

PENERAPAN NAÏVE BAYES DALAMS PREDIKSI KETERCAPAIAN NILAI KRITERIA KETUNTASAN MINIMAL SISWA

Dwi Hartanti¹, Kusri², Emha Luthfi Taufiq³

Universitas AMIKOM Yogyakarta

Jalan Ring Road Utara, Condongcatur, Depok, Condongcatur, Daerah Istimewa Yogyakarta
Email: dwihartanti048@gmail.com¹, kusri@amikom.ac.id², emhataufiqluthfi@amikom.ac.id³

ABSTRAK

Data mining adalah proses menemukan bagaimana hubungan dalam data yang tidak diketahui oleh pengguna menyajika dengan cara yang dapat dipahami sehingga hubungan menjadi dasar dalam pengambilan keputusan. Setiap sekolah memiliki sebuah masalah dalam nilai KKM siswa. KKM merupakan Kriteria Ketuntasan Minimal nilai siswa. KKM akan menjadi patokan nilai minimal suatu nilai matapelajaran. Banyak siswa yang nilai tidak tercapai nilai KKM akan menjadi sebuah pekerjaan sekolah dalam hal ini. Dalam hal ini sekolah membutuhkan sebuah penerapan datamining yang dapat melakukan prediksi ketercapain nilai KKM siswa. Agar sekolah mendapatkan informasi siswa yang diperkirakan tidak tercapai nilai mata pelajaran. Sehingga sekolah dapat melakukan sebuah tindakan untuk masalah ini. Tujuan dari penelitian ini adalah melakukan prediksi ketercapaian nilai KKM siswa dengan menggunakan algoritma Naïve Bayes. Dalam melakukan pengukuran tingkat akurasi ketepatan hasil naïve bayes dalam melakukan prediksi ketercapain nilai KKM dalam mengikuti matapelajaran. Atribut yang digunakan adalah nilai raa-rata UTS gasal, pendidikan orang tua, bimbingan belajar dan presensi kehadiran. Hasil pengujian dengan 600 record dengan porsi 75% data latih dan 25% data uji. Sistem menghasilkan akurasi sebesar 78%.

Kata kunci: *Prediski, Naïve Bayes, Data Mining, Confusion Matriks*

1. PENDAHULUAN

Dalam dunia pendidikan sekarang ini nilai menjadi salah satu penentu ketercapain nilai KKM siswa. Adanya nilai Kriteria Ketuntasan Minimal (KKM) setiap mata pelajaran merupakan salah satu muatan penting Kurikulum Tingkat Satuan Pendidikan. Kriteria Ketuntasan Minimal menjadi acuan bersama antara pendidik, peserta didik, dan orang tua peserta didik. Sehingga pihak-pihak yang berkepentingan terhadap penilaian di sekolah berhak untuk mengetahuinya. Setiap satuan pendidikan perlu melakukan sosialisasi agar informasi dapat diakses dengan mudah oleh peserta didik dan atau orang tuanya.

Kriteria Ketuntasan Minimal (KKM) ditetapkan pada setiap awal tahun pelajaran. Guru menetapkan Kriteria Ketuntasan Minimal (KKM) mata pelajaran dengan mempertimbangkan tiga aspek kriteria, yaitu kompleksitas, daya dukung dan kemampuan (intake) peserta didik. Dari ketiga aspek tersebut diberi skor antara 0-100, kemudian dihitung nilai rata-rata untuk setiap indikator, rata-rata setiap kompetensi dasar, rata-rata setiap standar kompetensi. Untuk menetapkan nilai Kriteria Ketuntasan Minimal (KKM) mata pelajaran diperoleh dari rata-rata semua standar kompetensi.

Pendidikan sangat berpengaruh positif terhadap ketercapaian nilai siswa karena dengan tingginya pendidikan orang tua maka akan memberikan perhatian yang khusus kepada pendidikan anak dan nilai anak.

Bimbingan belajar salah satu sarana yang banyak diikuti siswa yang ingin mendapatkan pelajaran diluar sekolah. Dengan siswa yang mengikuti bimbingan belajar akan membuat nilai siswa kan lebih meningkat dengan mengikuti bimbingan belajar.

