



Perbandingan Algoritma C 4.5 dan Naïve Bayes Dalam Memprediksi Kelulusan Mahasiswa

Muhammad Ahya Fajri Firmansyah¹, Alfian Adam Setiawan¹

¹Fakultas Sains dan Teknologi, Universitas IVET, Indonesia

DOI: <https://doi.org/10.31331/jsitee.v1i1.kodeartikel>

Info Articles

Sejarah Artikel:

Disubmit 2 April 2023

Direvisi 14 Mei 2023

Disetujui 1 Juni 2023

Keywords:

C4.5 algorithm ; Naïve Bayes ;

Graduation Prediction; Data
Mining

Abstrak

Penelitian ini dilakukan untuk membandingkan akurasi dua metode algoritma, yaitu algoritma C4.5 dan Naïve Bayes pada sejumlah datasets. Berdasarkan hasil penelitian Perbandingan Algoritma C4.5 Dan Naïve Bayes untuk memprediksi kelulusan tepat waktu Mahasiswa diperoleh hasil nilai akurasi C4.5 adalah 90.79%, sedangkan nilai akurasi Naïve Bayes 75.00% dapat disimpulkan bahwa dengan menggunakan model algoritma C 4.5 lebih tinggi tingkat akurasinya dalam memprediksi kelulusan mahasiswa.

Abstract

This study was conducted to compare the accuracy of two algorithm methods, namely the C4.5 algorithm and Naïve Bayes on a number of datasets. Based on the results of the C4.5 and Naïve Bayes Algorithm Comparison research to predict student graduation on time, the C4.5 accuracy score was 90.79%, while the Naïve Bayes accuracy value of 75.00% can be concluded that using the C 4.5 algorithm model the accuracy rate is higher in predicting student graduation.

✉ Alamat Korespondensi:

E-mail: muhammadahyaff@gmail.com

p-ISSN 2721-8341

e-ISSN 2963-4660

PENDAHULUAN

Prediksi merupakan proses atau tindakan meramalkan atau memperkirakan apa yang mungkin terjadi di masa depan berdasarkan informasi yang tersedia saat ini. Hal ini melibatkan penggunaan data dan pengetahuan yang ada untuk membuat perkiraan tentang hasil atau kejadian yang akan datang. Prediksi dapat dilakukan dalam berbagai bidang, termasuk pendidikan, ekonomi, bisnis, ilmu pengetahuan, cuaca, olahraga, dan banyak lagi. Menurut (Aulia, dkk.,2022) Prediksi merupakan suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin akan terjadi berdasarkan informasi sebelumnya dan sekarang yang dimiliki, agar kesalahan prediksi (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diminimalisir. Kelulusan adalah pencapaian seseorang dalam menyelesaikan persyaratan atau tugas yang diberikan oleh lembaga pendidikan, organisasi, atau institusi tertentu untuk memperoleh gelar, sertifikat, atau pengakuan resmi lainnya. Menurut (Hermawanti & Sarah, 2019) Lulus tepat waktu merupakan salah satu indikator keberhasilan mahasiswa dalam memperoleh gelar sarjana. Mahasiswa dikatakan lulus tepat waktu apabila mampu menyelesaikan studinya di perguruan tinggi selama kurang dari atau sama dengan empat tahun, sedangkan mahasiswa dikatakan tidak lulus tepat waktu apabila menyelesaikan studinya lebih dari empat tahun.

Berdasarkan pembahasan diatas, kriteria kelulusan dalam perguruan tinggi salah satunya adalah IPK, IPK ini menjadi salah satu faktor penting dalam menentukan apakah seorang mahasiswa dapat lulus atau tidak. Biasanya, perguruan tinggi menetapkan batas minimum IPK yang harus dicapai oleh mahasiswa untuk lulus. Dalam memprediksi suatu hal, banyak metode yang dapat digunakan, salah satunya data mining. Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data yang ada yang biasanya berupa data yang sangat besar dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga (Kusrini dan Luthfi, 2009). Data mining memiliki beberapa metode algoritma salah satunya algoritma klasifikasi. Beberapa algoritma klasifikasi yang sering digunakan diantaranya adalah Support Vector Machine, Decision Tree, dan Naive Bayes (Agus Heri, 2020).

