

Analisis Klasifikasi Kelulusan Siswa Baru SMKS Bina Satria Medan Memakai Data Mining Serta Metode Naive Bayes

Analysis Of The Graduation Classification Of New Students Of SMKS Bina Satria Medan Using Informasi Mining And The Naïve Bayes Method

Dhea Ananda Pratiwi^{a,1}, Dedek Indra Gunawan Hts^{b,2}

^aMahasiswa Universitas Potensi Utama, Jl.K.L Yos Sudarso KM 6.5 Tj.Mulia, Medan, 20241, Indonesia

^bDosen Universitas Potensi Utama, Jl.K.L Yos Sudarso KM 6.5 Tj.Mulia, Medan, 20241, Indonesia

dheaananda@gmail.com¹ dedek.indra@gmail.com²

ABSTRAK

Klasifikasi kelulusan siswa di SMKS Bina Satria Medan adalah salah satu syarat kelulusan siswa siswinya untuk ke jenjang pendidikan yang lebih tinggi. Saat ini Pendidikan diakui sebagai kunci kemajuan bangsa, dengan tujuan membentuk generasi berkualitas. Penelitian ini berfokus pada analisis faktor-faktor yang memprediksi tingkat kelulusan siswa, dengan nilai belajar di sekolah sebagai penentu utama. Penelitian menggunakan metode Data Mining, khususnya Naïve Bayes Classifier, dengan fokus pada faktor klasifikasi seperti Nilai Ujian Akhir Sekolah Berbasis Nasional (UASBN), Nilai Raport, dan Nilai Kehadiran. Kelebihan metode Naïve Bayes, seperti kesederhanaan dan kecepatan, diharapkan dapat membantu pihak sekolah memahami pola kelulusan siswa dan menganalisis faktor-faktor yang memengaruhi. Penerapan metode ini diharapkan tidak hanya memprediksi tingkat kelulusan siswa berdasarkan nilai belajar, tetapi juga mempersingkat waktu dalam proses pengambilan keputusan oleh pihak sekolah.

Kata Kunci : Klasifikasi siswa, data mining, naive bayes.

ABSTRACT

The graduation classification of students at SMKS Bina Satria Medan is one of the requirements for students to graduate to a higher level of education. Currently, education is recognized as the key to the nation's progress, with the aim of forming a quality generation. This research focuses on analyzing factors that predict student graduation rates, with the value of learning at school as the main determinant. The research uses Data Mining methods, especially the Naïve Bayes Classifier, with a focus on classification factors such as National School Final Examination Scores (UASBN), Report Card Scores, and Attendance Scores. The advantages of the Naïve Bayes method, such as simplicity and speed, are expected to help schools understand student graduation patterns and analyze influencing factors. The application of this method is expected to not only predict student graduation rates based on learning scores, but also shorten the time in the decision-making process by the school.

Keywords : Student classification, data mining, naive Bayes.

Info Artikel :

Disubmit: 09 Juli 2023

Direview: 16 Agustus 2023

Diterima : 27 September 2023

Copyright © 2024 – UPU Journal. All rights reserved.

1. PENDAHULUAN

Berkembangnya bermacam teknologi data pada masa ini, seluruh kebutuhan kehidupan sehari-hari membutuhkan data yang akurat, sehingga data hendak jadi sesuatu elemen berarti dalam pertumbuhan warga dikala ini sampai waktu mendatang. Kerap kali data yang didapat masih wajib digali ulang dari informasi yang jumlahnya sangat besar. Salah satunya dalam bidang pembelajaran pada dasarnya pembelajaran ialah salah satu aspek kemajuan maupun kemandirian bangsa. Lewat pembelajaran kita bisa membentuk generasi penerus bangsa yang bermutu. Pembelajaran Nasional berperan buat meningkatkan bermacam keahlian serta mutu siswa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa. [1]

Pada riset ini, dicoba analisis bermacam aspek yang hendak memprediksi tingkatan kelulusan siswa bersumber pada nilai belajar siswa sepanjang belajar di sekolah. Nilai belajar siswa sepanjang belajar di sekolah diyakini memiliki peranan yang besar serta berarti dalam memastikan tingkatan kelulusan siswa tersebut. Selaku penilaiannya, siswa hendak memakai nilai belajarnya sepanjang disekolah yang hendak dijadikan suatu acuan buat tingkatan kelulusan siswa tersebut. Dengan demikian, nilai belajar siswa tersebut bisa memastikan kenapa tingkatan kelulusan siswa yang satu dengan yang lain bisa berbeda. Oleh sebab itu, bisa dijabarkan aspek apa saja yang jadi bawah klasifikasi kelulusan merupakan Nilai Tes Akhir Sekolah Berbasis Nasional (UASBN), Nilai Raport, serta Nilai Kehadiran. Data yang hendak digunakan merupakan nilai-nilai siswa.

