

Klasifikasi Data Kelulusan Siswa SMK Muhammadiyah 1 Palembang Menggunakan Metode *Naive Bayes*

Kiki Ramadhani¹, Nazori Suhandi², Muhammad Haviz Irfani^{3*}

^{1,2,3} Teknik Informatika, Ilmu Komputer dan Sains, Universitas Indo Global Mandiri
Jalan Jenderal Sudirman KM.4 No. 62 Kota Palembang, Sumatera Selatan, Indonesia
m.haviz@uigm.ac.id

Abstrak

Perkembangan saat ini semakin pesat seiring dengan berkembangnya teknologi informasi, untuk membantu karyawan dalam menyelesaikan tugasnya dan menjamin efisiensi waktu. Didalam dunia pendidikan, penerapan data mining memberikan peluang besar untuk membantu sekolah dan perguruan tinggi, baik negeri maupun swasta dalam memperoleh wawasan yang berguna. Salah satu klasifikasi yang cocok digunakan dalam menyeleksi informasi alumni lebih lanjut adalah metode Naive Bayes karena memiliki tingkat kecepatan dan akurasi yang lebih tinggi sehingga mampu menangkap data lebih banyak dibandingkan metode lainnya. Penelitian ini bertujuan untuk menerapkan metode Naive Bayes dalam klasifikasi data lulusan siswa SMK Muhammadiyah 1 Palembang. Dataset diperoleh dari Wakil Kepala Sekolah SMK Muhammadiyah 1 Palembang 562 data. Hasil penelitian menunjukkan bahwa data training sebanyak 393 data dengan algoritma Naive Bayes berhasil memprediksi besarnya kelulusan mahasiswa dengan presentase keakuratan sebesar 85,80%. Data mining dan naive bayes mampu menampilkan informasi prediksi kelulusan siswa dengan menggunakan data siswa yang telah lulus sebagai data training dan data testing. Sebanyak 169 data testing yang dihasilkan penelitian ini bahwa siswa yang lulus sebanyak 110 siswa atau sekitar 65% dari jumlah data testing sebesar 85,80%. Berdasarkan hasil ini, metode naive bayes direkomendasikan untuk klasifikasi data dalam memprediksi kelulusan siswa.

Kata kunci: *Naive Bayes*, Data Mining, Klasifikasi, Akurasi

Abstract

Current developments are increasingly rapid along with the development of information technology, to assist employees in completing their tasks and ensuring time efficiency. In the world of education, the application of data mining provides a great opportunity to help schools and universities, both public and private, in gaining useful insights. One of the classifications suitable for further selection of alumni information is the Naive Bayes method because it has a higher level of speed and accuracy so it can capture more data than other methods. This study aims to apply the Naive Bayes method in classifying data on graduates of SMK Muhammadiyah 1 Palembang. The dataset was obtained from the Vice Principal of SMK Muhammadiyah 1 Palembang 562 data. The results showed that the training data of 393 data with the Naive Bayes algorithm successfully predicted the number of students graduating with an accuracy percentage of 85.80%. Data mining and naive bayes are able to display information on student graduation predictions by using data from students who have graduated as training data and testing data. This study yielded 169 test data points, with 110 students passing, or approximately 65% of the 85.80% test data. Based on these results, the naive Bayes method is recommended for data classification in predicting student graduation.

Keywords: *Naive Bayes*, Data Mining, Classification, Accuracy

I. PENDAHULUAN

Perkembangan teknologi informasi yang semakin pesat telah memberikan kontribusi signifikan dalam mendukung efisiensi kerja karyawan, khususnya dalam penyelesaian tugas secara lebih efektif dan efisien. Salah satu pemanfaatan teknologi ini melalui penerapan data mining, berupa proses yang bertujuan

untuk mengekstraksi informasi dan pengetahuan tersembunyi dari kumpulan data berukuran besar dengan memanfaatkan teknik statistik, algoritma machine learning, dan analisis data. Salah satu teknik penting dalam data mining adalah teknologi klasifikasi, yang digunakan untuk mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan pola atau karakteristik yang terdapat dalam data.

