Proses Klasifikasi K-NN, Naïve Bayes, dan Decision Tree

Untuk mengetahui apakah seorang mahasiswa S1 akan menyelesaikan studinya tepat waktu atau tidak, penelitian ini membandingkan data mining klasifikasi dengan menggunakan aplikasi RapidMiner. Proses model K-NN dapat dilihat pada Gambar 2 seperti yang terlihat di bawah ini.



Gambar 2. Proses Model K-NN

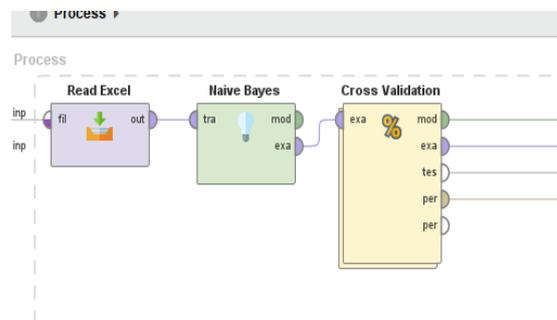
Pada gambar 2 proses model K-NN. *Read Excel* ini digunakan untuk membuat data yang berupa format xlsx Microsoft Excel. K-NN merupakan algoritma yang akan dilakukan pengujian akurasi, sedangkan *Cross Validation* digunakan sebagai subproses pelatihan dan subproses pengujian. Subproses pelatihan dapat digunakan untuk melatih model, model yang sudah dilatih dapat diterapkan dalam subproses pengujian. Model Validasi K-NN dapat dilihat pada Gambar 3.



Gambar 3. Model Validasi K-NN

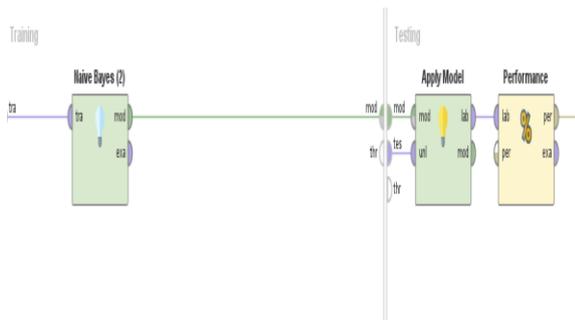
Model validasi metode klasifikasi yang digunakan pada penelitian ini dengan *K-Nearest Neighbor*. *Apply Model* merupakan operator yang digunakan sebagai penghubung metode K-NN. Sedangkan *Performance* digunakan untuk mengukur performa nilai akurasi dari model K-NN.

Proses model *Naïve Bayes* dapat dilihat pada Gambar 4 seperti dibawah ini.



Gambar 4. Proses Model *Naïve Bayes*

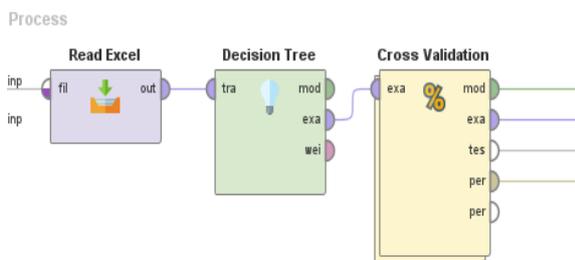
Pada gambar 4 proses model *Naïve Bayes*. *Read Excel* ini digunakan untuk membuat data yang berupa format xlsx Microsoft Excel. *Naïve Bayes* merupakan algoritma yang akan dilakukan pengujian akurasi, sedangkan *Cross Validation* digunakan sebagai subproses pelatihan dan subproses pengujian. Subproses pelatihan dapat digunakan untuk melatih model, model yang sudah dilatih dapat diterapkan dalam subproses pengujian. Sedangkan Model Validasi *Naïve Bayes* dapat dilihat pada Gambar 5.



Gambar 5. Model Validasi Naive Bayes

Model validasi metode klasifikasi yang digunakan pada penelitian ini dengan *Naive Bayes*. *Apply Model* merupakan operator yang digunakan sebagai penghubung metode *Naive Bayes*. Sedangkan *Performance* digunakan untuk mengukur performa nilai akurasi dari model *Naive Bayes*.

Model Proses *Decision Tree* dapat dilihat pada Gambar 6 sebagai berikut.