Data mining ialah proses ataupun aktivitas mengumpulkan sekumpulan informasi yang mempunyai kapasitas yang besar buat di ekstraksi sehingga jadi data yang bisa digunakan[2]. Informasi mining sendiri mempunyai watak metode salah satunya ialah klasifikasi. Metode klasifikasi terdiri dari sebagian tata cara salah satunya merupakan Naïve Bayes Clasifier. Dalam perihal ini tata cara informasi mining yang digunakan buat melaksanakan riset ini ialah Naïve Bayes dengan memakai aplikasi WEKA. Naïve Bayes ialah metode prediksi berbasis probabilistik yang simpel bersumber pada pada pelaksanaan teorema bayes.

Kelebihan dari tata cara Naïve Bayes ialah sangat sederhana gampang buat digunakan serta kilat memerlukan lebih sedikit informasi uji/pelatihan, menanggulangi informasi yang kontinyu ataupun diskrit, model ini pula bisa digunakan buat prediksi probabilistik[3]. Pelaksanaan tata cara Naïve Bayes ini diharapkan sanggup menolong pihak sekolah mengenali pola kelulusan siswa-siswinya dengan menggunakan informasi siswa dan diharapkan dapat menganalisa aspek yang mempengaruhi pada tingkatan kelulusan, pastinya yang hendak memprediksi tingkatan kelulusan siswa bersumber pada nilai belajar siswa dan bisa mempersingkat waktu.[4]

2. METODE

Istilah data mining dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

- a. Data Selection
- b. Pre- processing / Cleaning
- c. Transformation
- d. Data mining
- e. Interpretation / EvaluationMetodologi[5]

A. Bahan Penelitian

Penelitian ini dicoba pada data siswa kelas XII yang hendak lulus dari SMKS Bina Satria Medan. Desain riset ini merupakan riset deskriptif dan kuantitatif yang fokus pada analisis data supaya bisa diukur serta dinumerik dan data kualitatif yang digunakan buat menggali anggapan nilai, serta perilaku seorang ataupun kelompok dalam konteks sosial tertentu dengan mengaitkan wawancara,observasi partisipatif, riset permasalahan serta analisis dokumen.

B. Metode Pengumpulan Data

Dalam melaksanakan riset dibutuhkan sebagian metode buat mengumpulkan informasi yang dibutuhkan dalam aktivitas riset ini. Ada Pula metode dalam pengumpulan informasi merupakan :

1. Pengamatan (Observation)

Dalam tatacara pengamatan ini periset melaksanakan pengamatan secara langsung pada SMKS Bina Satria Medan.

2. Wawancara (Interview)

Pengumpulan informasi ataupun data pada tatacara ini bisa di jalani dengan wawancara ataupun mengajukan pertanyaan-pertanyaan langsung pada siswa-siswi kelas XII di SMKS Bina Satria Medan.

- a. Pertanyaan : Kriteria apa saja yang dicoba pihak sekolah dalam melaksanakan klasifikasi kelulusan siswa baru?

Jawaban : Yang dibutuhkan merupakan informasi siswa semacam informasi nilai, kedatangan siswa serta prestasi.

- b. Pertanyaan : Apakah ada kesusahan pihak sekolah dalam melaksanakan klasifikasi kelulusan siswa?

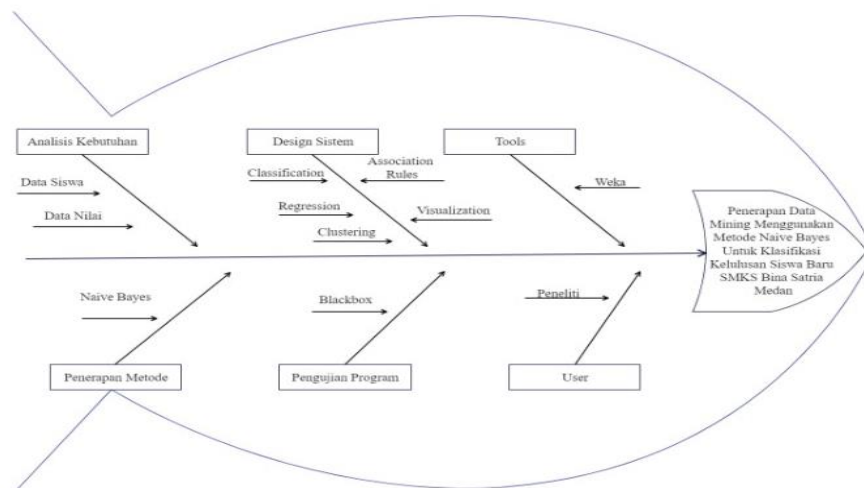
Jawaban : Melaksanakan klasifikasi perhitungan secara manual dari informasi nilai pula informasi kedatangan.

3. Sampel (*Sampling*)

Mempelajari serta memilah datadata yang ada serta cocok dengan bidang yang diseleksi selaku berkas lampiran. Data yang dibutuhkan merupakan data siswa-siswi kelas XII di SMKS Bina Satria Medan.