Proses ini dilakukan dengan membangun model prediktif menggunakan algoritma seperti *Decision Tree*, *Naive Bayes*, *Support Vector Machine* (SVM), atau *Artificial Neural Networks*, yang dilatih pada data historis dan diaplikasikan pada data baru untuk menentukan kategorinya. Dalam konteks pendidikan, penerapan data mining dan teknologi klasifikasi memberikan potensi besar dalam mendukung pengambilan keputusan di lingkungan sekolah maupun perguruan tinggi, baik negeri maupun swasta. Sebagai contoh, teknologi ini dapat digunakan untuk melakukan klasifikasi kelulusan siswa berdasarkan data profil dan akademik, seperti yang dapat diterapkan pada pengelolaan data siswa di SMK Muhammadiyah 1 Palembang guna memperoleh wawasan dalam proses seleksi masuk ke perguruan tinggi [1], [2].

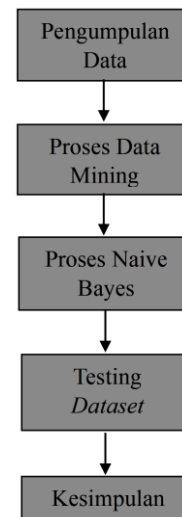
Salah satu klasifikasi yang cocok digunakan dalam menyeleksi informasi alumni lebih lanjut adalah metode Naive Bayes karena memiliki tingkat kecepatan dan akurasi yang lebih tinggi [3] sehingga mampu menangkap data lebih banyak dibandingkan metode lainnya. Naive bayes merupakan klasifikasi algoritma yang dibuat oleh ilmuwan Inggris bernama Thomas Bayes menggunakan metode probabilitas dan statistik. Algoritma ini memperkirakan peluang di masa depan berdasarkan data sebelumnya. Algoritma ini hanya memerlukan data yang kecil dalam menentukan parameter dalam proses klasifikasi [4].

Metode Naïve Bayes memiliki keunggulan dalam hal efisiensi, karena hanya memerlukan sejumlah kecil data pelatihan untuk memperkirakan parameter-parameter yang diperlukan dalam proses klasifikasi [5]. Proses pengelolaan data dalam penelitian ini mencakup tahapan preprocessing hingga klasifikasi, dengan menggunakan metode Naïve Bayes yang dipilih karena sifatnya yang sederhana, mudah diimplementasikan, serta memiliki kinerja yang cukup baik dalam berbagai kasus klasifikasi [6].

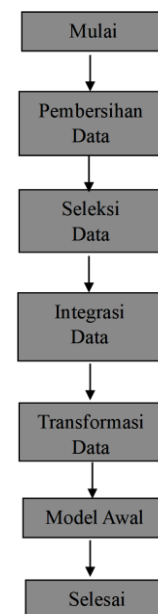
Penelitian sebelumnya yang dilakukan oleh [7] dengan judul “Klasifikasi Data Lulusan Siswa SMP Menggunakan Metode Naïve Bayes” menunjukkan bahwa metode ini mampu menghasilkan tingkat akurasi yang tinggi. Hasil penelitian tersebut menunjukkan bahwa Naïve Bayes Classification mampu mencapai tingkat akurasi hingga 80% pada dataset Admissions, yang menunjukkan efektivitas metode ini dalam mengklasifikasikan data kelulusan siswa. Berdasarkan temuan tersebut, penulis tertarik untuk mengangkat topik serupa dalam penelitian ini dengan judul “Penerapan Metode Naïve Bayes pada Klasifikasi Data Lulusan Siswa SMK Muhammadiyah 1 Palembang”.

II. METODE PENELITIAN

Tahapan penelitian yang dilakukan pada penelitian ini, yaitu Tahap Pengumpulan Data, Proses Data Mining, Proses Naïve Bayes, Testing Dataset, dan melakukan Kesimpulan dari tahapan-tahapan tersebut. Tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian



Gambar 2. Proses Data Mining

- Pengumpulan Data : pengumpulan data dilakukan dengan mengumpulkan data dari pihak sekolah SMK Muhammadiyah 1 Palembang dari berbagai jurusan mulai tahun 2022 sampai 2024 dengan total dataset sebanyak 562 buah.
- Proses Data Mining : pembersihan data, seleksi data, integrasi data, transformasi data, dan model

awal yang dibuat. Proses data mining dapat dilihat pada Gambar 2.

