

IMPLEMENTASI METODE BAYESIAN DALAM PENJURUSAN DI SMA BRUDERAN PURWOREJO

STUDI KASUS: SMA BRUDERAN PURWOREJO

Sendy Winanta
Yetli Oslan, Gunawan Santoso

Abstrak

Penjurusan siswa kelas X SMA yang akan naik ke kelas XI bertujuan mengarahkan peserta didik agar lebih fokus mengembangkan kemampuan dan minat yang dimiliki. Strategi ini diharapkan dapat memaksimalkan potensi, bakat atau talenta individu, sehingga juga akan memaksimalkan nilai akademisnya. Penentuan jurusan akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi nantinya. Jurusan yang tidak tepat bisa sangat merugikan siswa dan masa depannya.

Dalam skripsi ini dibuat sebuah sistem yang akan membantu dalam memberikan saran jurusan bagi siswa berdasarkan nilai yang dimiliki siswa dan dicocokkan dengan data-data nilai angkatan lama dengan menggunakan metode Bayesian Classification. Sistem dibuat dengan menggunakan program aplikasi Microsoft Visual FoxPro 9.0. Range nilai yang digunakan ada 2 yaitu range manual dan range dari box plot.

Keakuratan tertinggi dari hasil proses penjurusan dengan menggunakan range yang ditentukan secara manual berada pada range dengan interval 20 dan 25. Pada interval 20, angkatan 2009 mendapatkan sebesar 59 %, 2010 sebesar 66 %, 2011 sebesar 61 %. Sedangkan pada Interval 25, angkatan 2009 mendapatkan sebesar 62 %, 2010 sebesar 58 %, 2011 sebesar 54 %. Hasil proses penjurusan dengan menggunakan range yang dapatkan dari proses Box Plot rata-rata memiliki tingkat keakuratan lebih tinggi dibanding dengan range yang ditentukan secara manual. Pada range Box Plot, tingkat keakuratan angkatan 2009 adalah 62%, tingkat keakuratan angkatan 2010 adalah 63% dan angkatan 2011 adalah 66%. Untuk mengatasi siswa yang tidak mendapatkan saran jurusan karena terdapat atribut di tiap kelas yang probabilitasnya 0 dilakukan dengan cara mengambil nilai terendah pada mata pelajaran yang bersangkutan. Berdasarkan proses yang telah dilakukan, hasil penjurusan dengan cara tersebut rata-rata memiliki keakuratan lebih tinggi karena seluruh siswa mendapatkan saran jurusan tanpa terkecuali. Keakuratan tertinggi juga terdapat pada interval 20 dan 25. Pada interval 20, angkatan 2009 mendapatkan sebesar 58 %, 2010 sebesar 67 %, 2011 sebesar 58 %. Sedangkan pada Interval 25, angkatan 2009 mendapatkan sebesar 64 %, 2010 sebesar 64 %, 2011 sebesar 53 %

Kata Kunci : Penjurusan SMA, Bayesian, Klasifikasi, Box Plot

1. Pendahuluan

Penjurusan siswa kelas X SMA yang akan naik ke kelas XI bertujuan untuk mengarahkan peserta didik agar dapat lebih fokus mengembangkan kemampuan dan minat yang dimiliki. Strategi ini diharapkan dapat memaksimalkan potensi, bakat atau talenta individu, sehingga juga akan memaksimalkan nilai akademisnya. Penentuan jurusan akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi nantinya. Jurusan yang tidak tepat bisa sangat merugikan siswa dan masa depannya.