C. Kerangka Penelitian

Pengembangan sistem bisa berbentuk menyusun sesuatu sistem yang baru serta mengambil alih sistem yang lama secara totalitas ataupun membetulkan sistem yang sudah terdapat. Tiap sesi wajib di selesaikan terlebih dulu setelah itu diteruskan ke sesi selanjutnya buat menjauhi terbentuknya pengulangan sesi. Metodologi pengembangan sistem *Fishbone* merupakan diagram tulang ikan yang bisa menolong menciptakan pangkal pemicu permasalahan dalam proses pengembangan fitur lunak, di mana *Fishbone* diagram hendak mengenali bermacam karena potensial dari sesuatu dampak ataupun permasalahan serta menganalisis permasalahan tersebut lewat tahap *Brainstroming*. Berikut merupakan foto pengembangan fitur lunak menggunakan *Fishbone* diagram :



Gambar 1. Diagram Fishbone

1. HASIL DAN PEMBAHASAN

A. Pembahasan

Tahap ini merupakan tahap pelaksanaan tata cara pada sistem yang dirancang oleh periset yang dimana pelaksanaan tata cara pada sesuatu sistem merupakan buat memudahkan peroses penyelesaian dalam kasus yang terdapat pada sistem yang dirancang oleh periset membuat sistem pemesanan yang mempraktikkan tata cara *Naive Bayes*.

a. Studi Permasalahan

Menghitung probabilitas kelulusan memakai tata cara *Naive Bayes* dengan anggapan tanggapan kriteria-kriteria independen:

Tabel 1. Data Nilai Siswa

No	Nilai Kehadiran (X1)	Nilai UAS (X2)	Nilai Raport (X3)	Kelulusan (Y)
1.	70	85	Tinggi	Lulus
2.	70	80	Sedang	Lulus
3.	70	85	Tinggi	Lulus
4.	75	85	Sedang	Lulus
5.	80	80	Tinggi	Lulus
6.	80	85	Tinggi	Lulus
7.	90	85	Tinggi	Lulus
8.	75	75	Sedang	Lulus
9.	70	80	Sedang	Lulus
10.	90	90	Tinggi	Lulus

Untuk 190 data lainnya dapat dilihat dilampiran.

- Data pengujian ke-1:

Nilai Kehadiran (X1) = 70

Nilai UASBN (X2) = 85

Nilai Raport (X3) = Tinggi

a. Hitung probabilitas kelas lulus (Y=1):

$P(\text{Lulus}) = 0.6$

$P(X1=70|\text{Lulus}) = 0/6 = 0$ (tidak ada data lulus dengan X1=80)

$P(X2=85|\text{Lulus}) = 1/6 = 0.1667$

$P(X3=\text{Tinggi}|\text{Lulus}) = 3/6 = 0.5$

Probabilitas untuk kelas lulus = $P(\text{Lulus}) * P(X1=70|\text{Lulus}) * P(X2=85|\text{Lulus}) * P(X3=\text{Tinggi}|\text{Lulus}) = 0.6 * 0 * 0.1667 * 0.5 = 0$

b. Hitung probabilitas kelas tidak lulus (Y=0):

$P(\text{Tidak Lulus}) = 0.4$ (dari perhitungan sebelumnya)

$P(X1=70|\text{Tidak Lulus}) = 1/4 = 0.2$

$P(X2=85|\text{Tidak Lulus}) = 0/4 = 0$

$P(X3=\text{Tinggi}|\text{Tidak Lulus}) = 1/4 = 0.25$

Probabilitas untuk kelas tidak lulus = $P(\text{Tidak Lulus}) * P(X1=70|\text{Tidak Lulus}) * P(X2=85|\text{Tidak Lulus}) * P(X3=\text{Tinggi}|\text{Tidak Lulus}) = 0.4 * 0.25 * 0 * 0.25 = 0$

Karena probabilitas kelas lulus (0) $\leq$ probabilitas kelas tidak lulus (0), maka data pengujian ke-1 diklasifikasikan sebagai "Tidak Lulus" (Y=0).

Dengan demikian, berdasarkan perhitungan menggunakan metode *Naive Bayes*, hasil klasifikasi kelulusan siswa baru adalah sebagai berikut:

Tabel 2. Hasil Kelulusan

No.	Nilai Kehadiran (X1)	Nilai UAS (X2)	Nilai Raport (X3)	Kelulusan (Y)
1	70	85	Tinggi	Lulus
2	70	80	Sedang	Lulus
3	70	85	Tinggi	Lulus

1. Pengujian Program

Tahapan ini dapat dikatakan final dalam pembuatan suatu sistem. Sehabis melaksanakan analisa, desain serta pengkodean hingga sistem yang telah peruntukan digunakan oleh user. Pengujian secara blackbox (interface) ialah pengujian fitur lunak yang uji fungsionalitas dari sistem yang berlawanan dengan struktur internal ataupun kerja.