Gambar 6. Model Proses Decision Tree

Pada gambar 6 proses model *Decision Tree*. *Read Excel* ini digunakan untuk membuat data yang berupa format *xlsx* Microsoft Excel. *Decision Tree* merupakan algoritma yang akan dilakukan pengujian akurasi, sedangkan *Cross Validation* digunakan sebagai subproses pelatihan dan subproses pengujian. Subproses pelatihan dapat digunakan untuk melatih model, model yang sudah dilatih dapat diterapkan dalam subproses pengujian. Sedangkan Model Validasi *Decision Tree* dapat dilihat pada Gambar 7.

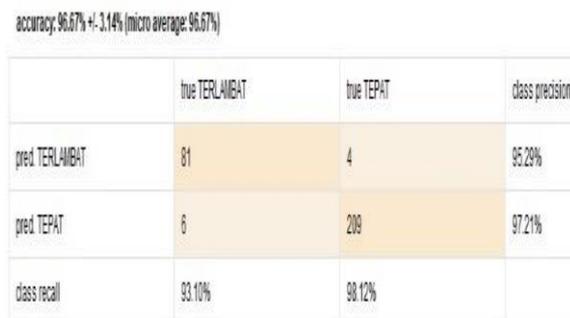


Gambar 7. Model Validasi Decision Tree

Model validasi metode klasifikasi yang digunakan pada penelitian ini dengan *Decision Tree*. *Apply Model* merupakan operator yang digunakan sebagai penghubung metode *Decision Tree*. Sedangkan *Performance* digunakan untuk mengukur performa nilai akurasi dari model *Decision Tree*.

Analisis Hasil Model K-Nearest Neighbor dengan Cross Validation dan Confusion Matrix

Hasil akurasi K-NN dapat dilihat pada Gambar 8.



Gambar 8. Hasil Akurasi K-NN

Hasil model K-NN menunjukkan tingkat akurasi 96.67%. Perhitungan nilai akurasi menggunakan *Confusion Matrix* sebagai berikut [5]

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} + \frac{209+81}{209+81+6+4} + \frac{290}{300} = 0,966 = 96.67\%$$

Analisis Hasil Model Naive Bayes dengan Cross Validation dan Confusion Matrix

Hasil akurasi algoritma *Naive Bayes* dapat dilihat pada Gambar 9.

Table View Plot View

accuracy: 77.33% +/- 6.81% (micro average: 77.33%)

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	67	48	58.26%
pred. TEPAT	20	165	89.19%
class recall	77.01%	77.46%	

Gambar 9. Hasil Akurasi *Naïve Bayes*

Hasil model *Naïve Bayes* menunjukkan tingkat akurasi 77.33%. Perhitungan nilai akurasi menggunakan *Confusion Matrix* adalah sebagai berikut [5]

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} + \frac{165+67}{165+67+20+48} + \frac{232}{300} = 0,773 = 77.33\%$$

Analisis Hasil Model *Decision Tree* dengan *Cross Validation* dan *Confusion Matrix*

Hasil akurasi algoritma *Naive Bayes* dapat dilihat pada Gambar 10.

Table View Plot View

accuracy: 94.00% +/- 3.70% (micro average: 94.00%)

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	78	9	89.66%
pred. TEPAT	9	204	95.77%
class recall	89.66%	95.77%	

Gambar 9. Hasil Akurasi *Naïve Bayes*

Model *Decision Tree* menghasilkan tingkat akurasi 94.00%. Perhitungan nilai akurasi menggunakan *Confusion Matrix* adalah sebagai berikut [5]

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} + \frac{204+78}{204+78+9+9} + \frac{282}{300} = 0,94 = 94.00\%$$

Analisis Hasil Komparasi Algoritma *Klasifikasi*

Berdasarkan pengujian komparasi algoritma klasifikasi tersebut dengan *Confusion Matrix*, maka didapatkan hasil

perbandingan hasil *Accuracy* pada Tabel 2.

Tabel 2. Hasil Komparasi

No	Metode	Accuracy
1	K-NN	96.67%.
2	<i>Naïve Bayes</i>	77.33%.
3	<i>Decision Tree</i>	94.00%.

Berdasarkan pada tabel 2 dapat dilihat bahwa kinerja hasil akurasi algoritma *K-Nearest Neighbor* lebih baik dari algoritma *Naïve Bayes* dan *Decision Tree* untuk memprediksi status kelulusan Mahasiswa.

