

PENERAPAN DATA MINING UNTUK PREDIKSI TINGKAT KELULUSAN SISWA MENGGUNAKAN METODE NAÏVE BAYES (STUDI KASUS SMP DHARMA WANITA 17 WONOAYU)

Oleh:

Tya Sukmawati

Arif Senja Fitrani

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Juni, 2023



Pendahuluan

Pendidikan merupakan suatu usaha untuk menyiapkan manusia dengan berbagai kegiatan, pelatihan, pengajaran yang dapat berpengaruh terhadap sikap dan tingkah laku untuk meningkatkan kualitas hidup manusia. Melalui pendidikan manusia dapat memperluas wawasan dan ilmu pengetahuan serta dapat meningkatkan tingkat spiritual dan keterampilan yang sangat diperlukan dalam bermasyarakat.

Memprediksi prestasi akademik dan tingkat kelulusan siswa sangatlah penting bagi suatu lembaga pendidikan karena hal tersebut dapat direncanakan dalam meningkatkan tingkat kelulusan siswa. Dalam meningkatkan angka kelulusan siswa di sekolah sistem ini akan membantu pihak sekolah untuk mengetahui pola kelulusan siswa dan siswinya. Dari pola tersebut diharapkan dapat membantu menganalisa faktor penyebab yang mempengaruhi tingkat kelulusan.

Saat ini di sekolah SMP Dharma wanita 17 Wonoayu masih belum ada sistem yang dapat memprediksi tingkat kelulusan siswa. Berdasarkan permasalahan tersebut penulis akan merancang aplikasi prediksi tingkat kelulusan siswa di SMP Dharma Dharma wanita 17 Wonoayu dengan menggunakan metode naïve bayes. Dengan adanya aplikasi data mining tersebut diharapkan dapat memprediksi tingkat kelulusan setiap tahunnya dan dapat merancang suatu pola pendidikan yang berdasarkan dari hasil prediksi tingkat kelulusan sebelumnya.

Pertanyaan Penelitian (Rumusan Masalah)

Dari latar belakang yang dikemukakan diatas, maka yang menjadi rumusan masalah dalam penelitian ini adalah bagaimana penerapan data mining dalam prediksi tingkat kelulusan siswa dengan menggunakan metode naive bayes classifier pada aplikasi Rapid Miner.

Metode

- **Teknik Pengumpulan Data**

Dalam memperoleh data yang dapat menunjang penelitian, maka peneliti melakukan pencarian dan pengumpulan data. Beberapa metode yang dibutuhkan dalam pengumpulan data ialah sebagai berikut:

1. Studi Pustaka (Literatur)
2. Observasi

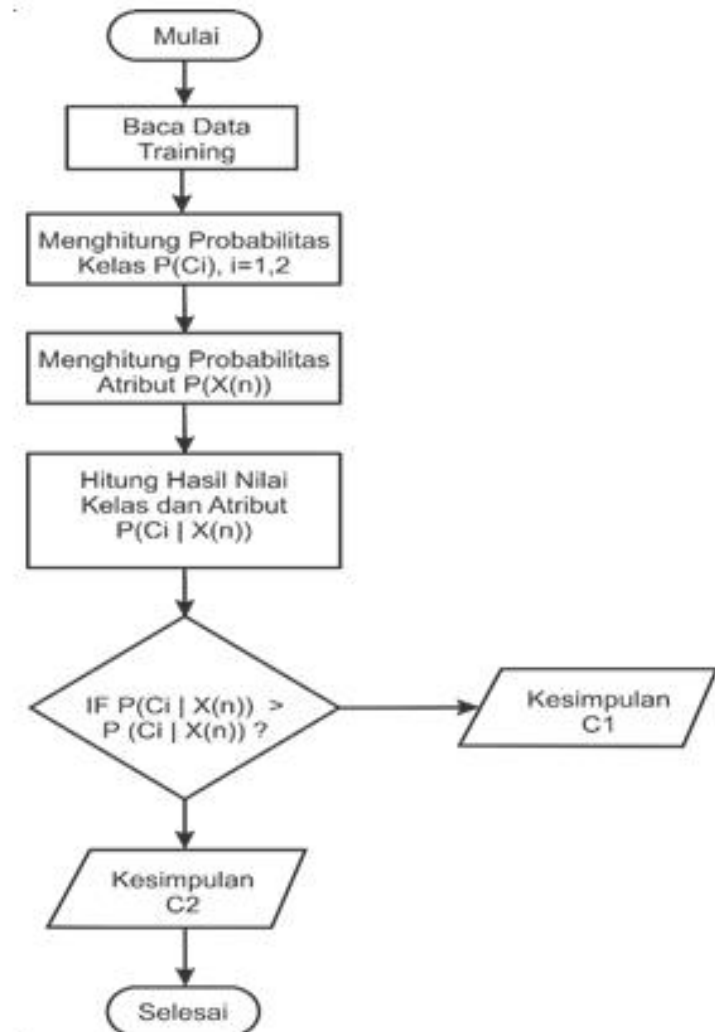
Hasil

- **Flowchart Sistem Aplikasi**

Berikut ini adalah gambaran dari flowchart metode *naive bayes* pada data testing dan data training untuk mengklasifikasi tingkat kelulusan SMP DHARMA WANITA 17 WONOAYU yaitu sebagai berikut:

Berdasarkan gambaran flowchart diatas, secara umum tahapan dari klasifikasi algoritma *naive bayes* pada data training ialah sebagai berikut :

1. Membaca data training.
2. Menghitung probabilitas Kelas.
3. Menghitung probabilitas atribut.
4. Menghitung nilai kelas dan atribut
5. Mendapatkan output dari hasil klasifikasi
6. Menghitung data pada variabel C1 menggunakan hasil perhitungan probabilitas pada data training
7. Mendapatkan hasil output prediksi dari hasil klasifikasi.