2. User

Pada sesi sistem penunjang keputusan buat klasifikasi kelulusan siswa baru SMKS Bina Satria Medan telah melewati sesi process serta siap buat digunakan, tidak menutup mungkin sistem ini hadapi pergantian kala telah digunakan oleh user. Pengguna aplikasi ini merupakan pihak sekolah SMKS Bina Satria Medan.

Pengolahan data menarangkan prosedur pengolahan informasi serta analisis data cocok dengan pendekata yang dicoba sebab riset ini memakai tata cara kuantitatif, data yang digunakan merupakan data siswa kelas XII yang hendak lulus dari SMKS Bina Satria Medan bersumber pada kriteria yang berlaku semacam informasi nilai serta data kehadiran.

Hasil dari data pilih hendak digunakan buat proses *Data Mining*, kemudian ditaruh pada sesuatu berkas yang terpisah dari database operasional. Pengolahan data riset yang dicoba oleh penulis dengan mengumpulkan informasi siswa-siswi yang baru lulus yang diperoleh lewat Sekolah SMKS Bina Satria Medan.

Pada tahapan setelah melaksanakan pilih data peneliti melaksanakan tahapan *preprocessing* saat sebelum dikerjakannya proses *Data Mining*. Pada tahapan *preprocessing* ini ada 2 proses yang dicoba ialah proses *cleaning* dengan melaksanakan membuang data yang duplikat, mengecek informasi yang tidak tidak berubah-ubah serta membetulkan kesalahan pada data. Proses kedua adalah proses *enrichment* yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan yang diperlukan untuk *Knowledge Discovery in Database (KDD)*.

Pada tahapan *transformation* ini informasi diganti ataupun digabung ke dalam format yang cocok buat diproses dalam *Data Mining*. Sebagian tata cara *Data Mining* memerlukan format data yang spesial saat sebelum dapat di aplikasikan..

Pada tahapan *Data Mining* ini merupakan proses mencari pola ataupun data yang menarik pada idata yang terpilih dengan memakai metode serta tata cara tertentu. Pemilihan metode serta tata cara tertentu yang pas sangat tergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan. Metode yang penulis gunakan pada penelitian ini adalah Metode *Naive Bayes*.

Pada tahap terakhir ini peneliti menyusun hasil dari data ataupun data pada proses yang dihasilkan dari *data mining* kedalam wujud yang gampang di paham oleh pihak yang berkepentingan. Tahap ini

tercantum pengecekan dari pola informasi ataupun data yang ditemui berlawanan dengan kenyataan ataupun hipotesis yang terdapat lebih dahulu.

Pada tahapan ini diperoleh data nilai dan data absensi siswa-siswi yang sesuai dengan kriteria kelulusan di SMKS Bina Satria Medan.

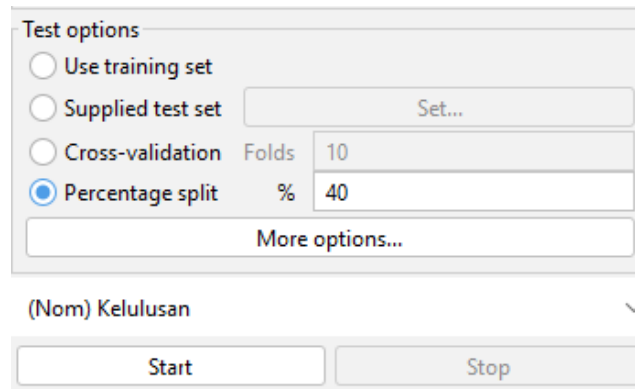
B. Hasil

Hasil Penerapan Perhitungan Model dengan Algoritma *Naive Bayes*.

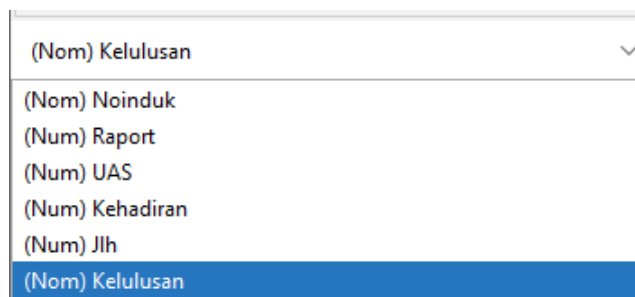
1. Split 40%

Tahapan Perhitungan dengan Split 40% data *train* :

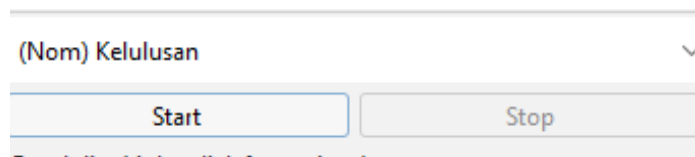
- Menentukan persentase jumlah data yang akan dilatih.
- Memilih kelas target uji, dalam hal ini adalah “Kelulusan”.
- Menekan tombol ‘start’ untuk memulai *test*.
- Sistem melakukan pembagian data secara otomatis 40% dari 200 yaitu 80 sebagai data latih dan data uji sebanyak 60% dari 200 yaitu 120.
- Muncul hasil pengujian sebanyak 120 data.