Kehadiran atau presensi siswa dalam mengikuti pembelajaran menjadi salah satu faktor ketercapaian nilai siswa. Siswa yang rajin mengikuti pembelajaran tingkat pemahaman akan suatu materi kna lebih baik daripada siswa yang jarang mengikuti pembelajaran.

Nilai merupakan indeks atau petunjuk bahwa siswa tercapai atau tidak tercapainya anak dalam mengikuti mata pelajaran di sekolah. Nilai akan diolah dengan beberapa matapelajaran dan akan didapatkan rata-rata nilai siswa.

Tujuan penelitian ini adalah : 1) Untuk mengetahui akurasi yang diperoleh dari penerapan metode naïve bayes pada prediksi ketercapain nilai KKM siswa. 2) Untuk membangun sistem yang membantu dalam prediksi ketercapain nilai KKM siswa. Dengan sistem ini akan membantu sekolah dalam mengetahui siswa mana saja yang tidak tercapai dan tercapai. Sekolah akan melakukan tindakan atas informasi yang didapatkan salah satunya melakukan pembelajaran tambahan.

Berdasarkan pengamatan penelitian terdahulu sebagai bahan rujukan yaitu penelitian yang dilakukan oleh Tutus Pranigki pada jurnal tersebut

Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan K-NN. Penelitian dilakukan dengan berbagai metode algoritma. Penelitian ini menghasilkan naïve ayes memiliki nilai akurasi yang tertinggi.

Masalah yang muncul yang berhubungan dengan ini adalah banyaknya siswa yang tidak tuntas pada mata pelajaran tertentu. Maka sekolah harus melakukan sebuah tindakan ntuk mengatasi banyak siiswa yang tidak tuntas seperti memberikan pelajaran tambahan, menata penempatan tempat duduk siswa. Berdasarkan penjabaran diatas maka di perlukan sebuah sistem prediksi dengan data mining yang dapat melakukan prediksi ketercapaian KKM siswa. Dengan informasi yang dihasilkan dapat membantu sekolah dalam memprediksi siswa yang tidak tuntas.

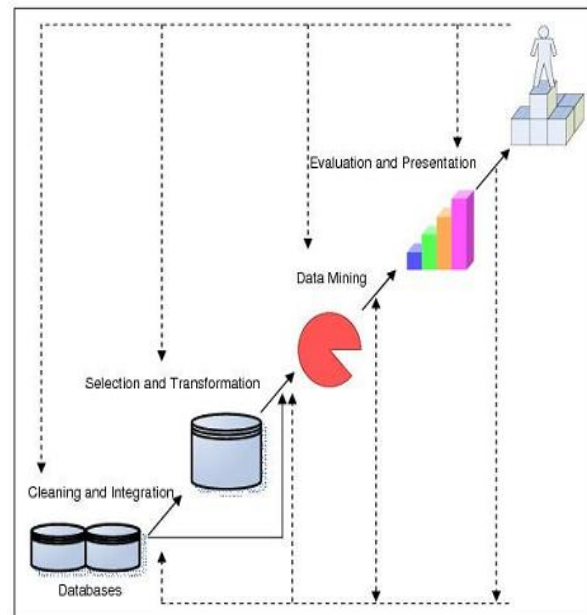
Sedangkan Penelitian yang dilakukan penulis adalah mengetahui akursi sistem prediksi ketercapain nilai KKM siswa menggunakan metode Naive Bayes. Suatu sistem harus memiliki akurasi tinggi dalam melakukan prediksi.

2. ISI PENELITIAN

2.2 Definisi

2.2.1 Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database.. Data mining juga merupakan proses dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining merupakan bagian yang terintegrasi dengan proses *Knowledge Discovery in Databases* (KDD). KDD adalah menggali dan menganalisis informasi dan pengetahuan yang tersembunyi dari sejumlah besar himpunan data.[7]. Tahapan proses *Knowledge Discovery in Databases* ditunjukkan pada gambar 1.



Gambar 1. Tahapan Data mining

2.2.2 Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi. Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining juga merupakan proses dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil.

2.2.3 Naïve Bayes

Algoritma *Naïve Bayes Classifier* (NBC) merupakan algoritma yang memanfaatkan teori probabilitas, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Klasifikasi-klasifikasi *Bayes* adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Untuk klasifikasi *Bayes* sederhana yang lebih dikenal sebagai *naïve Bayesian Classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain.

