

SISTEM PENDUKUNG KEPUTUSAN UNTUK MEMPREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES

Diana Laily Fithri, Eko Darmanto

Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus
Gondangmanis, PO Box 53, Bae, Kudus 59352
Email : dila_fitri@yahoo.com

Abstrak

Sistem Pendukung Keputusan untuk menentukan tingkat prediksi kelulusan mahasiswa digunakan untuk menentukan tingkat kelulusan mahasiswa di Progdil Sistem Informasi. Prediksi tersebut nantinya digunakan sebagai sumber informasi untuk menghasilkan sebuah keputusan. Penelitian ini menghasilkan sistem pendukung keputusan dengan menggunakan metode Naïve Bayes dengan menggunakan beberapa parameter, yaitu jenis kelamin, alamat, umur, status pekerjaan mahasiswa, status pernikahan mahasiswa, rata-rata IPK, jumlah SKS dan status mahasiswa. Pengolahan data mining mahasiswa dengan menggunakan metode naïve Bayes dimulai dari proses Data Gathering, Data Preprocessing, Proposed Model/Method, Method Test and Experiment, Result Evaluation and Validation. Dalam penelitian ini hasil yang dicapai memiliki akurasi untuk tepat waktu sebesar 93% dan akurasi untuk terlambat sebesar 71%. dengan menggunakan metode Naïve Bayes yang semakin optimal dengan menentukan mahasiswa lulus tepat waktu atau terlambat

Kata kunci: akurasi, data mining, naïve bayes, tepat waktu, terlambat

1. PENDAHULUAN

Memprediksi prestasi akademik siswa sangat penting bagi suatu lembaga pendidikan karena program strategis tersebut dapat direncanakan dalam meningkatkan atau mempertahankan kinerja siswa selama masa studi di lembaga (Ibrohim, 2010). Perguruan tinggi saat ini dituntut untuk memiliki keunggulan bersaing dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana, dan manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing (Kovačić, 2006). Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari-hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Berdasarkan penelitian yang dibuktikan bahwa untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tingkat atau tahun ajaran pertama. Ada beberapa faktor yang mempengaruhi prediksi kelulusan mahasiswa yang sesuai dengan waktu studi, diantaranya : NEM SMA, status mahasiswa, jenis kelamin, umur mahasiswa, gaji orang tua, pekerjaan orang tua. Dalam menentukan Prediksi kinerja mahasiswa dengan akurasi yang tinggi bermanfaat untuk mengidentifikasi siswa dengan prestasi akademik (Suhartinah, 2010).

Lembaga pendidikan perguruan tinggi menghadapi masalah siswa yang terkait dengan tingkat kelulusan, perguruan tinggi dengan mahasiswa lebih tinggi tingkat retensinya dan cenderung memiliki tingkat kelulusan yang lebih tinggi dalam waktu empat tahun. Rata – rata tingkat retensi nasional mendekati 55% dan di beberapa perguruan tinggi kurang dari 20% dari mahasiswa pascasarjana (Nandeshwar, 2009).

Tujuan dari Penelitian ini untuk menghasilkan sistem pendukung keputusan untuk memprediksi kelulusan mahasiswa di tingkat perguruan tinggi.

2. METODOLOGI

2.1. Metode Penelitian

Metode penelitian yang dilakukan adalah metode penelitian eksperimen, dengan tahapan penelitian seperti berikut (Santoso, 2007):

- a. Pengumpulan Data (*Data Gathering*)

Penelitian ini menggunakan data mahasiswa Progdil Sistem Informasi Fakultas Teknik Universitas Muria Kudus untuk memprediksi kelulusan mahasiswa.

- b. Pengolahan Awal Data (*Data Preprocessing*)
Data perolehan di transformasi untuk mendapatkan atribut yang relevan dan sesuai dengan format input algoritma *naïve bayes*. Atribut yang digunakan meliputi umur, alamat, jenis Kelamin, status pekerjaan mahasiswa, status pernikahan mahasiswa, rata – rata IPK (Indeks Prestasi Kumulatif), jumlah SKS (Satuan Kredit Semester) dan status mahasiswa.
- c. Model/Metode Yang Diusulkan (*Proposed Model/Method*)
Metode yang diusulkan adalah metode perbandingan tingkat akurasi dari algoritma *naïve bayes* yang dapat digunakan untuk memprediksi kelulusan mahasiswa.
- d. Eksperimen dan Pengujian Metode (*Method Test and Experiment*)
Pengujian dilakukan dengan menggunakan sebagian data mahasiswa untuk *training* dan sebagian lagi sebagai data *testing*. Perhitungan dengan masing-masing algoritma akan diulang beberapa kali untuk mendapatkan besaran parameter terbaik.
- e. Evaluasi dan Validasi Hasil (*Result Evaluation and Validation*)
Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan Algoritma *Naïve Bayes*. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal. Pengukuran kinerja dilakukan dengan membandingkan nilai error hasil prediksi masing-masing algoritma sehingga dapat diketahui algoritma yang lebih akurat

