

ALGORITMA BAYESIAN CLASSIFICATION UNTUK MEMPREDIKSI HEREGISTRASI MAHASISWA BARU DI STMIK WIDYA PRATAMA

Devi Sugianti

Program Studi Sistem Informasi, STMIK WidyaPratama

Jl. Patriot 25 Pekalongan

Telp (0285)427816

email : dv1398@yahoo.com

ABSTRAK

Setiap tahun akademik STMIK Widya Pratama melakukan kegiatan pendaftaran. Pada tahun 2011 jumlah pendaftar sebanyak 913 dengan jumlah heregristrasi sebanyak 658 maka 255 mahasiswa tidak melakukan heregristrasi. Bayesian classification merupakan salah satu metode dari data mining yang digunakan untuk mengklasifikasi data. Metode Bayesian classification ini digunakan untuk mengetahui kemungkinan pengunduran diri seorang calon mahasiswa dapat diketahui lebih dini agar pihak perguruan tinggi dapat melakukan tindakan yang perlu untuk mempertahankan calon mahasiswa. Dalam memprediksi heregristrasi mahasiswa baru dengan atribut asal kota, gelombang, progd, dan status tes. Keakuratan yang telah dilakukan dengan menggunakan metode bayesian classification 78 %. Pengetesan data dengan kota asal Pemalang, gelombang 3 Progd TI, status tes adalah bebas tes. Terdapat 9 mahasiswa yang melakukan pendaftaran, dengan yang melakukan heregristrasi 7 mahasiswa.

Kata Kunci: memprediksi, heregristrasi, Bayesian classification

1. PENDAHULUAN

1.1 Latar Belakang

Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Widya Pratama Pekalongan berdiri tahun 2002 berdasarkan SK Mendiknas Nomor : 149/D/O/2002 yang diselenggarakan oleh Yayasan Widya Pratama. STMIK Widya Pratama mempunyai 4 Program Studi yaitu Teknik Informatika, Sistem Informasi, Manajemen Informatika dan Komputerisasi Akuntansi. Setiap tahun akademik STMIK Widya Pratama melakukan kegiatan penerimaan mahasiswa baru. Pada tahun akademik 2011/2012 terdapat 913 pendaftar dengan data seperti berikut:

Tabel 1. Data Pendaftaran mahasiswa tahun 2011

Program Studi	Jumlah Pendaftar
MI	171
KA	62
SI	148
TI	532
Total	913

Dari jumlah pendaftar 913, akan tetapi yang melakukan heregristrasi mahasiswa baru 658 mahasiswa. Ada 255 mahasiswa tidak melakukan heregristrasi. Adapun rincian untuk heregristrasi mahasiswa baru adalah sebagai berikut:

Tabel 2. Data Heregristrasi mahasiswa baru tahun 2011

Program Studi	Jumlah Heregristrasi
MI	123
KA	40
SI	103
TI	392

Jika kemungkinan pengunduran diri seseorang calon mahasiswa baru dapat diketahui lebih dini, maka pihak manajemen Perguruan Tinggi dapat melakukan tindakan-tindakan yang diperlukan untuk mempertahankan calon-calon mahasiswa tersebut (Kusrini dan Luthfi, 2009). Data mining dapat melakukan analisa dari kasus lama. Menurut Gratner Group data mining adalah sesuatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti statistic dan matematika (Iorse, 2005). Bayesian classification merupakan pengklasifikasi *statistic* yang dapat digunakan untuk memprediksi probabilitas kenggotaan suatu *class*. Bayesian classification didasarkan pada teorema bayes yang memiliki kemampuan kalsifikasi serupa

dengan *decision tree dan neural network* (kusrini,2009). Penelitian ini dilakukan untuk mengetahui klasifikasi mahasiswa yang melakukan herregistrasi atau tidak melakukan herregistrasi dengan algoritma Bayesian classification

1.2 Landasan Teori

1.2.1 Data Mining

data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Turban. 2005). Pengelompokan data mining berdasarkan tugas yang dapat dilakukan, antara lain (larose:2005) :

1. Deskripsi

Deskripsi adalah Menggambarkan pola dan kecenderungan yang terdapat dalam data yang memungkinkan memberikan penjelasan dari suatu pola atau kecenderungan tersebut.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, akan tetapi variabel target estimasi lebih kearah numeric dari pada kearah kategori.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, akan tetapi dalam prediksi nilai dari hasilakan terwujud di masa yang akan datang.

4. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan memprediksikan kelas untuk data yang tidak diketahuikelasnya.

5. Clustering

Clustering atau analisis cluster adalah proses pengelompokan satu set benda-benda fisik atau abstrak kedalam kelas objek yang sama.