Teorema *Bayes* memiliki bentuk umum sebagai berikut

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \dots\dots\dots(1)$$

Dalam hal ini:

- X : data dengan class yang belum diketahui
- H : hipotesis data X merupakan suatu class spesifik
- P(H|X) : probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
- P(H) : Probabilitas hipotesis H (*priorprobability*)
- P(X|H) : probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) probabilitas dari X

Hasil dari rumus diatas untuk mencari probabilitas yang paling tinggi dengan membandingkan seluruh data yang ada yaitu P(H|X).

Teorema *Bayesian* dengan asumsi bahwa setiap variabel bersifat bebas (*independence*). Sehingga dengan asumsi seperti itu menyimpulkan bahwa setiap atribut tidak saling terkait (*conditionally independence*) maka

$$P(X|Ci) = \sum_{k=1}^n P(Xk|Ci) \dots\dots\dots(2)$$

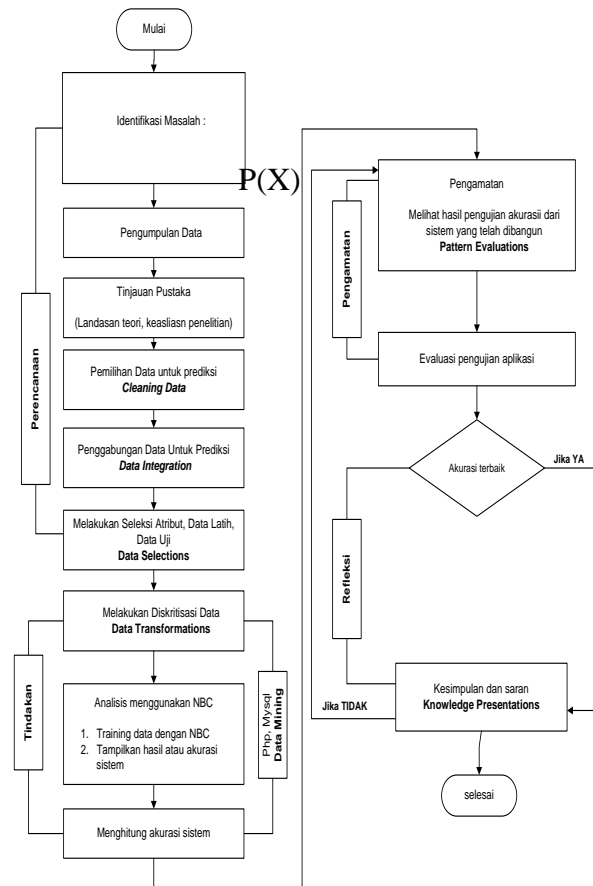
Setelah diperoleh hasil dari seluruh data pada setiap class, maka hasil akhirnya dapat menggunakan rumus:

$$P(X|Ci) = \arg \max P(Xi|Ci) * P(Ci) \dots\dots\dots(3)$$

2.2 Alur Penelitian

2.2.1 Metode Penelitian

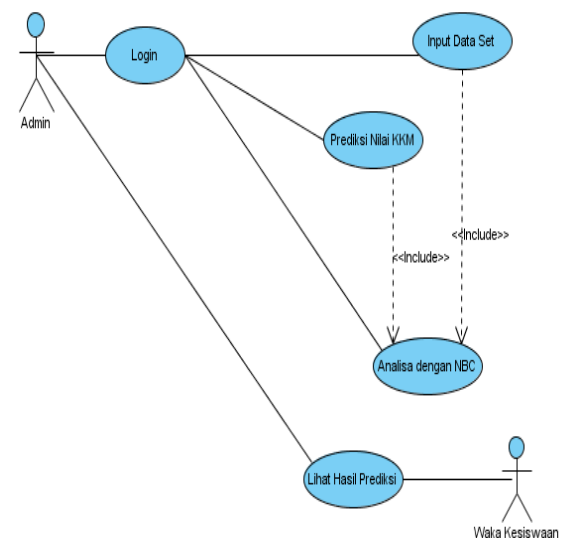
Metode yang digunakan pada penelitian ini adalah menggunakan metode Naïve Bayes karena metode ini sesuai digunakan untuk prediksi ketercapaian nilai KKM siswa. Tahapan penelitian mengikuti alur proses yang telah dibuat agar mempermudah dalam melakukan penelitian. Alur penelitian terlihat pada gambar 2.