- c. Pembersihan data (*Data Cleaning*) : proses pembersihan noise (kotoran) dan data yang tidak konsisten dari kumpulan data, dan agar tidak ada missing value.
- d. Seleksi Data (*Data Selection*) : proses pemilihan data yang relevan yang diambil dari database sehingga dilakukan seleksi. Seleksi data ini melakukan pemilihan atribut yang relevan dan menghilangkan atribut yang tidak digunakan.
- e. Integrasi Data (*Integration Data*) : data yang telah dilakukan pembersihan lalu disatukan data satu dengan data lainnya yang berkaitan.
- f. Transformasi Data (*Transformation Data*) : data diubah dan digabung dengan memiliki format yang sesuai untuk diproses di data mining.

III. HASIL DAN PEMBAHASAN

A. Transformasi Data

Proses transformasi data untuk dapat mengubah bentuk data yang dilakukan agar dataset siap diproses untuk dapat menghasilkan analisis yang lebih baik.

Row No.	Prediksi	JENIS KELA...	JURUSAN	NILAI UJIAN ...	NILAIKOMP...	PRESTA
1	Tidak Lulus	Perempuan	TKJ 1	77.730	79	Ada
2	Lulus	Laki-Laki	TKJ 1	81.300	82	Ada
3	Lulus	Perempuan	AKL	85.880	87	Tidak Ada
4	Lulus	Laki-Laki	TKJ 2	83.090	85	Tidak Ada
5	Lulus	Laki-Laki	TKJ 2	82.640	83	Tidak Ada
6	Lulus	Laki-Laki	TKJ 1	83.900	84	Ada
7	Tidak Lulus	Laki-Laki	TKJ 2	83.910	75	Tidak Ada
8	Lulus	Laki-Laki	TKJ 3	83.400	84	Tidak Ada
9	Lulus	Laki-Laki	AKL	83.300	86	Tidak Ada
10	Lulus	Perempuan	AKL	88.310	84	Tidak Ada
11	Lulus	Perempuan	AKL	83.190	85	Tidak Ada
12	Lulus	Perempuan	AKL	89.560	89	Tidak Ada
13	Tidak Lulus	Laki-Laki	TKJ 2	82.500	77	Tidak Ada
14	Tidak Lulus	Laki-Laki	TKJ 2	75.770	75	Tidak Ada
15	Lulus	Laki-Laki	TKJ 1	82.360	80	Ada

Gambar 3. Dataset Sebelum Transformasi Data

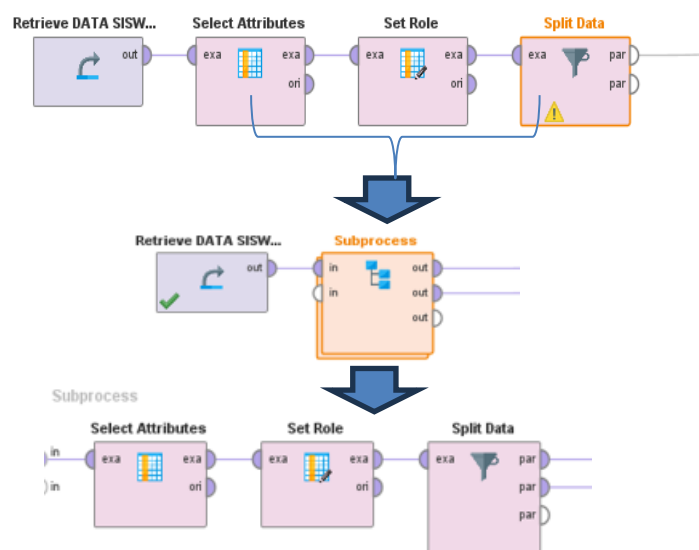
Berdasarkan Gambar 3, proses normalisasi dilakukan pada atribut nilai ujian dan nilai kompetensi. Mengubah nilai asli menjadi dalam skala tertentu, terlihat dari nilai-nilai yang berkisar antara -1 hingga 1. Normalisasi ini dilakukan untuk menyamakan skala nilai dari atribut tersebut, sehingga tidak ada atribut yang mendominasi atau memiliki bobot lebih besar dalam analisis data (Gambar 4). Nilai ujian dari nilai asli 77.73 berubah menjadi -1.717 dan nilai kompetensi keahlian dari nilai asli 79.00 berubah menjadi -0.327. Ini menunjukkan bahwa proses transformasi telah selesai.