Hasil

- Input

Pada penelitian ini, peneliti menggunakan data siswa dan siswi dari sekolah smp dharma wanita 17 wonoayu. Data yang digunakan dari angkatan 2015 – 2020 dengan total 450 data siswa dan siswi.

DATA SEBELUM PREPROCESSING													
DATASET SMP DHARMA WANITA 17 WONOAYU													
TAHUN ANGKATAN 2015-2020													
NO	ANGKATAN	NAMA SISWA	JENIS KELAMIN	PENDIDIKAN AYAH	PENDIDIKAN IBU	ALAMAT RUMAH	JUMLAH SAUDARA	PEKERJAAN AYAH	PEKERJAAN IBU	TINGKAT EKONOMI	NILAI RAPOR	NILAI UAS	STATUS KELULUSAN
1	2015	Dika Ari	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	77	80	LULUS
2	2015	Kurniawan Adi	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	74	73	LULUS
3	2015	Feri Permana	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	81	83	LULUS
4	2015	Agus Prasetyo	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	86	88	LULUS
5	2015	M. Tomi	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	71	80	LULUS
6	2015	Bagus Adi Wijaya	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	69	74	LULUS
7	2015	Kurniawan Agung	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	89	86	LULUS
8	2015	M. Dicky	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	77	71	LULUS
9	2015	Hendra Setiawan	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	79	81	LULUS
10	2015	Mashuri	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	76	88	LULUS
11	2015	Taufik Hidayat	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	84	90	LULUS
12	2015	Doni Firmansyah	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	66	60	TIDAK LULUS
13	2015	Agus Dwi Setia	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Menengah	72	89	LULUS
14	2015	Muhammad Roni	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Bawah	63	71	LULUS
15	2015	Muhammad Anas	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Bawah	61	75	LULUS
16	2015	M. Mukhorobin	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Bawah	88	94	LULUS
17	2015	Lukman Hakim	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	74	87	LULUS
18	2015	Danar Dharmawan	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	70	76	LULUS
19	2015	Vika Tanjung	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	62	70	LULUS
20	2015	Ahsana Maswa	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	81	86	LULUS
21	2015	Rahmat Hariyanto	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	73	79	LULUS
22	2015	Adi Saputra	Laki-Laki	SMA	SMA	Wonoayu	1	PNS	Swasta	Atas	70	72	LULUS
23	2015	Bayu Adi Nugroho	Laki-Laki	SMA	SMA	Wonoayu	3	PNS	Swasta	Atas	66	70	LULUS
24	2015	Hendra Harja	Laki-Laki	SMA	SMA	Wonoayu	3	PNS	Swasta	Atas	55	61	TIDAK LULUS
25	2015	Akbar Maulana	Laki-Laki	SMA	SMA	Wonoayu	3	PNS	Swasta	Atas	89	98	LULUS
26	2015	Agus Wahyu	Laki-Laki	SMA	S1	Wonoayu	3	PNS	PNS	Atas	71	79	LULUS
27	2015	Hendri Kurniawan	Laki-Laki	SMA	S1	Wonoayu	3	Swasta	PNS	Atas	69	78	LULUS
28	2015	Danang Prasetya	Laki-Laki	SMA	S1	Wonoayu	3	Swasta	PNS	Atas	75	64	LULUS

Hasil

- **Data Preprocessing**

A. Atribut Selection

Dalam preprocessing atribut selection merupakan salah satu tahap untuk memilih atribut yang dianggap relevan dan menghilangkan atribut yang memiliki sifat privasi atau tidak relevan.

Atribut yang dianggap relevan	Atribut yang tidak relevan
Jenis kelamin	Nama
Pendidikan ayah	Alamat
Pendidikan ibu	Angkatan
Jumlah saudara	
Pekerjaan ayah	
Pekerjaan ibu	
Tingkat ekonomi	
Nilai rapor	
Nilai UAS	
Status Kelulusan	

B. Data Cleaning

Dalam langkah *data cleaning*, data mentah akan dibersihkan melalui beberapa proses seperti mengisi nilai yang hilang, memperbaiki data karena ada simbol seperti (@ # : ;) dan lain-lain. Data juga bisa dibersihkan dan dirapikan menggunakan segmen-segmen yang memiliki ukuran serupa atau dengan mengelompokkannya ke dalam kelompok data yang serupa (*grouping*)[10].

C. Data Tranformation

Pada tabel 2. data nilai rapor dan nilai UAS memiliki tipe data numeric, agar data dapat diolah maka harus dilakukan proses tranformasi data kedalam bentuk kategori, sebagai berikut:

NILAI RAPOR	NILAI UAS
77	80
74	73
81	83
86	88
71	80

Nilai Rata-Rata Rapor dan UAS	
Kategori	Keterangan
TINGGI	81-100
SEDANG	70-80
KURANG	<=69

NILAI RAPOR	NILAI UAS
SEDANG	SEDANG
SEDANG	SEDANG
TINGGI	TINGGI
TINGGI	TINGGI
SEDANG	SEDANG

Hasil

Data Random

Pada proses ini adalah mengacak data record agar data record yang memiliki nilai yang sama tidak berurutan. Fungsi tahap ini untuk mengacak data record sehingga data record tersebut acak dan tidak memiliki nilai sama yang berurutan. Agar dapat meningkatkan akurasi perhitungan data mining naïve bayes.