Gambar 2. Alur Penelitian

2.2.2 Use case diagram

Use case diagram adalah suatu bentuk diagram yang akan menggambarkan fungsionalitas yang diharapkan dari sebuah sistem dilihat dari sudut pandang user. Diagram use case menggambarkan aktivitas-aktivitas yang dapat dilakukan oleh user. Untuk gambar Use case terlihat pada gambar 3.



Gambar 3. Use Case Diagram

Dari use case yang terlihat pada gambar 3 maka dapat di deskripsikan sebagai berikut

1. Deskripsi use case input data set

Tabel 1. Deskripsi Use Case Diagram Input Data Set

Nama Use Case	Input data set
Aktor	Admin
Deskripsi	Melakukan input data set
Kondisi Awal	Sistem menampilkan menu utama petugas
Aktor	Sistem
1. Aktor memilih menu import data set	2. Sistem akan menampilkan form import data set
3. Aktor memilih tombol cari data	4. Sistem akan mengarahkan ke direktori yang akan dituju.
5. Aktor memilih data file csv yang telah ditentukan.	6. Sistem akan mengarahkan file csv ke partisi didalam sistem
7. Aktor memilih tombol impot data	8. Sistem akan memasukan data dari csv kedalam database mysql

2. Deskripsi use case proses NBC

Tabel 2. Deskripsi use case diagram proses NBC

Nama Use Case	Proses NBC
Aktor	Admin
Deskripsi	Melakukan prediksi ketercapaian nilai KKM siswa dengan komposisi data training dan data testing yang telah diinputkan oleh aktor
Kondisi Awal	Melakukan training dan testing data set siswa
Aktor	Sistem
1. Aktor memilih menu analisa naïve bayes	2. Sistem melakukan proses analisa data training dan data testing menggunakan metode Naïve Bayes
	3. Apabila sistem selesai melakukan pemrosesan. Sistem akan menampilkan

	pemberitahuan bahwa pemrosesan telah selesai
--	--

3. Deskripsi use case prediksi data baru

Tabel 3. Deskripsi use case diagram prediksi data baru

Nama Use Case	Prediksi Data Baru
Aktor	
Deskripsi	Melakukan prediksi data baru
Kondisi Awal	Melakukan data baru belum diprediksi
Aktor	Sistem
1. Aktor memilih menu prediksi	2. Sistem menampilkan form prediksi baru
3. Aktor mengisi form prediksi data baru	4. Jika klik tombol Input maka sistem akan merespon dengan menampilkan keputusan hasil prediksi

2.3 Pengolahan Data

2.3.1 Pengambilan Data

Pengambilan data dalam penelitian ini penulis mendapatkan data langsung ke obyek penelitian. Pengambilan data langsung ke objek akan mendapatkan data yang valid. Data yang diperoleh adalah data siswa, data orang tua dan data pendukung lainnya. Data yang digunakan adalah data pada tahun ajaran 2015/2016 dan data tahun ajaran 2016/2017. Data yang di peroleh berjumlah 560 record. Atribut digunakan untuk masukan dalam melakukan prediksi. Adapun atribut yang digunakan dapat dilihat pada tabel 5.

Tabel 5. Atribut data siswa

Atribut	Keterangan
Nama	Nama lengkap siswa yang bersangkutan
Kelas	tingkatan kelas siswa yang bersangkutan
Nilai UTS	Nilai UTS semester ganjil siswa yang bersangkutan
Bimbingan Belajar	Keikutsertaan siswa bimbingan belajar
Pendidikan Orang Tua	Pendidikan orang tua dari siswa tersebut
Keterangan	Keterangan tuntas atau tidak tuntas siswa dalam mengikuti mata pelajaran.

2.3.2 Pembersihan Data

Pembersihan data merupakan salah satu tahap yang dilakukan dalam penelitian ini. Proses pembersihan data dari missing valuedan redundansi

data. Data disebut missing value jika pada record tertentu pada salah satu atribut tidak mempunyai nilai. Record yang missing value akan dihapus dan tidak digunakan. Pada dataset yang sama jika record berisi nilai yang sama maka record tersebut juga akan dihapus dan tidak digunakan.