2.2. Penerapan Metode Naïve Bayes

Metode Naïve Bayes yang digunakan untuk memprediksi kelulusan mahasiswa dengan menggunakan beberapa variable, antara lain jenis kelamin, umur, status pekerjaan mahasiswa, status pernikahan mahasiswa, rata-rata IPK, jumlah SKS dan status mahasiswa. Tabel 1 menunjukkan tabel untuk perhitungan prediksi kelulusan mahasiswa .

Tabel 1. Perhitungan prediksi kelulusan mahasiswa

Variabel	Sub Variabel	Perhitungan prediksi tepat waktu	Perhitungan prediksi terlambat
Jenis kelamin	Laki – Laki	0.35	0.60
	Perempuan	0.99	0.38
Umur	18 - 20	0.13	0.87
	21-22	0.32	0.19
	>22	0.28	0.51
Status Pekerjaan mahasiswa	Mahasiswa	0.49	0.14
	Bekerja	0.25	0.18
Status pernikahan mahasiswa	Menikah	0.28	0.18
	Belum menikah	0.44	0.15
Rata – rata IPK	< 2.00	0.05	0.24
	2.00 – 3.00	0.26	0.11
	>3.00	0.72	0.15
Jumlah SKS	<100	0.27	0.73
	100-110	0.33	0.66
	>110	0.53	0.47
Status Mahasiswa	Aktif	0.49	0.14
	Cuti	0.07	0.93

2.3. Evaluasi

Evaluasi kinerja *K-Means clustering* ini menggunakan *confusion matrix*. Evaluasi tersebut bertujuan untuk menilai berapa persen kinerja dari sistem dapat mengenali dua kelas yang terdapat pada data uji yang telah disediakan sewaktu proses *pengclusteran*. Kinerja sistem yang dievaluasi dengan menghitung nilai akurasi. Dari perhitungan akurasi akan diketahui sejauh mana algoritma *K-Means* dapat mendeteksi Kolesterol manusia berdasarkan segmentasi citra iris mata.

3. HASIL DAN PEMBAHASAN

3.1. Tahapan – tahapan dalam preprocessing data mining

- a. *Cleaning* atau Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*)
Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak relevan dengan hipotesa data mining yang kita miliki. Pembersihan data yang tidak relevan akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
- b. Integrasi data (penggabungan data dari beberapa sumber)
Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas - entitas yang unik. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya. Hasil integrasi data sering diwujudkan dalam sebuah data *warehouse*.
- c. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di *mining*)
Beberapa teknik data *mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Karenanya data berupa angka numeric yang berlanjut perlu dibagi – bagi menjadi beberapa interval. Proses ini sering disebut *binning*. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data *mining* yang nantinya karena ada beberapa karakteristik dari teknik – teknik data *mining* tertentu yang tergantung pada tahapan ini.
- d. Aplikasi teknik data mining
Aplikasi teknik data *mining* sendiri hanya merupakan salah satu bagian dari proses data *mining*. Beberapa teknik data *mining* sudah umum dipakai. Ada kalanya teknik-teknik data *mining* umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data *mining* di bidang tertentu atau untuk data tertentu.
- e. Evaluasi pola yang ditemukan (untuk menemukan yang menarik/bernilai)
Dalam tahap ini hasil dari teknik data *mining* berupa polapola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.
- f. Presentasi pola yang ditemukan untuk menghasilkan aksi
Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data *mining*. Karenanya presentasi hasil data *mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data *mining* (Vercellis,2009).

3.2. Metode Naïve Bayes

Teorema keputusan *bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan probabilitas dan ongkos yang ditimbulkan dalam keputusan keputusan tersebut. Ide dasar dari *bayes* adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek. Misalkan terdapat dua jenis objek dengan kemungkinan kemunculan random, selanjutnya ingin diprediksi objek apa yang akan lewat selanjutnya(Santoso, 2007).