KESIMPULAN

Pada penelitian ini untuk memprediksi kelulusan mahasiswa dengan menggunakan 3 metode yaitu K-NN, *Naïve Bayes* dan *Decision Tree*. Atribut yang digunakan untuk klasifikasi data mining terdiri dari NIM, Jenis Kelamin, IPS1, IPS2, IPS3, IPS4, IPS5, IPS6, IPS7, IPK dan status. Dari hasil proses pengujian dan prediksi dengan aplikasi RapidMiner menggunakan tiga metode yang telah dilakukan. Metode *K-Nearest Neighbor* (KNN) memperoleh hasil akurasi sebesar 96.67%, pada pengujian prediksi dengan metode *Naïve Bayes* memperoleh nilai akurasi sebesar 77.33%, sedangkan metode *Decision Tree* memperoleh hasil akurasi sebesar 94.00%. Sehingga metode K-NN menjadi metode terbaik dalam klasifikasi komparasi dalam memprediksi kelulusan mahasiswa tepat waktu dengan prediksi nilai akurasi 96.67%. Oleh karena itu, disarankan pada penelitian selanjutnya untuk membuat perbandingan antara metode klasifikasi dan metode *clustering* untuk mengetahui nilai akurasi dari penggunaan kedua metode algoritma tersebut.

DAFTAR PUSTAKA

- [1] E. Etriyanti, D. Syamsuar, and Y. N. Kunang, "Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier Dan C4.5 Untuk Memprediksi Kelulusan Mahasiswa," *Telematika*, vol. 13, no. 1, pp. 56-57, 2020.
- [2] Azahari, Yulindawati, D. Rosita, and S. Mallala, "Komparasi Data Mining Naive Bayes dan Neural Network Memprediksi Studi Mahasiswa S1," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 7, no. 3, pp. 443-452, 2018.
- [3] E. F. Wati, and B. Rudianto, "Penerapan Algoritma KNN, Naive Bayes dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa," *Jurnal Format*, vol. 11, no. 2, pp. 168-175, 2022.
- [4] I. A. Nikmatun, and I. Waspada, "Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Simetris*, vol. 10, no. 2, pp. 421-432, 2019.
- [5] A. Budiyantra, Irwansyah, E. Prengki, P. A. Pratama, and N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," *Jurnal Ilmu Pengetahuan dan Teknologi Komputer*, vol. 5, no. 2, pp. 265-270, 2020.
- [6] R. P. S. Putri, and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Jurnal Ilmu Komputer dan Informatika Khazanah Informatika*, vol. 4, no. 1, pp. 1-7, 2018.
- [7] J. A. Samudra, S. Anraeni, and Herman, "Penerapan Metode K-Nearest Neighbor untuk Memprediksi Tingkat Kelulusan Mahasiswa Berbasis Web pada Fakultas Ilmu Komputer UMI," *Busiti Buletin Sistem Informasi dan Teknologi Islam*, vol. 1, no. 4, pp. 230-237, 2020.
- [8] L. Setiayani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naive Bayes : Systematic Review," *Faktor Exacta*, vol. 13, no. 1, pp. 35-43, 2020.
- [9] Irmayansyah, and M. T. Kastrilia, "Penerapan Algoritma C4.5 untuk Prediksi Mahasiswa Berpotensi Lulus Tidak Tepat Waktu," *Jurnal Ilmiah Teknologi - Informasi dan Sains*, vol. 10, no. 2, pp. 9-18, 2020.
- [10] D. Safitri, S.S. Hilabi, and F. Nurapriani, "Analisis Penggunaan Algoritma Klasifikasi Dalam Prediksi Kelulusan Menggunakan Orange Data Mining," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 8, no. 1, pp. 75-81, 2023.
- [11] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *Technologia Jurnal Ilmiah*, vol. 13, no. 3, pp. 207-214, 2022.
- [12] P. S. C. Moonallika, K. Q. Fredlina, and I. B. K. Sudiatmika, "Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara)," *Progresif : Jurnal Ilmiah Komputer*, vol. 16, no. 1, pp. 47-56, 2020.
- [13] A. Wibowo, and A. Rohman,

- "Prediksi Predikat Kelulusan Mahasiswa Menggunakan Naive Bayes dan Decision Tree pada Universitas XYZ," *Expert Jurnal Manajemen Sistem Informasi dan Tteknologi*, vol. 12, no. 2, pp. 104-112, 2022.
- [14] N. purwati, and A. D. Januanti, "Prediksi Tingkat Kelulusan Mahasiswa dengan Algoritma Naive Bayes," *Jurnal Pepaduan Jurnal Ilmiah, Ilmu Komputer dan Sistem Informasi*, vol. 2, no. 1, pp. 126-137, 2021.
- [15] I. Iskandar, L. Hiryanto, and J. Hendryli, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5 Dengan Teknik Pruning," *JIKSI Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 6, no. 1 , pp. 64-68, 2018.