

Application of Data Mining to Predict Student Graduation Rate Using Naive Bayes Method (Case Study of Smp Dharma Wanita 17 Wonoayu)

[Penerapan Data Mining Untuk Prediksi Tingkat Kelulusan Siswa Menggunakan Metode Naive Bayes (Studi Kasus Smp Dharma Wanita 17 Wonoayu)]

Tya Sukmawati ¹⁾, Arif Senja Fitriani ²⁾, Ade Eviyanti ³⁾, Mochamad Alfian Rosid ⁴⁾

1)Program Studi Teknik Informatika,, Universitas Muhammadiyah Sidoarjo, Indonesia

2)Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

3)Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

4)Program Studi Teknik Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: tyasukmawati98@gmail.com ¹⁾, afsjim@umsida.ac.id ²⁾

Abstract.. Data mining is a way to find new information extracted from large amounts of data that helps in decision making. By utilizing student main data, student graduation and student average scores as data sources, it is expected to produce information about student graduation rates through data mining techniques. The graduation prediction category is measured from the attributes of gender, report card scores, test scores, etc. This study used a dataset of students from class 2015-2020 at SMP Dharma Wanita 17 Wonoayu, which totaled 450 data. The data is divided into two, namely training data by 70% and data testing by 30%. From testing the data on the rapid miner application, the calculation results obtained an accuracy rate of 99.26%.

Keywords – data mining, classification methods, naïve Bayes algorithms.

Abstrak. Data mining merupakan cara untuk menemukan informasi baru yang diambil dari data dalam jumlah besar yang membantu dalam pengambilan keputusan. Dengan memanfaatkan data induk siswa, kelulusan siswa dan nilai rata-rata siswa sebagai sumber datanya, diharapkan dapat menghasilkan informasi tentang tingkat kelulusan siswa melalui teknik data mining. Kategori prediksi kelulusan di ukur dari atribut jenis kelamin, nilai rapor, nilai uas dll. Penelitian ini menggunakan dataset siswa angkatan tahun 2015-2020 smp dharma wanita 17 wonoayu yang berjumlah 450 data. Data tersebut dibagi menjadi dua yaitu data training sebesar 70% dan data testing sebesar 30%. Dari pengujian data tersebut pada aplikasi rapid miner didapatkan hasil perhitungan tingkat akurasi sebesar 97.78%.

Kata Kunci – data mining, metode klasifikasi, algoritma naïve bayes.

I. PENDAHULUAN

A. Latar Belakang

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. Keberhasilan sistem pendidikan dapat dilihat dari meningkatnya kualitas pendidikan dari tahun ke tahun. Salah satu faktor yang menjadi indikator kualitas pendidikan adalah tingkat kelulusan. Dengan tingginya tingkat kelulusan maka dapat disimpulkan bahwa sistem pendidikan berjalan dengan baik.

Klasifikasi adalah tugas utama dalam data mining, dalam klasifikasi sebuah pengklasifikasian yang dibuat dari sekumpulan data dengan kelas yang telah ditentukan sebelumnya [1]. Metode naïve bayes merupakan klasifikasi statistik yang dapat memprediksi kelas suatu anggota probabilitas. Algoritma ini akan memprediksi suatu probabilitas baru berdasarkan hasil probabilitas yang sudah ada.

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 menggunakan *software* rapid miner untuk membantu memprediksi tingkat kelulusan siswa di smp dharma wanita 17 wonoayu dengan menggunakan metode naïve bayes. Dengan adanya penerapan *software* rapid miner diharapkan dapat membantu untuk memprediksi tingkat kelulusan siswa.

B. Penelitian Terdahulu

Berdasarkan Beberapa penelitian terdahulu yang diambil oleh peneliti untuk dijadikan bahan pertimbangan dan sumber referensi yang berhubungan dengan judul dari penelitian ini, diantaranya sebagai berikut:

Tabel 1. Penelitian Terdahulu

No.	Judul	Author	tahun	Metode	Publisher
1.	Penerapan data mining untuk analisis karakteristik DPT non participate sebagai prediksi partisipan pemilu dengan menggunakan metode naive bayes classifier	Uky dzatalini rojaby [2]	2018	Naïve bayes	Knowledge-based systems
2.	Implementasi klasifikasi bayesian untuk strategi menyerang jarak dekat pada npc (non player character)menggunakan unity 3d	Siti asmiatun, Latus hermawan, Tri daryanti [3]	2016	Naïve bayes	Jurnal transformatika
3.	Prediction of study period students (bachelor degree) muhammadiyah university of sidoarjo based on decision tree method using c4. 5 algorithm	Arif senja fitrani [4]	2019	C4.5	Conference series

Keterangan:

