



**PENERAPAN DATA MINING DENGAN NAÏVE BAYES UNTUK
KLASIFIKASI SISWA SEKOLAH MENENGAH ATAS DALAM PENENTUAN
PERGURUAN TINGGI**

Muchammad Farrell Putro, Edhi Prayitno, Juarni Siregar, Muhammad Muharrom
Fakultas Teknologi Informasi Program Studi Sistem Informasi Universitas Nusa
Mandiri, Fakultas Teknik dan Informatika Program Studi Teknologi Informasi
Universitas Bina Sarana Informatika
(Naskah diterima: 1 Maret 2021, disetujui: 30 April 2021)

Abstract

This study aims to determine the classification of public or private tertiary institutions based on the dataset that has been obtained along with the attributes of UN scores, non-academic scores, UAS scores and psychological tests that will be used for datamining using the naïve Bayes method and the calculation process using excel manual with a predetermined formula. The number of data is 22 which is used as a reference in the results of the determination, 2 data objects as testing data where this research gives the results that the prediction of data mining classification using naïve Bayes is in accordance with the classification of training data.

Keyword: naïve bayes, classification, college determination

Abstrak

Penelitian ini bertujuan untuk mengetahui klasifikasi penentuan perguruan tinggi negeri atau swasta berdasarkan dataset yang telah didapatkan beserta atribut nilai UN, Nilai non akademis, nilai UAS dan Psikotes yang akan digunakan datamining menggunakan metode naïve bayes dan proses perhitungannya menggunakan manual excel dengan rumus yang sudah ditentukan. Jumlah data ada 22 yang dijadikan rujukan dalam hasil penentuan, 2 data objek sebagai data testing dimana Penelitian ini memberikan hasil bahwa prediksi klasifikasi data mining menggunakan naïve bayes sesuai dengan klasifikasi data training.

Kata Kunci: naïve bayes, klasifikasi, penentuan perguruan tinggi

I. PENDAHULUAN

Penentuan klasifikasi kelulusan sekolah adalah merupakan salah satu hal yang sangat penting karena pilihan sekolah swasta maupun negeri tentu akan berdampak bagi motivasi dan hasil belajar siswa

dan tentunya akan mempengaruhi kualitas belajar. menimbang hal ini tentu saja kualitas pendidikan menjadi prioritas utama guna menciptakan lulusan - lulusan yang kompeten. Dikarenakan keinginan siswa untuk meneruskan jenjang yang lebih tinggi, pihak se-

kolah harus mampu mengklasifikasikan mana siswa yang layak ke perguruan tinggi negeri dan perguruan tinggi swasta, tentu hal tersebut ada data yang mengacu untuk keputusan yang dapat diperoleh dan tentunya akan memudahkan pengklasifikasian. oleh karena itu setiap siswa tingkat akhir harus didapati dari hasil ujian UN, UAS Psikotes yang akan layak masuk ke perguruan tinggi apa. Maka proses tersebut di implementasi menggunakan metode klasifikasi Naive Bayes diharapkan mampu mengklasifikasikan penentuan para siswanya dan tepat sasaran.

Pada penelitian sebelumnya metode naïve bayes juga digunakan dalam memprediksi penyakit yang terkait dengan Dermatologi yang menyebabkan kematian di mana metode Naive Bayes digunakan untuk mengenal pola data untuk mengungkap kemungkinan apa saja yang dapat terjadi pada penyakit dermatologi. Metode Naive Bayes juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi.

II. KAJIAN TEORI

2.1 Data Mining

Data Mining merupakan proses pengestraksian informasi dari sekumpulan data

yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data. 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. Definisi lain mengatakan Data Mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. 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.

2.2 Tahapan-Tahapan Data Mining

Sebagai suatu rangkaian proses, Data Mining dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base. Tahap-tahap Data Mining adalah sebagai berikut: a. Pembersihan data (Data Cleaning) Pembersihan data merupakan proses menghilangkan 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 kedalam knowledge based yang ditemukan. g. Presentasi pengetahuan (*Knowledge Presentation*) Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

2.3 Metode Naïve 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. 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.

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. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

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

III. METODE PENELITIAN

Teknik Pengumpulan data yang penulis lakukan adalah sebagai berikut:

a. Observasi;

Observasi dilakukan penulis dengan cara mengamati secara langsung Data yang terdapat pada sekolah Yadika 13, guna untuk menentukan data training dan testing sesuai dengan apa yang penulis teliti.

b. Studi Pustaka;

Adalah tahapan yang penulis lakukan dalam mengumpulkan material atau bahan-bahan yang akan dipakai dalam menganalisa dan sumber referensi dalam mendukung analisa, serta dengan cara mengumpulkan melalui buku-buku literatur, jurnal dan internet yang sesuai dengan judul penelitian yaitu buku-buku yang terkait dengan materi yang akan dibahas.

IV. HASIL PENELITIAN

A. Teknik dengan Naive Bayes

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.

