

## Penggunaan *Rapidminer* Untuk Memprediksi Kelulusan Mahasiswa Dengan Algorithm Naive Bayes

Fauziah, Azzahra Sri Ramadhantya

Universitas Islam As-syafi'iyah, Teknik Informatika  
fauziah.fst@uia.ac.id

### Abstrak

Data mining dengan metode Algoritma Naïve Bayes mampu memprediksi masa depan dengan menggunakan perhitungan probabilitas. Penelitian ini bertujuan untuk mengukur akurasi dan mengetahui penggunaan Algoritma Naïve Bayes untuk memprediksi kelulusan mahasiswa. Penelitian ini menggunakan penelitian kuantitatif dengan subyek penelitian yaitu data lulusan dari mahasiswa Teknik Informatika tahun 2016, 2017, 2018, dan 2019 menjadi data training dan data mahasiswa Teknik Informatika yang masih aktif tahun angkatan 2020, 2021, dan 2022 menjadi data testing. Algoritma Naïve Bayes memprediksi masa depan dengan menggunakan pengalaman sebelumnya. Perhitungan menggunakan RapidMiner Studio 10.1. Hasil penelitian pengujian menggunakan RapidMiner Studio 10.1 dengan data training sebanyak 45 data diperoleh keakurasian sebesar 94.12%. Data training sebanyak 42 data di uji dan di dapatkan hasil sebanyak 3 mahasiswa diprediksi tepat waktu dan 39 mahasiswa diprediksi terlambat

**Kata kunci:** Algoritma Naïve Bayes, Klasifikasi, Prediksi Kelulusan

### Abstract

*Data mining using the Naïve Bayes Algorithm method is able to predict the future using probability calculations. This research aims to measure accuracy and determine the use of the Naïve Bayes algorithm to predict student graduation. This research uses quantitative research with research subjects, namely graduate data from Informatics Engineering students in 2016, 2017, 2018, and 2019 as training data and data from Informatics Engineering students who are still active in the 2020, 2021, and 2022 classes. become testing data. The Naïve Bayes algorithm predicts the future using previous experience. Calculations using RapidMiner Studio 10.1. The results of testing research using RapidMiner Studio 10.1 with 45 training data obtained an accuracy of 94.12%. 42 training data were tested and the results obtained were that 3 students were predicted to be on time and 39 students were predicted to be late*

*Keywords: Naïve Bayes Algorithm, Classification, Graduation Prediction*

## PENDAHULUAN

Tahun ajaran baru seringkali ditandai dengan masuknya mahasiswa baru pada suatu universitas. Data yang diperoleh bagian akademik Fakultas Sains dan Teknologi terhadap mahasiswa yang lulus pada Prodi Teknik Informatika berjumlah 87 data mahasiswa yang lulus pada tahun 2019-2023. Jumlah mahasiswa yang masuk lebih banyak dibandingkan mahasiswa yang lulus. Hal ini menyebabkan banyaknya informasi mahasiswa yang menumpuk. Jumlah mahasiswa baru dan mahasiswa yang lulus tepat waktu tidak sebanding, sehingga perlu dilakukan kajian metode yang memungkinkan prediksi awal kelulusan mahasiswa untuk membantu perguruan tinggi dalam mengambil keputusan, mengurangi jumlah mahasiswa yang lulus tidak tepat waktu. Prediksi ini menggunakan metode klasifikasi. Tugas metode klasifikasi adalah memprediksi hasil suatu variable (kelas) yang memiliki nilai polinomial 2 (kategoris). Prediksi waktu kelulusan mahasiswa kedalam kelompok tepat waktu dan terlambat yang dilihat dari data master bagian akademik fakultas. Dalam penelitian ini peneliti menggunakan metode Algoritma Naïve Bayes. Naïve Bayes adalah algoritma yang digunakan untuk klasifikasi memprediksi probabilitas suatu peristiwa dengan menggunakan teknik probabilistic dan statistic yang lebih dikenal dengan Teorema Bayes. **DATA MINING**

Perkembangan teknologi informasi dan arus informasi yang semakin pesat memerlukan pengelolaan yang tepat agar dapat menghasilkan informasi yang dapat digunakan dalam pengambilan Keputusan. Saat ini berkembang bidang ilmu yang mempelajari penggalian data (data mining). Pada dasarnya data mining

adalah tentang mencari atau mengekstraksi kumpulan data untuk menemukan informasi yang tidak terduga. Data mining dapat menangani data berskala besar, serta data mining dapat menggunakan pengalaman 7 masa lampau untuk meningkatkan proses model pembelajarannya seperti pada penerapan klasifikasi (Ni Luh Wiwik Sri Rahayu Ginantra *et al.*, 2021).