Pengambilan keputusan penjurusan oleh sekolah dipertimbangkan dengan melihat beberapa faktor, antara lain nilai akademis, hasil test IQ, minat siswa, saran orang tua, dan lain sebagainya. Pihak sekolah yang dalam hal ini adalah guru BP dituntut sebijaksana mungkin dalam memutuskan jurusan yang tepat. Menentukan jurusan dengan memperhatikan banyak faktor yang kompleks dan dilakukan secara manual mempunyai banyak kelemahan. Data yang banyak cukup menyita waktu dan menguras tenaga, serta menuntut ketelitian ekstra. Selain itu, cara seperti ini memungkinkan terjadinya kesalahan baik yang manusiawi maupun yang disengaja. Penilaian subjektif dengan memberikan keistimewaan tersendiri kepada pihak tertentu acap kali menimbulkan ketidakadilan.

Perkembangan teknologi yang pesat diiringi dengan kebutuhan akan informasi yang cepat guna meningkatkan efektifitas pelayanan dan daya saing memungkinkan dibangunnya sebuah sistem informasi untuk membantu mengambil keputusan penjurusan dengan menerapkan metode Bayesian Classification. Pengolahan berbagai data, berbagai informasi dan berbagai metode dengan kemampuan teknologi yang canggih akan sangat membantu meminimalisasi kesalahan, sehingga dapat memutuskan jurusan secara cepat, tepat dan adil.

2. Landasan Teori

a. Klasifikasi

Klasifikasi merupakan salah satu tugas yang penting dalam *Data Mining*. Tujuan utama dari klasifikasi adalah mengorganisasikan dan mengelompokkan data ke dalam kelas-kelas yang berbeda. Definisi pengklasifikasian adalah sebuah fungsi yang bersifat prediksi dan menggolongkan data item tertentu ke dalam sebuah kelas. Sebuah pengklasifikasian dibuat dari sekumpulan data latih dengan kelas yang telah ditentukan dan dikenal ciri-cirinya sebelumnya. Performa pengklasifikasian biasanya diukur dengan ketepatan.

b. Bayesian Klasifikasi

Bayesian Classification didasarkan pada Teorema Bayes yang memiliki kemampuan hampir serupa dengan *Decision Tree* dan *Neural Network*. Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis. Teorema Bayes ini dibuat oleh Thomas Bayes yang ditulis pada *paper* yang berjudul “*An Essay toward Solving a Problem in The Doctrine of Chance*”. Inti dari Teorema Bayes tersebut adalah memprediksi probabilitas dimasa yang akan datang berdasarkan pengalaman dimasa sebelumnya.

Bayesian Classification adalah suatu metode pengklasifikasian data dengan model statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan pada suatu kelas. Metode *Bayesian Classification* digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari sejumlah alternatif. *Bayesian Classification* merupakan salah satu metode yang sederhana yang dapat digunakan untuk data yang tidak konsisten dan data bias. Metode Bayes juga merupakan metode yang baik dalam mesin pembelajaran berdasarkan *data training* dengan berdasarkan pada probabilitas bersyarat.

Bentuk umum Teorema Bayes adalah :

$$P(H | X) = \frac{P(X | H).P(H)}{P(X)} \quad [1]$$

Keterangan:

X : Sampel data dengan kelas yang belum diketahui

H : Hipotesis data X merupakan suatu kelas spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (*posterior probability*)

$P(H)$: Probabilitas hipotesis H (*prior probability*)

$P(X|H)$: Probabilitas X berdasar kondisi pada hipotesis H
 $P(X)$: Probabilitas dari X

Cara kerja Bayesian Classification :

- 1) Misalkan ada satu set sampel $S = \{S_1, S_2, \dots, S_m\}$ (m adalah total sampel) dimana setiap sampel S_i diwakili vektor n dimensi $\{X_1, X_2, \dots, X_n\}$. Nilai X_i sesuai dengan atribut A_1, A_2, \dots, A_n masing-masing. Juga ada kelas k C_1, C_2, \dots, C_k , dan setiap sampel milik salah satu kelas.
- 2) Data tambahan sampel X tidak diketahui kelasnya. Untuk memprediksi kelas dari sampel baru tersebut menggunakan probabilitas kondisional $P(C_i|X)$, dimana i adalah kelas-kelas yang ada. Probabilitas ini dihitung dengan menggunakan Teorema Bayes :