	JENIS KELAMIN	PENDIDIKAN AYAH	PENDIDIKAN IBU	ALAMAT RUMAH	JUMLAH SAUDARA	PEKERJAAN AYAH	PEKERJAAN IBU	TINGKAT EKONOMI	NILAI RAPOR	NILAI UAS	STATUS KELULUSAN
5	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	77	80	LULUS
6	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	74	73	LULUS
7	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	81	83	LULUS
8	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	86	88	LULUS
9	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	71	80	LULUS
10	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	69	74	LULUS
11	Laki-Laki	SMA	SMA	Wonoayu	2	Wiraswasta	Ibu Rumah Tangga	Menengah	89	86	LULUS
12	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	77	71	LULUS
13	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	79	81	LULUS
14	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	76	88	LULUS
15	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	84	90	LULUS
16	Laki-Laki	SMA	SMA	Wonoayu	1	Wiraswasta	Ibu Rumah Tangga	Menengah	66	60	TIDAK LULUS
17	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Menengah	72	89	LULUS
18	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Bawah	63	71	LULUS
19	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Bawah	61	75	LULUS
20	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Ibu Rumah Tangga	Bawah	88	94	LULUS
21	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	74	87	LULUS
22	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	70	76	LULUS
23	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	62	70	LULUS
24	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	81	86	LULUS
25	Laki-Laki	SMA	SMA	Wonoayu	1	Swasta	Swasta	Bawah	73	79	LULUS
26	Laki-Laki	SMA	SMA	Wonoayu	1	PNS	Swasta	Atas	70	72	LULUS

	JENIS KELAMIN	PENDIDIKAN AYAH	PENDIDIKAN IBU	JUMLAH SAUDARA	PEKERJAAN AYAH	PEKERJAAN IBU	TINGKAT EKONOMI	NILAI RAPOR	NILAI UAS	STATUS KELULUSAN
1	Laki-Laki	SMA	SMA	2	Wiraswasta	Ibu Rumah Tangga	Menengah	KURANG	SEDANG	LULUS
2	Perempuan	SMA	SMA	2	Swasta	Wiraswasta	Menengah	SEDANG	TINGGI	LULUS
3	Perempuan	S1	S1	2	Wiraswasta	Wiraswasta	Bawah	KURANG	SEDANG	LULUS
4	Perempuan	D3	SMA	3	Swasta	Wiraswasta	Bawah	SEDANG	SEDANG	LULUS
5	Laki-Laki	SMA	SMA	2	Swasta	Swasta	Bawah	SEDANG	SEDANG	LULUS
6	Laki-Laki	SMA	SMK	2	Swasta	PNS	Atas	SEDANG	TINGGI	LULUS
7	Perempuan	D3	D3	3	PNS	PNS	Atas	SEDANG	TINGGI	LULUS
8	Laki-Laki	SMA	SMA	2	PNS	Swasta	Menengah	SEDANG	TINGGI	LULUS
9	Laki-Laki	SMA	SMA	3	Swasta	Swasta	Menengah	TINGGI	SEDANG	LULUS
10	Perempuan	D3	SMA	1	Swasta	Wiraswasta	Menengah	TINGGI	SEDANG	LULUS
11	Laki-Laki	SMA	SMA	1	Swasta	Swasta	Menengah	TINGGI	SEDANG	LULUS
12	Perempuan	S1	S1	3	Wiraswasta	PNS	Menengah	KURANG	SEDANG	LULUS
13	Laki-Laki	S1	D3	3	PNS	PNS	Atas	SEDANG	SEDANG	LULUS
14	Perempuan	SMK	SMA	1	Wiraswasta	Swasta	Bawah	SEDANG	SEDANG	LULUS
15	Laki-Laki	SMA	SMK	2	Swasta	Wiraswasta	Atas	SEDANG	TINGGI	LULUS
16	Laki-Laki	SMA	SMA	2	Swasta	Swasta	Atas	TINGGI	TINGGI	LULUS
17	Laki-Laki	S1	SMA	2	PNS	Wiraswasta	Atas	SEDANG	TINGGI	LULUS
18	Perempuan	SMK	SMA	1	Wiraswasta	Wiraswasta	Bawah	SEDANG	SEDANG	LULUS
19	Laki-Laki	D3	SMA	2	Swasta	PNS	Menengah	SEDANG	SEDANG	LULUS
20	Laki-Laki	S1	SMA	2	PNS	Wiraswasta	Menengah	SEDANG	TINGGI	LULUS
21	Laki-Laki	SMA	SMA	3	PNS	Swasta	Atas	SEDANG	SEDANG	LULUS
22	Perempuan	SMA	SMA	1	Swasta	Wiraswasta	Menengah	KURANG	SEDANG	LULUS
23	Laki-Laki	SMA	SMA	2	Swasta	Wiraswasta	Atas	SEDANG	SEDANG	LULUS
24	Laki-Laki	S1	D3	3	PNS	PNS	Menengah	KURANG	SEDANG	LULUS
25	Laki-Laki	SMA	S1	2	Wiraswasta	PNS	Menengah	TINGGI	SEDANG	LULUS
26	Perempuan	S1	D3	3	Wiraswasta	PNS	Menengah	SEDANG	TINGGI	LULUS
27	Laki-Laki	SMA	SMA	1	Swasta	Ibu Rumah Tangga	Bawah	KURANG	SEDANG	LULUS
28	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Bawah	KURANG	SEDANG	LULUS
29	Laki-Laki	S1	D3	1	PNS	Wiraswasta	Menengah	SEDANG	SEDANG	LULUS
30	Laki-Laki	SMA	SMK	2	Swasta	PNS	Atas	SEDANG	SEDANG	LULUS
31	Perempuan	SMA	SMA	1	Wiraswasta	Ibu Rumah Tangga	Atas	SEDANG	SEDANG	LULUS
32	Laki-Laki	SMA	SMA	1	Wiraswasta	Ibu Rumah Tangga	Menengah	SEDANG	TINGGI	LULUS
33	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Atas	SEDANG	SEDANG	LULUS

Split Data

Sebelum melakukan proses data mining naïve bayes. Dataset yang telah melalui proses data random akan dibagi menjadi dua bagian yaitu data training dan data testing. Pembagian data ini disebut split data, dimana alokasi data sebesar 70% digunakan untuk data training dan 30% untuk data testing. Jadi dari 450 dataset akan dibagi menjadi data training sebanyak 315 data dan data testing sebanyak 135 data.

