

DATA MINING UNTUK MEMPREDIKSI JUMLAH KELULUSAN MAHASISWA

Yuniar Andi Astuti¹

Sondang²

Manajemen Informatika, Politeknik Unggul LP3M

Abstrak

Dalam mengikuti perkuliahan, setiap mahasiswa tentu mengharapkan dirinya berhasil dan lulus. Oleh karena itu setiap mahasiswa haruslah paham dan berusaha keras agar berhasil lulus. Penerapan metode naïve bayes diharapkan mampu untuk memprediksi besarnya kelulusan mahasiswa agar lebih mudah meningkatkan jumlah lulusan yang berkualitas, dari 60 data mahasiswa yang diuji dengan metode naïve bayes, maka diperoleh hasil persentase 85% untuk keakuratan prediksi, di mana dari 60 data jumlah kehadiran, jumlah tugas, biaya kuliah dan lokasi tempat tinggal yang diuji terdapat 51 data mahasiswa yang berhasil diklasifikasikan dengan benar.

Kata Kunci — Data Mining, Naïve Bayes, Kelulusan mahasiswa

I. PENDAHULUAN

Tujuan pendidikan menurut UU No. 2 Tahun 1985 adalah untuk mencerdaskan kehidupan bangsa dan mengembangkan manusia yang seutuhnya, yaitu bertakwa terhadap Tuhan Yang Maha Esa, memiliki pengetahuan, sehat jasmani dan rohani, memiliki budi pekerti luhur, mandiri, kepribadian yang mantap, dan bertanggungjawab terhadap bangsa. Berdasarkan tujuan diatas maka pendidikan merupakan salah faktor agar dapat mewujudkan bangsa yang maju dan berkembang.

Menurut Suyata, fungsi utama sekolah pada awalnya adalah pengajaran namun dalam perkembangannya sekolah berfungsi majemuk dengan pendidikan sebagai intinya. Persoalan jumlah dan siapa yang perlu memperoleh pendidikan kiranya cukup jelas, yaitu semua rakyat pembentuk bangsa kita, sedangkan yang perlu dipikirkan dan di usahakan adalah kualitas pendidikan atau mutu kecerdasannya, serta cara mencapainya merupakan implikasi pesan utama cita-cita yang diletakkan oleh bapak-bapak pendiri Republik Indonesia dan pengisian pesan tersebut perlu dicari, dikaji, dan terus dikembangkan.

Memasuki abad ke- 21 gelombang globalisasi dirasakan kuat dan terbuka. Kemajuan teknologi dan perubahan yang terjadi memberikan kesadaran baru bahwa Indonesia tidak lagi berdiri sendiri. Indonesia berada di tengah-tengah dunia yang baru, dunia terbuka sehingga orang bebas membandingkan kehidupan dengan negara lain.

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. *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.

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:

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. 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 di masa 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.*

II. METODE PENELITIAN

Adapun metode penelitian yang diterapkan pada penelitian ini adalah sebagai berikut:

1. Analisis Masalah dan Studi Literatur

Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian. Dalam hal ini mengamati permasalahan yang berhubungan dengan faktor-faktor penyebab kegagalan mahasiswa dalam mengikuti perkuliahan. Permasalahan-permasalahan yang ada, selanjutnya dianalisa untuk mengetahui bagaimana cara penyelesaian terhadap masalah tersebut dan menentukan ruang lingkup permasalahan yang akan diteliti. Mempelajari dasar teori dari berbagai literatur mengenai penerapan metode Naive Bayes, konsep dan teori data mining dan pengeluaran biaya rumah tangga, melalui jurnal-jurnal dan agar mendapatkan dasar pengetahuan untuk melakukan penelitian selanjutnya.

2. Mengumpulkan Data

Prosedur sistematis yang digunakan untuk mengumpulkan data yaitu dengan mengajukan pertanyaan-pertanyaan kepada responden. Hasil kuesioner yang didapatkan akan digunakan untuk bahan penganalisisan data terhadap metode Naive Bayes. Setelah data dikumpulkan dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada metode Naive Bayes.

3. Implementasi dan Pengujian

Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah tools. Tools yang akan digunakan dalam implementasi penelitian ini adalah dengan menggunakan Software Weka. Selanjutnya Pengujian dilakukan untuk mengetahui apakah penelitian yang dilakukan telah sesuai dengan tujuan yang diharapkan.