$$P(C_i | X) = \frac{P(X | C_i).P(C_i)}{P(X)} \quad [2]$$

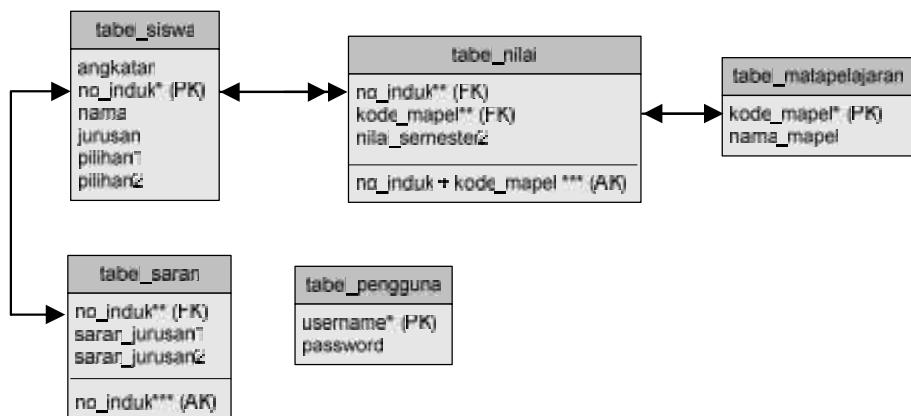
- 3) $P(X)$ adalah konstan dari semua kelas, hanya $P(X|C_i).P(C_i)$ yang perlu dimaksimalkan. Dihitung prior probability kelas.
- 4) Karena perhitungan dari $P(X|C_i)$ sangat kompleks, terutama untuk set data yang besar, asumsi Naive dari *conditional independence* antara atribut dibuat. Menggunakan asumsi kita dapat mengekspresikan $P(X|C_i)$ sebagai produk:

$$P(X | C_i) = \prod_{t=1}^n P(X_t | C_i) \quad [3]$$

- dimana X_t adalah nilai-nilai untuk atribut dalam sampel X . Probabilitas $P(X_t|C_i)$ dapat diperkirakan dari training set data.
- 5) Untuk mengklasifikasikan contoh X , $P(X|C_i).P(C_i)$ dievaluasi tiap kelas C_i . Data contoh X untuk kelas C_i jika dan hanya jika:
$$P(C_i|X) \cdot P(C_i) > P(C_j|X) \cdot P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i$$
 - 6) Bayesian Classification memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode ini memiliki tingkat kesalahan minimal dibandingkan dengan semua classifier lain yang dikembangkan dalam Data Mining. Tetapi, bagaimanapun juga tidak ada metode yang dapat memprediksi secara tepat sempurna.

3. Perancangan Basis Data

Atribut bukan kunci berfungsi sebagai informasi yang dibutuhkan untuk menjelaskan suatu entitas. Berikut merupakan tampilan :



Gambar 1. Rancangan Basis Data

4. Implementasi dan Analisa Sistem

a. Implementasi Sistem

1) Form Penjurusan Angkatan

Form penjurusan angkatan merupakan form inti dari penelitian skripsi ini. Form ini berfungsi untuk memilih angkatan yang hendak dijuruskan dan kemudian dilakukan proses mencari saran jurusan. Yang memiliki hak untuk mengakses form ini adalah BP.