Dalam penelitian Aulia, dkk (2022) bertujuan untuk membandingkan akurasi dua metode algoritma, yaitu algoritma C4.5 dan Naïve Bayes pada sejumlah datasets. Adapun sumber datasets yang digunakan yaitu data mahasiswa Program Studi Teknik Informatika (STMIK Palangkaraya) yang mana setiap datasets memiliki jumlah data (instances) dan jumlah atribut (attributes) yang berbeda. Hasil penelitian perbandingan algoritma C4.5 Dan Naïve Bayes untuk memprediksi kelulusan tepat waktu Mahasiswa STMIK Palangkaraya, diperoleh hasil akurasi dari kedua algoritma tersebut menunjukkan tingkat akurasi Algoritma C4.5 sebesar 90% lebih baik dari pada algoritma Naive Bayes yakni yang hanya sebesar 85%. Algoritma C4.5 juga memberikan nilai pada recall dan precision sebesar 92% dan 94% lebih baik dibandingkan algoritma Naive Bayes dengan nilai 86% dan 93%.

Penelitian Gunawan, dkk (2021) bertujuan untuk membandingkan hasil dua metode algoritma yaitu algoritma K-Nearest Neighbour (K-NN) dan Naïve Bayes. Penelitian ini menghasilkan kesimpulan bahwa algoritma Naïve Bayes memiliki tingkat akurasi yang sama dengan algoritma KNN dalam memprediksi kelulusan mahasiswa program studi Pendidikan Kedokteran yaitu sebesar 90 %.

Dalam Penelitian Eko, dkk (2019) mencoba membandingkan tingkat akurasi antara algoritma C4.5 dan K-NN dalam prediksi kelulusan tepat waktu, dan sumber datasets yang digunakan yaitu data mahasiswa Program Studi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto. Penelitian ini menghasilkan kesimpulan bahwa algoritma K-NN memiliki tingkat akurasi yang lebih tinggi yaitu 89,14% dan algoritma C 4.5 memiliki akurasi 84,75%.

Penelitian (Anam dan Santoso, 2018) mencoba membandingkan tingkat akurasi antara algoritma C4.5 dan Naive Bayes dalam prediksi persediaan stok barang, dengan hasil pengujian menunjukkan algoritma C4.5 memiliki tingkat akurasi sebesar 96.40% lebih baik dari tingkat akurasi algoritma Naive Bayes sebesar 95.11%. Pengujian divalidasi dengan 10-fold cross validation dan evaluasi menggunakan confusion matrix.

Berdasarkan pembahasan diatas, peneliti akan membandingkan tingkat akurasi dari dua metode data mining, yaitu algoritma C 4.5 dan algoritma Naïve Bayes dalam memprediksi keuluan mahasiswa. Pemilihan penggunaan Algoritma C4.5 dan Naïve Bayes pada penelitian ini didasarkan pada beberapa alasan, yaitu kedua algoritma memiliki kesamaan mudah diimplementasikan dan dapat memberikan hasil yang baik dalam kasus klasifikasi.

METODE

Penelitian ini menggunakan pendekatan penelitian klasifikasi, Pendekatan klasifikasi adalah salah satu jenis pendekatan dalam Machine Learning. Klasifikasi merupakan sebuah metode dengan membagi suatu objek menjadi beberapa kelas berdasarkan kelas yang diinginkan (Matondang, dkk., 2022). Klasifikasi dalam data mining merupakan metode yang bertujuan memprediksi class suatu objek yang namanya tidak diketahui (R. Fitra & I. Rusdi, 2018).