Pembahasan

- **Proses Naïve Bayes**

Pengujian ini menggunakan data training sebanyak 315 dan data testing sebanyak 135 dari data siswa dengan nomor urut 1 sampa 315.

Menghitung Probabilitas Label / Class Status Kelulusan pada data training : Rumus probabilitas class : $P(C) = \text{Jumlah Data C} / \text{Jumlah Total Data}$

$$P(\text{Class} = \text{LULUS}) = 302 / 315 = 0.958$$

$$P(\text{Class} = \text{TIDAK LULUS}) = 13 / 315 = 0.041$$

Menghitung Probabilitas antar atribut pada data training :

Rumus Probabilitas Atribut : $P(X | C) = \text{Kondisi Atribut X} / \text{Total C}$

$$P(\text{Jenis Kelamin} = \text{Laki - Laki} | \text{Lulus}) = 142 / 302 = 0.470$$

$$P(\text{Jenis Kelamin} = \text{Laki - Laki} | \text{Tidak Lulus}) = 6 / 13 = 0.461$$

DATA TRAINING PENGUJIAN										
NO	JENIS KELAMIN	PENDIDIKAN AYAH	PENDIDIKAN IBU	JUMLAH SAUDARA	PEKERJAAN AYAH	PEKERJAAN IBU	TINGKAT EKONOMI	NILAI RAPOR	NILAI UAS	STATUS KELULUSAN
1	Laki-Laki	SMA	SMA	2	Wiraswasta	Ibu Rumah Tangga	Menengah	KURANG	SEDANG	LULUS
2	Perempuan	SMA	SMA	2	Swasta	Wiraswasta	Menengah	SEDANG	TINGGI	LULUS
3	Perempuan	S1	S1	2	Wiraswasta	Wiraswasta	Bawah	KURANG	SEDANG	LULUS
4	Perempuan	D3	SMA	3	Swasta	Wiraswasta	Bawah	SEDANG	SEDANG	LULUS
5	Laki-Laki	SMA	SMA	2	Swasta	Swasta	Bawah	SEDANG	SEDANG	LULUS
6	Laki-Laki	SMA	SMK	2	Swasta	PNS	Atas	SEDANG	TINGGI	LULUS
7	Perempuan	D3	D3	3	PNS	PNS	Atas	SEDANG	TINGGI	LULUS
8	Laki-Laki	SMA	SMA	2	PNS	Swasta	Menengah	SEDANG	TINGGI	LULUS
9	Laki-Laki	SMA	SMA	3	Swasta	Swasta	Menengah	TINGGI	SEDANG	LULUS
10	Perempuan	D3	SMA	1	Swasta	Wiraswasta	Menengah	TINGGI	SEDANG	LULUS
11	Laki-Laki	SMA	SMA	1	Swasta	Swasta	Menengah	TINGGI	SEDANG	LULUS
12	Perempuan	S1	S1	3	Wiraswasta	PNS	Menengah	KURANG	SEDANG	LULUS
13	Laki-Laki	S1	D3	3	PNS	PNS	Atas	SEDANG	SEDANG	LULUS
14	Perempuan	SMK	SMA	1	Wiraswasta	Swasta	Bawah	SEDANG	SEDANG	LULUS
15	Laki-Laki	SMA	SMK	2	Swasta	Wiraswasta	Atas	SEDANG	TINGGI	LULUS
16	Laki-Laki	SMA	SMA	2	Swasta	Swasta	Atas	TINGGI	TINGGI	LULUS
17	Laki-Laki	S1	SMA	2	PNS	Wiraswasta	Atas	SEDANG	TINGGI	LULUS
18	Perempuan	SMK	SMA	1	Wiraswasta	Wiraswasta	Bawah	SEDANG	SEDANG	LULUS
19	Laki-Laki	D3	SMA	2	Swasta	PNS	Menengah	SEDANG	SEDANG	LULUS
20	Laki-Laki	S1	SMA	2	PNS	Wiraswasta	Menengah	SEDANG	TINGGI	LULUS
21	Laki-Laki	SMA	SMA	3	PNS	Swasta	Atas	SEDANG	SEDANG	LULUS
22	Perempuan	SMA	SMA	1	Swasta	Wiraswasta	Menengah	KURANG	SEDANG	LULUS
23	Laki-Laki	SMA	SMA	2	Swasta	Wiraswasta	Atas	SEDANG	SEDANG	LULUS
24	Laki-Laki	S1	D3	3	PNS	PNS	Menengah	KURANG	SEDANG	LULUS
25	Laki-Laki	SMA	S1	2	Wiraswasta	PNS	Menengah	TINGGI	SEDANG	LULUS
26	Perempuan	S1	D3	3	Wiraswasta	PNS	Menengah	SEDANG	TINGGI	LULUS
27	Laki-Laki	SMA	SMA	1	Swasta	Ibu Rumah Tangga	Bawah	KURANG	SEDANG	LULUS
28	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Bawah	KURANG	SEDANG	LULUS
29	Laki-Laki	S1	D3	1	PNS	Wiraswasta	Menengah	SEDANG	SEDANG	LULUS
30	Laki-Laki	SMA	SMK	2	Swasta	PNS	Atas	SEDANG	SEDANG	LULUS
31	Perempuan	SMA	SMA	1	Wiraswasta	Ibu Rumah Tangga	Atas	SEDANG	SEDANG	LULUS
32	Laki-Laki	SMA	SMA	1	Wiraswasta	Ibu Rumah Tangga	Menengah	SEDANG	TINGGI	LULUS
33	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Atas	SEDANG	SEDANG	LULUS
...
315	Perempuan	SMK	SMA	2	Swasta	Swasta	Menengah	SEDANG	TINGGI	LULUS