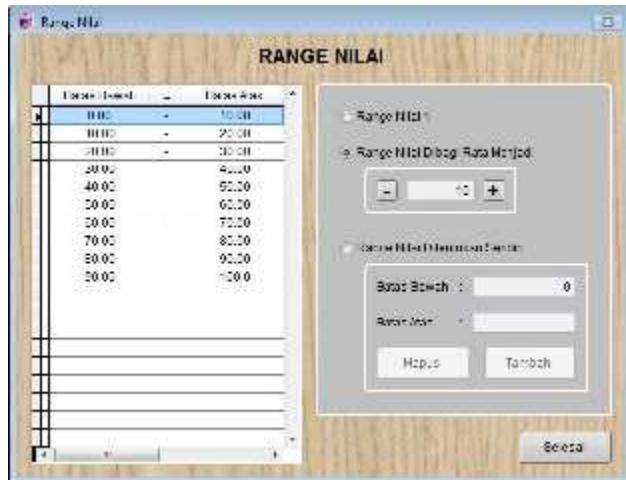


Gambar 2. Form Penjurusan Angkatan

Langkah pertama untuk melakukan proses penjurusan ini adalah dengan memilih tahun angkatan yang akan dijuruskan terlebih dahulu. Ketika angkatan terpilih, maka akan muncul keterangan berapa banyak siswa angkatan tersebut. Selanjutnya pilih cara menentukan range yang akan digunakan dalam proses penjurusan, apakah dengan merange nilai secara manual (biasa) atau dengan memanfaatkan metode box plot. Jika range sudah ditentukan, maka tinggal klik tombol proses untuk melakukan penjurusan. Terdapat komponen *loading* untuk menampilkan berapa persen proses penjurusan telah selesai. Jika proses penjurusan sudah selesai 100% maka sistem akan mengarahkan pada form saran jurusan untuk menampilkan hasil proses penjurusan.

2) Range Nilai Manual

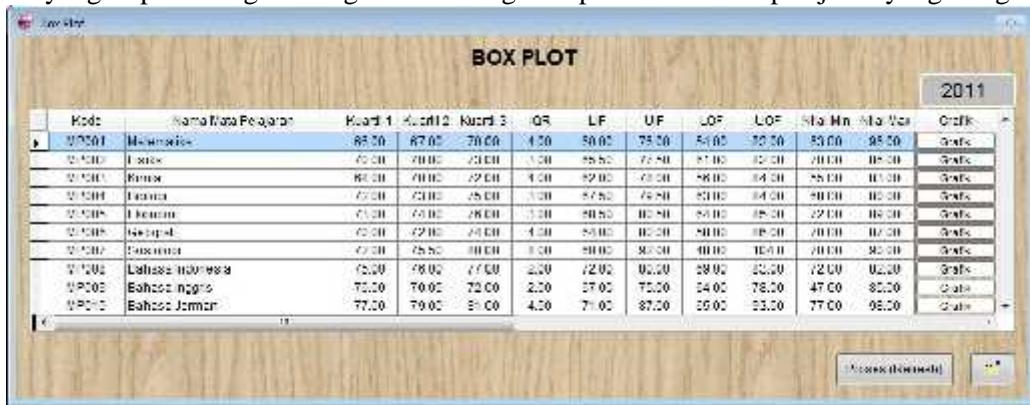
Form range nilai merupakan form yang digunakan untuk menentukan range nilai untuk penjurusan. Dalam range ini kita dapat menentukan range dengan 3 cara. Cara pertama, range nilai 1 yaitu nilai dari 1 sampai 100 berdiri sendiri-sendiri. Cara kedua adalah dengan membagi rata menjadi beberapa bagian sesuai yang ditentukan *user*. Cara ketiga memungkinkan *user* menentukan range secara manual, sesuai yang dikehendaki.



Gambar 3. Range Nilai Manual

3) Range Box Plot

Form box plot merupakan form yang digunakan untuk mencari titik-titik range sesuai dengan data-data nilai yang dimiliki suatu angkatan berdasarkan tiap mata pelajaran. Form ini diakses lewat tombol box plot pada form penjurusan. Form ini akan menampilkan box plot berdasarkan angkatan yang dipilih pada form penjurusan. Untuk mencari titik-titik box plot dilakukan dengan mengklik tombol proses, maka kemudian sistem akan memproses dan menampilkan hasil perhitungan box plot. *User* dapat melihat grafik box plot setiap mata pelajaran angkatan yang terpilih dengan mengklik tombol grafik pada baris mata pelajaran yang diinginkan.