Decision Tree

Pohon keputusan adalah algoritma yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon. Keuntungan decision tree yaitu mudah dibaca tanpa perlu pengetahuan statistik, mudah disiapkan tanpa harus menghitung dengan perhitungan yang rumit. Metode C4.5 mengubah fakta besar menjadi pohon keputusan yang mewakili aturan, sehingga aturan tersebut dapat dengan mudah dipahami oleh manusia (Farmawati & Narti, 2022)

Algoritma C 4.5

Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan yang merupakan perkembangan dari algoritma ID3, algoritma tersebut digunakan untuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut node internal atau test node, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu class (Maimon & Rokack, 2010). Langkah-langkah membangun pohon keputusan menggunakan algoritma C4.5 adalah sebagai berikut (Kusrini & Luthfi, 2009):

1. Pilih atribut sebagai akar. Pemilihan atribut sebagai akar berdasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung nilai gain tertinggi digunakan persamaan berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan: S : himpunan kasus
A : atribut n : jumlah partisi atribut A
|S_i| : jumlah kasus pada partisi ke-i
|S| : jumlah kasus dalam S
Nilai entropi dapat dihitung dengan cara berikut:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Dimana: S : himpunan kasus
n : jumlah partisi S
Pi : proporsi dari S_i terhadap S

2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Algoritma Naïve Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. (Kusrini, 2009).

Teorema Bayes memiliki bentuk umum sebagai berikut :

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

X = Data dengan class yang belum diketahui
H = Hipotesis data X merupakan suatu class spesifik
P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)
 $P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut
 $P(X)$ = Probabilitas dari X