Pembahasan

Menghitung Probabilitas $P(C | X)$ pada data testing dengan menggunakan pengetahuan hasil probabilitas class dan atribut pada data training. Sebagai contoh menghitung probabilitas data record no.2 pada data testing di gambar no 7.

$$P(\text{Class} = \text{'Lulus'} | X(i)) = 0.958 * (0.529 * 0.367 * 0.516 * 0.519 * 0.417 * 0.304 * 0.433 * 0.552 * 0.658) \\ = 0.958 * 0.00103 = 0.00099$$

$$P(\text{Class} = \text{'Tidak Lulus'} | X(i)) = 0.041 * (0.538 * 0.307 * 0.384 * 0.461 * 0.384 * 0.230 * 0.384 * 0 * 0) \\ = 0.041 * 0 = 0$$

Jadi hasil perhitungan probabilitas no.1 menghasil class **LULUS** dengan nilai probabilitas terbesar. Maka dapat disimpulkan untuk record no.1 mempunyai status kelulusan yaitu **LULUS**.

1	JENIS KELAMIN	PENDIDIKAN AYAH	PENDIDIKAN IBU	JUMLAH SAUDARA	PEKERJAAN AYAH	PEKERJAAN IBU	TINGKAT EKONOMI	NILAI RAPOR	NILAI UAS	STATUS KELULUSAN
2	Perempuan	SMA	SMA	2	Swasta	Swasta	Menengah	SEDANG	SEDANG	LULUS
3	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Menengah	SEDANG	SEDANG	LULUS
4	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Menengah	KURANG	SEDANG	LULUS
5	Perempuan	D3	S1	2	PNS	Wiraswasta	Menengah	TINGGI	SEDANG	LULUS
6	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Atas	TINGGI	TINGGI	LULUS
7	Perempuan	D3	SMK	1	Wiraswasta	Wiraswasta	Atas	SEDANG	SEDANG	LULUS
8	Perempuan	SMA	SMA	3	Wiraswasta	PNS	Menengah	SEDANG	SEDANG	LULUS
9	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Menengah	KURANG	SEDANG	LULUS
10	Laki-Laki	S1	D3	3	PNS	PNS	Atas	SEDANG	TINGGI	LULUS
11	Perempuan	D3	S1	2	Swasta	Wiraswasta	Bawah	SEDANG	SEDANG	LULUS
12	Perempuan	S1	S1	3	Wiraswasta	Wiraswasta	Bawah	TINGGI	TINGGI	LULUS
13	Perempuan	SMK	SMA	2	Wiraswasta	Swasta	Atas	SEDANG	SEDANG	LULUS
14	Laki-Laki	SMA	SMA	1	Swasta	Swasta	Atas	TINGGI	TINGGI	LULUS
15	Laki-Laki	S1	SMA	2	PNS	Wiraswasta	Atas	SEDANG	TINGGI	LULUS
16	Perempuan	SMA	SMA	1	Wiraswasta	Ibu Rumah Tangga	Atas	SEDANG	SEDANG	LULUS
17	Laki-Laki	S1	D3	4	PNS	Swasta	Atas	SEDANG	SEDANG	LULUS
18	Perempuan	D3	D3	3	PNS	Swasta	Atas	SEDANG	SEDANG	LULUS
19	Laki-Laki	SMA	SMK	2	PNS	PNS	Menengah	TINGGI	SEDANG	LULUS
20	Laki-Laki	S1	D3	2	Swasta	PNS	Menengah	SEDANG	SEDANG	LULUS
21	Perempuan	D3	SMK	1	Wiraswasta	Ibu Rumah Tangga	Atas	SEDANG	TINGGI	LULUS
22	Perempuan	SMK	SMA	2	Swasta	PNS	Atas	SEDANG	SEDANG	LULUS
23	Laki-Laki	SMA	SMK	2	Swasta	Wiraswasta	Atas	KURANG	SEDANG	LULUS
24	Laki-Laki	SMA	SMA	3	PNS	Swasta	Atas	KURANG	KURANG	TIDAK LULUS
25	Laki-Laki	SMA	SMK	2	Swasta	Swasta	Bawah	TINGGI	TINGGI	LULUS
26	Perempuan	S1	S1	4	Wiraswasta	Wiraswasta	Menengah	TINGGI	TINGGI	LULUS
27	Perempuan	SMA	SMA	3	Wiraswasta	PNS	Menengah	SEDANG	SEDANG	LULUS

Pembahasan

Hasil dan Output

Menghitung akurasi perhitungan naïve bayes pada data training berjumlah 315 dan data testing berjumlah 135. Berikut ini adalah rumus perhitungan akurasi :

$$\text{Akurasi} = \frac{\text{Jumlah Data Prediksi Benar atau Salah}}{\text{Jumlah Data yang diprediksi}} \times 100\%$$

Dibawah ini adalah perhitungan akurasi naïve bayes pada hasil perhitungan manual di tabel nomor 6.