## METODE

### ALGORITMA NAÏVE BAYES

Algoritma Naïve Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. Naïve Bayes adalah klasifikasi yang menggunakan metode probabilitas dan statistik yang ditemukan oleh ilmuwan Inggris Thomas Bayes, untuk memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu, oleh karena itu dikenal dengan Teorema Bayes. Teorema ini dikombinasikan dengan Naïve Bayes yang mengasumsikan bahwa suku-suku antar atribut adalah independen. Klasifikasi Naïve Bayes mengasumsikan bahwa ada atau tidaknya ciri-ciri tertentu suatu kelas tidak ada hubungannya dengan ciri kelas lain. Banyak penelitian sebelumnya yang menggunakan algoritman data mining Naïve Bayes untuk menyelesaikannya (Sadewo *et al.*, 2019).

Teorema Bayes memiliki bentuk umum sebagai berikut:

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

### CONFUSION MATRIX

Confusion matrix ialah tabel yang memberikan output pekerjaan berdasarkan klasifikasi. Menghitung angkayang akurat merupakan hal terpenting dalam penelitian agar bisa diketahui tingkat keberhasilan dan

Kegagalan terhadap penelitian(Malelak, Ardiada and Feoh, 2021)

Table 1. Confusion Matrix

Aktual	Prediksi	
	Cepat	Terlambat
Positif	True Positive (TN)	False Positive (FN)
Negatif	False Negative (FN)	True Negative (TN)

Confusion matrix Tabel 2.1 terdapat 4 istilah untuk menunjukkan hasil klasifikasi, antara lain:

1. True Positive (TP) yaitu data yang bersifat positif dan terdeteksi positif
2. False Positive (FP) yaitu data yang bersifat positif namun terdeteksi negative
3. False Negative (FN) yaitu data yang bersifat negatif dan terdeteksi positif
4. True Negative (TN) yaitu data yang bersifat negatif dan terdeteksi negative

**JAVA**

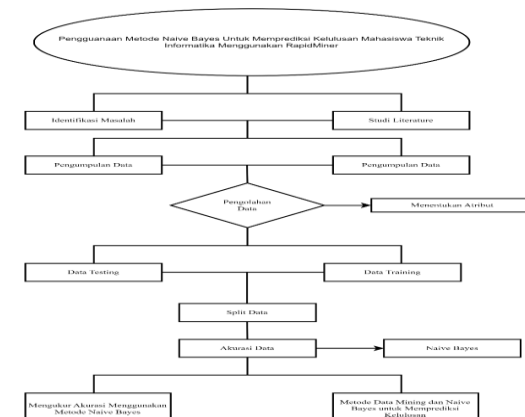
Setiap *software* memiliki fitur dan manfaatnya masing-masing. Java adalah bahasa pemrograman tingkat tinggi (high level). Namun, pemrograman ini memiliki bahasa yang mudah dipahami karena menggunakan bahasa sehari-hari. Java diciptakan *James Gosling, Patrick Naughton, dan Mike Sheridan* pada perusahaan Sun Microsystems di tahun 1991. Perusahaan ini memerlukan waktu sekitar 18 bulan untuk mengembangkan versi pertamanya yang awalnya disebut Oak menjadi nama Java. Java diakuisi oleh perusahaan *Oracle* selama tahap pengembangannya. Sejak perkembangannya hingga saat ini Java sudah beberapa kali mengalami perubahan. Java mempunyai banyak keunggulan. Salah satunya adalah mudah dipelajari oleh siapa saja karna

syntax (tata bahasa) yang mirip dengan bahasa manusia. Di sisi lain, banyak artikel atau tutorial yang membahas tentang pemrograman Java. Hal ini sangat memungkinkan pemulayang belum pernah mempelajari bahasa pemrograman apa pun untuk mempelajari pemrograman (Yuni Sugiarti, 2018).