Untuk menentukan data yang nanti-nya 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	Nilai UN	Non-Akademis	Nilai UAS	Psikotes	Klasifikasi
1	Baik	Ada	Cukup	Tinggi	Negeri
2	Baik	Ada	Cukup	Rendah	Negeri
3	Baik	Tidak	Kurang	Rendah	Negeri
4	Cukup	Tidak	Baik	Rendah	Negeri
5	Cukup	Tidak	Kurang	Rendah	Swasta
6	Kurang	Ada	Baik	Rendah	Negeri
7	Cukup	Tidak	Cukup	Tinggi	Negeri
8	Kurang	Tidak	Cukup	Rendah	Swasta
9	Kurang	Ada	Baik	Rendah	Negeri

10	Cukup	Tidak	Kurang	Rendah	Negeri
11	Baik	Tidak	Kurang	Tinggi	Negeri
12	Cukup	Tidak	Kurang	Rendah	Swasta
13	Baik	Ada	Kurang	Rendah	Negeri
14	Baik	Ada	Baik	Rendah	Negeri
15	Baik	Ada	Baik	Rendah	Negeri
16	Kurang	Tidak	Cukup	Rendah	Swasta
17	Kurang	Tidak	Cukup	Rendah	Swasta
18	Kurang	Tidak	Kurang	Tinggi	Swasta
19	Baik	Tidak	Cukup	Tinggi	Negeri
20	Cukup	Ada	Kurang	Tinggi	Negeri
21	Cukup	Ada	Kurang	Tinggi	Negeri
22	Kurang	Ada	Kurang	Rendah	Swasta

1) Menghitung jumlah class/label

Tabel 2. Tabel Jumlah class

Label	jumlah data	jumlah seluruh data	hasil
P(Y=Negeri)	15	22	0,68
P(Y=Swasta)	7	22	0,32

2) Mencari Peluang

Tabel 3. Tabel Mencari Peluang

Nilai UN	Negeri	Swasta	P(Negeri)	P(Swasta)
Baik	8	0	0,533	0,000
Cukup	5	2	0,333	0,286
Kurang	2	5	0,133	0,714
Total	15	7		
Non-Aka	Negeri	Swasta	P(Negeri)	P(Swasta)
Ada	9	1	0,600	0,143
Tidak	6	6	0,400	0,429
Total	15	7		
Nilai UAS	Negeri	Swasta	P(Negeri)	P(Swasta)
Baik	5	0	0,357	0,000
Cukup	4	3	0,286	0,428571429
Kurang	5	4	0,357	0,571
Total	14	7		
Psikotes	Negeri	Swasta	P(Negeri)	P(Swasta)
Tinggi	6	1	0,400	0,143
Rendah	9	6	0,600	0,857
Total	15	7		

3) Test dan sesuaikan pada data testing

nilai UN	nilai UAS	nono akademis	psikotest	kualifikasi
cukup	baik	ada	rendah	?
kurang	cukup	tidak	rendah	?

- 1) Jika nilai UN cukup maka masukan nilai dari cukup
- 2) Jika nilai UAS baik maka masukan nilai dari baik
- 3) Jika nilai non akademis ada maka masukan nilai dari ada
- 4) Jika nilai psikotes rendah maka masukan nilai dari rendah dan seterusnya.

Maka menghasilkan nilai sebagai berikut:

Tabel 4. Penyesuaian data testing

1	0	Nilai	Non	Nilai	Psikotes
	Negeri	0,333	0,600	0,357	0,6000
	Swasta	0,286	0,143	0,000	0,8571
2	1	Nilai	Non	Nilai	Psikotes
	Negeri	0,133	0,400	0,286	0,6000
	Swasta	0,714	0,429	0,429	0,8571

4) kalikan semua kelas dan kesimpulannya

- 1) Kalikan semua value class dari nilai un,non akademis,nilai uas,dan psikotes.
- 2) Jika jumlah hasil perkalian lebih besar pada nilai negeri, maka ditetapkan negeri pada data testing menurut naïve bayes.
- 3) Jika jumlah hasil perkalian lebih besar pada nilai swasta maka ditetapkan swasta pada data testing menurut naïve bayes.

Maka menghasilkan nilai akhir sebagai berikut:

Tabel 5. Nilai akhir testing

1	0	Nilai	Non	Nilai	Psikotes	
	Negeri	0,333	0,600	0,357	0,6000	0,0429
	Swasta	0,286	0,143	0,000	0,8571	0,0000
2	1	Nilai	Non	Nilai	Psikotes	
	Negeri	0,133	0,400	0,286	0,6000	0,0091
	Swasta	0,714	0,429	0,429	0,8571	0,1125

Maka kualifikasi yang didapat untuk data testing dari hasil nilai yang tertinggi adalah

nilai UN	nilai UAS	nono akademis	psikotest	kualifikasi
cukup	baik	ada	rendah	negeri
kurang	cukup	tidak	rendah	swasta

V.KESIMPULAN

Berdasarkan penelitian mengenai penentuan masuk perguruan tinggi adalah sebagai berikut:

- 1) Berdasarkan data siswa yang diperoleh, proses *Data Mining* membantu dalam penerapan metode *Naive Bayes* dalam mendapatkan informasi dari hasil klasifikasi masuk perguruan tinggi.
- 2) Berdasarkan data siswa yang diperoleh, proses penerapan metode *Naive bayes* menghasilkan prediksi dari 2 data testing adalah pertama negeri dan kedua swasta
- 3) Berdasarkan hasil tersebut memiliki prediksi yang sesuai dengan data training yang sebagai acuan perhitungan

DAFTAR PUSTAKA

- 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.
- 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.
- Angga Ginanjar Maburur, Riani Lubis, (2012). *Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit*, Jurnal Komputer dan Informatik (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.
- 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.

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