HASIL DAN PEMBAHASAN

Data Set

Data yang digunakan dalam penelitian ini adalah data set kelulusan romi yaitu 379 mahasiswa terdiri dari IP semester 1-8 dan IPK yang digunakan untuk mengetahui status kelulusan, berikut adalah data responden:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	NAMA	JENIS KELAMIN	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
2	ANIK WIDAYANTI	PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07	TERLAMBAT
3	HESTYNA PRIHASTA	PEREMPUAN	MAHASISWA	32	BELUM MENIKAH	3	3.3	3.14	3.14	2.84	3.13	3.25	0	3.17	TERLAMBAT
4	MURYA ARIEF BASUKI	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	3.5	3.3	3.7	3.29	3.53	3.72	3.73	0	3.54	TERLAMBAT
5	NANIK SUSANTI	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	3.17	3.41	3.61	3.36	3.48	3.63	3.46	0	3.41	TERLAMBAT
6	RIFKA ISTIQFARINA	PEREMPUAN	BEKERJA	29	BELUM MENIKAH	2.9	2.89	3.3	2.85	2.98	3	3.08	0	3.09	TERLAMBAT
7	SUHARYONO	LAKI - LAKI	BEKERJA	27	BELUM MENIKAH	2.95	2.82	3.09	3.1	2.78	3.16	3.23	0	3.07	TERLAMBAT
8	FARIKHATUN NAZLI	PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	2.76	3.14	2.6	2.95	3.23	3.33	3.3	3.3	3.06	TEPAT
9	FIFI SUNALISA	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	2.62	2.89	2.32	2.5	2.5	2.86	3.05	2.5	2.91	TEPAT
10	HENDRIK MULIYANTO	PEREMPUAN	BEKERJA	25	MENIKAH	3.6	3.54	3.52	3.39	3.52	3.68	3.15	0	3.4	TERLAMBAT
11	MAM AGUNG RIBOWO	PEREMPUAN	BEKERJA	28	BELUM MENIKAH	2.71	2.55	1.77	2.11	1.93	2.13	1.78	0.2	2.2	TERLAMBAT
12	IMAM SANTOSA	PEREMPUAN	BEKERJA	27	BELUM MENIKAH	3.14	3.46	3.4	3.43	3.27	3.15	3.81	4	3.44	TERLAMBAT
13	IRFAN EKO WAHYUDI	PEREMPUAN	BEKERJA	32	BELUM MENIKAH	2.67	2.3	1.57	1.44	1.58	1.68	1.13	0.94	2.4	TERLAMBAT
14	IWAN HAMBALI	PEREMPUAN	BEKERJA	26	BELUM MENIKAH	2.57	2.82	2.2	2.45	2.1	2.42	1	1.42	2.45	TERLAMBAT
15	M SYAIFULLAH	PEREMPUAN	BEKERJA	31	BELUM MENIKAH	2.71	3	2.65	2.27	2.13	3.34	2.5	0	2.57	TERLAMBAT
16	DIANA LAILY FITRI	PEREMPUAN	MAHASISWA	26	BELUM MENIKAH	3.24	3.38	3.44	3.3	3.56	3.45	3.35	3.35	3.45	TERLAMBAT
17	DONNY PRABOWO	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2.86	2.86	2.45	1.86	3.19	3.14	2.98	2	3.04	TERLAMBAT
18	EDI JATMIKO	PEREMPUAN	MAHASISWA	27	BELUM MENIKAH	2.71	3.27	2.54	3.36	3.23	3.28	3.15	2	3.19	TERLAMBAT
19	ADI SUBEKTI	LAKI - LAKI	BEKERJA	29	BELUM MENIKAH	2.67	2.2	1.45	3	3.17	3.08	3.33	0	2.95	TERLAMBAT
20	AHMAD IBROZI	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	2.67	2.68	1.95	1.61	2.36	2.21	0.78	0.92	2.11	TERLAMBAT
21	ANDI HARDIANTO	LAKI - LAKI	MAHASISWA	27	BELUM MENIKAH	3.1	3.71	2.96	3.4	3.4	3.61	3.73	3.4	3.38	TEPAT
22	ANDI SUNARWAN	LAKI - LAKI	BEKERJA	28	BELUM MENIKAH	3	3.33	2.88	2.8	3.36	3.07	3.43	4	3.29	TERLAMBAT
23	ANNY AZIZAH	LAKI - LAKI	BEKERJA	26	BELUM MENIKAH	3.12	3.23	2.96	3	3.31	3	3.08	3	3.19	TERLAMBAT
24	ARIEF PRAM WAHYUD	LAKI - LAKI	MAHASISWA	26	BELUM MENIKAH	2.9	3.32	2.83	2.91	3.3	2.83	3.27	4	3.11	TERLAMBAT
25	INY SELA PUTROWIBO	LAKI - LAKI	BEKERJA	29	BELUM MENIKAH	2.95	3.41	3.48	3.24	2.95	3	2.75	0	3.2	TERLAMBAT
274	ABDUL ROSYID	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	2.29	2.89	2.48	3.14	2.9	2.4	2.73	0	2.85	TEPAT
375	ARY JULI SETIYANTO	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	1.98	2.5	2.14	2.77	2.61	2.93	2.82	2.5	0.99	TEPAT
377	INA ZAHROTUL UMAN	PEREMPUAN	BEKERJA	23	BELUM MENIKAH	2.74	2.75	2.55	3	2.98	2.8	3.14	3	2.97	TEPAT
378	A WAHYUADI KRISNA	PEREMPUAN	MAHASISWA	23	BELUM MENIKAH	2.74	2.75	2.55	3	2.98	2.8	3.14	3	3.03	TEPAT
379	NIMATUL JANNAH	PEREMPUAN	MAHASISWA	23	BELUM MENIKAH	3.02	2.94	3.25	2.87	3	2.94	3.09	3	3.16	TEPAT
380	NDU SETYO WICAKSO	LAKI - LAKI	MAHASISWA	23	BELUM MENIKAH	3.1	3.06	3	3.23	2.79	3	2.41	3	2.16	TEPAT

Gambar 1. Data Responden

Pada tahap ini menentukan data yang akan di proses. Berikut adalah data hasil pengolahan yang sudah melalui tahap seleksi.