**HASIL DAN PEMBAHASAN**

**RAPIDMINER**

*RapidMiner* merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* dan Ralf Klittenberg dari *rapid-i.com* dan mereka memiliki layar GUI (*Graphical User Interface*) yang memudahkan penggunaan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibangun dengan menggunakan program Java di bawah lisensi *GNU Public Licence* dan *RapidMiner* dapat dijalankan pada sistem operasi apapun. Tidak diperlukan keahlian coding khusus untuk menggunakan *RapidMiner* karena semua model sudah disediakan. *RapidMiner* dikhususkan untuk penggunaan data mining. Model yang ditawarkan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-



lain(Sudarsono et al., 2021)

Gambar 1. Diagram Rapidminer

*RapidMiner* adalah perangkat lunak analisis data penambangan data mandiri yang dapat diintegrasikan ke dalam berbagai bahasa pemrograman.

*RapidMiner* ditulis dalam bahasa pemrograman Java, sehingga dapat dijalankan di beberapa sistem operasi. *RapidMiner* menyediakan UI untuk mendesain pipa analisis, di mana akan menghasilkan file XML yang dapat menjelaskan proses analisis yang ingin diterapkan oleh pengguna ke data. *Rapid Miner* akan membaca file ini untuk menjalankan analisis secara otomatis (Prasetyo et al., 2021)

metode penelitian kuantitatif.

Penerapan metode penelitian kuantitatif berfokus pada penyajian hasil informasi yang diperoleh melalui gambar, table, diagram, dan grafik, serta digunakan sebagai bahan analisis data dengan menggunakan metode *Naïve Bayes*. Setelah pengumpulan data, maka dilakukan analisis data untuk menyesuaikan proses data yang ditangani

dengan metode *Naïve Bayes*.

### IMPLEMENTASI

Dalam hal pengolahan data, tahap implementasi adalah bagaimana pengolahan data diimplementasikan pada *tools*. *Tools* yang digunakan untuk melakukan penelitian ini adalah *Software RapidMiner*. Selanjutnya diuji apakah penelitian tersebut sesuai dengan tujuan yang telah ditetapkan, yaitu untuk memprediksi kelulusan mahasiswa khususnya pada prodi Teknik Informatika.

#### 5. Proses Seleksi Pembersihan dan Transformasi Data

##### a. Seleksi Data

Seleksi data akan melakukan pemilihan atribut yang dibutuhkan dan menghilangkan atribut yang tidak digunakan. Pada dataset mahasiswa

Gambar 1 Diagram Alir

alumni dipergunakan untuk data training, atribut yang digunakan meliputi:

##### 1) Jenis Kelamin

Variable jenis kelamin hanya terdiri dari dua

kemungkinan yaitu laki-laki dan perempuan.

##### 2) Kota Kelahiran

Variabel kota kelahiran dikelompokkan menjadi dari dalam kota Jakarta dan Bekasi atau dari luar Jakarta dan Bekasi. Untuk kota kelahiran Jakarta dan Bekasi dikelompokkan data dalam kota sedangkan untuk kota kelahiran luar Jakarta

##### 1) ataupun Bekasi dikelompokkan Keputusan

Variable keputusan merupakan data yang berfungsi untuk menentukan hasil keputusan. Dalam data keputusan hanya memiliki 2 nilai yaitu tepat waktu dan tidak tepat waktu.

Table 2. Keputusan

Lama Studi	Keputusan
< 4 Tahun	Tepat Waktu
> 4 Tahun	Terlambat

b. Pembersihan Data Pembersihan data dari noise dan atribut yang tidak digunakan sehingga akan dibersihkan agar memperoleh atribut yang relevan dan tidak missing value sesuai dengan atribut yang digunakan.