Row No.	Prediksi	NILAI UJIAN ...	NILAIKOMP...	JENIS KELA...	JURUSAN	PRESTASI N...
1	Tidak Lulus	-1.717	-0.327	Perempuan	TKJ 1	Ada
2	Lulus	-0.747	0.309	Laki-Laki	TKJ 1	Ada
3	Lulus	0.498	1.370	Perempuan	AKL	Tidak Ada
4	Lulus	-0.260	0.946	Laki-Laki	TKJ 2	Tidak Ada
5	Lulus	-0.383	0.522	Laki-Laki	TKJ 2	Tidak Ada
6	Lulus	-0.040	0.734	Laki-Laki	TKJ 1	Ada
7	Tidak Lulus	-0.037	-1.176	Laki-Laki	TKJ 2	Tidak Ada
8	Lulus	-0.176	0.734	Laki-Laki	TKJ 3	Tidak Ada
9	Lulus	-0.203	1.158	Laki-Laki	AKL	Tidak Ada
10	Lulus	1.158	0.734	Perempuan	AKL	Tidak Ada
11	Lulus	-0.233	0.946	Perempuan	AKL	Tidak Ada
12	Lulus	1.498	1.795	Perempuan	AKL	Tidak Ada
13	Tidak Lulus	-0.421	-0.751	Laki-Laki	TKJ 2	Tidak Ada
14	Tidak Lulus	-2.250	-1.176	Laki-Laki	TKJ 2	Tidak Ada
15	Lulus	-0.459	-0.115	Laki-Laki	TKJ 1	Ada

Gambar 4. Dataset Sesudah Transformasi Data

B. Splitting Data

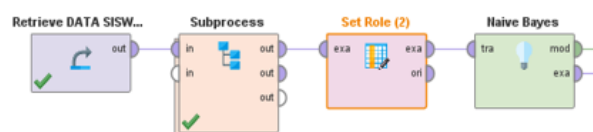
Setelah data selesai dilakukan Preprocessing data, data akan dicari akurasi. Namun, terlebih dahulu dilakukan pembagian data atau split data yang dibagi menjadi 70% untuk data training dan 30% untuk data testing. Untuk operator yang digunakan masih tetap sama, selanjutnya masukan operator *Set Role* untuk menentukan Label pada dataset (Gambar 5).



Gambar 5. Sub Proses dan Split Data

Pada Gambar 5 memperlihatkan tampilan operator yang telah menjadi Subprocess yang berfungsi sebagai ruang kerja terpisah dan menyimpan serangkaian operator digunakan untuk memproses data secara lebih terstruktur.

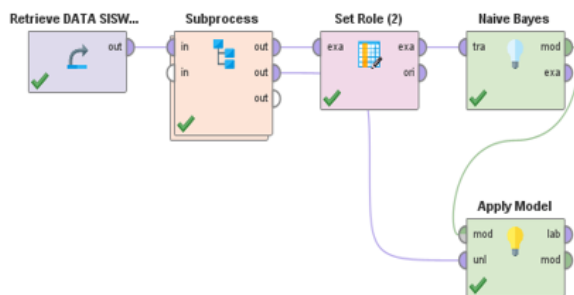
C. Proses Naïve Bayes



Gambar 6. Proses Naïve Bayes by Set Role

Langkah pada Gambar 6 yang dilakukan untuk mengatasi masalah Missing Label dengan menggunakan operator Set Role. Operator "Set Role" digunakan untuk mengidentifikasi peran setiap atribut dalam dataset, seperti membedakan antara atribut prediktor (input) dan atribut target (output). Hal ini penting agar algoritma klasifikasi seperti Naïve Bayes dapat mengetahui variabel mana yang harus diprediksi. Selanjutnya, data dibagi dengan operator "Split Data", yang berfungsi untuk membagi dataset menjadi dua subset, yaitu data pelatihan dan data pengujian. Rasio pembagian biasanya disesuaikan, misalnya 70:30 atau 80:20 untuk memastikan bahwa model yang dilatih cukup representatif namun tetap dapat diuji pada data yang tidak dikenali sebelumnya (*unseen data*).