III. HASIL DAN 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.

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 berikut:

Tabel 1. Data Training

No	Jumlah Kehadiran	Jml Tugas	Biaya Kuliah	Lokasi TT
1	Tinggi	Banyak	Tinggi	Jauh
2	Tinggi	Banyak	Tinggi	Jauh
3	Tinggi	Banyak	Tinggi	Jauh
4	Tinggi	Banyak	Sedang	Standar
5	Sedang	Kecil	Kecil	Standar
6	Sedang	Kecil	Kecil	Standar
7	Sedang	Kecil	Sedang	Standar
8	Sedang	Kecil	Sedang	Sedang
9	Sedang	Kecil	Sedang	Sedang

10	Sedang	Kecil	Sedang	Standar
11	Sedang	Kecil	Kecil	Standar
12	Sedang	Sedang	Kecil	Dekat
13	Rendah	Kecil	Kecil	Dekat
14	Sedang	Sedang	Sedang	Sedang
15	Tinggi	Banyak	Sedang	Sedang
60	Sedang	Kecil	Kecil	Dekat

Kriteria dan Probabilitas

Adapun nilai probabilitas setiap kriteria didapatkan dari data latih pada tabel 1. Adapun nilai probabilitas setiap kriteria sebagai berikut.

Probabilitas Kriteria Jumlah Kehadiran

Berdasarkan data training pada tabel 1, diketahui jumlah data latih sebanyak 60 data, dimana dari 60 data tersebut terdapat 5 mahasiswa dengan jumlah kehadiran tinggi dan jumlah tugas sedikit, 1 mahasiswa dengan jumlah kehadiran sedang dan tugas sedikit, 6 mahasiswa dengan jumlah kehadiran tinggi dan tugas sedang kemudian 7 mahasiswa dengan jumlah kehadiran sedang dan tugas sedang, 18 mahasiswa dengan jumlah kehadiran tinggi dan tugas banyak, 14 mahasiswa dengan jumlah kehadiran sedang dan tugas banyak, dan 7 mahasiswa dengan jumlah kehadiran rendah dan tugas banyak.

Tabel 2. Probabilitas Kriteria Jumlah Kehadiran

Jumlah Kehadiran	Jumlah Tugas			Probabilitas		
	Sedikit	Sedang	Banyak	Sedikit	Sedang	Banyak
Tinggi	5	6	18	0.83	0.40	0.46
Sedang	1	7	14	0.17	0.46	0.36
Rendah	0	2	7	0	0.13	0.18
Jumlah	6	15	39	0.10	0.25	0.65

Probabilitas Kriteria Biaya Kuliah

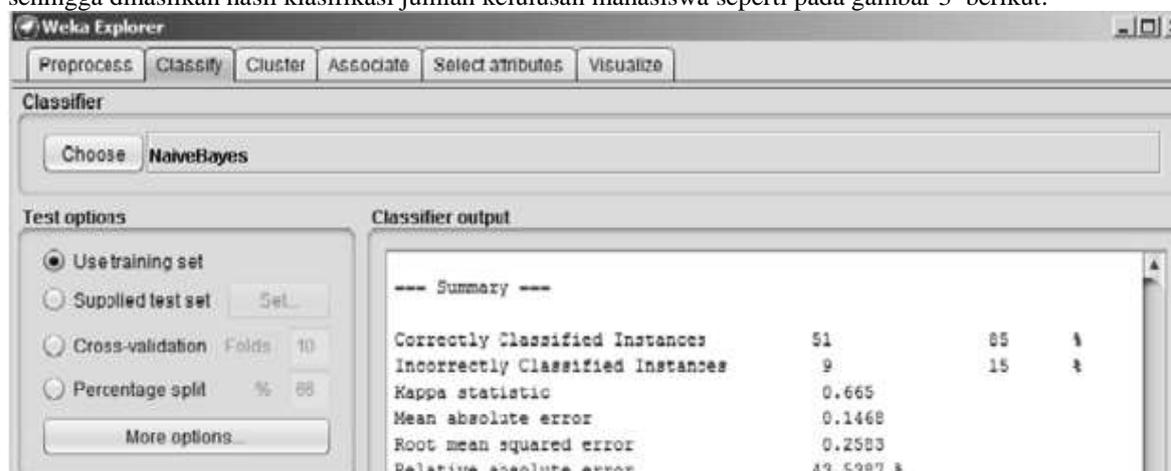
Berdasarkan data training pada tabel 1, diketahui jumlah data latih sebanyak 60 data, dimana dari 60 data tersebut terdapat 4 mahasiswa dengan biaya kuliah tinggi dan lokasi tempat tinggal dekat, 1 mahasiswa dengan biaya kuliah sedang dan lokasi tempat tinggal dekat, 3 mahasiswa dengan biaya kuliah kecil dan lokasi tempat tinggal dekat, 9 mahasiswa dengan biaya kuliah tinggi dan lokasi tempat tinggal sedang, 14 mahasiswa dengan biaya kuliah sedang dan lokasi tempat tinggal sedang, 17 mahasiswa dengan biaya kuliah rendah dan lokasi tempat tinggal dekat dan 5 mahasiswa dengan biaya kuliah tinggi dan lokasi tempat tinggal jauh serta 7 mahasiswa dengan biaya kuliah sedang dan lokasi tempat tinggal jauh.