Objek pertama diwakili oleh h_1 dan objek kedua diwakili oleh h_2 . karena apa yang akan muncul bersifat probablistik maka h adalah suatu variabel yang harus di deskripsikan secara probabilistik. Selanjutnya *probabilitas a priori*, $(P h_1)$ dan $(P h_2)$ masing-masing melihat peluang munculnya objek 1 dan objek 2. Walaupun probabilitas kemunculan kedua objek tersebut tidak diketahui dengan pasti tapi setidaknya dapat diestimasi dari data yang tersedia. Misalkan N adalah jumlah total kedua objek, kemudian N_1 dan N_2 masing masing menyatakan jumlah objek 1 dan objek 2, selanjutnya

$$P(h_1) \approx \frac{N_1}{N} \text{ dan } P(h_2) \approx \frac{N_2}{N} \dots\dots\dots(1)$$

Untuk mengestimasi *probabilitas* digunakan rumus, yang sering disebut m-estimate :

$$P(a_i|v_j) = \frac{c_{ij} + m \cdot P(a_i)}{n_j + m} \dots\dots\dots(2)$$

Prediksi kelulusan mahasiswa di Program studi Sistem Informasi digunakan untuk meningkatkan mutu dan kualitas mahasiswa, juga memberikan motivasi dan dukungan kepada mahasiswa agar dapat menempuh gelar sarjana dalam waktu yang telah ditentukan, yaitu sekitar 4 tahun dalam 8 semester. Penelitian yang berjudul Prediksi kelulusan mahasiswa menggunakan metode *Naïve Bayes* dengan 8 parameter telah mendapatkan akurasi yang cukup tinggi.

3.3. Hasil Prediksi

a. Perhitungan tepat waktu:

$$1) P (mahasiswa \text{ tepat waktu} / \text{ total mahasiswa data training }) * P (\text{ Laki - Laki} | \text{ Tepat waktu}) * P (\text{ Dalam kota} | \text{ Tepat waktu}) * P (18-20 | \text{ Tepat waktu}) * P (\text{ Mahasiswa} | \text{ Tepat waktu}) * P (\text{ Belum menikah} | \text{ Tepat waktu}) * P (2.80 | \text{ Tepat waktu}) * P (116 | \text{ Tepat waktu}) * P (\text{ aktif} | \text{ Tepat waktu})$$

$$2) \text{ Jadi, } 0.43 * 0.35 * 0.44 * 0.13 * 0.49 * 0.44 * 0.37 * 0.51 * 0.49 = 0.00017$$

b. Perhitungan terlambat:

$$1) P (mahasiswa \text{ terlambat dalam data training} / \text{ total mahasiswa ata training }) * P (\text{ Laki - Laki} | \text{ Terlambat}) * P (\text{ Dalam kota} | \text{ Terlambat}) * P (18-20 | \text{ Terlambat}) * P (\text{ Mahasiswa} | \text{ Terlambat}) * P (\text{ Belum menikah} | \text{ Terlambat}) * P (2.80 | \text{ Terlambat}) * P (116 | \text{ Terlambat}) * P (\text{ aktif} | \text{ Terlambat})$$

$$2) \text{ Jadi, } 0.57 * 0.60 * 0.52 * 0.87 * 0.14 * 0.15 * 0.47 * 0.14 = 0.0003$$