1	STATUS MAHASISWA	UMUR	STATUS NIKAH	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK	STATUS KELULUSAN
2	BEKERJA	28	BELUM MENIKAH	2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07	TERLAMBAT
3	MAHASISWA	32	BELUM MENIKAH	3	3.3	3.14	3.14	2.84	3.13	3.25	0	3.17	TERLAMBAT
4	BEKERJA	29	BELUM MENIKAH	3.5	3.3	3.7	3.29	3.53	3.72	3.73	0	3.54	TERLAMBAT
5	MAHASISWA	27	BELUM MENIKAH	3.17	3.41	3.61	3.36	3.48	3.63	3.46	0	3.41	TERLAMBAT
6	BEKERJA	29	BELUM MENIKAH	2.9	2.89	3.3	2.85	2.98	3	3.08	0	3.09	TERLAMBAT
7	BEKERJA	27	BELUM MENIKAH	2.95	2.82	3.09	3.1	2.78	3.16	3.23	0	3.07	TERLAMBAT
8	MAHASISWA	26	BELUM MENIKAH	2.76	3.14	2.6	2.95	3.23	3.33	3.3	3.3	3.06	TEPAT
9	MAHASISWA	27	BELUM MENIKAH	2.62	2.89	2.32	2.5	2.5	2.86	3.05	2.5	2.91	TEPAT
10	BEKERJA	25	MENIKAH	3.6	3.54	3.52	3.39	3.52	3.68	3.15	0	3.4	TERLAMBAT
11	BEKERJA	28	BELUM MENIKAH	2.71	2.55	1.77	2.11	1.93	2.13	1.78	0.2	2.2	TERLAMBAT
12	BEKERJA	27	BELUM MENIKAH	3.14	3.46	3.4	3.43	3.27	3.15	3.81	4	3.44	TERLAMBAT
13	BEKERJA	32	BELUM MENIKAH	2.67	2.3	1.57	1.44	1.58	1.68	1.13	0.94	2.4	TERLAMBAT
14	BEKERJA	26	BELUM MENIKAH	2.57	2.82	2.2	2.45	2.1	2.42	1	1.42	2.45	TERLAMBAT
15	BEKERJA	31	BELUM MENIKAH	2.71	3	2.65	2.27	2.13	3.34	2.5	0	2.57	TERLAMBAT
16	MAHASISWA	26	BELUM MENIKAH	3.24	3.38	3.44	3.3	3.56	3.45	3.35	3.35	3.45	TERLAMBAT
17	MAHASISWA	27	BELUM MENIKAH	2.86	2.86	2.45	1.86	3.19	3.14	2.98	2	3.04	TERLAMBAT
18	MAHASISWA	27	BELUM MENIKAH	2.71	3.27	2.54	3.36	3.23	3.28	3.15	2	3.19	TERLAMBAT
19	BEKERJA	29	BELUM MENIKAH	2.67	2.2	1.45	3	3.17	3.08	3.33	0	2.95	TERLAMBAT
20	MAHASISWA	27	BELUM MENIKAH	2.67	2.68	1.95	1.61	2.36	2.21	0.78	0.92	2.11	TERLAMBAT
21	MAHASISWA	27	BELUM MENIKAH	3.1	3.71	2.96	3.4	3.4	3.61	3.73	3.4	3.38	TEPAT
22	BEKERJA	28	BELUM MENIKAH	3	3.33	2.88	2.8	3.36	3.07	3.43	4	3.29	TERLAMBAT
23	BEKERJA	26	BELUM MENIKAH	3.12	3.23	2.96	3	3.31	3	3.08	3	3.19	TERLAMBAT
24	MAHASISWA	26	BELUM MENIKAH	2.9	3.32	2.83	2.91	3.3	2.83	3.27	4	3.11	TERLAMBAT
25	BEKERJA	29	BELUM MENIKAH	2.95	3.41	3.48	3.24	2.95	3	2.75	0	3.2	TERLAMBAT
375	MAHASISWA	23	BELUM MENIKAH	2.29	2.89	2.48	3.14	2.9	2.4	2.73	0	2.85	TEPAT
376	MAHASISWA	23	BELUM MENIKAH	1.98	2.5	2.14	2.77	2.61	2.93	2.82	2.5	0.99	TEPAT
377	BEKERJA	23	BELUM MENIKAH	2.74	2.75	2.55	3	2.98	2.8	3.14	3	2.97	TEPAT
378	MAHASISWA	23	BELUM MENIKAH	2.74	2.75	2.55	3	2.98	2.8	3.14	3	3.03	TEPAT
379	MAHASISWA	23	BELUM MENIKAH	3.02	2.94	3.25	2.87	3	2.94	3.09	3	3.16	TEPAT
380	MAHASISWA	23	BELUM MENIKAH	3.1	3.06	3	3.23	2.79	3	2.41	3	2.16	TEPAT