Tabel 2. Probabilitas Kriteria Biaya Kuliah

Biaya Kuliah	Lokasi Tempat Tinggal			Probabilitas		
	Dekat	Sedang	Jauh	Rendah	Sedang	Tinggi
Tinggi	4	9	5	0.50	0.22	0.42
Sedang	1	14	7	0.13	0.35	0.58
Kecil	3	17	0	0.37	0.43	0.00
Jumlah	8	40	12	0.14	0.66	0.20

Pengujian Metode Naïve Bayes

Dari nilai probabilitas di atas akan diuji sebanyak 60 data dan diselesaikan dengan menggunakan tools weka sehingga dihasilkan hasil klasifikasi jumlah kelulusan mahasiswa seperti pada gambar 3 berikut:



Gambar 3. Hasil Klasifikasi Metode Naïve Bayes

Berdasarkan gambar 3 di atas dapat dilihat persentase untuk *Correctly Classified Instance* adalah sebesar 85% sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 15%. Di mana dari 60 data jumlah kehadiran, jumlah tugas, biaya kuliah dan lokasi tempat tinggal ada sebanyak 51 data berhasil diklasifikasikan dengan benar dan sebanyak 9 data tidak berhasil diklasifikasikan dengan benar.

IV. KESIMPULAN

Berdasarkan penelitian tentang memprediksi kelulusan mahasiswa dapat ditarik beberapa kesimpulan sebagai berikut :

1. 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 kelulusan mahasiswa berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive Bayes* itu sendiri.
2. Berdasarkan data mahasiswa yang dijadikan data *training*, metode *Naive Bayes* berhasil mengklasifikasikan 51 data dari 60 data yang diuji. Sehingga metode *Naive Bayes* berhasil memprediksi besarnya kelulusan mahasiswa dengan persentase keakuratan sebesar 85%.

V. DAFTAR PUSTAKA

- Elmasri, Ramez and Shamkant B. Navathe, 2000, "*Fundamentals of Database Systems. Third Edition*", Addison Wesley Publishing Company, New York. Han, J. and Kamber, M, 2006, "*Data Mining Concept and Techniques Second Edition*".
- Morgan Kauffman, San Fransisco. Kadir, Abdul, 1999, "*Konsep dan Tuntunan Praktis Basis Data*", Penerbit Andi, Yogyakarta. Kusrini dan Emha Taufik Luthfi, 2009, "*Algoritma Data Mining*", Penerbit Andi Yogyakarta
- Manjusha K. K., Sankaranarayanan, K., Seena P., 2014, Prediction of Different Dermatological Conditions Using Naive Bayesian Classification, *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol 4, No 1, Hal 864- 868.
- Ting, S. L., Ip, W. H., Tsang, A. H.C., 2011, Is Naive Bayes a Good Classifier for Document Classification?, *International Journal of Software Engineering and Its Applications*, Vol. 5, No. 3, Hal 37-46.
- Taruna R., S., Hiranwal, S., 2013, Enhanced Naive Bayes Algorithm for Intrusion Detection in Data Mining, *International Journal of Computer Science and Information Technologies*, Vol.6, No. 4, Hal 960-962.
- Mabrur, A. G., Lubis, R., 2012, Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit, *Jurnal Komputer dan Informatika (KOMPUTA)*, Vol.1, No. 1, Hal 53-57.
- Mujiasih, S., 2011, Pemanfaatan Data Mining Untuk Prakiraan Cuaca, *Jurnal Meteorologi dan Geofisika*, Vol. 12, No. 2, Hal 189 - 195.
- Ridwan, M., Suyono, H., Sarosa, M., 2013, Penerapan Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier, *Jurnal EECCIS*, Vol 1, No. 7, Hal. 59-64.
- Patil, T. R., Sherekar, M. S., 2013, Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification, *International Journal of Computer Science and Applications*, Vol. 6, No. 2, Hal 256-261.
- Bustami., 2013, Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi, *TECHSI : Jurnal Penelitian Teknik Informatika*, Vol. 3, No.2, Hal. 127-146.
- Pattekari, S. A., Parveen, A., 2012, Prediction System for Heart Disease Using Naive Bayes, *International Journal of Advanced Computer and Mathematical Sciences*, ISSN 2230-9624, Vol. 3, No 3, Hal 290-294.