1. “Penerapan data mining untuk analisis karakteristik dpt nonparticipate sebagai prediksi partisipan pemilu dengan menggunakan metode naive bayes classifier” ditulis oleh uky dzatalini rojaby pada tahun 2018. Penelitian ini untuk menganalisis penyebab dari rendahnya tingkat partisipasi dalam pemilu. Dengan penerapan data mining ini dapat mengetahui penyebab dari rendahnya tingkat partisipasi dapat meningkatkan partisipasi masyarakat dalam pemilu berikutnya. Metode yang digunakan yaitu naive bayes classifier, dan desain penelitian dengan menggunakan crisp-dm. Data penelitian adalah data dpt pilpres desa pasekaran kab. Batang tahun 2014 yang akan menggunakan confusion matrix dan divalidasi dengan teknik split validasi. Akurasi tertinggi didapatkan setelah data penelitian hasil yaitu 88,46%.
2. “Implementasi klasifikasi bayesian untuk strategi menyerang jarak dekat pada npc (non player character)menggunakan unity 3d” di tulis oleh siti asmiatun, latus hermawan, dan tri daryanti pada tahun 2016. Penelitian ini menerapkan algoritma bayesian untuk klasifikasi perilaku penyerangan npc

tersebut. Klasifikasi tersebut diharapkan dapat meningkatkan strategi menyerang musuh. Klasifikasi penyerangan npc dibagi menjadi dua perilaku penyerangan yaitu perilaku memukul dan menggigit. Sedangkan untuk variabel yang digunakan dalam klasifikasi bayesian adalah *health point*, *attack point player* dan jarak yang diperoleh dari kondisi npc. Dari hasil pengujian klasifikasi bayesian menggunakan pengujian *confusion matrix*, dengan percobaan permainan sebanyak 10 kali, telah menghasilkan presentasi tingkat akurasi pada *matriks kebingungan* mencapai nilai persentase sebesar 80 %.

3. “prediction of study period students (bachelor degree) muhammadiyah university of sidoarjo based on decision tree method using c4. 5 algorithm” ditulis oleh arif senja fitrani tahun 2019. Penelitian ini menggunakan algoritma c4.5 untuk memprediksi periode kelulusan mahasiswa universitas muhammadiyah sidoarjo. Dimana dalam penelitian ini untuk mengetahui masalah atau hambatan apa yang menghambat dalam kelulusan tepat waktu mahasiswa. Dalam penelitian ini menghasilkan faktor-faktor yang mempengaruhi kelulusan mahasiswa dengan persentase 85%.

II. METODE PENELITIAN

A. Alat dan Bahan Penelitian

berikut ini adalah alat dan bahan penelitian untuk menunjang dalam penelitian ini:

1. Hardware
 - a. Sebuah laptop dengan spesifikasi
 - b. Procesor : intel core i3
Memory : 4 gb
Motherboard : acer
Harddisk : 500 gb
Os : windows 8
 - c. Mouse
2. Software
 - a. Rapid minner.
 - b. Microsoft word 2016.
 - c. Microsoft excel 2016.
3. Data siswa smp dharma wanita 17 wonoayu angkatan 2015-2020.
4. Jurnal tentang metode naïve bayes.

B. Teknik Pengumpulan Data

1. Studi Literatur
Yaitu melalui berbagai artikel dan jurnal yang didapat di internet dan buku-buku referensi yang menunjang dalam penelitian ini.
2. Observasi
Yaitu melakukan observasi secara langsung ke lokasi penelitian yaitu smp dharma wanita 17 wonoayu untuk melakukan pengumpulan data siswa angkatan 2015-2020.

C. Metode Naïve bayes

Metode naive bayes adalah algoritma dari klasifikasi yang bekerja dengan cara memprediksi peluang yang akan terjadi berdasarkan data yang telah didapatkan sebelumnya dengan hasil akhir atau class yang ditentukan secara independen oleh kombinasi dari keberagaman nilai atribut pada data yang sedang diklasifikasi [5].

Rumus perhitungan metode naïve bayes [6]:

$$P(C | X(n)) P(c) \times \prod_{i=1}^n P(X(i) | c)$$

Gambar 1. Rumus naïve bayes

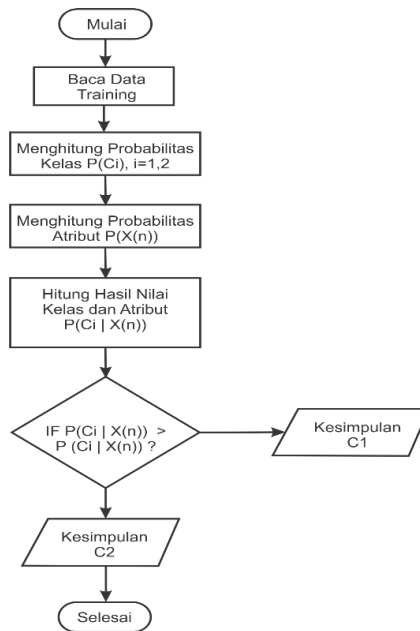
Keterangan:

$P(c | X(n))$ = Probabilitas dari kondisi X.