Tabel 6. Data siswa

Nama	Rata UTS	Pend. Ortu	Bimbel	Absen
Adimas	6.5	SMA		50
Aan Adi		S1	Tidak Ikut	50
Agus	6.2		Tidak Ikut	50
Ahmad	7.1	S1	Ikut	50
Andrianto	6.9	S1	Tidak Ikut	50
AnggikSetyawan	7.5	SMA		50
Ardianto Fajar	7	S1	Tidak Ikut	50
Bayu Aji Saputro	6.5	SMA	Tidak Ikut	50
Bayu Prastiyo	6.9	S1	Tidak Ikut	50
Budiyanto	6.3	SD	Tidak Ikut	50
Candra Ari	5.9		Tidak Ikut	50
Danny Aji	7.7	SMA	Ikut	50
Dimas Anjasmara	7.1	S1	Tidak Ikut	50
Dyangga Farista	6.3	SMA	Tidak Ikut	50
Eko Pujiyanto	7.1	SMP	Tidak Ikut	50
Ferri Pradia	6.8	S1		50
Galang Prihhatoro	7.9	SMA	Tidak Ikut	50
Ibnu Kurniawan Aji	5.5	SMP	Ikut	50
Muhamad Ridwan	5.9	SMP	Ikut	50
Muhammad Suro	6.6	SD	Tidak Ikut	50

Data pada tabel 3 ini data dimana yang sudah melalui tahap pembersihan data atau cleaning data. Penulis mengambil data yang memiliki value atau nilai yang lengkap. Untuk data yang memiliki missing value tidak digunakan dan dihapus. Hasil dari pembersihan data ini akan berguna dalam prediksi ketercapaian nilai KKM siswa.

Hasil pembersihan data atau cleaning data dapat dilihat pada tabel 7.

Tabel 7. Hasil Pembersihan data

Nama	Rata UTS	Pend. Ortu	Bimbel	Absen
Aan Adi		S1	Tidak Ikut	50
Ahmad	7.1	S1	Ikut	50
Andrianto	6.9	S1	Tidak Ikut	50
Ardianto Fajar	7	S1	Tidak Ikut	50
Bayu Aji Saputro	6.5	SMA	Tidak Ikut	50
Bayu Prastiyo	6.9	S1	Tidak Ikut	50
Budiyanto	6.3	SD	Tidak Ikut	50
Danny Aji	7.7	SMA	Ikut	50
Dimas Anjasmara	7.1	S1	Tidak Ikut	50
Dyangga Farista	6.3	SMA	Tidak Ikut	50
Eko Pujiyanto	7.1	SMP	Tidak Ikut	50
Galang Prihhatoro	7.9	SMA	Tidak Ikut	50
Ibnu Kurniawan Aji	5.5	SMP	Ikut	50
Muhamad Ridwan	5.9	SMP	Ikut	50
Muhammad Suro	6.6	SD	Tidak Ikut	50

2.2.3 Integrasi Data

Integrasi Data adalah tahap yang dilakukan dalam penelitian ini peneliti menggabungkan semua data yang dikumpulkan dan yang akan digunakan. Peneliti melakukan penggabungan dari beberapa data yaitu data siswa dan data pendidikan orang tua yang berkaitan dengan aspek-aspek penilaian untuk ketercapaian nilai KKM siswa. Sehingga dengan melakukan penggabungan data ini diperoleh data set yang sesuai dengan perancangan sistem yang dibangun.

2.2.5 Seleksi Data Atribut

Dalam data mining dibutuhkan data latihan dan data uji. Maka dalam tahap ini penulis melakukan seleksi data atribut untuk melakukan prediksi ketercapaian nilai KKM siswa. Dalam seleksi data penulis menyeleksi data menjadi data latihan dan data uji yang akan digunakan. Data latihan merupakan suatu data yang akan digunakan untuk melakukan penelusuran pola dalam data mining atau sebagai data yang digunakan data sumber. Data uji merupakan data yang akan diujikan atau data yang akan dicoba pada sistem. Data yang sudah selesai melalui tahap seleksi data digunakan kemudian akan di bagi menjadi data latihan dan data uji dengan proporsi pembagian sebesar 75% dari data

keseluruhan sebagai data latih dan 25% dari data keseluruhan sebagai data uji.

