

KLASIFIKASI METODE NAIVE BAYES DALAM DATA MINING UNTUK MENENTUKAN KONSENTRASI SISWA (STUDI KASUS DI MAS PAB 2 MEDAN)

Alfa Saleh

*Teknik Informatika STMIK Potensi Utama
Jl K.L. Yos Sudarso KM 6.5 No.3-A, Tanjung Mulia, Medan
Email : alfasoleh1@gmail.com*

Abstrak

Determining the concentration of students is essential in developing students' skills in terms of science and also to focus the student in a particular area of concentration, this is done so that every student can learn more in the subjects in accordance with a predetermined concentration for each student. To determine the concentration of these students, in order to be applied the Naive Bayes method in classifying the concentration of students based on the data which are trained to determine the probability of each of the criteria that are used in nominal criteria as well as numerical criteria and from the probability values can be predicted concentrations appropriate for the student. Based on the results of research conducted to determine the concentration of students with Naive Bayes method, the accuracy of the classification results obtained in accordance with the data tested by 90.8333%, where the data of 120 students were tested, there were as many as 109 students of data successfully classified and predicted correctly.

Keywords: Data Mining, Naive Bayes, Concentration of Students.

1. Pendahuluan

Latar Belakang

Perkembangan ilmu pengetahuan dan teknologi telah membawa perubahan di hampir semua aspek kehidupan manusia di mana berbagai permasalahan hanya dapat dipecahkan kecuali dengan upaya penguasaan dan peningkatan ilmu pengetahuan dan teknologi. Penentuan konsentrasi bagi siswa SMA/MA sederajat merupakan proses untuk memfokuskan siswa dalam bidang konsentrasi tertentu, hal ini dilakukan agar setiap siswa dapat mempelajari lebih dalam mata pelajaran – mata pelajaran yang sesuai dengan konsentrasi yang telah ditentukan untuk setiap siswa. Yang menjadi masalah ialah penulis ingin mendapatkan informasi dari histori nilai akademik siswa kelas 11 dan 12 Aliyah Swasta PAB 2 Helvetia sehingga setiap siswa kelas 10 dapat diklasifikasikan dalam kategori

konsentrasi yang sesuai berdasarkan nilai yang mereka peroleh.

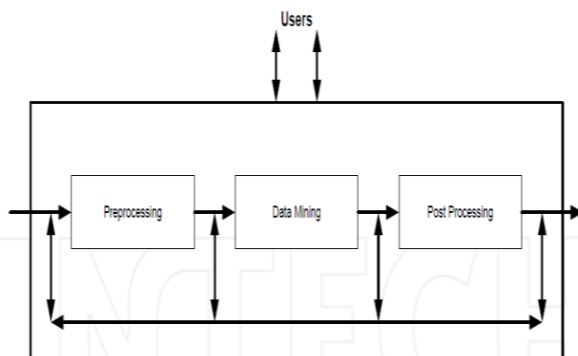
Hal ini juga kiranya telah menjadi bahan penelitian untuk kategori Sistem pendukung keputusan dalam menentukan jurusan di SMA yang sesuai dengan kemampuan siswa dengan dasar yang digunakan dalam penentuan jurusan adalah nilai semester, nilai potensi dan nilai pilihan siswa (Tresna dan Dimara,2011). Penelitian lainnya seputar pemilihan jurusan juga penulis temukan, di mana dalam proses pemilihan jurusan ini digunakanlah metode 360 derajat (Stefanie dan Natalia,2013). Untuk melakukan perhitungan yang ada dalam penelitian ini maka digunakan teknik pengklasifikasian dengan metode *Naive Bayes*. Metode *Naive Bayes* juga digunakan dalam memprediksi penyakit Dermatologi yang diabaikan tapi bahkan dapat menyebabkan kematian di mana metode *Naive Bayes* digunakan untuk mengenal pola data untuk mengungkap kemungkinan penyakit dermatologi (Manjusa et al, 2014). Metode

Naive Bayes juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi (Ting et al, 2011).