$P(c)$ = Probabilitas kondisi class.

$P(x_i | c)$ = Probabilitas kondisi data yang di cari dari attribute.

Dibawah ini flowchart metode naïve bayes [7]:



Gambar 2. Flowchart naïve bayes

D. Data Preparation

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 ANGGATAN 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 Anes	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

Gambar 3. Data asli siswa

E. Data Preprocessing

1. 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.

Tabel 2. Atribut selection

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	

2. 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*) [8].

3. Data Transformation

Pada tabel 3. 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:

Tabel 3. Data asli

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

Tabel 4. Tranformasi data

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

Tabel 5. Setelah proses tranformasi

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

4. 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.

4	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

Gambar 4. Dataset sebelum proses random

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

Gambar 5. Dataset setelah proses random

5. 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.

F. Proses Naïve Bayes

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

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

Gambar 6. Data training

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 : $P(X | C) = \text{kondisi atribut X} / \text{total C}$

$$P(\text{Jenis kelamin} = \text{laki - laki} | \text{lulus}) = 142 / 302 = 0.470$$

$$P(\text{Jenis kelamin} = \text{laki - laki} | \text{tidak lulus}) = 6 / 13 = 0.461$$

Tabel 7. Perhitungan probabilitas data training

No.	Atribut	LULUS	TIDAK LULUS
1.	Jenis kelamin	Laki-Laki	$142 / 302 = 0.470$
		Perempuan	$6 / 13 = 0.461$
2.	Pendidikan ayah	SMA	$111 / 302 = 0.367$
		SMK	$4 / 13 = 0.307$
		D3	$36 / 302 = 0.119$
		S1	$2 / 13 = 0.153$
3.	Pendidikan ibu	D3	$62 / 302 = 0.205$
		S1	$5 / 13 = 0.384$
		SMA	$93 / 302 = 0.307$
3.	Pendidikan ibu	SMA	$156 / 302 = 0.516$
		SMK	$5 / 13 = 0.384$
315	Perempuan	SMK	$30 / 302 = 0.099$
		SMA	$3 / 13 = 0.230$

		D3	$64 / 302 = 0.211$	$1 / 13 = 0.076$
		S1	$52 / 302 = 0.172$	$4 / 13 = 0.307$
4.	Jumlah saudara	1	$72 / 302 = 0.238$	$6 / 13 = 0.461$
		2	$157 / 302 = 0.519$	$6 / 13 = 0.461$
		3	$66 / 302 = 0.218$	$1 / 13 = 0.076$
		4	$7 / 302 = 0.023$	$0 / 13 = 0$
5	Pekerjaan ayah	Wiraswasta	$91 / 302 = 0.301$	$2 / 13 = 0.153$
		PNS	$85 / 302 = 0.281$	$6 / 13 = 0.461$
		Swasta	$126 / 302 = 0.417$	$5 / 13 = 0.384$
6.	Pekerjaan ibu	Ibu Rumah Tangga	$28 / 302 = 0.092$	$3 / 13 = 0.230$
		Wiraswasta	$81 / 302 = 0.268$	$2 / 13 = 0.153$
		PNS	$101 / 302 = 0.334$	$5 / 13 = 0.384$
7.	Tingkat ekonomi	Swasta	$92 / 302 = 0.304$	$3 / 13 = 0.230$
		Atas	$114 / 302 = 0.377$	$5 / 13 = 0.384$
		Menengah	$131 / 302 = 0.433$	$5 / 13 = 0.384$
8.	Nilai rapor	Bawah	$57 / 302 = 0.188$	$3 / 13 = 0.230$
		Tinggi	$92 / 302 = 0.304$	$0 / 13 = 0$
		Sedang	$167 / 302 = 0.552$	$0 / 13 = 0$
9.	Nilai UAS	Kurang	$43 / 302 = 0.142$	$13 / 13 = 1$
		Tinggi	$101 / 302 = 0.334$	$0 / 13 = 0$
		Sedang	$199 / 302 = 0.658$	$0 / 13 = 0$
		Kurang	$2 / 302 = 0.006$	$13 / 13 = 1$