Pada perhitungan akurasi diatas menunjukkan bahwa perhitungan naïve bayes pada dataset siswa smp dharma wanita 17 wonoayu yang menggunakan 315 data siswa sebagai data training dan 135 data siswa sebagai data testing menunjukkan tingkat akurasi sebesar 99,25%.

$$\text{Akurasi Prediksi Benar} = \frac{134}{135} \times 100\% = 99,25\%$$

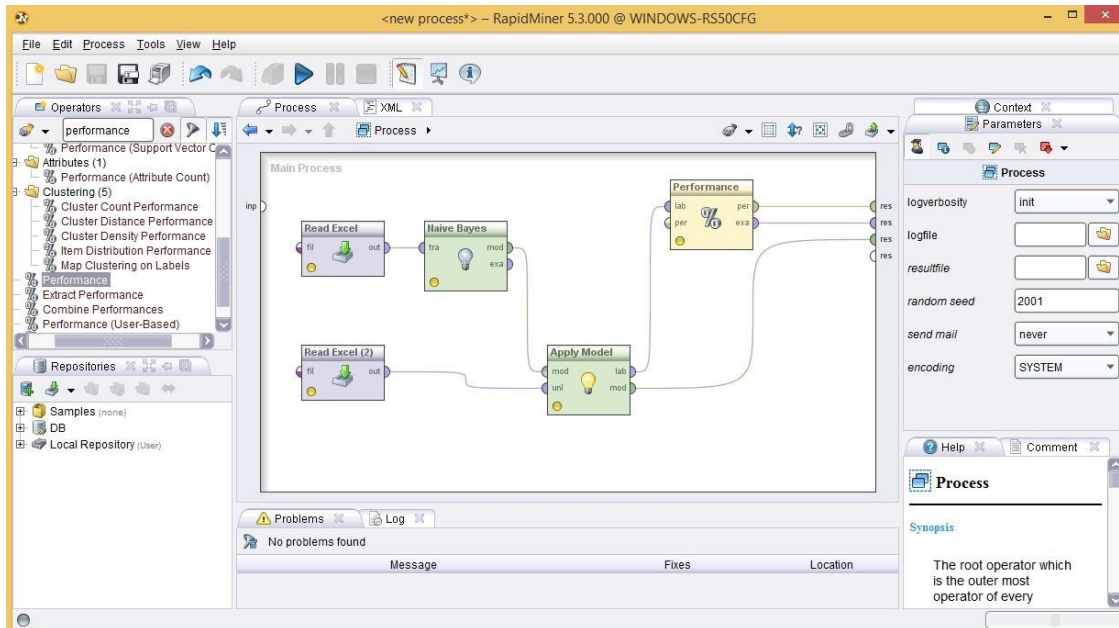
$$\text{Akurasi Prediksi Salah} = \frac{1}{135} \times 100\% = 0.74\%$$

Pembahasan

- **Pengujian Naïve Bayes di Rapid Miner**

1. Langkah Awal Pemodelan Data Mining

Model dibuat dengan dua data yaitu data training sebanyak 315 dan data testing sebanyak 135. Data tersebut diimport kedalam rapid miner. Setelah itu membuat pemodelan algoritma dengan menggunakan algoritma naïve bayes yang kemudian dihubungkan ke *tools apply* model dan performance.