Knowledge Discovery in Database (KDD)

Adapun beberapa definisi dari *Knowledge Discovery in Database (KDD)* sebagai berikut: *Knowledge Discovery in Database (KDD)* merupakan sebuah proses dengan beberapa tingkatan, tidak sepele, interaktif dan berulang untuk identifikasi pola yang dipahami, sah, baru dan secara potensial berguna mulai dari sekumpulan data yang sangat besar (Silvia dan Germano, 2009). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Mujib dkk, 2013), selain itu KDD juga didefinisikan sebagai proses yang terdiri dari urutan interaktif mulai dari pembersihan data, integrasi data, penyeleksian data, transformasi data, *Data Mining*, evaluasi pola dan presentasi pengetahuan. Dari beberapa definisi di atas bisa ditarik kesimpulan bahwa *Knowledge Discovery in Database (KDD)* merupakan sebuah kegiatan yang bertahap guna menggali informasi dari sekumpulan data yang sangat besar.

KDD dikarakteristikan sebagai sebuah proses yang terdiri dari beberapa tahap operasional : *Preprocessing*, *Data Mining* dan *Post Processing*. Pada gambar 1 akan menunjukkan tahapan karakteristik KDD (Silvia dan Germano, 2009).



Gambar 1. Tahap Knowledge Discovery in Database (KDD)

a. Tahap *Preprocessing*.

Pada tahap *Preprocessing* memahami fungsi – fungsi yang berhubungan dengan penerimaan, organisasi dan untuk perawatan data, tahap ini sebagai tujuan penyusunan data untuk mengikuti tahap dalam *Data Mining*.

b. Tahap *Data Mining*.

Pada tahap *Data Mining* mendefinisikan teknik – teknik dan algoritma – algoritma yang digunakan oleh masalah di dalam pertanyaan, sebagai contoh dari teknik yang bisa digunakan dalam tahap ini seperti *Neural Network*, *Rough Set*, *Genetic Algorithms*, Statistik dan Probabilistik. Pemilihan teknik tergantung, dalam banyak hal, pada jenis tugas yang harus dikembangkan.

c. Tahap *Post Processing*.

Dalam tahap *Post Processing* pemeliharaan dari pengetahuan diperoleh selama tahap *Data Mining*. tahap ini tidak selalu dibutuhkan namun hal itu memungkinkan validasi kegunaan dari pengetahuan yang ditemukan.

Data Mining

Data Mining merupakan proses pengekstraksian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data (Shyara dan Saroj, 2013). *Data Mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran, atau bahkan keduanya (Angga dan Riani, 2012). Definisi lain mengatakan *Data Mining* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar (Subekti, 2011). Dari beberapa definisi di atas dapat ditarik kesimpulan bahwa *Data Mining* merupakan proses ataupun kegiatan untuk mengumpulkan data yang berukuran besar kemudian mengekstraksi data tersebut menjadi informasi – informasi yang nantinya dapat digunakan.

Tahap-tahap *Data Mining*

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantara *knowledge base*.

Tahap-tahap *Data Mining* adalah sebagai berikut (Mujib dkk, 2013):

a. Pembersihan data (*Data Cleaning*)

Pembersihan data merupakan proses menghilangkan-kan *noise* dan data yang tidak konsisten atau data tidak relevan.

- b. Integrasi data (*Data Integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
- c. Seleksi data (*Data Selection*)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.
- d. Transformasi data (*Data Transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *Data Mining*.
- e. Proses *Mining*
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *Data Mining*.
- f. Evaluasi pola (*Pattern Evaluation*)
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.
- g. Presentasi pengetahuan (*Knowledge Presentation*)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

Metode Naive Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Patil dan Sherekar, 2013). Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya (Bustami, 2013)

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Mujib dkk, 2013). Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja

jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan (Pattekari dan Parveen, 2012).

Persamaan Metode Naive Bayes

Persamaan dari teorema *Bayes* adalah (Bustami, 2013) :

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

Di mana :

- X :Data dengan *class* yang belum diketahui
- H : Hipotesis data merupakan suatu *class* spesifik
- $P(H|X)$:Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)
- $P(H)$: Probabilitas hipotesis H (prior probabilitas)
- $P(X|H)$:Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut :

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)}$$

Di mana Variabel C merepresentasikan kelas, sementara variabel $F1 \dots Fn$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{prior \times likelihood}{evidence}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan.

Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= (C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing masing petunjuk ($F_1, F_2 \dots F_n$) saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C)$$

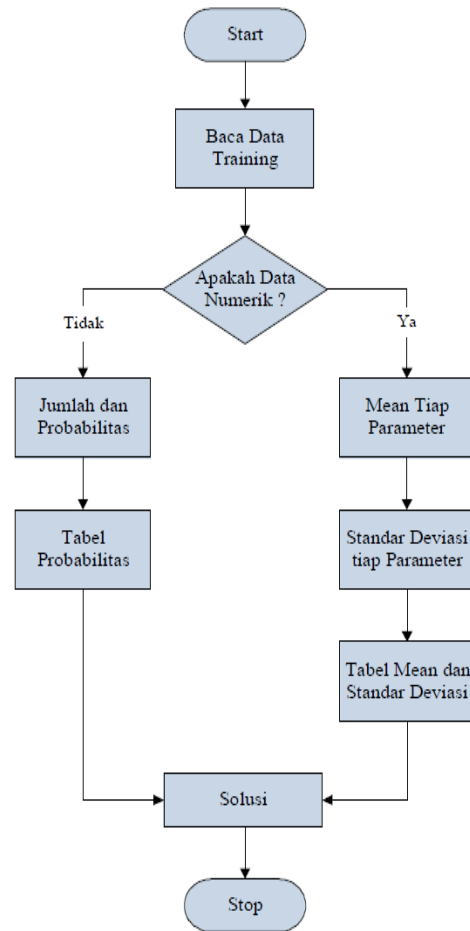
Persamaan di atas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* :

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Di mana :

- P : Peluang
- X_i : Atribut ke i
- x_i : Nilai atribut ke i
- Y : Kelas yang dicari
- y_i : Sub kelas Y yang dicari
- μ : *mean*, menyatakan rata - rata dari seluruh atribut
- σ : *Deviasi standar*, menyatakan varian dari seluruh atribut.

Alur dari metode *Naive Bayes* dapat dilihat pada gambar 2 sebagai berikut :



Gambar 2. Alur Metode Naive Bayes

1. Baca *data training*
2. Hitung Jumlah dan probabilitas, namun apabila data numerik maka :
 - a. Cari nilai *mean* dan standar deviasi dari masing masing parameter yang merupakan data numerik. Adapun persamaan yang digunakan untuk menghitung nilai rata - rata hitung (*mean*) dapat dilihat sebagai berikut :

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

di mana :

- μ : rata - rata hitung (*mean*)
- x_i : nilai sample ke - i
- n : jumlah sampel

dan persamaan untuk menghitung nilai simpangan baku (standar deviasi) dapat dilihat sebagai berikut :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}}$$

di mana :

σ : standar deviasi

x_i : nilai x ke -i

μ : rata-rata hitung

n : jumlah sampel

- b. Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
3. Mendapatkan nilai dalam tabel *mean*, standart deviasi dan probabilitas.
4. solusi kemudian dihasilkan.

2. Pembahasan

Penerapan Metode Naive Bayes

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam metode Naive Bayes data String yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinyu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria dengan nilai data numerik. Adapun penerapan metode Naive Bayes sebagai berikut.

a. Baca *Data Training*

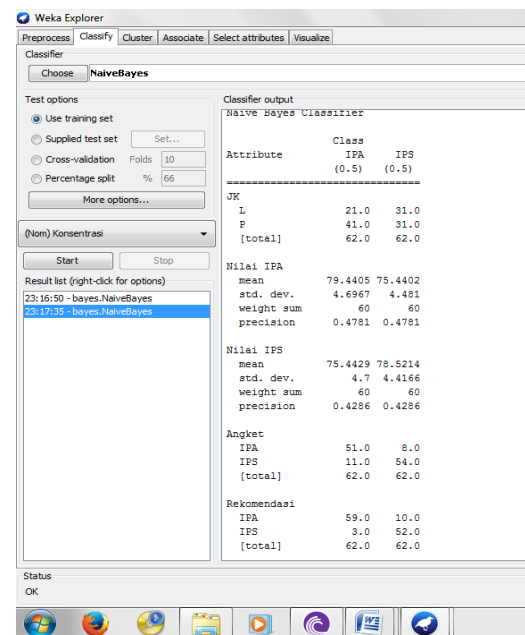
Untuk menentukan data yang nantinya akan dianalisis dengan metode *Naive Bayes* maka langkah pertama yang dilakukan adalah membaca data latih. Adapun data latih yang digunakan dapat dilihat pada tabel 1 berikut :