2.2.6 Pembahasan

Dalam tahap ini peneliti melakukan pengujian dengan menggunakan data 2 tahun pelajaran dari tahun 2015/2016 sampai 2016/2017 dengan menggunakan rincian data sebagai berikut:

Data keseluruhan dalam 2 tahun ajaran 2015/2016 sampai 2016/2017 sebanyak 645 record data siswa. Dengan atribut yang telah ditentukan. Oleh peneliti dalam melakukan prediksi kecerpaian nilai KKM siswa.

Dari data yang ada beberapa data yang tidak memiliki data yang lengkap atau missing value sebanyak 15 record data siswa

Setelah data yang didapatkan yang digunakan maka tahap selanjutnya melakukan integrasi data atau penggabungan data yang akan digunakan atau data yang sudah melalui pembersihan data. Data yang akan digunakan sebanyak 630 record.

Tahapan selanjutnya penulis melakukan pembagian data atribut yaitu data latih dan data uji. Data latih 75% dan data uji 25%. Untuk pembagian data latih dan data uji terlihat pada tabel 8.

Tabel 8. Pembagian Data Latih & Data Uji

Jenis Data	Prosentase	Jumlah
Data Latih	75%	473
Data Uji	25%	157

2.2.6 Prediksi Dengan Naïve Bayes

Pada tahapan ini akan mengolah data latih dan data uji menggunakan metode naive bayes. Setelah pembagian data latih dan data uji dengan hasil 473 record sebagai data latih dan 157 record sebagai data uji yang akan menentukan nilai tercapai dan tidak tercapai nilai KKM siswa menggunakan metode naive bayes. Dari hasil prediksi yang sesuai dengan data asli sebanyak 145 record. Sehingga data yang tidak sesuai dengan data asli sebanyak 12 record. Langkah selanjutnya menguji kinerja sistem. Tujuan dari menguji kinerja sistem adalah untuk mengetahui kinerja sistem dalam melakukan prediksi ketercapaian nilai KKM siswa.

2.2.7 Pengujian Dengan Confusion Matrix

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini

melakukan perhitungan dengan 3 keluaran, yaitu: recall, precision dan accuracy

1. Recall adalah proporsi kasus positif yang diidentifikasi dengan benar. Rumus dari recall = $TP/(TP+FN)$
2. Precision adalah proporsi kasus dengan hasil positif yang benar. Rumus dari Precision = $TP/(TP+FP)$
3. Accuracy adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus Rumus dari accuracy = $(TP+TN)/(TP+TN+FP+FN)$.

Pada tahap ini melakukan pengujian terhadap sistem yang dibangun. Pada tahap ini akan menguji sistem menggunakan metode pengujian Confusion Matrix. Pengujian dilakukan agar sistem tersebut layak untuk digunakan dalam memprediksi ketercapaian nilai KKM siswa. Adapun rincian pengujian adalah terlihat pada tabel 5.

Rumus perhitungan Confusion Matrix terlihat pada tabel 5.

Tabel 5. Rumus Confusion Matrix

Confusion Matrix		Data Sebenarnya	
		TRUE	FALSE
Data Prediksi	TRUE	TP (True Positiv Corect Result)	FP (False Positiv Unexpecte d Result)
	FALSE	FN (False Negativ Missing Result)	TN (True Negative Corect Absence Of Result)

Rumus untuk perhitungan confusion matrix jika di tuliskan seperti dibawah ini:

$$\text{Nilai Precision} = TP / (TP+FP)$$

$$\text{Nilai Recall} = TP / (TP+FN)$$

$$\text{Nilai Accuracy} = TP + TN / (TP+TN+FP+FN)$$

Mendefinisikan sebagai berikut

Hasil dari pengujian dengan *confusion matrix* terlihat pada tabel 6.

Tabel 6. Data Hasil Prediksi

Dari data yang terlihat pada tabel 6 didapatkan hasil perhitungan *confusion matrix* sebagai berikut

$$\begin{aligned} \text{Nilai Precision} &= 145/(145+12) = 0,92 \\ \text{Nilai Recall} &= 145/(145+157) = 0,48 \\ \text{Nilai Accuracy} &= (145+473)/(145+473+157+12) \\ &= 0,78 \end{aligned}$$

Apabila hasil dari perhitungan *confusion matrix* di konversikan ke nilai yang lebih mudah yaitu nilai persen maka akan didapatkan hasil sebagai berikut

$$\begin{aligned} \text{Nilai Precision} &= 0,90 \times 100 = 90\% \\ \text{Nilai Recall} &= 0,47 \times 100 = 47\% \\ \text{Nilai Accuracy} &= 0,715 \times 100 = 71\% \end{aligned}$$

Setelah dilakukan perhitungan metode naïve bayes dengan data sebanyak 600 record dengan menggunakan data latih sebanyak 473 data dan data uji 157 data maka didapatkan bahwa nilai precision sebesar 92%, nilai recall sebesar 48% dan nilai accuracy sebesar 78%.