Gambar 2. Menentukan Pembagian Data Latih



Gambar 3. Memilih *Class Target*



Gambar 4. Menekan tombol ‘Start’

```
Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      119      99.1667 %
Incorrectly Classified Instances    1        0.8333 %
Kappa statistic                    0.6629
Mean absolute error                0.0083
Root mean squared error            0.0913
Relative absolute error            20.7071 %
Root relative squared error        71.178 %
Total Number of Instances          120

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
1.000  0.500  0.992    1.000  0.996    0.704  0.750  0.992    Lulus
0.500  0.000  1.000    0.500  0.667    0.704  1.000  1.000    Tidak Lulus
Weighted Avg.  0.992  0.492  0.992    0.992  0.990    0.704  0.754  0.992

=== Confusion Matrix ===
  a  b  <-- classified as
118  0  |  a = Lulus
 1  1  |  b = Tidak Lulus
```

Gambar 5. Hasil Uji Split 40f%

Berdasarkan gambar 2. data disajikan ulang ke dalam tabel berikut :

Tabel 2. Detail Akurasi *Class*

Metriks	Kelas "Lulus" (a)	Kelas "Tidak Lulus" (b)	Keterangan
TP Rate (Recall)	1.000	0,500	Tingkat identifikasi "Lulus" yang benar oleh model.
FP Rate	0,500	0	Tingkat kesalahan identifikasi "Tidak Lulus" oleh model.
Precision	0,992	1.000	Tingkat ketepatan model dalam mengklasifikasikan "Lulus."
F-Measure	0,996	0,667	Ukuran keseimbangan antara <i>Precision</i> dan <i>Recall</i> .
MCC	0,704	0,704	Ukuran korelasi antara hasil prediksi dan data aktual.
ROC Area	0,750	1.000	Area di bawah kurva ROC (<i>Receiver Operating Characteristic</i>).
PRC Area	0,992	1.000	Area di bawah kurva <i>Precision-Recall</i> .

Tabel 3. *Confusion Matrix* Hasil Uji Split 40%

	Hasil Sebenarnya "Lulus"	Hasil Sebenarnya "Tidak Lulus"	Keterangan
Hasil Prediksi "Lulus"	118	0	Kasus yang benar-benar "Lulus" dan diprediksi "Lulus" oleh model.
Hasil Prediksi "Tidak Lulus"	1	1	Kasus yang sebenarnya "Tidak Lulus" tetapi diprediksi "Lulus" oleh model.

Hasil Perhitungan Manual Confusion Matrix Split 40% :

1. Akurasi $= ((TP + TN) / (TP + TN + FP + FN)) * 100\%$
 $= ((118 + 0) / (118 + 0 + 0 + 1)) * 100\%$
 $= (118 / 119) * 100\%$
 $\approx 99.16\%$
 Jadi, akurasi sekitar 99.16%. (99.2%) atau 0.992
2. Presisi $= ((TP / (TP + FP)) * 100\%$
 $= ((118 / (118 + 0)) * 100\%$
 $= (118 / 118) * 100\%$
 $= 100\%$
 Jadi, presisi adalah 100%.
3. Recall $= ((TP / (TP + FN)) * 100\%$
 $= ((118 / (118 + 1)) * 100\%$
 $= (118 / 119) * 100\%$
 $= 99.16\%$
 Jadi, recall adalah 99.16%. (99.2%) atau 0.992

Penjelasan dari hasil evaluasi data yang tertera diatas adalah sebagai berikut :

Penilaian prestasi model terhadap data uji mengungkapkan temuan yang substansial. Dari keseluruhan 120 sampel yang terdapat dalam dataset uji, sekitar 119 diantaranya berhasil diklasifikasikan secara akurat oleh model, menciptakan tingkat akurasi sebesar 99.1667%, sebuah pencapaian yang luar biasa dalam konteks klasifikasi. Selain itu, statistik *Kappa* dengan nilai sekitar 0.6629 menggambarkan tingkat kesesuaian yang signifikan antara prediksi yang diberikan oleh model dengan label yang sebenarnya pada data uji. Penilaian ini juga diperkuat oleh pengukuran *Mean Absolute Error* dan *Root Mean Squared Error* yang sangat kecil, menunjukkan kemampuan model dalam memberikan prediksi yang sangat mendekati nilai sebenarnya pada data uji.

Ketika mempertimbangkan akurasi berdasarkan kelas, model ini terutama unggul dalam mengklasifikasikan kelas "Lulus," dengan nilai *Precision*, *Recall*, dan *F-Measure* yang signifikan. Namun, penting untuk dicatat bahwa untuk kelas "Tidak Lulus," meskipun model memiliki nilai *Precision* yang tinggi, terjadi penurunan nilai *Recall*, yang menandakan bahwa model lebih cenderung mengidentifikasi siswa yang lulus daripada yang tidak lulus.