Gambar 2. Hasil Pengolahan Data

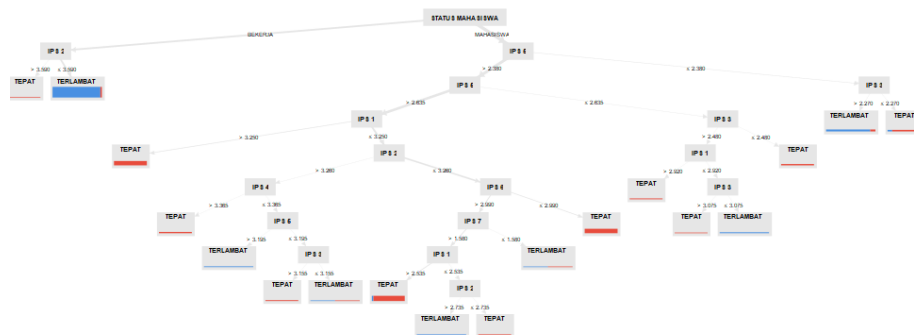
Implementasi Algoritma C4.5 dengan Tools Rapid Miner

Pada tahap ini penulis menggunakan dua operator yaitu Retrieve Data yang digunakan untuk memasukan data yang telah di import dari excel dan C4.5 yang digunakan untuk mengolah data menjadi informasi.



Gambar 3. Klasifikasi Algoritma C 4.5

Pada gambar diatas merupakan bentuk desain model dari dataset kelulusan mahasiswa yang direlasikan dengan Algoritma C4.5 menghasilkan sebuah pohon keputusan sebagai berikut:

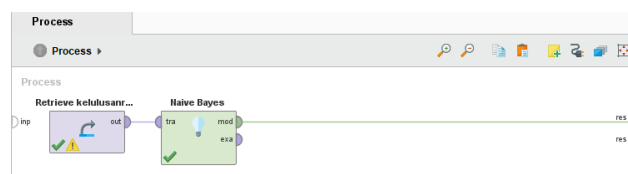


Gambar 4. Hasil Pohon Keputusan Kelulusan Mahasiswa

Pada gambar diatas bahwa node tertinggi dari pohon keputusan adalah Status Mahasiswa.

Implementasi algoritma Naïve Bayes dengan Tools RapidMiner

Desain model algoritma Naïve Bayes menggunakan Tools RapidMiner:



Gambar 5. Klasifikasi Naïve Bayes

Pada gambar diatas merupakan bentuk desain model dari dataset kelulusan mahasiswa yang direlasikan dengan Algoritma Naïve Bayes sehingga menghasilkan model aturan sebagai berikut:

SimpleDistribution

```
Distribution model for label attribute STATUS KELULUSAN

Class TERLAMBAT (0.430)
14 distributions

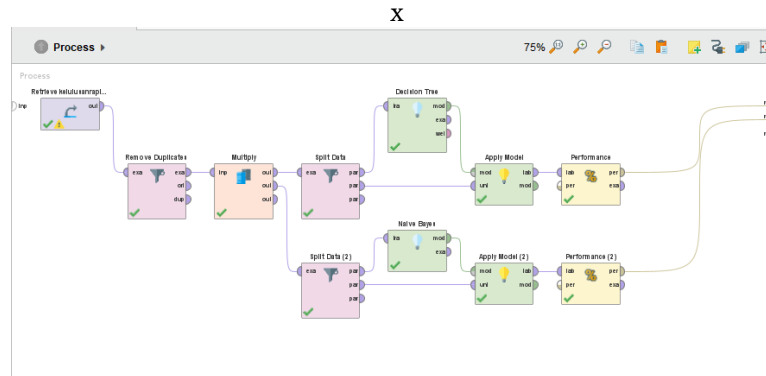
Class TEPAT (0.570)
14 distributions
```

Gambar 6. Hasil Algoritma Naïve Bayes

Pada gambar diatas dapat dilihat bahwa hasil dari Algoritma Naïve Bayes diperoleh class terlambat dengan nilai 0.430, dan class tepat dengan nilai 0.570. Maka dapat disimpulkan bahwa class tepat merupakan class yang tertinggi pada kelulusan mahasiswa

Evaluasi dan Validasi

Penerapan modeling Algoritma C4.5 dan Naïve Bayes dapat dilihat pada Gambar 7.