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

Gambar 7. Data Testing

Tabel 8. Hasil perhitungan manual data testing

No.	Class Prediksi	Status Kelulusan	P (Lulus)	P (Tidak Lulus)	Status
1.	LULUS	LULUS	0.001007164	0	BENAR
2	LULUS	LULUS	0.000102408	0	BENAR
3	LULUS	LULUS	0.000156922	0	BENAR
4	LULUS	LULUS	6.13571E-05	0	BENAR
5	LULUS	LULUS	2.49176E-05	0	BENAR
6	LULUS	LULUS	2.74537E-05	0	BENAR
7	LULUS	LULUS	0.000335698	0	BENAR
8	LULUS	LULUS	0.000156922	0	BENAR
9	LULUS	LULUS	4.22486E-05	0	BENAR
10	LULUS	LULUS	7.18371E-05	0	BENAR
11	LULUS	LULUS	9.14733E-06	0	BENAR
12	LULUS	LULUS	0.000205298	0	BENAR
13	LULUS	LULUS	9.97412E-05	0	BENAR
14	LULUS	LULUS	0.000196461	0	BENAR
15	LULUS	LULUS	8.83502E-05	0	BENAR
16	LULUS	LULUS	8.04201E-06	0	BENAR
17	LULUS	LULUS	5.69575E-05	0	BENAR
18	LULUS	LULUS	7.01324E-05	0	BENAR
19	LULUS	LULUS	0.000337301	0	BENAR
20	LULUS	LULUS	4.81661E-06	0	BENAR
21	LULUS	LULUS	0.000312066	0	BENAR
22	LULUS	LULUS	3.39116E-05	0	BENAR
23	TIDAK LULUS	TIDAK LULUS	5.70852E-07	7.10318E-06	BENAR
24	LULUS	LULUS	2.09126E-05	0	BENAR
25	LULUS	LULUS	2.22969E-06	0	BENAR
26	LULUS	LULUS	0.000335698	0	BENAR
27	LULUS	LULUS	0.00034759	0	BENAR
28	LULUS	LULUS	7.28276E-05	0	BENAR
29	LULUS	LULUS	0.000453667	0	BENAR
30	LULUS	LULUS	8.71852E-06	0	BENAR
31	LULUS	LULUS	3.43347E-05	0	BENAR
32	LULUS	LULUS	0.000312066	0	BENAR
33	LULUS	LULUS	6.5487E-05	0	BENAR
34	LULUS	LULUS	0.000118928	0	BENAR
35	LULUS	LULUS	6.52394E-05	0	BENAR
36	LULUS	LULUS	0.000178363	0	BENAR
37	LULUS	LULUS	5.85643E-05	0	BENAR
38	LULUS	LULUS	4.98706E-05	0	BENAR
39	LULUS	LULUS	3.68244E-05	0	BENAR
40	LULUS	LULUS	0.000191487	0	BENAR
41	LULUS	LULUS	0.000215511	0	BENAR
42	LULUS	LULUS	0.001007164	0	BENAR
43	LULUS	LULUS	0.000292134	0	BENAR
44	LULUS	LULUS	3.70854E-06	0	BENAR
45	LULUS	LULUS	0.000208315	0	BENAR
46	LULUS	LULUS	0.000214261	0	BENAR
47	LULUS	LULUS	0.00012253	0	BENAR
48	LULUS	LULUS	0.00056118	0	BENAR
49	LULUS	LULUS	0.000118725	0	BENAR
50	LULUS	LULUS	6.61825E-07	0	BENAR
51	LULUS	LULUS	0.000312066	0	BENAR
52	TIDAK LULUS	TIDAK LULUS	3.48348E-07	5.91932E-06	BENAR
53	LULUS	LULUS	9.23348E-05	0	BENAR
54	LULUS	LULUS	3.85005E-05	0	BENAR
55	TIDAK LULUS	TIDAK LULUS	5.31267E-07	8.28705E-06	BENAR
56	LULUS	LULUS	0.000135115	0	BENAR