Pada pemeriksaan melalui matriks kebingungan (*confusion matrix*), dapat dilihat bahwa sebagian besar sampel yang termasuk dalam kategori "Lulus" berhasil diklasifikasikan dengan benar oleh model. Walaupun ada satu sampel yang termasuk dalam kategori "Tidak Lulus" yang keliru diklasifikasikan sebagai "Lulus." Namun, secara keseluruhan, hasil evaluasi model terhadap data uji menunjukkan kinerja yang signifikan dan dapat diandalkan dalam memprediksi hasil siswa.

Dalam analisis kinerja model, terdapat temuan yang mencerminkan kemampuan model dalam mengklasifikasikan kategori "Lulus" dengan sangat baik. Terdapat 118 kasus yang berhasil diklasifikasikan dengan benar sebagai "Lulus" oleh model (*True Positives*), menunjukkan bahwa model mampu mengidentifikasi siswa yang sebenarnya lulus dengan tingkat akurasi yang tinggi. Namun, perlu diperhatikan bahwa dalam satu kasus (*False Positives*), model salah mengklasifikasikan siswa sebagai "Lulus" padahal sebenarnya mereka adalah "Tidak Lulus," mengindikasikan adanya satu prediksi positif yang tidak akurat.

Untuk mengukur tingkat ketepatan prediksi positif, nilai *Precision* yang sekitar 0.992 sangat tinggi, mencerminkan bahwa sekitar 99.2% dari prediksi "Lulus" yang dibuat oleh model benar-benar sesuai dengan kenyataan. Sementara itu, nilai *Recall* yang mencapai 1.000 menunjukkan bahwa model mampu mengidentifikasi semua siswa yang lulus dengan benar. Ini berarti model berhasil menemukan semua kasus positif yang sebenarnya.

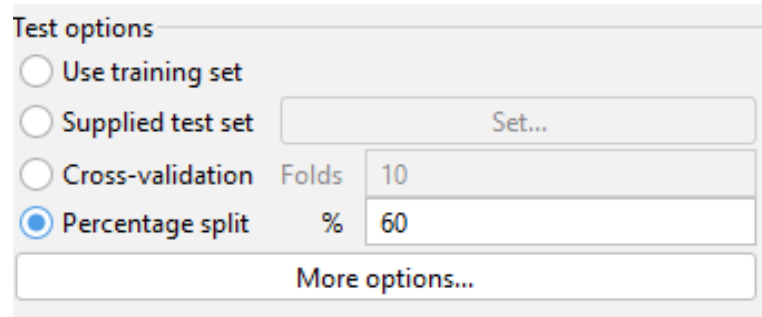
Selain itu, dengan nilai *F-Measure* sekitar 0.996, dapat disimpulkan bahwa model memiliki keseimbangan yang baik antara kemampuan untuk memprediksi dengan benar (*Precision*) dan kemampuan untuk mengidentifikasi semua kasus positif yang sebenarnya (*Recall*). Hasil ini

menggambarkan kinerja model yang sangat baik dalam mengklasifikasikan siswa yang lulus dalam penelitian ini.

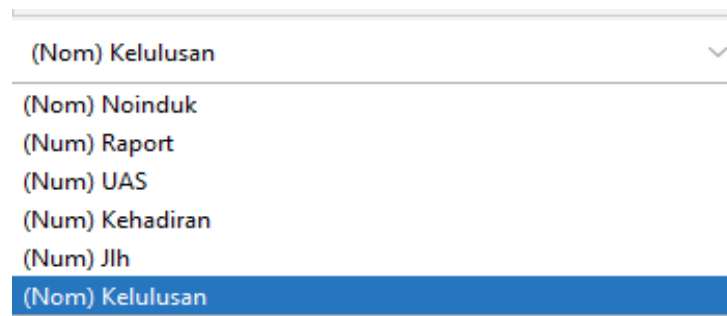
2. Split 60%

Tahapan Perhitungan dengan Split 60% data *train* :

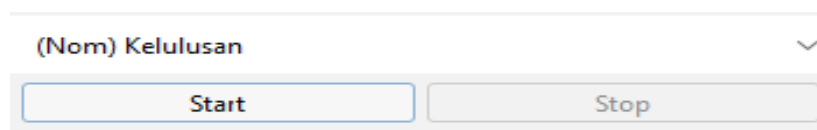
- Menentukan persentase jumlah data yang akan dilatih.
- Memilih kelas target uji, dalam hal ini adalah “Kelulusan”.
- Menekan tombol ‘start’ untuk memulai *test*.
- Sistem melakukan pembagian data secara otomatis 60% dari 200 yaitu 120 sebagai data latih dan data uji sebanyak 40% dari 200 yaitu 80.
- Muncul hasil pengujian sebanyak 80 data.