2.2.7 Studi Kasus Perhitungan Naïve Bayes

Pada tahap ini akan memberikan contoh perhitungan dalam menggunakan metode naïve bayes. Tujuan dalam tahap ini adalah agar pembaca mengetahui bagaimana perhitungan naïve bayes secara jelas dan mudah untuk dipahami.

Dalam tahap perhitungan ini akan menentukan nilai tercapai dan tidak tercapai nilai KKM siswa menggunakan metode naïve bayes. Dengan atribut yang telah di tentukan yaitu nilai rata-rata UTS gasal, pendidikan orang tua, keikutsertaan siswa dalam Bimbel dan kehadiran siswa.

Berikut adalah data baru yang akan digunakan sebagai contoh perhitungan. Berikut adalah data baru adalah

- a Rata-rata nilai UTS Gasal = 7.5
- b Pendidikan Orang Tua = SMA
- c Keikutsertaan Bimbel = Tidak Ikut
- d Kehadiran = 49

Maka untuk menentukan hasil akhir dari data baru yang diatas maka perhitungan dilakukan secara jelas dan mudah dipahami sebagai berikut:

- a. Menghitung $P(C_i)$

$$P(\text{Keterangan}=\text{'Tercapai'}) = \frac{460}{473} = 0,87$$

$$P(\text{Keterangan}=\text{'Tidak Tercapai'}) = \frac{13}{473} = 0,02$$

- Precision = Merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem
- Recall = Merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.
- Accuracy = Sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual

b. Menghitung $P(X|C_i)$ untuk setiap kelas i

$$P(\text{Rata-rata nilai UTS}=\text{'7.5'} | \text{keterangan}=\text{'Tercapai'}) = \frac{133}{135} = 0.98$$

$$P(\text{Rata-rata nilai UTS}=\text{'7.5'} | \text{keterangan}=\text{'Tidak Tercapai'}) = \frac{2}{135} = 0.02$$

$$P(\text{Pendidikan Orang Tua} = \text{'SMA'} | \text{keterangan}=\text{'Tercapai'}) = \frac{343}{360} = 0,95$$

$$P(\text{Pendidikan Orang Tua} = \text{'SMA'} | \text{keterangan} = \text{'Tidak Tercapai'}) = \frac{17}{360} = 0,04$$

$$P(\text{Keikutsertaan Bimbel} = \text{'Tidak Ikut'} | \text{Keterangan}=\text{'Tercapai'}) = \frac{463}{473} = 0.97$$

$$P(\text{Tidak Melaksanakan Piket} = \text{'Tidak Ikut'} | \text{keterangan}=\text{'Tidak Tercapai'}) = \frac{13}{473} = 0.02$$

$$P(\text{Kehadiran} = \text{'49'} | \text{keterangan} = \text{'Tercapai'}) = \frac{46}{50} = 0.92$$

$$P(\text{Kehadiran} = \text{'49'} | \text{keterangan} = \text{'Tidak Disiplin'}) = \frac{4}{50} = 0.08$$

Confusion Matrix		Data Sebenarnya	
		TRUE	FALSE
Data Prediksi	TRUE	145	12
	FALSE	157	473

b. Menghitung $P(X|C_i)$ untuk setiap kelas

$$P(X| \text{level} = \text{'Tercapai'}) = 0.98 \times 0.95 \times 0.97 \times 0.92 = 0.8308244$$

$$P(X| \text{level} = \text{'Tidak Tercapai'}) = 0.02 \times 0.04 \times 0.07 \times 0.08 = 0.000672$$

c. Menghitung $P(X|C_i) \cdot P(C_i)$:

$$P(X| \text{keterangan} = \text{'Tercapai'}) \cdot P(\text{level} = \text{'Tercapai'}) = 0.87 \times 0.8308244 = 0.722817228$$

$$P(X| \text{keterangan} = \text{'Tidak Tercapai'}) \cdot P(\text{level} = \text{'Tidak Tercapai'}) = 0.02 \times 0.000672 = 0.0001344$$