Tabel 1. Data Training

NO	Jenis Kelamin	Nilai IPA	Nilai IPS	Angket	Rekomendasi	Konsentrasi
1	L	75,8	73,2	IPA	IPA	IPA
2	L	80,5	74,4	IPS	IPA	IPS
3	L	75	85	IPS	IPA	IPS
4	P	78	80	IPA	IPS	IPS
5	P	81,6	80,8	IPS	IPS	IPS
6	P	78	70,5	IPS	IPA	IPA
7	L	75	73	IPA	IPA	IPA
8	P	85	75	IPA	IPA	IPA
9	P	73,3	75	IPS	IPS	IPS
10	L	81,6	80,1	IPA	IPA	IPA
11	P	81	74,5	IPA	IPA	IPA
12	P	85	70	IPA	IPA	IPA
13	P	85,3	84	IPA	IPA	IPA
14	P	83,5	77,4	IPA	IPA	IPA
15	P	83,75	82,6	IPS	IPS	IPS
16	L	80	80	IPS	IPS	IPS
17	P	80,1	78,8	IPS	IPS	IPS
18	L	79,75	79,1	IPS	IPS	IPS
119	P	85	81,5	IPA	IPA	IPA
120	P	73,3	80	IPS	IPS	IPS

b. Kriteria dan Probabilitas

Adapun nilai probabilitas setiap kriteria didapatkan dari data latih pada tabel 1. Adapun nilai probabilitas setiap kriteria dapat dilihat pada pengujian dengan tools Weka di bawah ini.



Gambar 3. Nilai Probabilitas Setiap Kriteria

Dari nilai probabilitas di atas akan diuji data sebanyak 120 data siswa dan diselesaikan dengan penyelesaian sebagai berikut :

Tabel 2. Sampel Data Uji Pertama

No	Nama Siswa	Jenis Kelamin	Nilai IPA	Nilai IPS	Angket	Rekomendasi	Konsentrasi
1	Abdul Rahmad	L	75,8	73,2	IPA	IPA	IPA

Sebelum menentukan probabilitas setiap kriteria, terlebih dahulu dilakukan pemisahan

antara *likelihood* konsentrasi IPA dan *likelihood* konsentrasi IPS.

Likelihood Konsentrasi IPA :

- Jenis Kelamin : L (Laki - Laki)
Berdasarkan nilai probabilitas pada tabel 4.3, didapat nilai probabilitas untuk kriteria Jenis Kelamin = Perempuan dengan konsentrasi IPA adalah 20/60.
 $P(\text{Jenis Kelamin} = P \mid \text{Konsentrasi} = \text{IPA}) = 20 / 60 = 1/3$
- Nilai IPA = 75,8
Untuk menghitung probabilitas kriteria nilai IPA akan digunakan rumus *Densitas Gauss* pada persamaan (7) dikarenakan data pada nilai IPA merupakan data yang bernilai kontinu sebagai berikut :
 $P(\text{Nilai IPA} = 75,8 \mid \text{Konsentrasi} = \text{IPA})$

$$= \frac{1}{\sqrt{2 \times 3.14 \times 4.6913}} e^{-\frac{(75,8-79.4205)^2}{2 \times (4.6913)^2}}$$

$$= \frac{1}{5,4278} e^{-\frac{13.108}{44.0166}}$$

$$= 0.1842 \times 0.7425$$

$$= 0.1368$$

Dari perhitungan di atas digunakan nilai rata – rata hitung dan simpangan baku pada tabel 4.6. sehingga dihasilkan nilai *Densitas Gauss* sebagai nilai probabilitas kriteria nilai IPA.

- Nilai IPS = 73,2
Untuk menghitung probabilitas kriteria nilai IPS akan digunakan rumus *Densitas Gauss* pada persamaan (7) dikarenakan data pada nilai IPS merupakan data yang bernilai kontinu sebagai berikut :

$$P(\text{Nilai IPS} = 73,2 \mid \text{Konsentrasi} = \text{IPA})$$

$$= \frac{1}{\sqrt{2 \times 3.14 \times 4.7080}} e^{-\frac{(73,2-75.4697)^2}{2 \times (4.7080)^2}}$$

$$= \frac{1}{5.4375} e^{-\frac{5.1515}{44.3305}}$$

$$= 0.1839 \times 0.8903$$

$$= 0.1637$$

Dari perhitungan di atas digunakan nilai rata – rata hitung dan simpangan

baku pada tabel 4.9. sehingga dihasilkan nilai *Densitas Gauss* sebagai nilai probabilitas kriteria nilai IPS.