Gambar 6. Menentukan Pembagian Data Latih



Gambar 7. Memilih *Class* Target



Gambar 8. Menekan tombol ‘Start’

```

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances      79          98.75 %
Incorrectly Classified Instances    1           1.25 %
Kappa statistic                    0.661
Mean absolute error                 0.0125
Root mean squared error             0.1118
Relative absolute error             30.8081 %
Root relative squared error         71.5029 %
Total Number of Instances          80

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          1.000    0.500    0.987     1.000    0.994     0.703    0.750    0.987    Lulus
          0.500    0.000    1.000     0.500    0.667     0.703    1.000    1.000    Tidak Lulus
Weighted Avg.    0.988    0.488    0.988     0.988    0.985     0.703    0.756    0.988

=== Confusion Matrix ===
  a  b  <-- classified as
 78  0  |  a = Lulus
  1  1  |  b = Tidak Lulus

```

Gambar 9. Hasil Uji Split 60%

Berdasarkan gambar 9. data disajikan ulang ke dalam tabel sebagai berikut:

Tabel 4. Detail Akurasi *Class*

Metriks	Kelas "Lulus" (a)	Kelas "Tidak Lulus" (b)	Keterangan
TP Rate (Recall)	1,000	0,500	Tingkat identifikasi "Lulus" yang benar oleh model.
FP Rate	0,500	0,000	Tingkat kesalahan identifikasi "Tidak Lulus" oleh model.
Precision	0,987	1,000	Tingkat ketepatan model dalam mengklasifikasikan "Lulus."
F-Measure	0,994	0,667	Ukuran keseimbangan antara <i>Precision</i> dan <i>Recall</i> .
MCC	0,703	0,703	Ukuran korelasi antara hasil prediksi dan data aktual.
ROC Area	0,750	1,000	Area di bawah kurva ROC (<i>Receiver Operating Characteristic</i>).
PRC Area	0,987	1,000	Area di bawah kurva <i>Precision-Recall</i> .

Tabel 5. *Confusion Matrix* Hasil Uji Split 60

	Hasil Sebenarnya "Lulus"	Hasil Sebenarnya "Tidak Lulus"	Keterangan
Hasil Prediksi "Lulus"	78	0	Kasus yang benar-benar "Lulus" dan diprediksi "Lulus" oleh model.
Hasil Prediksi "Tidak Lulus"	1	1	Kasus yang sebenarnya "Tidak Lulus" tetapi diprediksi "Lulus" oleh model.

Hasil Perhitungan Manual Confusion Matrix Split 60% :

- $$\begin{aligned} \text{Akurasi} &= ((\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})) * 100\% \\ &= ((78 + 0) / (78 + 0 + 0 + 1)) * 100\% \\ &= (78 / 79) * 100\% \end{aligned}$$

$\approx 98.7342\%$

Jadi, akurasi sekitar 98.7342%(98.7%) atau 0.987

$$\begin{aligned} 2. \text{ Presisi} &= ((TP / (TP + FP)) * 100\% \\ &= ((78 / (78 + 0)) * 100\% \\ &= (78 / 78) * 100\% \\ &= 100\% \end{aligned}$$

Jadi, presisi adalah 100%.

$$\begin{aligned} 3. \text{ Recall} &= ((TP / (TP + FN)) * 100\% \\ &= ((78 / (78 + 1)) * 100\% \\ &= (78 / 79) * 100\% \\ &= 98.7342\% \end{aligned}$$

Jadi, recall adalah 98.7342%. (98.7%) atau 0,987

Penjelasan dari hasil evaluasi data yang tersaji pada tabel sebagai berikut :

Hasil pengujian model menunjukkan pencapaian yang signifikan dalam mengklasifikasikan status kelulusan siswa. Dari total 80 sampel yang dievaluasi, model berhasil mengklasifikasikan 79 di antaranya dengan benar, mencapai tingkat akurasi sekitar 98.75%. Hasil ini mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam memprediksi kelulusan siswa. Selanjutnya, tingkat kesepakatan antara prediksi model dan label sebenarnya diukur menggunakan nilai *Kappa Statistic*, yang mencapai sekitar 0.661. Hal ini menunjukkan bahwa model memiliki tingkat kesepakatan yang baik dalam memprediksi kelulusan.

Dalam hal ketepatan prediksi, analisis *Mean Absolute Error* (MAE) dengan nilai sekitar 0.0125 dan *Root Mean Squared Error* (RMSE) sekitar 0.1118 mengindikasikan bahwa prediksi model mendekati nilai sebenarnya dengan baik. Model juga menunjukkan tingkat akurasi yang sangat tinggi untuk kelas "Lulus" dengan *Precision*, *Recall*, dan *F-Measure* yang tinggi. Namun, perlu dicatat bahwa untuk kelas "Tidak Lulus," model memiliki nilai *Precision* yang tinggi tetapi *Recall* lebih rendah.