57	LULUS	LULUS	1.20264E-06	0	BENAR
58	LULUS	LULUS	0.000115389	0	BENAR
59	LULUS	LULUS	0.000240649	0	BENAR
60	LULUS	LULUS	7.82987E-05	0	BENAR
61	LULUS	LULUS	0.000524748	0	BENAR
62	LULUS	LULUS	8.91182E-05	0	BENAR
63	LULUS	LULUS	0.000265176	0	BENAR
64	LULUS	LULUS	0.000165964	0	BENAR
65	LULUS	LULUS	4.17716E-05	0	BENAR
66	LULUS	LULUS	9.56556E-05	0	BENAR
67	LULUS	TIDAK LULUS	2.06364E-08	0	SALAH
68	LULUS	LULUS	2.12007E-05	0	BENAR
69	LULUS	LULUS	7.28276E-05	0	BENAR
70	LULUS	LULUS	0.000165786	0	BENAR
71	LULUS	LULUS	0.00018488	0	BENAR
72	LULUS	LULUS	4.85904E-05	0	BENAR
73	LULUS	LULUS	2.51931E-05	0	BENAR
74	LULUS	LULUS	0.000178363	0	BENAR
75	LULUS	LULUS	4.14462E-05	0	BENAR
76	LULUS	LULUS	6.60528E-05	0	BENAR
77	LULUS	LULUS	0.000123253	0	BENAR
78	LULUS	LULUS	0.000125133	0	BENAR
79	LULUS	LULUS	0.000224028	0	BENAR
80	LULUS	LULUS	0.000684856	0	BENAR
81	TIDAK LULUS	TIDAK LULUS	1.88462E-07	2.84127E-06	BENAR
82	LULUS	LULUS	0.000684856	0	BENAR
83	LULUS	LULUS	4.14462E-05	0	BENAR
84	LULUS	LULUS	0.000391651	0	BENAR
85	LULUS	LULUS	6.68443E-05	0	BENAR
86	LULUS	LULUS	0.000118605	0	BENAR
87	LULUS	LULUS	0.000111376	0	BENAR
88	LULUS	LULUS	0.000165786	0	BENAR
89	LULUS	LULUS	2.53905E-05	0	BENAR
90	LULUS	LULUS	4.17716E-05	0	BENAR
91	LULUS	LULUS	0.00014672	0	BENAR
92	LULUS	LULUS	0.000126152	0	BENAR
93	LULUS	LULUS	4.42606E-05	0	BENAR
94	LULUS	LULUS	3.62775E-05	0	BENAR
95	LULUS	LULUS	0.001007164	0	BENAR
96	LULUS	LULUS	0.000337301	0	BENAR
97	LULUS	LULUS	5.70713E-05	0	BENAR
98	LULUS	LULUS	9.72198E-05	0	BENAR
99	LULUS	LULUS	4.49845E-05	0	BENAR
100	TIDAK LULUS	TIDAK LULUS	9.72393E-07	5.91932E-06	BENAR
101	LULUS	LULUS	9.41493E-05	0	BENAR
102	LULUS	LULUS	0.000103961	0	BENAR
103	LULUS	LULUS	0.00014672	0	BENAR
104	LULUS	LULUS	0.000227544	0	BENAR
105	LULUS	LULUS	0.000367167	0	BENAR
106	LULUS	LULUS	4.61035E-05	0	BENAR
107	LULUS	LULUS	0.000684856	0	BENAR
108	LULUS	LULUS	8.09877E-05	0	BENAR
109	LULUS	LULUS	0.000151676	0	BENAR
110	LULUS	LULUS	6.58516E-05	0	BENAR
111	LULUS	LULUS	7.82987E-05	0	BENAR
112	LULUS	LULUS	2.33274E-05	0	BENAR
113	LULUS	LULUS	9.09529E-05	0	BENAR
114	LULUS	LULUS	0.000128859	0	BENAR

115	LULUS	LULUS	0.000158385	0	BENAR
116	LULUS	LULUS	0.000225758	0	BENAR
117	LULUS	LULUS	2.89081E-05	0	BENAR
118	LULUS	LULUS	0.000106218	0	BENAR
119	LULUS	LULUS	2.53905E-05	0	BENAR
120	LULUS	LULUS	0.00010185	0	BENAR
121	LULUS	LULUS	0.000220595	0	BENAR
122	LULUS	LULUS	2.71828E-05	0	BENAR
123	LULUS	LULUS	3.0539E-05	0	BENAR
124	LULUS	LULUS	2.41401E-05	0	BENAR
125	LULUS	LULUS	0.000158385	0	BENAR
126	LULUS	LULUS	1.5178E-05	0	BENAR
127	LULUS	LULUS	0.000208315	0	BENAR
128	LULUS	LULUS	0.00034759	0	BENAR
129	LULUS	LULUS	3.43347E-05	0	BENAR
130	TIDAK LULUS	TIDAK LULUS	2.30619E-07	3.94621E-07	BENAR
131	LULUS	LULUS	0.001007164	0	BENAR
132	LULUS	LULUS	2.9126E-05	0	BENAR
133	LULUS	LULUS	8.04201E-06	0	BENAR
134	LULUS	LULUS	0.000218388	0	BENAR
135	LULUS	LULUS	8.32422E-05	0	BENAR

G. Hasil atau 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 8.

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

$$\text{Akurasi prediksi salah} = \frac{1}{135} \times 100 \% = 0.74\%$$

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%.

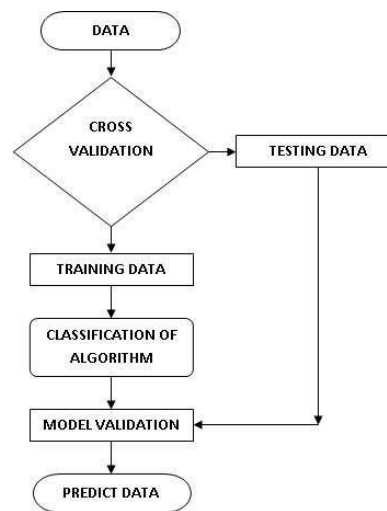
III. HASIL DAN PEMBAHASAN

A. Rapid Minner

Rapidminer merupakan salah satu tools yang dipakai dalam data mining. Rapidminer memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. Rapidminer merupakan software yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri [9].