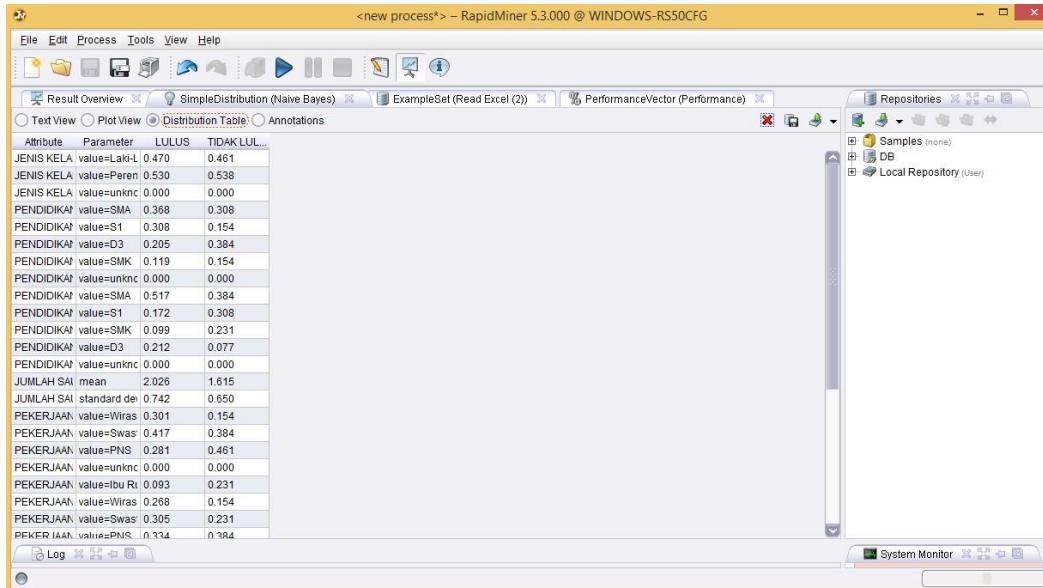


2. Hasil Perhitungan Naïve Bayes Dengan Rapid Minner

Row No.	STATUS KE...	confidence...	confidence...	prediction(S...	JENIS KEL...	PENDIDIK...	PENDIDIK...	JUMLAH SA...	PEKERJAA...	PEKERJAA...	TINGKAT E...	NILAI RA...
1	LULUS	1	0	LULUS	Perempuan	SMA	SMA	2	Swasta	Swasta	Menengah	SEDAN
2	LULUS	1	0	LULUS	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Menengah	SEDAN
3	LULUS	1.000	0.000	LULUS	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Menengah	KURAN
4	LULUS	1.000	0.000	LULUS	Perempuan	D3	S1	2	PNS	Wiraswasta	Menengah	TINGGI
5	LULUS	1	0	LULUS	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Atas	TINGGI
6	LULUS	1.000	0.000	LULUS	Perempuan	D3	SMK	1	Wiraswasta	Wiraswasta	Atas	SEDAN
7	LULUS	1	0	LULUS	Perempuan	SMA	SMA	3	Wiraswasta	PNS	Menengah	SEDAN
8	LULUS	1.000	0.000	LULUS	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Menengah	KURAN
9	LULUS	1	0	LULUS	Laki-Laki	S1	D3	3	PNS	PNS	Atas	SEDAN
10	LULUS	1.000	0.000	LULUS	Perempuan	D3	S1	2	Swasta	Wiraswasta	Bawah	SEDAN
11	LULUS	1	0	LULUS	Perempuan	S1	S1	3	Wiraswasta	Wiraswasta	Bawah	TINGGI
12	LULUS	1	0	LULUS	Perempuan	SMK	SMA	2	Wiraswasta	Swasta	Atas	SEDAN
13	LULUS	1.000	0.000	LULUS	Laki-Laki	SMA	SMA	1	Swasta	Swasta	Atas	TINGGI
14	LULUS	1	0	LULUS	Laki-Laki	S1	SMA	2	PNS	Wiraswasta	Atas	SEDAN
15	LULUS	1.000	0.000	LULUS	Perempuan	SMA	SMA	1	Wiraswasta	Ibu Rumah T	Atas	SEDAN
16	LULUS	1	0	LULUS	Laki-Laki	S1	D3	4	PNS	Swasta	Atas	SEDAN
17	LULUS	1	0	LULUS	Perempuan	D3	D3	3	PNS	Swasta	Atas	SEDAN
18	LULUS	1.000	0.000	LULUS	Laki-Laki	SMA	SMK	2	PNS	PNS	Menengah	TINGGI
19	LULUS	1	0	LULUS	Laki-Laki	S1	D3	2	Swasta	PNS	Menengah	SEDAN
20	LULUS	1.000	0.000	LULUS	Perempuan	D3	SMK	1	Wiraswasta	Ibu Rumah T	Atas	SEDAN

Pembahasan

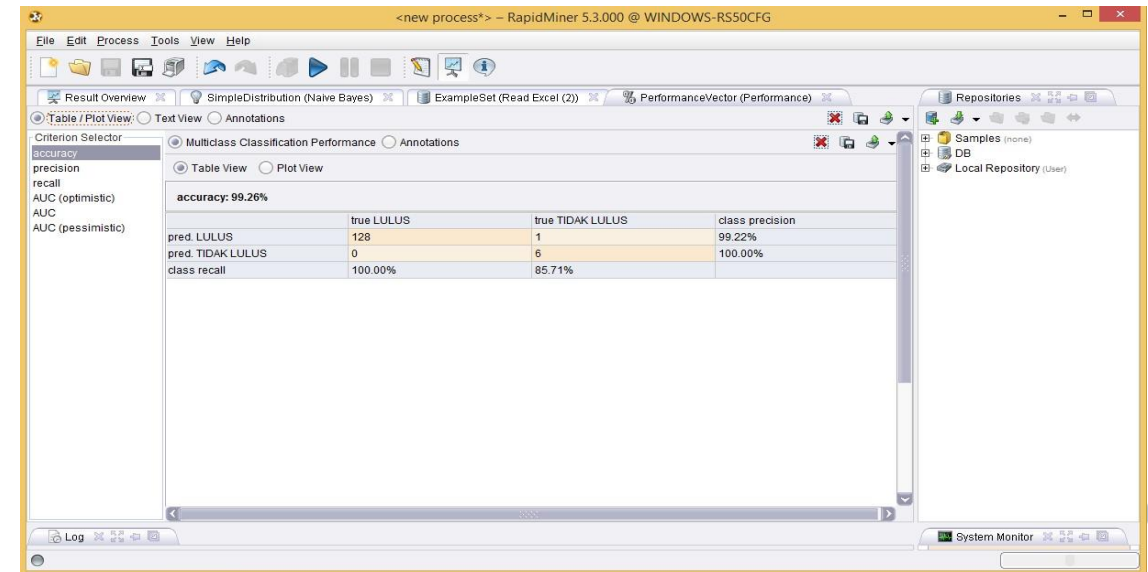
3. Hasil Perhitungan probabilitas Atribut



Attribute	Parameter	LULUS	TIDAK LULUS
JENIS KELA	value=Laki-L	0.470	0.461
JENIS KELA	value=Perem	0.530	0.538
JENIS KELA	value=unknc	0.000	0.000
PENDIDIKAN	value=SMA	0.368	0.308
PENDIDIKAN	value=S1	0.308	0.154
PENDIDIKAN	value=D3	0.205	0.384
PENDIDIKAN	value=SMK	0.119	0.154
PENDIDIKAN	value=unknc	0.000	0.000
PENDIDIKAN	value=SMA	0.517	0.384
PENDIDIKAN	value=S1	0.172	0.308
PENDIDIKAN	value=SMK	0.099	0.231
PENDIDIKAN	value=D3	0.212	0.077
PENDIDIKAN	value=unknc	0.000	0.000
JUMLAH SAI	mean	2.026	1.615
JUMLAH SAI	standard de	0.742	0.650
PEKERJAAN	value=Wiras	0.301	0.154
PEKERJAAN	value=Swas	0.417	0.384
PEKERJAAN	value=PNS	0.281	0.461
PEKERJAAN	value=unknc	0.000	0.000
PEKERJAAN	value=ibu Ru	0.093	0.231
PEKERJAAN	value=Wiras	0.268	0.154
PEKERJAAN	value=Swas	0.305	0.231
PEKERJAAN	value=PNS	0.334	0.384