Melalui matriks konfusi, dapat dilihat bahwa mayoritas sampel dalam kategori "Lulus" berhasil diklasifikasikan dengan benar oleh model. Meskipun demikian, terdapat satu kasus *False Negative*, yang merupakan siswa "Tidak Lulus" yang salah diklasifikasikan sebagai "Lulus." Selain itu, dalam konteks analisis variabilitas, *Relative Absolute Error* sekitar 30.81%, mengindikasikan adanya variasi yang signifikan antara prediksi model dan nilai sebenarnya pada beberapa sampel. Variasi ini perlu menjadi fokus dalam analisis lebih lanjut.

Hasil pengujian ini memiliki implikasi penting dalam pengambilan keputusan terkait kelulusan siswa. Meskipun model menunjukkan potensi besar, perlu diperhatikan adanya satu kasus *False Negative* dan variasi dalam *Relative Absolute Error* yang memerlukan investigasi lebih lanjut untuk memastikan keandalan prediksi model.

4. KESIMPULAN

Software WEKA merupakan *software* yang dapat digunakan untuk membantu pihak sekolah dalam pengelompokan data kelulusan siswa. *Software* yang diterapkan ini menggunakan metode *Naive Bayes*. Berdasarkan uraian tersebut diatas maka dapat ditarik kesimpulan sebagai berikut:

1. Penerapan *data mining* dapat membantu dalam mengklasifikasi data kelulusan siswa di SMKS Bina Satria Medan.
2. Data siswa diolah menggunakan metode *naive bayes* untuk menentukan faktor-faktor kelulusan siswa berdasarkan data nilai uasbn, kehadiran, dan nilai raport.
3. Hasil dari penelitian ini diharapkan dapat membantu pihak sekolah dalam mengklasifikasi siswa baru dengan efisien dan akurat.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Potensi Utama yang sudah banyak memberikan masukan serta anjuran dalam penyelesaian skripsi ini.

REFERENSI

- [1] R Amalia, 2020. "*Penerapan Data Mining untuk Memprediksi Hasil Kelulusan Siswa Menggunakan Metode Naïve Bayes*".
- [2] Tumini dan Lisma Damayanti, 2018. "Penerapan Data Mining Menggunakan Metode Naïve Bayes Untuk Memprediksi Kelulusan Siswa di SMP Negeri 11 Kotabumi Utara
- [3] Moonallika, KQ Fredlina, 2020. "*Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara)*".
- [4] Y Septiani, PF Ariyani, 2022. "*Penerapan Algoritma Naive Bayes Menentukan Klasifikasi Tingkat Kelulusan Siswa SMK Media Informatika Jakarta*".
- [5] N Yustira, D Witarsyah, 2021. "*Implementasi Algoritma Naïve Bayes Classification untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu*".
- [6] Asri, Saffinah Indah, 2021. "*PENERAPAN DATA MINING SISTEM PENILAIAN PRESTASI MAHASISWA MENGGUNAKAN METODE ALGORITMA NAÏVE BAYES CLASSIFICATION (STUDI KASUS STMIK AKAKOM YOGYAKARTA)*".
- [7] SAbdurrohman, AWibowo, 2022. "*Klasifikasi Data Mining Untuk Memprediksi Status Penerimaan Di Perguruan Tinggi Negeri Bagi Lulusan Bimbel NF Dengan Algoritme Naive Bayes*".
- [8] Julianto Simatupang dan Setiawan Sianturi (2019). Perancangan Sistem Informasi Pemesanan Tiket Bus Pada PO. Handoyo Berbasis Online. Jurnal Intra-Tech.
- [9] Luwis H. Laisina, dkk (2018). Sistem Informasi Data Jemaat GPM Gidion Waiyari Ambon Dan Jemaat GPM Halong Anugerah Ambon. Jurnal SIMETRIK.
- [10] Maydianto dan Muhammad Rasid Ridho (2021). Rancang Bangun Sistem Informasi Point Of Sale Dengan Framework Codeigniter Pada CV Powershop. Jurnal Comasie.
- [11] Nadia Ayu Agustri Anggara, dkk (2022). Penerapan Customer Relationship Management (CRM) Dalam Sistem Informasi Penjualan Kosmetik Berbasis Web. Jurnal Building of Informatics, Technology and Science (BITS).
- [12] Ricki dan Joni Devitra (2019). Analisis Dan Perancangan Sistem Pendukung Keputusan Pemilihan Karyawan Terbaik Pada PT. Wahyu Perdana Persada Dengan Metode Simple Additive Weighting (SAW). Jurnal Manajemen Sistem Informasi.
- [14] Salman Wajdi (2022). Pengaruh Kepuasan Pelanggan Dan Kualitas Pelayanan Terhadap Loyalitas Pelanggan Bengkel Resmi Motor Honda (AHASS) "Safari Service Padang". Jurnal Akademi Akuntansi Indonesia Padang (JAAIP).
- [15] Wirhan Fahrozi dan Charles Bronson Harahap (2018). Sistem Informasi Transparansi Nilai Mata Kuliah Berbasis Web. Jurnal JURIKOM PRIMA.