##### C. Transformasi Data

Tahap transformasi data yaitu data yang sudah dibersihkan akan ditransformasikan formatnya sehingga bisa diolah menggunakan *data mining*. Format yang

Table 3. Data Training

Jenis Kela min	Tempat Lahir	IP K	Tipe Sekol ah	Ekonomi	Keputu san
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Umu m	Sedan g	Terlamb at
LK	Dalam Kota	3	Kejuruan	Rendah	Terlamb at
LK	Luar Kota	3	Umu m	Sedan g	Terlamb at
LK	Dalam Kota	3	Umu m	Tinggi	Terlamb at
PR	Dalam Kota	3	Umu m	Tinggi	Terlamb at
PR	Luar Kota	3	Umu m	Tinggi	Terlamb at
LK	Luar Kota	3	Umu m	Tinggi	Terlamb at
LK	Dalam Kota	2	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Sedan g	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Sedan g	Terlamb at
LK	Dalam Kota	3	Umu m	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Tepat Waktu
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Umu m	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at

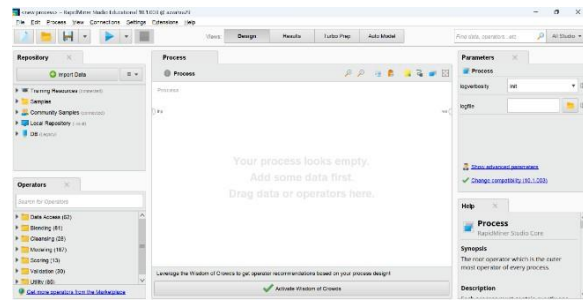
LK	Dalam Kota	3	Kejuruan	Sedan g	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
PR	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	2	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Umu m	Sedan g	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
PR	Luar Kota	3	Kejuruan	Sedan g	Terlamb at
LK	Dalam Kota	2	Kejuruan	Sedan g	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Umu m	Tinggi	Terlamb at
LK	Dalam Kota	2	Umu m	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Umu m	Tinggi	Terlamb at
PR	Dalam Kota	3	Umu m	Tinggi	Terlamb at
PR	Dalam Kota	3	Umu m	Tinggi	Tepat Waktu
PR	Dalam Kota	3	Umu m	Tinggi	Tepat Waktu
LK	Dalam Kota	3	Umu m	Tinggi	Terlamb at
PR	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Dalam Kota	3	Kejuruan	Tinggi	Terlamb at
LK	Luar Kota	3	Kejuruan	Tinggi	Terlamb at

Berikut adalah proses pengujian data dan tampilan RapidMiner Studio 10.1



Gambar 2. Tampilan RapidMiner Studio 10.1

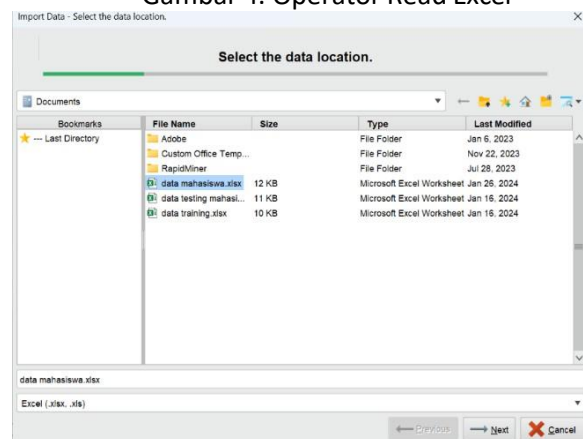
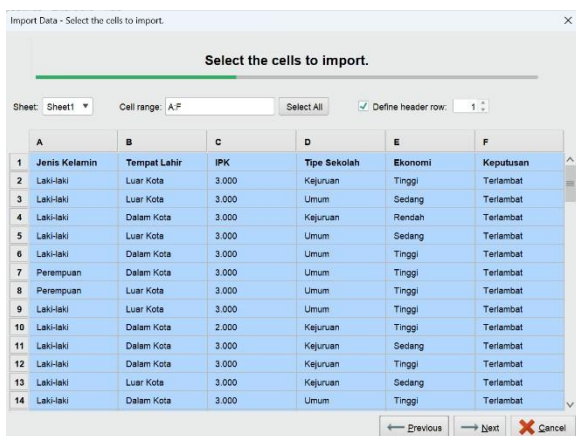
Buka aplikasi Rapidminer, setelah muncul seperti maka akan muncul tampilan start with kemudian pilih blank process untuk membuka lembar kerja baru/ membuka file yang telah tersimpan seperti Gambar 2