Gambar 4. Range Box Plot

4) Form Saran Jurusan

Form saran jurusan adalah form yang akan menampilkan hasil dari proses penjurusan. Untuk melihat detail perhitungan proses penjurusan setiap siswa dapat dilihat dengan mengklik tombol detail pada setiap baris sesuai dengan siswa yang ingin dilihat prosesnya.



Gambar 5. Form Saran Jurusan

5) Laporan Saran Jurusan

SMA Bruderan

Jl. K.H. Wahid Hasyim no.6 Purworejo - 54717
Telp. (0275) 321504 Fax. (0275) 323463

Detail Proses

Angkatan	: 2010	Pilihan 1	: PA
No. Induk	: 3703	Pilihan 2	: Bantara
Nama Lengkap	: PATRICK BONARDO S. S.		
<hr/>			
Proses Penjurusan		Rangka Disain	
3710 - PAIRICE DOMARDO S. S.		Range Disain	
$P(\text{Kelas} = \text{IPS})$ $P(\text{Kelas} = \text{PA})$ $P(\text{Matematika} \leq 70 \wedge \text{PA} < 100) = 400 / 748$ $P(\text{Matematika} \leq 70 \wedge \text{PA} \geq 100) = 118$ $P(\text{Matematika} \leq 70 \wedge \text{PA} \leq 100) = 400 / 748$ $P(\text{Fisika} \leq 65 \wedge \text{Kelas} = \text{IPS})$ $P(\text{Kimia} \leq 65 \wedge \text{PA} < 100) = 903 / 143$ $P(\text{Kimia} \leq 65 \wedge \text{PA} \geq 100) = 6.914685814685915000$ $P(\text{Kimia} \leq 65 \wedge \text{Kelas} = \text{PA})$ $P(\text{Biology} \leq 70 \wedge \text{PA} < 100) = 277 / 143$ $P(\text{Biology} \leq 70 \wedge \text{Kelas} = \text{IPS})$ $P(\text{Kemampuan} \leq 70 \wedge \text{PA} < 100) = 277 / 143$ $P(\text{Kemampuan} \leq 70 \wedge \text{Kelas} = \text{PA})$	$P(\text{Kelas} = \text{IPS})$ $P(\text{Kelas} = \text{PA})$ $P(\text{Matematika} \leq 70 \wedge \text{PA} < 100) = 400 / 748$ $P(\text{Matematika} \leq 70 \wedge \text{PA} \geq 100) = 118$ $P(\text{Matematika} \leq 70 \wedge \text{PA} \leq 100) = 400 / 748$ $P(\text{Fisika} \leq 65 \wedge \text{Kelas} = \text{IPS})$ $P(\text{Kimia} \leq 65 \wedge \text{PA} < 100) = 903 / 143$ $P(\text{Kimia} \leq 65 \wedge \text{PA} \geq 100) = 6.914685814685915000$ $P(\text{Kimia} \leq 65 \wedge \text{Kelas} = \text{PA})$ $P(\text{Biology} \leq 70 \wedge \text{PA} < 100) = 277 / 143$ $P(\text{Biology} \leq 70 \wedge \text{Kelas} = \text{IPS})$ $P(\text{Kemampuan} \leq 70 \wedge \text{PA} < 100) = 277 / 143$ $P(\text{Kemampuan} \leq 70 \wedge \text{Kelas} = \text{PA})$		

Gambar 6. Laporan Saran Jurusan

b. Analisa Sistem

Berikut ini adalah tabel keakuratan metode Bayesian dalam kasus penjurusan di SMA Bruderan Purworejo.