D. Pembuatan Model



Gambar 7. Pembentukan Model

Pada Gambar 7, terlihat bahwa atribut yang ditampilkan berupa attribute name dan target role. Attribute name diatur sebagai prediksi, sedangkan target role ditetapkan sebagai label. Selanjutnya, operator Apply Model untuk menghubungkan Subprocess dan Naive Bayes ke Apply Model.

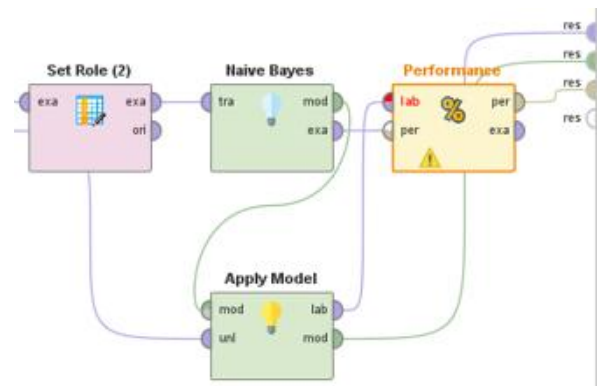
Setelah data diproses dalam subprocess, operator "Set Role (2)" digunakan untuk mendefinisikan atribut target yang akan diprediksi, dalam hal ini kemungkinan kelulusan siswa. Atribut target diatur dengan peran (role) sebagai label, sementara atribut lainnya diposisikan sebagai prediktor. Penetapan peran ini penting untuk memastikan bahwa algoritma klasifikasi mengenali variabel mana yang digunakan sebagai acuan dalam pembelajaran model.

Tahap selanjutnya adalah penggunaan algoritma Naïve Bayes berbasis probabilitas yang bekerja berdasarkan Teorema Bayes dan asumsi independensi antar fitur. Operator Naïve Bayes melakukan pelatihan (training) terhadap data pelatihan, menghasilkan sebuah model klasifikasi yang siap digunakan. Efisiensinya dalam menangani data berukuran besar dan kemampuannya dalam memberikan hasil yang kompetitif, meskipun dengan asumsi sederhana.

Setelah model selesai dibangun, dilakukan proses "Apply Model", yaitu penerapan model Naïve Bayes yang telah dilatih pada data pengujian atau data baru yang tidak dikenali sebelumnya. Operator ini menerima dua masukan utama, yaitu model klasifikasi dan data uji, kemudian menghasilkan keluaran berupa prediksi kelas (label) serta skor kepercayaannya. Proses ini memungkinkan evaluasi

performa model dengan membandingkan hasil prediksi terhadap label sesungguhnya dalam data pengujian.

E. Pengujian Model



Gambar 8. Uji Performance Model

Melakukan uji performance dengan menarik operator Performance dan sambungkan operator-operator yang relevan satu sama lain agar diperoleh hubungan. Mengukur tingkat keakuratan prediksi model terhadap data yang diuji dan mengetahui apakah parameter model sudah optimal untuk menghasilkan kinerja terbaik (Gambar 8). Hubungan antara Apply Model dan Performance sangat penting dalam proses validasi model. Operator Performance menerima dua masukan utama yaitu label aktual dan label hasil prediksi. Di sinilah dilakukan perbandingan antara nilai prediksi yang dihasilkan oleh model dengan nilai sebenarnya dalam dataset. Output dari operator ini berupa metrik evaluasi seperti akurasi, presisi, recall, F-measure, dan lain-lain, tergantung pada konfigurasi yang digunakan.