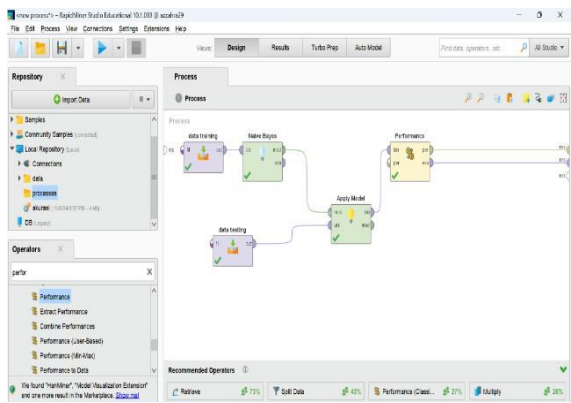
Gambar 3. Tampilan Lembar Kerja  
 Selanjutnya proses input data dan membutuhkan operator *read excel* dengan *drag and drop* kedalam lembar kerja seperti Proses pengambilan data dengan menggunakan *double click* pada operator *read excel* lalu pilih file data seperti Gambar 4 dan 5



Gambar 4. Operator Read Excel



Gambar 5. Import Data Training  
 Selanjutnya klik next pada data yang dipilih seperti Setelah itu tentukan tipe data lalu pada kelas data diberi atribut label dengan pilihan *change role* dan klik finish maka tampilannya seperti Gambar 6 dan 7



Gambar 6 Import data

Row No.	Keputusan	predicted	confidence	confidence	Jenis Kelamin	Tempat Lahir	Tipe Sekolah	Eksistensi	IPK
20	Terlambat	0.999	0.001	0.001	Laki-laki	Dalen Kota	Hajulan	Selang	3
21	Terlambat	0.800	0.199	0.199	Laki-laki	Dalen Kota	Uluhan	Tinggi	3
22	Terlambat	0.900	0.099	0.099	Laki-laki	Dalen Kota	Hajulan	Selang	3
24	Terlambat	0.800	0.199	0.199	Laki-laki	Dalen Kota	Uluhan	Tinggi	3
25	Terlambat	0.900	0.100	0.100	Laki-laki	Dalen Kota	Uluhan	Tinggi	3
26	Terlambat	0.999	0.001	0.001	Laki-laki	Dalen Kota	Hajulan	Reman	3
27	Terlambat	0.970	0.030	0.030	Laki-laki	Dalen Kota	Hajulan	Tinggi	3
28	Terlambat	0.700	0.300	0.300	Laki-laki	Dalen Kota	Uluhan	Tinggi	3
29	Terlambat	0.800	0.199	0.199	Laki-laki	Dalen Kota	Uluhan	Tinggi	3
30	Terlambat	0.900	0.099	0.099	Laki-laki	Dalen Kota	Hajulan	Reman	3
31	Terlambat	0.970	0.030	0.030	Laki-laki	Dalen Kota	Hajulan	Tinggi	3

Gambar 7 Hasil prediksi pengujian data

Jumlah data testing yang digunakan sebanyak 42 mahasiswa teknik informatika angkatan tahun 2020, 202, 2022 dengan menggunakan metode *Naïve Bayes* didapatkan hasil prediksi penelitian bahwa mahasiswa lulus tepat waktu sebanyak 3 mahasiswa.



Gambar 8. Grafik Presentase Kelulusan

**PEMBAHASAN**

Penelitian ini menerapkan metode *Naïve Bayes*. Data yang didapatkan berasal dari bagian akademik fakultas sains dan teknologi Universitas Islam As- syafi'iyah dan juga menggunakan survei kepada mahasiswa aktif teknik informatika. Data yang diperoleh

dibersihkan kemudian digunakan. Adapun atribut yang digunakan yaitu jenis kelamin, asal sekolah, ekonomi, tempat lahir dan indeks prestasi kumulatif (IPK). Data yang diambil dibagi menjadi dua kelompok yaitu data mahasiswa yang telah lulus sebagai *data training* diambil sebanyak 45 data dan data mahasiswa aktif tahun angkatan 2020, 2021 dan 2022 sebagai *data testing* berjumlah 42 data. Pada penelitian ini memiliki dua kelas dat yaitu tepat waktu dan terlambat. Selanjutnya pengujian dengan *RapidMiner Studio 10.1* sebagai alat bantu untuk memperoleh tingkat akurasi dari metode *Naïve Bayes*. Berikut ini adalah perhitungan untuk mendapatkan nilai akurasi dengan persamaan berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+FN+F} \times 100\% \\
 &= \frac{13+3}{13+0+1+3} \times 100\% \\
 &= 94.12\%
 \end{aligned}$$

Nilai 94.12% membuktikan bahwasannya metode *Naïve Bayes* dapat digunakan untuk melakukan klasifikasi prediksi kelulusan mahasiswa. Nilai akurasi 94.12% bisa diakibatkan oleh atribut saling bebas dengan atribut lainnya.