Gambar 7. Model Klasifikasi Algoritma C4.5 dan Naïve Bayes

Pada Gambar 7, tahapan pertama dalam membangun model Algoritma yaitu pemanggilan dataset dengan menggunakan retrieve pada RapidMiner, dan gambar diatas menjelaskan model training yang digunakan sebagai operator algoritma C4.5 dan Naïve Bayes serta testing validasi. Berikut hasil akurasi kedua algoritma:

Table View Plot View

accuracy: 90.79%

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	30	4	88.24%
pred. TEPAT	3	39	92.86%
class recall	90.91%	90.70%	

Gambar 8. Hasil Akurasi Algoritma C 4.5

Table View Plot View

accuracy: 75.00%

	true TERLAMBAT	true TEPAT	class precision
pred. TERLAMBAT	26	12	68.42%
pred. TEPAT	7	31	81.58%
class recall	78.79%	72.09%	

Gambar 9. Hasil Akurasi Algoritma Naïve Bayes

Berdasarkan pada Gambar 7 diatas bahwa tingkat akurasi menggunakan Algoritma C4.5 adalah 90.79% sedangkan pada Gambar.. bahwa tingkat akurasi menggunakan Algoritma Naïve Bayes adalah 75.00% maka tingkat akurasi menggunakan algoritma C 4.5 lebih tinggi dibandingkan dengan Naïve Bayes dengan selisih 15.79%.

SIMPULAN

Berdasarkan hasil pengujian dan analisis bahwa pengujian ini bertujuan untuk mengetahui diantara model algoritma C4.5 dan Naive Bayes yang memiliki akurasi paling tinggi untuk memprediksi kelulusan mahasiswa. Berdasarkan hasil pengukuran tingkat akurasi kedua algoritma tersebut, diketahui bahwa nilai akurasi C4.5 adalah 90.79%, sedangkan nilai akurasi Naive Bayes 75.00% dapat disimpulkan bahwa dengan menggunakan model algoritma C 4.5 lebih tinggi tingkat akurasinya dalam memprediksi kelulusan mahasiswa.

DAFTAR PUSTAKA

- Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal Energy*, 13-19.
- Pratama, R., Huda, B., Novalia, E., dan Kabir, H. (2022). Perbandingan Algoritma C4.5 dan Naïve Bayes dalam Menentukan Persediaan Stok. vol. 6, no 2, 115-122
- Gunawan, M., Zarlis, M., dan Roslina (2021). Analisis Komparasi Algoritma Naïve Bayes dan K- Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 513-523
- Rahmayanti, A., Rusdiana, L., Suratno, S. (2022). Perbandingan Metode Algoritma C4.5 dan Naive Bayes Untuk Memprediksi Kelulusan Mahasiswa. *Walisongo Journal of Information Technology*. vol. 4, no. 1, 11-22
- Budiman, A., Setyanto, A., Wibowo, F, W. (2019). Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma C4.5. *TEKNOMATIKA*, vol.11, no.2, 83-93
- Turban, E., Aronson, J. E., dan Liang, T.-P. (2005). *Decision Support Systems and Intelligent Systems* (7th ed.). USA: Pearson Education
- Kusrini, dan Luthfi, E. T. (2009). *Agoritma Data Mining*. Yogyakarta: Penerbit Andi.
- F. Narti, "JTIM : Jurnal Teknologi Informasi dan Multimedia Perbandingan Algoritma C4 . 5 dan Naive Bayes dalam Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Daring," vol. 4, no. 1, 2022.
- Maimon, o., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook Second Edition*. New York:Springer.
- M. R. Matondang, M. R. Lubis, and H. S. Tambunan, "Analisis Data mining dengan Metode C.45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap," *Brahmana J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 74–81, 2021, doi: 10.30645/brahmana.v2i2.68
- R. Fitra and I. Rusdi, "Penerapan Metode Algoritma K-Nearest Neighbor Menggunakan Rapidminer Studio pada Klasifikasi Status Sosial Ekonomi Studi Kasus : Kelurahan Kapuk Muara Rt,"pp. 653–660, 2018
- <https://romisatriawahono.net/dm/>