Penggunaan operator Performance memungkinkan peneliti untuk mengetahui sejauh mana kemampuan model dalam mengklasifikasikan data secara tepat. Dalam konteks pendidikan, seperti klasifikasi kelulusan siswa, nilai akurasi yang tinggi dari operator ini akan menunjukkan bahwa metode Naïve Bayes cukup efektif dalam memprediksi status kelulusan berdasarkan atribut-atribut yang dimiliki siswa. Sebaliknya, jika nilai performa rendah, perlu dilakukan evaluasi ulang terhadap pemilihan fitur, proporsi data, atau bahkan metode yang digunakan.

accuracy: 85.80%

	true Tidak Lulus	true Lulus	class precision
pred. Tidak Lulus	43	16	72.88%
pred. Lulus	8	102	92.73%
class recall	84.31%	86.44%	

Gambar 9. Uji Performance Model

Gambar 9 memperlihatkan hasil performance vector menunjukkan nilai akurasi sebesar 85,80%. Dari 393 data training, sebanyak 102 data yang dinyatakan lulus diprediksi benar, sebanyak 43 data yang dinyatakan tidak lulus diprediksi benar, sementara 8 data yang dinyatakan lulus diprediksi salah, dan 16 data yang diprediksi tidak lulus diprediksi salah.

Total data testing sebanyak 169 data didapatkan persentase siswa lulus sebanyak 65% atau sebanyak 110 siswa sedangkan siswa yang tidak lulus sebanyak 35% atau sebanyak 59 siswa.

Row No.	Prediksi	prediction(P...	confidence(...	confidence(...
1	Tidak Lulus	Tidak Lulus	0.946	0.054
2	Lulus	Lulus	0.092	0.908
3	Lulus	Lulus	0.002	0.998
4	Tidak Lulus	Tidak Lulus	0.864	0.136
5	Lulus	Lulus	0.001	0.999
6	Lulus	Lulus	0.000	1.000
7	Tidak Lulus	Tidak Lulus	0.999	0.001
8	Lulus	Lulus	0.363	0.637
9	Tidak Lulus	Lulus	0.402	0.598
10	Lulus	Lulus	0.463	0.537
11	Tidak Lulus	Tidak Lulus	0.882	0.118
12	Tidak Lulus	Tidak Lulus	0.639	0.361
13	Lulus	Lulus	0.369	0.631

ExampleSet (169 examples, 4 special attributes, 6 regular attributes)

Gambar 10. Data testing yang digunakan

Tabel 1. Data Uji Kelulusan Berdasarkan Jenis Kelamin

Prediksi Kelulusan Berdasarkan Jenis Kelamin		
Jenis Kelamin	Lulus	Tidak Lulus
Laki-Laki	53	24
Perempuan	58	35

$P(\text{Laki-Laki} | \text{Lulus})$ yaitu probabilitas siswa yang lulus berjenis kelamin laki-laki. Jumlah siswa laki-laki yang lulus = 53. Total siswa yang lulus = 111, sehingga

$$P(\text{Laki-Laki} | \text{Lulus}) = \frac{53}{111} \approx 0,477$$

$P(\text{Perempuan} | \text{Lulus})$ yaitu probabilitas siswa yang lulus berjenis kelamin Perempuan. Jumlah siswa perempuan yang lulus = 58. Total siswa yang lulus = 111, maka

$$P(\text{Perempuan} | \text{Lulus}) = \frac{58}{111} \approx 0,523$$

$P(\text{Laki-Laki} | \text{Tidak Lulus})$ yaitu jumlah siswa laki-laki yang tidak lulus = 24. Total siswa yang tidak lulus = 59.

$$P(\text{Laki-Laki} | \text{Tidak Lulus}) = \frac{24}{59} \approx 0,407$$

$P(\text{Perempuan} | \text{Tidak Lulus})$ yaitu jumlah siswa perempuan yang tidak lulus = 35. Total siswa yang tidak lulus = 59.

$$P(\text{Perempuan} | \text{Tidak Lulus}) = \frac{35}{59} \approx 0,593$$

Berdasarkan persentase ini, siswa Perempuan mendominasi dalam kategori lulus maupun tidak lulus.

Tabel 2. Prediksi Kelulusan Berdasarkan Jurusan

Prediksi Kelulusan Berdasarkan Jurusan		
Jurusan	Lulus	Tidak Lulus
TKJ	64	41
BDP	14	7
AKL	32	11

Dari total 105 siswa TKJ (Laki-Laki dan Perempuan), sebanyak 64 siswa lulus dan 41 siswa tidak lulus. Sedangkan dari 21 siswa BDP (Laki-Laki dan Perempuan), 14 siswa lulus dan 7 siswa tidak lulus. Selain itu, dari 43 siswa AKL (Laki-Laki dan Perempuan), 32 siswa lulus dan 11 siswa tidak lulus. Berdasarkan persentase tersebut, jurusan TKJ memiliki persentase kelulusan tertinggi di antara kategori lainnya (Tabel 2).