Prediksi dilakukan dengan 42 *data training* menggunakan *Naïve Bayes* mendapatkan hasil sebanyak 3 mahasiswa diprediksi tepat waktu, 39 mahasiswa diprediksi terlambat.

**KESIMPULAN DAN SARAN**

**KESIMPULAN**

Berdasarkan hasil penelitian yang dilakukan menggunakan metode *Naïve Bayes* dengan alat bantu *RapidMiner Studio 10.1*, maka diperoleh:

- 6. Implementasi data training sebanyak 45 data dengan menggunakan metode algoritma

- Naïve Bayes berhasil memprediksi kelulusan mahasiswa dengan presentase keakuratan 94.12%.
7. Data mining mampu untuk mengolah dan mengetahui pola data dalam jumlah besar secara otomatis, yang dimana bertujuan untuk mendukung hasil agar sesuai dengan yang diharapkan.
  8. Naive Bayes mampu menampilkan informasi prediksi kelulusan mahasiswa dengan menggunakan data mahasiswa yang telah lulus sebagai data traing dan mahasiswa aktif sebagai data testing. Sebanyak 42 data uji yang dihasilkan penelitian ini bahwa mahasiswa yang diprediksi lulus tepat waktu sebanyak 3 mahasiswa dari jumlah data testing dengan keakurasian sebesar 94.12%

#### SARAN

Penelitian ini masih banyak kekurangan dan kelemahan. Peneliti menyarankan untuk penelitin selanjutnya, yaitu:

1. Menggunakan tambahan data- data dengan jumlah yang lebih besar dan variable atau atribut dalam penentuan prediksi kelulusan.
2. Pengembangan metode klasifikasi diharapkan menggunakan dua

aatau lebih metode sehingga dapat membandingkan keakuratan.

3. Pada penelitian ini menggunakan alat bantu *RapidMiner Studio* 10.1 diharapkan pada penelitian selanjutnya dapat menggunakan aplikasi tambahan untuk memperkuat dan mendukung hasil akurasi.

#### DAFTAR PUSTAKA

Malelak, K.H.L., Ardiada, I.M.D. and Feoh, G. (2021) 'Implementasi Klasifikasi Naive Bayes Dalam Memprediksi Lama Studi Mahasiswa (Studi Kasus : Universitas Dhyana Pura)', *SINTECH (Science and Information Technology) Journal*, 4(2), pp. 202–209. Available at: <https://doi.org/10.31598/sintechjournal.v4i2.964>.

Ni Luh Wiwik Sri Rahayu Ginantra, F.N.A. et al. (2021) 'DATA MINING DAN PENERAPAN ALGORITMA'.

Prasetyo, V.R. et al. (2021) 'Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression', *Jurnal Nasional Teknologi dan Sistem Informasi*, 7(1), pp. 8–17. Available at: <https://doi.org/10.25077/teknosi.v7i1.2021.8-17>.

Fauziah (2022) 'Sistem Informasi pada Persediaan Barang Berbasis Web di Frozen Food', *SATESI: Jurnal Sains Teknologi dan Sistem Informasi*, 2(2), pp. 177–181. Available at: <https://doi.org/10.54259/satesi.v2i2.1211>.



Sudarsono, B.G. *et al.* (2021) 'Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner', *JBASE - Journal of Business and Audit Information Systems*, 4(1), pp. 13–21. Available at: <https://doi.org/10.30813/jbase.v4i1.2729>.

Yuni Sugiarti, S.T.M.K. (2018) 'Dasar-Dasar Pemograman Java Netbeans: Database, UML, dan Interface', in. 'Alfian, D. and 'Fatih, W. (2024) 'Aplikasi Perancangan Berbasis Android Pada Bengkel Jasa Service AC', *INSIT- Informatika Sains Teknologi*, 2(1), pp. 30–37.