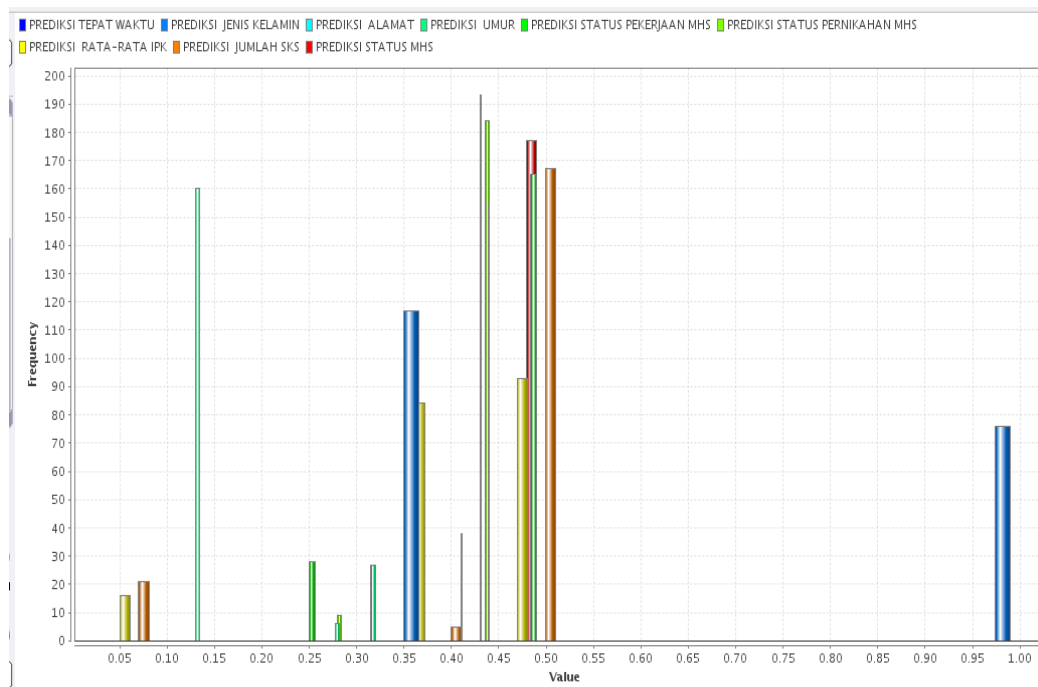
c. Prediksi tepat waktu = 0.00017

d. Prediksi terlambat = 0.0003

e. Penentuan prediksi $0.00017 > 0.0003$

Jadi, prediksi mahasiswa tersebut adalah **Tepat waktu**

Dari gambar grafik dibawah ini ditunjukkan bahwa variabel yang paling tinggi nilai prediksinya adalah variabel status pekerjaan mahasiswa dan status mahasiswa. Jadi, alasan mahasiswa dapat lulus tepat waktu jika status pekerjaan tersebut adalah mahasiswa dan status mahasiswa tersebut adalah aktif. Gambar 1 dibawah ini menunjukkan grafik perbandingan variable naïve bayes.



Gambar 1. Grafik perbandingan variabel naïve bayes

4. KESIMPULAN

Dari hasil penelitian yang dilakukan dari tahap awal sampai dengan tahap pengujian penerapan metode naïve bayes untuk proses prediksi kelulusan mahasiswa, didapatkan kesimpulan bahwa:

- (1) Penerapan metode algoritma naïve bayes menggunakan 8 (delapan) parameter yaitu jenis kelamin, umur, alamat, status pekerjaan mahasiswa, status pernikahan mahasiswa, rata – rata IPK, jumlah SKS dan status mahasiswa dengan tingkat akurasi 93%.
- (2) Metode Naïve Bayes digunakan untuk menghitung probabilitas dengan kemungkinan tepat waktu atau terlambat dalam menentukan prediksi kelulusan mahasiswa.
- (3) Dalam pengolahan data mining seharusnya dilakukan tahapan-tahapan dalam data mining, yaitu cleaning data, integrasi data, transformasi data, aplikasi data mining dan evaluasi pola yang ditemukan dalam data mining.

DAFTAR PUSTAKA

- Christos Pierrakeas Dimitris Kalles, "Analyzing student performance in distance learning with genetic algorithms and decision trees," *Hellenic Open University*.
- Daliela Rusli Zaidah Ibrohim, "PREDICTING STUDENTS' ACADEMIC PERFORMANCE: COMPARING ARTIFICIAL," *Annual SAS Malaysia Forum*, September 2006.
- Ernastuti Marselina Silvia Suhartinah, 2010, "GRADUATION PREDICTION OFGUNADARMA UNIVERSITY STUDENTS USING ALGORITHM AND NAIVE BAYES C4.5 ALGORITHM,".

- M. Ramaswami and R. Bhaskaran, 2010, "A CHAID Based Performance Prediction Model in Educational Data Mining," *IJCSI International Journal of Computer Science*, vol. Vol. 7, no. Issue 1, p. 1, January.
- Subodh Chaudhari Ashutosh Nandeshwar, April 2009, "Enrollment Prediction Models Using Data Mining,".
- Zlatko J. Kovačić, 2010, "Early Prediction of Student Success; Mining Students Enrolment Data," *Proceedings of Informing Science & IT Education Conference (InSITE)*.