$P(\text{TKJ} | \text{Lulus})$ yaitu jumlah siswa TKJ yang lulus = 64. Total siswa yang lulus = 110.

$$P(\text{TKJ} | \text{Lulus}) = \frac{64}{110} \approx 0.582$$

$P(\text{TKJ} | \text{Tidak Lulus})$ yaitu jumlah siswa TKJ yang tidak lulus = 41. Total siswa yang tidak lulus = 59.

$$P(\text{TKJ} | \text{Tidak Lulus}) = \frac{41}{59} \approx 0,695$$

$P(\text{BDP} | \text{Lulus})$ yaitu jumlah siswa BDP yang lulus = 14. Total siswa yang lulus = 110.

$$P(\text{BDP} | \text{Lulus}) = \frac{14}{110} \approx 0,127$$

$P(\text{BDP} | \text{Tidak Lulus})$ yaitu jumlah siswa BDP yang tidak lulus = 7. Total siswa yang tidak lulus = 59.

$$P(\text{BDP} | \text{Tidak Lulus}) = \frac{7}{59} \approx 0,119$$

$P(\text{AKL} | \text{Lulus})$ yaitu jumlah siswa AKL yang lulus = 32. Total siswa yang lulus = 110.

$$P(\text{AKL} | \text{Lulus}) = \frac{32}{110} \approx 0,291$$

$P(\text{AKL} | \text{Tidak Lulus})$ yaitu jumlah siswa AKL yang tidak lulus = 11. Total siswa yang tidak lulus = 59.

$$P(\text{AKL} | \text{Tidak Lulus}) = \frac{11}{59} \approx 0,186$$

Seorang siswa dari jurusan TKJ, akan diprediksi apakah akan Lulus atau Tidak Lulus:

$$\begin{aligned} \text{a. } P(\text{Lulus} | \text{TKJ}) &\propto P(\text{Lulus}) \times P(\text{TKJ} | \text{Lulus}) \\ &= 0.651 \times 0.582 \approx 0.3789 \end{aligned}$$

$$\begin{aligned} \text{b. } P(\text{Tidak Lulus} | \text{TKJ}) &\propto P(\text{Tidak Lulus}) \times P(\text{TKJ} | \text{Tidak Lulus}) \\ &= 0.349 \times 0.695 \approx 0.2425 \end{aligned}$$

Karena $P(\text{Lulus} | \text{TKJ}) > P(\text{Tidak Lulus} | \text{TKJ})$ maka prediksi = Lulus.

Begitu juga untuk Jurusan BDP dan AKL,

$$\text{a. } P(\text{Lulus} | \text{BDP}) \propto 0.651 \times 0.127 \approx 0.0827$$

$$\text{b. } P(\text{Tidak Lulus} | \text{BDP}) \propto 0.349 \times 0.119 \approx 0.0415$$

Prediksi = Lulus

$$\text{a. } P(\text{Lulus} | \text{AKL}) \propto 0.651 \times 0.291 \approx 0.1895$$

b. $P(\text{Tidak Lulus} | \text{AKL}) \propto 0.349 \times 0.186 \approx 0.0649$
 Prediksi = Lulus.

Selain itu, diperkuat dengan Tabel 3 sebelumnya menunjukkan terdapat 77 siswa dengan nilai kompetensi keahlian ≤ 80 , di mana 28 siswa dinyatakan lulus dan 49 siswa tidak lulus. Di sisi lain, ada 89 siswa dengan nilai kompetensi keahlian ≥ 80 , dengan 82 siswa lulus dan 7 siswa tidak lulus. Nilai kompetensi keahlian di atas 80 menunjukkan peluang kelulusan yang jauh lebih tinggi dibandingkan dengan nilai di bawah 80.