- Angket = IPA
Berdasarkan nilai probabilitas pada tabel 4.10, didapat nilai probabilitas untuk kriteria Angket = IPA dengan konsentrasi IPA adalah 50/60.
 $P(\text{Angket} = \text{IPA} \mid \text{Konsentrasi} = \text{IPA}) = 50 / 60 = 5/6$
- Rekomendasi = IPA
Berdasarkan nilai probabilitas pada tabel 4.11, didapat nilai probabilitas untuk kriteria Rekomendasi = IPS dengan konsentrasi IPA adalah 2/60.
 $P(\text{Rekomendasi} = \text{IPS} \mid \text{Konsentrasi} = \text{IPA}) = 58/60$
- Konsentrasi = IPA
Berdasarkan nilai probabilitas pada tabel 4.12, didapat nilai probabilitas konsentrasi IPA adalah 60/120.
 $P(\text{Konsentrasi} = \text{IPA}) = 60/120 = 1/2$

Likelihood Konsentrasi IPS :

- Jenis Kelamin : L (Laki - Laki)
Berdasarkan nilai probabilitas pada tabel 4.3, didapat nilai probabilitas untuk kriteria Jenis Kelamin = Perempuan dengan konsentrasi IPS adalah 30/60.
 $P(\text{Jenis Kelamin} = P \mid \text{Konsentrasi} = \text{IPS}) = 30 / 60 = 3/6 = 1/2$
- Nilai IPA = 75,8
Untuk menghitung probabilitas kriteria nilai IPA akan digunakan rumus *Densitas Gauss* pada persamaan (7) dikarenakan data pada nilai IPA merupakan data yang bernilai kontinu sebagai berikut :

$$= \frac{1}{\sqrt{2 \times 3.14 \times 4.4702}} e^{-\frac{(75.8-75.4525)^2}{2 \times (4.4702)^2}}$$

$$= \frac{1}{5,2984} e^{-\frac{0.1208}{39.9654}}$$

$$= 0.1887 \times 0.9970$$

$$= 0.1882$$

Dari perhitungan di atas digunakan nilai rata – rata hitung dan simpangan baku pada tabel 4.6. sehingga dihasilkan nilai *Densitas Gauss* sebagai nilai probabilitas kriteria nilai IPA.

- Nilai IPS = 73,2
Untuk menghitung probabilitas kriteria nilai IPS akan digunakan rumus *Densitas Gauss* pada persamaan (7) dikarenakan data pada nilai IPS merupakan data yang bernilai kontinu sebagai berikut :
 $P(\text{Nilai IPS} = 73,2 \mid \text{Konsentrasi} = \text{IPS})$

$$= \frac{1}{\sqrt{2 \times 3.14 \times 4.4666}} e^{-\frac{(73,2-78.5233)^2}{2 \times (4.4666)^2}}$$

$$= \frac{1}{5.2962} e^{-\frac{28.3375}{39.9010}}$$

$$= 0.1888 \times 0.4915$$

$$= 0.0928$$

Dari perhitungan di atas digunakan nilai rata – rata hitung dan simpangan baku pada tabel 4.9. sehingga dihasilkan nilai *Densitas Gauss* sebagai nilai probabilitas kriteria nilai IPS.

- Angket = IPA
Berdasarkan nilai probabilitas pada tabel 4.10, didapat nilai probabilitas untuk kriteria Angket = IPA dengan konsentrasi IPS adalah 7/60.
 $P(\text{Angket} = \text{IPA} \mid \text{Konsentrasi} = \text{IPS}) = 7 / 60$
- Rekomendasi = IPA
Berdasarkan nilai probabilitas pada tabel 4.11, didapat nilai probabilitas untuk kriteria Rekomendasi = IPS dengan konsentrasi IPS adalah 51/60.
 $P(\text{Rekomendasi} = \text{IPA} \mid \text{Konsentrasi} = \text{IPS}) = 9/60$
- Konsentrasi = IPS
Berdasarkan nilai probabilitas pada tabel 4.12, didapat nilai probabilitas konsentrasi IPS adalah 60/120.
 $P(\text{Konsentrasi} = \text{IPS}) = 60/120 = 1/2$