Rapidminer ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi. Rapidminer sebelumnya bernama yale (yet another learning environment), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh ralfklinkenberg, ingo mierswa, dan simon fischer di artificial intelligence unit dari university of dortmund.

Flowchar Rapidminer [10]:

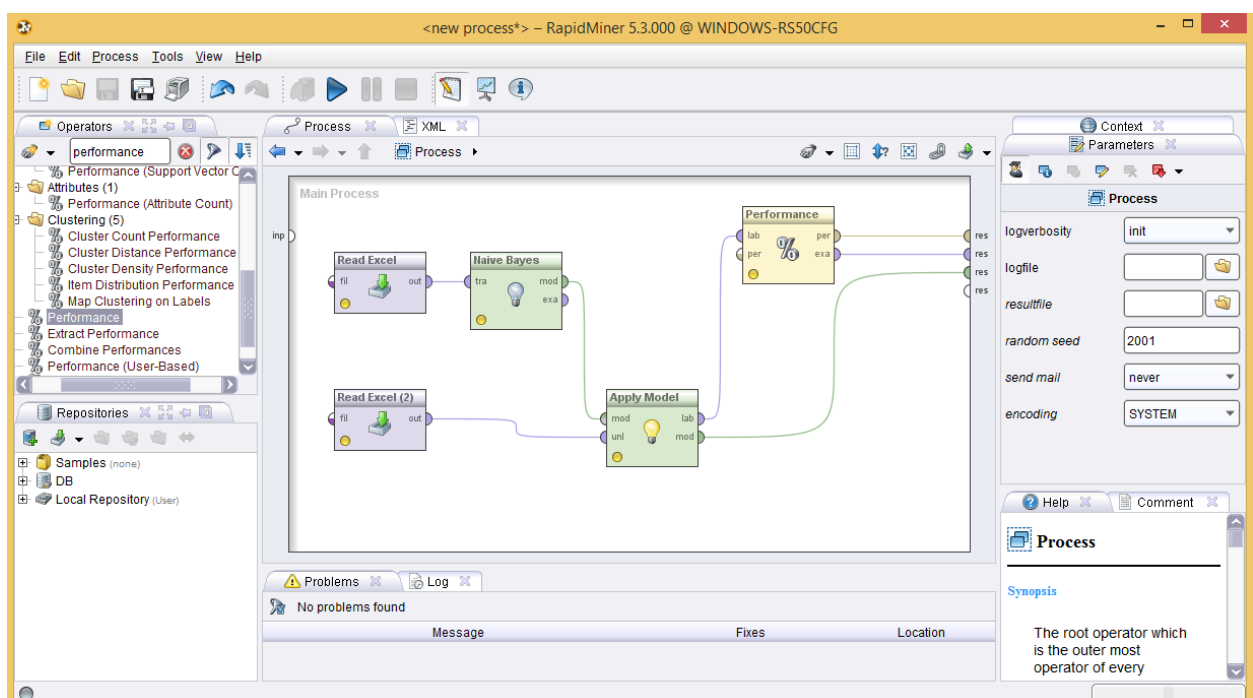


Gambar 8. Flowchart rapidminer

B. 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.



Gambar 9. Pemodelan algoritma naïve bayes

2. Hasil perhitungan naïve bayes dengan rapid minner

Row No.	STATUS KE...	confidence(...)	confidence(...)	prediction(S...	JENIS KEL...	PENDIDIKA...	PENDIDIKA...	JUMLAH SA...	PEKERJAA...	PEKERJAA...	TINGKAT E...	NILAI RA...
1	LULUS	1	0	LULUS	Perempuan	SMA	SMA	2	Swasta	Swasta	Menengah	SEDANG
2	LULUS	1	0	LULUS	Laki-Laki	S1	D3	3	Wiraswasta	PNS	Menengah	SEDANG
3	LULUS	1.000	0.000	LULUS	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Menengah	KURANG
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	SEDANG
7	LULUS	1	0	LULUS	Perempuan	SMA	SMA	3	Wiraswasta	PNS	Menengah	SEDANG
8	LULUS	1.000	0.000	LULUS	Perempuan	S1	SMA	2	Wiraswasta	Swasta	Menengah	KURANG
9	LULUS	1	0	LULUS	Laki-Laki	S1	D3	3	PNS	PNS	Atas	SEDANG
10	LULUS	1.000	0.000	LULUS	Perempuan	D3	S1	2	Swasta	Wiraswasta	Bawah	SEDANG
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	SEDANG
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	SEDANG
15	LULUS	1.000	0.000	LULUS	Perempuan	SMA	SMA	1	Wiraswasta	Ibu Rumah T	Atas	SEDANG
16	LULUS	1	0	LULUS	Laki-Laki	S1	D3	4	PNS	Swasta	Atas	SEDANG
17	LULUS	1	0	LULUS	Perempuan	D3	D3	3	PNS	Swasta	Atas	SEDANG
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	SEDANG
20	LULUS	1.000	0.000	LULUS	Perempuan	D3	SMK	1	Wiraswasta	Ibu Rumah T	Atas	SEDANG