Tabel 3. Prediksi Kelulusan Berdasarkan Nilai Kompetensi Keahlian

Prediksi Kelulusan Berdasarkan Nilai Kompetensi Keahlian		
Nilai Kompetensi Keahlian	Lulus	Tidak Lulus
≤ 80	28	49
≥ 80	82	7

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \\
 &= \frac{43+102}{43+8+16+102} \times 100\% \\
 &= 85,80\%
 \end{aligned}$$

Nilai akurasi sebesar 85,80% menunjukkan bahwa metode Naive Bayes efektif dalam melakukan klasifikasi prediksi kelulusan siswa. Tingkat akurasi ini tergolong sangat baik [7], [8] untuk memprediksi hasil pada data testing.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, prediksi kelulusan siswa SMK Muhammadiyah 1 Palembang menggunakan data mining dengan metode naïve bayes didapatkan implementasi data training sebanyak 393 data dengan algoritman Naïve Bayes berhasil memprediksi besarnya kelulusan mahasiswa dengan presentase keakuratan sebesar 85,80%.

Jurusan TKJ memiliki kontribusi kelulusan tertinggi, dengan probabilitas sekitar 58,2% dari total siswa yang lulus. Ini menunjukkan bahwa sebagian besar siswa yang berhasil lulus berasal dari jurusan Teknik Komputer dan Jaringan. Sedangkan jurusan AKL menunjukkan risiko ketidaklulusan yang lebih tinggi dibanding BDP, dengan probabilitas sebesar 27% lebih tinggi dari BDP sebesar 12%. Hal ini bisa menjadi perhatian dalam evaluasi program pembelajaran pada jurusan AKL.

Sebaran siswa tidak lulus paling banyak berasal dari jurusan TKJ dengan probabilitas sekitar 34,5%, yang meskipun masih memiliki jumlah lulusan tinggi, namun tetap menunjukkan bahwa jurusan ini juga perlu memperhatikan upaya peningkatan keberhasilan belajar bagi siswa yang berisiko tidak lulus.

REFERENSI

- [1] D. A. Punkastyo, F. Septian, and A. Syaripudin, "Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Untuk Prediksi Kelulusan Siswa," *Journal of System and Computer Engineering (JSCE)*, vol. 5, no. 1, pp. 24–35, Jan. 2024, doi: 10.61628/jsce.v5i1.1073.
- [2] Jefri Jefri and Zaehol Fatah, "Klasifikasi Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes," *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 2, no. 1, pp. 29–37, Feb. 2025, doi: 10.69714/mhjqlv85.
- [3] K. Sinaga and K. Handoko, "Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes," 2021.
- [4] A. Z. M. Sigid Widodo, A. Pandu Kusuma, and W. Dwi Puspitasari, "Analisis Algoritma Naive Bayes Classifier (Nbc) Pada Klasifikasi Tingkat Minat Barang Di Toko Violet Cell," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 87–94, Jan. 2023, doi: 10.36040/jati.v7i1.5692.
- [5] A. Andika, S. Syarli, and C. R. Sari, "Data Mining Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naïve Bayes," *Journal Pegguruang: Conference Series*, vol. 4, no. 1, p. 423, May 2022, doi: 10.35329/jp.v4i1.2358.
- [6] Y. Septiani and F. Ariyani, "Penerapan Algoritma Naive Bayes Menentukan Klasifikasi Tingkat Kelulusan Siswa SMK Media Informatika Jakarta," 2022. [Online]. Available: <https://senafti.budiluhur.ac.id/index.php>
- [7] L. Lia Regitaningtyas, T. Maharani, B. Hikmahwan Akademi Komunitas Negeri Pacitan Jl Walanda Maramis No, and K. Pacitan Kab Pacitan, "Klasifikasi Data

Lulusan Siswa Smp Menggunakan Metode Naïve Bayes,” *Kumpulan jurnal Ilmu Komputer (KLIK)*, vol. 09, no. 1, 2022.

- [8] M. Arifat, Wardiana Adinda Putri, and A. S. Mufida, “Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks Pembangunan Manusia Di Provinsi Jawa Timur,” *Jurnal Statistika dan Komputasi*, vol. 2, no. 1, pp. 31–43, Jun. 2023, doi: 10.32665/statkom.v2i1.1661.