Setelah diperoleh nilai probabilitas setiap kriteria, kemudian dilakukan perkalian nilai kriteria sesuai dengan *likelihood* konsentrasi sebagai berikut :

Likelihood Konsentrasi IPA
 $= (1/3) \times (0.1368) \times (0.1637) \times (5/6) \times (58/60) \times (1/2)$
 $= 0.006013$
Likelihood Konsentrasi IPS
 $= (1/2) \times (0.1882) \times (0.0928) \times (7/60) \times (9/60) \times (1/2)$
 $= 0.000153$

Untuk menghasilkan nilai Probabilitas maka dilakukan normalisasi terhadap *likelihood* tersebut sehingga jumlah nilai yang diperoleh sama dengan 1.

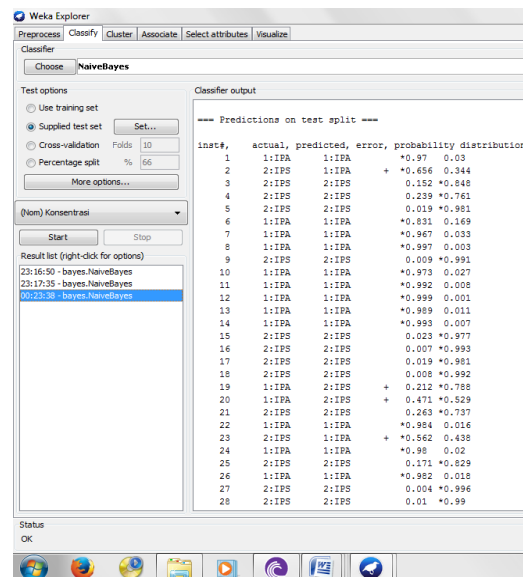
Probabilitas Konsentrasi IPA
 $= \frac{0.006013}{0.006013+0.000153} = 0,97$
 Probabilitas Konsentrasi IPS
 $= \frac{0.000153}{0.006013+0.000153} = 0,03$

Dari nilai probabilitas yang diperoleh maka dapat disimpulkan dari data siswa yang diuji di atas diketahui probabilitas siswa tersebut masuk ke konsentrasi IPS lebih besar dari pada probabilitas masuk ke konsentrasi IPA. Adapun nilai persentase yang dihasilkan dari nilai probabilitas kedua konsentrasi tersebut dapat dilihat pada tabel 3 di bawah ini :

Tabel 3 Hasil Persentase Probabilitas Konsentrasi

Konsentrasi	Probabilitas	Presentasi
IPA	0,97	97,00%
IPS	0,03	3,00%

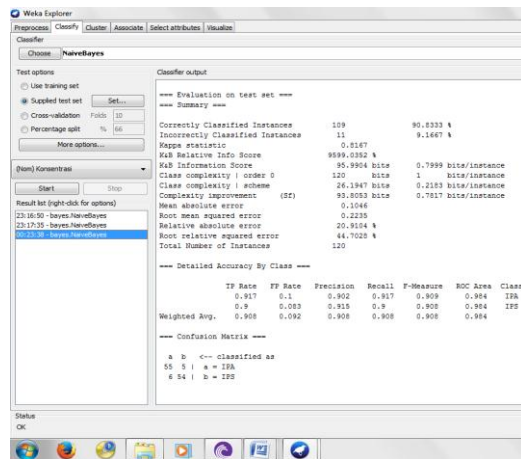
Hasil uji coba dengan tools Weka untuk melihat seberapa akurat klasifikasi metode *Naive Bayes* dalam menentukan konsentrasi siswa dapat dilihat pada gambar 4 sebagai berikut.



Gambar 4. Hasil Klasifikasi Naive Bayes

Berdasarkan gambar 4 di atas, dapat diketahui dari 120 data siswa yang diuji dengan 6 buah kriteria sebagai pendukung pengklasifikasian di mana setiap kriteria memiliki nilai probabilitas tersendiri untuk setiap *class*-nya terdapat 109

data siswa yang berhasil diklasifikasikan dengan benar sementara sebanyak 11 data siswa tidak berhasil diklasifikasikan dengan benar. Dengan begitu keakuratan metode Naive Bayes dalam mengklasifikasikan 120 data siswa adalah sebesar 90,8333 %. Persentasi ini dapat dilihat pada gambar 5 berikut :



Gambar 5. Presentase Keakuratan Metode Naive Bayes

dilihat persentase untuk *Correctly Classified Instance* adalah sebesar 90,8333 % sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 9,1667 %. Dengan *Confusion Matrix* untuk *class* IPA sebanyak 55 data siswa yang berhasil diklasifikasikan dan sebanyak 6 data siswa yang tidak berhasil diklasifikasikan. Sedangkan untuk *class* IPS sebanyak 54 data siswa yang diklasifikasikan dengan benar dan 5 data siswa yang tidak berhasil diklasifikasikan.