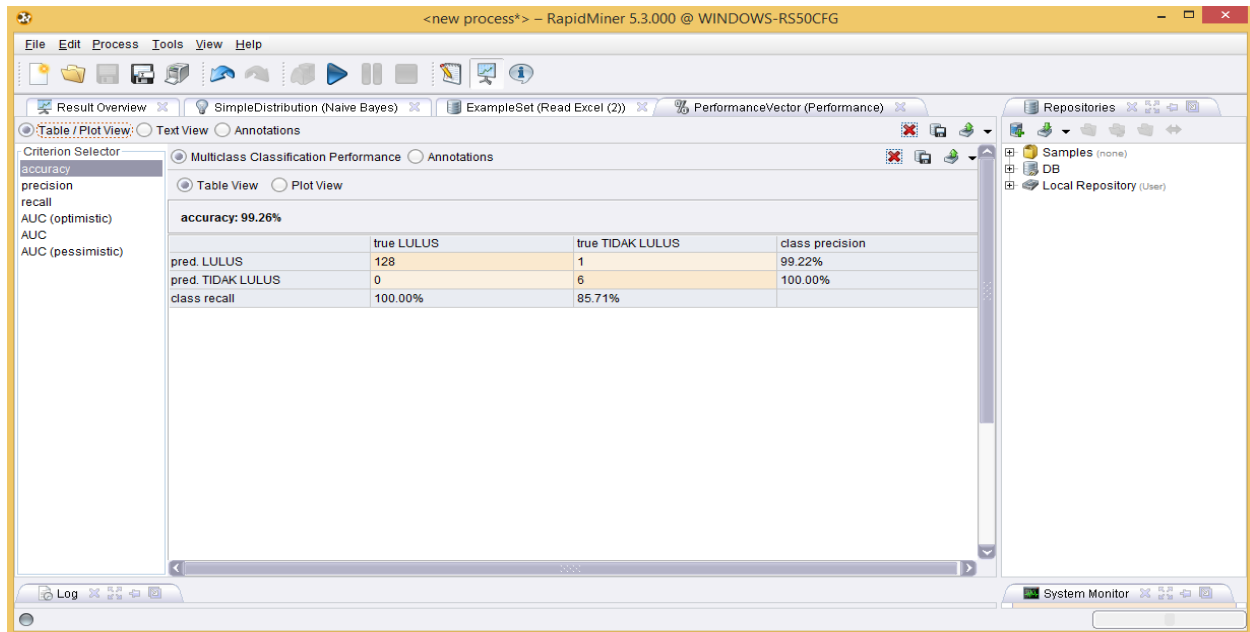
Gambar 10. Hasil perhitungan rapidminer

3. Hasil Perhitungan probabilitas Atribut

Attribute	Parameter	LULUS	TIDAK LUL...
JENIS KELA	value=Laki-L	0.470	0.461
JENIS KELA	value=Peren	0.530	0.538
JENIS KELA	value=unknc	0.000	0.000
PENDIDIKA	value=SMA	0.368	0.308
PENDIDIKA	value=S1	0.308	0.154
PENDIDIKA	value=D3	0.205	0.384
PENDIDIKA	value=SMK	0.119	0.154
PENDIDIKA	value=unknc	0.000	0.000
PENDIDIKA	value=SMA	0.517	0.384
PENDIDIKA	value=S1	0.172	0.308
PENDIDIKA	value=SMK	0.099	0.231
PENDIDIKA	value=D3	0.212	0.077
PENDIDIKA	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 Ri	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