4. Hasil Performance Dan Akurasi

Pada gambar Nilai akurasi menunjukkan hasil akurasi perhitungan naïve bayes di aplikasi rapid minner dengan menggunakan dataset sebanyak 450 dibagi menjadi dua yaitu data training sebanyak 315 dan data testing sebanyak 135. Dari hasil pengujian pada data tersebut menghasilkan akurasi sebesar 99.26%.

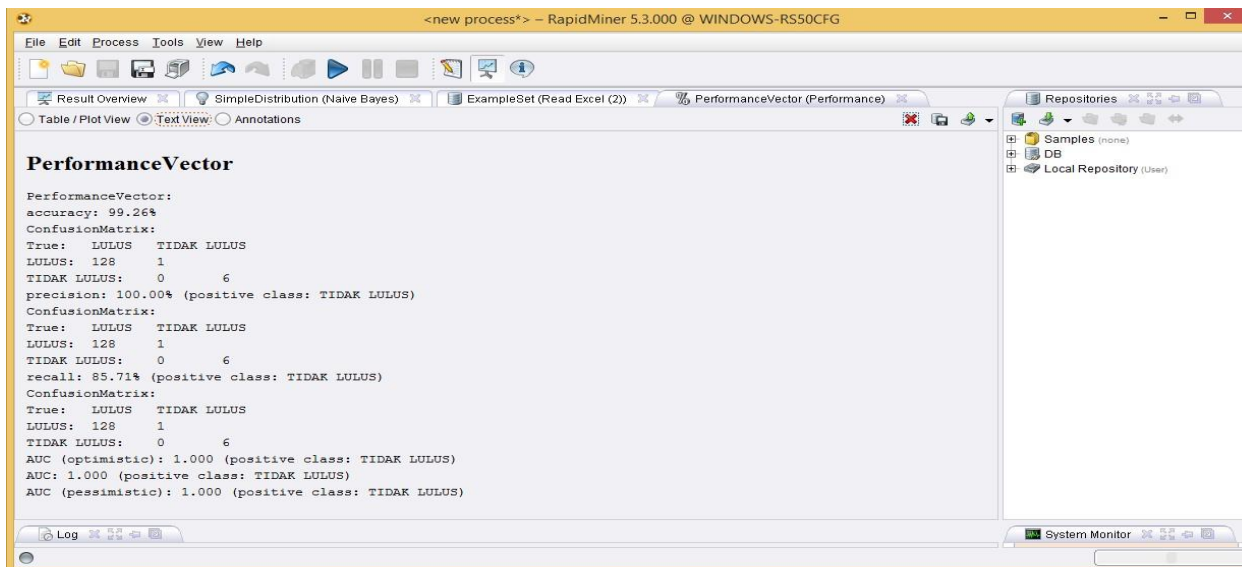


	true LULUS	true TIDAK LULUS	class precision
pred. LULUS	128	1	99.22%
pred. TIDAK LULUS	0	6	100.00%
class recall	100.00%	85.71%	

Pembahasan

5. Tampilan Performance Vektor

Pada gambar Performance vector menunjukkan hasil performance vector perhitungan naïve bayes di aplikasi rapid minner dengan menggunakan dataset sebanyak 450 dibagi menjadi dua yaitu data training sebanyak 315 dan data testing sebanyak 135. Dari hasil pengujian pada data tersebut menghasilkan akurasi sebesar 99.26%.



KESIMPULAN

Pada gambar Performance vector menunjukkan hasil performance vector perhitungan naïve bayes di aplikasi rapid minner dengan menggunakan dataset sebanyak 450 dibagi menjadi dua yaitu data training sebanyak 315 dan data testing sebanyak 135. Dari hasil pengujian pada data tersebut menghasilkan akurasi sebesar 99.26%.

Referensi

1. Zhang. Harry, "FLAIRS conference," *"The Optimality of Naïve Bayes"*, 2004.
2. G. P.-S. a. P. S. U. Fayyad, "AI Magazine Volume 17 Number 3," *From Data Mining*, vol. vol. 17, 2010.
3. Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi Data*, 2017.
4. J. d. M. K. Han, *Data Mining Concepts and Techniques Second*, 2006.
5. N. S. F. K. H. Susana, "JURSISTEKNI (*Jurnal Sistem Informasi dan Teknologi Informasi*), vol. Vol 4, January 2022.
6. Z. e. a. Ahmad, "" Transactions on Emerging Telecommunications Technologies," *"Network intrusion detection system: A systematic study of machine learning and deep learning approaches.*, 2021.
7. N. R. Irawati, "JURTEKSI 5," *"Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer*, 199-204..
8. M. Zed, *Yayasan Pustaka Obor Indonesia," Metode Penelitian Kepustakaan*, 2008.
9. C. S. C. T. K. L. K. J. N. A. A. L. F. D. I. M. T. D. R. S. H. G. G. A. S. N. N. N. M. Duha, Bookchapter: *Pengalaman Observasi di SMPN 2 Toma*, 2022.
10. Satria, "IlmuKomputer.com J. Softw. Eng.," *"Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naive Bayes*, Vols. vol. 1, no. 2.