Hasil dari perhitungan dengan data rata-rata nilai UTS, pendidikan Orang Tua, Keikutsertaan Bimbel dan Kehadirandengan menggunakan algoritma Naïve Bayes diatas didapatkan hasil keterangan = "Tercapai"

3. PENUTUP

3.1 Kesimpulan

Dari penelitian yang telah dilakukan maka dapat diambil kesimpulan bahwa dengan menggunakan naïve bayes dapat melakukan prediksi terhadap ketercapaian nilai KKM siswa. Dengan menggunakan naïve bayes dengan data sebanyak 473 data record untuk data latih dan 157record untuk data uji. Dari pengujian didapatkan nilai accuracy sebesar 78%, precion sebesar 92% dan recall sebesar 48%.

3.2 Saran

Agar penelitian berikutnya yang akan melakukan penelitian yang sama, agar mendapatkan hasil yang lebih baik sebagai berikut

1. Pada penelitian ini dalam hal pengukuran kinerja sistem menggunakan metode Confusion Matrix. Sebagai perbandingan dalam hal pengujian kinerja sistem prediksi dapat menggunakan metode lain seperti K-Fold Cross Validation.
2. Untuk mengetahui tingkat kinerja sistem yang berbeda, proses pelatihan dan pengujian data untuk klasifikasi kedisiplinan siswa menggunakan metode selain menggunakan algoritma Naïve Bayes Classifictaion.

DAFTAR PUSTAKA

- [1]. Mardapi, Dj. dan Ghofur, A, (2013). *Pedoman Umum Pengembangan Penilaian; Kurikulum Berbasis Kompetensi SMA*. Jakarta: Direktorat Pendidikan Menengah Umum
- [2]. Peraturan Menteri Pendidikan Nasional Nomor 14 tahun 2008 tentang *Organisasi dan Tata Kerja Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah Departemen Pendidikan Nasional*. Jakarta: Direktorat Jenderal Manajemen Pendidikan Dasar dan Menengah.
- [3]. Praningki Tutus, Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN. *Cite c Journal*, Vol. 4, No. 2, Februari 2017 – April 2017
- [4]. Deny Wiria Nugraha dkk. 2017. Prediksi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier . *SemanTIK*, Vol.3, No.2, Jul-Des 2017, pp. 13-22 ISSN : 2502-8928
- [5]. Husni Naparin.2016. Prediksi Peminatan Siswa Sma Menggunakan Metode Naive Bayes. *SYSTEMIC* Vol. 02, No. 01 ISSN 2460-8092.
- [6]. Jayanto, Arief. 2013. *Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa*. *Jurnal Teknologi Informasi DINAMIK* Volume 18, No.1, Januari 2013 : 09-16 ISSN : 0854-9524.
- [7]. Mandasari, Vina, dkk. 2011. *Analisis Kepuasan Konsumen Terhadap Restoran Cepat Saji Melalui Pendekatan Data Mining*. *Jurnal Generic* Vol.6 No.1 (Januari 2011)
- [8]. Ridwan, Mujib, dkk. 2013. *Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. *Jurnal EECCIS* Vol.7, No. 1, Juni 2013
- [9]. Untari, Dwi, "Data Mining untuk Analisa Prediksi Mahasiswa Non-Aktif menggunakan Metode Decision Tree C45", Universitas Dian Nuswantoro: Semarang, 2014.
- [10]. G. Subbalakshmi, K. Ramesh and M. C. Rao, "Decision Support in Heart Disease Prediction System Using Naive Bayes," *Indian Journal of Computer Science and Engineering*, vol. 2, no. No. 2, ISSN: 0976-5166, pp. 170-176, April-June 2011.
- [11]. M. Sarosa, H. Suryono and M. Ridwan, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Jurnal EECCIS*, vol. 7, no. No. 1, pp. 59-64, 2013.
- [12]. R. R. Patil, "Heart Disease Prediction System using Naive Bayes and Jelinek-Mercer Smoothing," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 5, ISSN (Online): 2278-1021, ISSN (Print): 2319-5940, pp. 6787-6789, 2014.