Gambar 11. Perhitungan probabilitas atribut

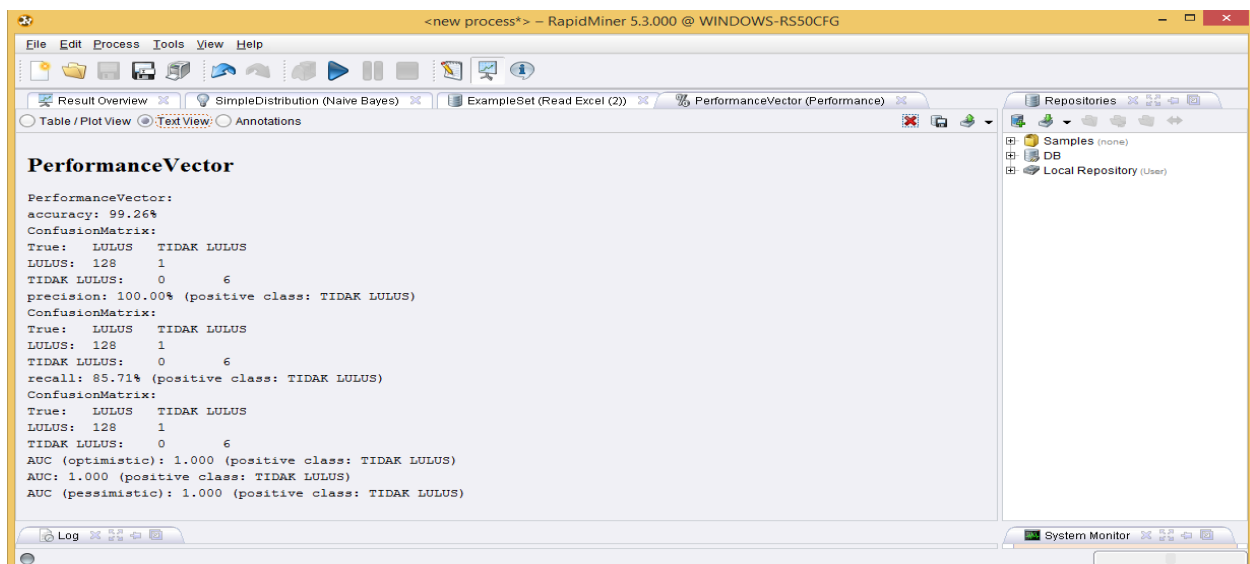
4. Hasil performance dan akurasi



Gambar 12. Nilai akurasi

Pada gambar 12. 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%.

5. Tampilan performance vektor



Gambar 13. Performance vector

Pada gambar 13. 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%.

IV. SIMPULAN

A. Kesimpulan

Berdasarkan penelitian yang telah dikerjakan, maka yang dapat disimpulkan pada penerapan data mining dengan menggunakan dataset smp dharma wanita 17 wonoayu dan algoritma naïve bayes yang diimplementasikan pada aplikasi rapid minner menunjukkan tingkat akurasi kelulusan yang tinggi yaitu sebesar 97.78%

B. Saran

Untuk saran kedepannya dari penulis yaitu mengembangkan aplikasi secara mandiri dan menggunakan metode data mining yang lain, sehingga bisa membandingkan antara metoda data mining.

UCAPAN TERIMA KASIH

Pada kali ini terima kasi kepala sekolah smp dharma wanita 17 wonoayu karena telah mengijinkan peneliti untuk merekap data siswa angkatan 2015-2020 dan membantu dalam proses pengerjaan skripsi. Terima kasih juga kepada universitas muhammadiyah sidoarjo dan dosen pembimbing bapak arif senja fitrani S.Pd., M.Kom. Yang telah memberika ilmu dan support kepada skripsi ini dan terima kasih kepada dekan fakultas sain dan teknologi telah menjadi pembimbing dalam menjalankan setiap pekerjaan skripsi ini.

REFERENSI

- [1] H. Zhang, "The optimality of naive bayes. In Valerie Barr and Zdravko Markov, editors, Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)". AAAI Press, 2004.
- [2] R. U. Dzatalini, "Knowledge-Based Systems", "Penerapan Data Mining Untuk Analisa Karakteristik DPT Non-Participate Sebagai Prediksi Partisipan Pemilu dengan Metode Naïve Bayes Classifier", 2015.
- [3] S. Asmiatun and A. Hendrawan, "Implementasi Klasifikasi Bayesian Untuk Strategi Menyerang Jarak Dekat Pada Npc (Non Player Character)Menggunakan Unity 3D," *J. Transform.*, vol. 13, no. 2, p. 42, 2016.
- [4] A. S. Fitrani, "Prediction of Study Period Students (Bachelor Degree) Muhammadiyah University of Sidoarjo Based on Decision Tree Method using C4.5 Algorithm," *J. Phys. Conf. Ser.*, vol. 1179, no. 1, 2019.
- [5] O. Nurdiawan dan N. Salim, "Penerapan Data Mining pada Penjualan Barang Menggunakan Metode Naive Bayes Classifier untuk Optimal Strategi Pemasaran , " *Jurnal Teknologi Informasi dan Komunikasi* ISSN; 2252-4517.
- [6] M. Hall, "A decision tree-based attribute weighting filter for naive Bayes," *Knowledge-Based Syst.*, vol. 20, no. 2, pp. 120–126, 2007.
- [7] J. Wu and Z. Cai, "Attribute weighting via differential evolution algorithm for attribute Weighted Naive Bayes (WNB)," *J. Comput. Inf. Syst.*, vol. 7, no. 5, pp. 1672–1679, 2011.
- [8] R. Satria, , " *IlmuKomputer.com J. Softw. Eng.*," "Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naive Bayes, vol. 1, no. 2, pp. 101–108, 2015.
- [9] N. R. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer," *JURTEKSI* 5, 199-204, 2019.
- [10] T. R. Wulandari, "Data Mining Teori dan Aplikasi Rapidminer," Yogyakarta : Penerbit Gava Media, 2017.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.