- 1) Keakuratan Bayesian dengan menggunakan range yang ditentukan secara manual.

Tabel 1.
Keakuratan Bayesian dengan Range Manual

Keakuratan	Angkatan 2009	Angkatan 2010	Angkatan 2011
Interval 1	26%	32%	21%
Interval 2	41%	43%	24%
Interval 4	57%	50%	24%
Interval 5	49%	47%	26%
Interval 10	59%	60%	28%
Interval 20	59%	66%	61%
Interval 25	62%	58%	54%
Interval 50	58%	57%	57%
Interval 100	59%	58%	57%

- 2) Keakuratan Bayesian dengan menggunakan range yang ditentukan dari proses BoxPlot

Tabel 2.
Keakuratan Bayesian dengan Range Boxplot

Tahun	Keakuratan Box Plot
2009	62%
2010	63%
2011	66%

5. Penutup

Beberapa kesimpulan yang dapat diambil dari penelitian menggunakan metode *Bayesian Classification* dalam kasus penjurusan di SMA Bruderan Purworejo antara lain :

- a. Semakin banyaknya data dan semakin bervariasinya data, maka akan memperbesar kemungkinan siswa dapat diberikan saran jurusan dan hasil penjurusan juga akan lebih bervariasi dan akurat.
- b. Keakuratan tertinggi dari hasil proses penjurusan dengan menggunakan range yang ditentukan secara manual berada pada range dengan interval 20 dan 25. Pada interval 20, angkatan 2009 mendapatkan sebesar 59 %, 2010 sebesar 66 %, 2011 sebesar 61 %. Sedangkan pada Interval 25, angkatan 2009 mendapatkan sebesar 62 %, 2010 sebesar 58 %, 2011 sebesar 54 %.
- c. Hasil proses penjurusan dengan menggunakan range yang dapatkan dari proses Box Plot rata-rata memiliki tingkat keakuratan lebih tinggi dibanding dengan range yang ditentukan secara manual. Pada range Box Plot, tingkat keakuratan angkatan 2009 adalah 62%, tingkat keakuratan angkatan 2010 adalah 63% dan angkatan 2011 adalah 66%.
- d. Untuk mengatasi siswa yang tidak mendapatkan saran jurusan karena terdapat atribut di tiap kelas yang probabilitasnya 0 dilakukan dengan cara mengambil nilai terendah pada mata pelajaran yang bersangkutan. Berdasarkan proses yang telah dilakukan, hasil penjurusan dengan cara tersebut rata-rata memiliki keakuratan lebih tinggi karena seluruh siswa mendapatkan saran jurusan tanpa terkecuali. Keakuratan tertinggi juga terdapat pada interval 20 dan 25. Pada interval 20, angkatan 2009 mendapatkan sebesar 58 %, 2010 sebesar 67 %, 2011 sebesar 58 %. Sedangkan pada Interval 25, angkatan 2009 mendapatkan sebesar 64 %, 2010 sebesar 64 %, 2011 sebesar 53 %.

Daftar Pustaka

- Han, Jiawei & Micheline Kamber. 2001. *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Kantardzic, Mehmed. 2003. *Data Mining: Concepts, Models, Methods and Algorithms*. United States of America: Wiley-Interscience.
- Kristanto, Harianto IR. 1996. *Konsep dan Perancangan Database*. Yogyakarta: Penerbit Andi.
- Satrianto, Mehmed. 2003. *Teorema Bayes*. Diakses tanggal 20 Maret 2013 dari http://Pdf.searchengine/teorema_bayes.pdf.
- Suryadi, Christine. 2003. *Probabilitas dan Statistika Teorema Bayes*. Diakses 20 Maret 2013 dari <http://IF ITB/CS/ProbabilitasdanStatistikaTeoremaBayes.pdf>.
- Sutedjo, Budi. 2002. *Perencanaan & Pembangunan Sistem Infromasi*. Yogyakarta: Andi Offset.