3. Kesimpulan

Berdasarkan penelitian tentang menentukan konsentrasi siswa dapat ditarik beberapa kesimpulan sebagai berikut :

1. Berdasarkan data akademik siswa yang diperoleh, proses *Data Mining* membantu dalam penerapan metode *Naive Bayes* dalam mendapatkan informasi dari hasil klasifikasi konsentrasi siswa.
2. Metode *Naive Bayes* memanfaatkan data *training* untuk menghasilkan probabilitas setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi konsentrasi siswa berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes* itu sendiri.

3. Berdasarkan data akademik siswa yang dijadikan data *training*, metode *Naive Bayes* berhasil mengklasifikasikan 109 data siswa dari 120 data yang diuji. Sehingga dengan demikian metode *Naive Bayes* ini berhasil memprediksi konsentrasi siswa dengan persentase keakuratan sebesar 90,8333 %.

Daftar Pustaka

Journal:

Tresna yudha Prawira, Dimara kusuma Hakim, (2011). *Sistem Pendukung Keputusan berbasis Web untuk Menentukan Penjurusan (IPA/IPS/Bahasa) pada SMA Islam Bumiayu*, JUITA ISSN : 2086-9398 Vol. 1 Nomor 4, November 2011.

Stefanie G.N.L. Worang, Natalia K. Toera, (2013), *Penerapan Metode 360 Derajat dalam Sistem Pendukung Keputusan penentuan Jurusan SMA Berbasis*, *Seminar nasional Aplikasi Teknologi Informasi (SNATI) 2013,15 Juni 2013, Yogyakarta*.

Shyara taruna R, Saroj Hiranwal, (2013). *Enhanced Naive Bayes Algorithm for Intrusion Detection in Data Mining*, *International Journal of Computer Science and information Technologies*, Vol. 4, 2013.

Manjusha K.K, et al, (2014). *Prediction of Different Dermatological Conditions Using Naive Bayesian Classification*, *International Journal of Advanced Research in Computer Science and Software Engineering*, 2014.

S.L. Ting , et al, (2011). *Is Naive Bayes a Good Classifier for Document Classification ?*, *International journal of Software Engineering and Its Applications*, Vol. 5, 3, July, 2011.

Mujib Ridwan, dkk, (2013), *Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*, *Jurnal EECCIS* Vol. 7, No. 1, Juni 2013.

Angga Ginanjar Mabur, Riani Lubis, (2012). *Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit*, *Jurnal Komputer dan Informatika (KOMPUTA)* Edisi 1, Vol. 1, Maret 2012.

Surbekti Mujiasih, (2011). *Pemanfaatan Data Mining Untuk Prakiraan Cuaca*, Jurnal Meteorologi dan Geofisika, Volume 12, Nomor 2, September 2011.

Tina R. Patil, S.S. Sherekar, (2013). *Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification*, International Journal of Computer Science and Applications, Vol. 6, No. 2, April 2013.

Bustami, (2013). *Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*, TECHSI : Jurnal Penelitian Teknik Informatika.

Shadab Adam Pattekari, Asma Parveen, (2012), *Prediction System for Heart Disease Using Naive Bayes*, International Journal of Advanced Computer and Mathematical Sciences, ISSN 2230-9624, Vol. 3, Issue 3, 2012.

Book:

Silvia Rissino, Germano Lambert-Torres, (2009). *Rough Set Theory-Fundamental Concepts, Principals, Data extraction, and Applications, Data Mining and Knowledge Discovery in Real Life Application, February 2009 I-Tech, Vienna, Austria.*

Jamie MacLennan, ZhaoHui Tang, Bogdan Crivat, (2009), *Data Mining with Microsoft SQL Server 2008*, Wiley Publishing, Inc, Indianapolis, Indiana.