6. Asosiasi

Asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu dalam dunia bisnis lebih umum disebut analisis keranjang belanja

1.2.2. Bayesian Classification

Bayesian Classification adalah Pengklasifikasian statistic yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi seperti dengan *decision tree dan neural network* (kusrini:2009)

Teorema Bayes memiliki bentuk umum seperti berikut:

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

Keterangan:

X = data dengan class yang belum diketahui

H = hipotesa data X merupakan satu class spesifik

P(H|X) = probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)

P(H) = probabilitas hipotesis H (*prior probability*)

P(X|H) = probabilitas X berdasarkan kondisi hipotesis H

P(X) = probabilitas dari X

2. METODE PENELITIAN

2.1 Jenis Penelitian

Langkah-langkah penelitian yang dilakukan adalah perumusan masalah, penentuan teknik yang akan digunakan, pemrosesan data, transformasi data, analisa hasil dan penarikan kesimpulan.

Preproses data dilakukan, karena data yang didapatkan masih dalam keadaan berantakan dan harus diolah lagi terlebih dahulu, sebelum memasuki proses data selanjutnya. Setelah data ditransormasi atau diolah, kemudian masuk kepada tahap analisis data. Dari data *training* yang telah dianalisis, kemudian dibuat prediksi klasifikasi untuk kemungkinan-kemungkinan yang akan datang bagi data *testing* jenis yang sama.

2.2 Populasi dan Sampel

Dari data pendaftaran tahun akademik 2011/2012 terdapat jumlah pendaftar 913 calon mahasiswa. STMIK Widya Pratama terdapat 2 jenjang yaitu D3, dan S1. Untuk Program Studi terdapat 4 program studi, dan 3 gelombang pendaftaran, serta ada 2 status yaitu tes dan tidak tes.

Dalam memprediksi heregistrasi mahasiswa baru dibutuhkan variable: kota, progdi, st daftar, gelombang.

Tabel 3. pendaftaran berdasarkan kota asal

Kota	Tidak Heregistrasi	Heregistrasi	Total
BANJARNEGARA		1	1
BATANG	57	156	213
KAB.PEKALONGAN	23	54	77
PEKALONGAN	145	393	538
PEMALANG	30	52	82
TEGAL		2	2
TOTAL	255	658	913

Tabel 4. Pendaftaran berdasarkan program studi

Progdi	Tidak Heregistrasi	Heregistrasi	Total
MI	48	123	171
KA	22	40	62
SI	45	103	148
TI	140	392	532
Total	255	658	913

Tabel 5. Pendaftaran berdasarkan status tes

Status Daftar	Tidak Heregistrasi	Heregistrasi	Total
BEBAS TEST	133	422	555
TEST	122	236	358
Total	255	658	913

Tabel 6. Pendaftaran berdasarkan gelombang

gelombang	Tidak Heregistrasi	Heregistrasi	Total
1	95	94	189
2	122	409	531
3	38	155	193
Total	255	658	913

2.3 Metode Pengumpulan Data

Dalam pengambilan data primer dari data pendaftaran yang telah terintegrasi dengan sistem pendaftaran dan pengambilan data juga dari data heregistrasi yang berada pada sistem BAAK.

3. HASIL DAN PEMBAHASAN

Terdapat 2 *class* yaitu Ya (heregistrasi) dan Tidak (tidak heregistrasi). Misal data X belum diketahui *class* nya

Kota : Pemalang
 ProgdI : Teknik Informatika
 Status daftar : Bebas test
 Gelombang : 3

Penyelesaian

P(Ci) merupakan *prior probability* untuk setiap *class* berdasar data contoh:

$$P(Ya) = 658/913 = 0,72$$

$$P(Tidak) = 255/913 = 0,28$$

Hitung P(X|Ci) untuk i=1,2

P(asal kota = "Pemalang" | heregristrasi = "Ya") =

$$\frac{52}{658} = 0,08$$

P(asal kota = "Pemalang" | heregristrasi = "Tidak") =

$$\frac{30}{255} = 0,12$$

P(program studi = "TI" | heregristrasi = "Ya") =

$$\frac{392}{658} = 0,59$$

P(program studi = "TI" | heregristrasi = "Tidak") =

$$\frac{22}{255} = 0,55$$

P(Status daftar = "bebas test" | heregristrasi = "Ya") =

$$\frac{422}{658} = 0,64$$

P(status daftar = "bebas test" | heregristrasi = "Tidak") =

$$\frac{133}{255} = 0,52$$

P(gelombang = "3" | heregristrasi = "Ya") =

$$\frac{155}{658} = 0,23$$

P(gelombang = "3" | heregristrasi = "Tidak") =

$$\frac{38}{255} = 0,15$$

P(X | heregristrasi = "Ya")

$$= 0,08 \times 0,59 \times 0,64 \times 0,23$$

$$= 0,00695$$

P(X | heregristrasi = "Tidak")

$$= 0,12 \times 0,55 \times 0,52 \times 0,15$$

$$= 0,00512$$

P(X | heregristrasi = "ya") P(heregristrasi = "ya")

$$0.00695 \times 0,72 = 0.005$$

$$P(X|\text{heregistrasi}=\text{"tidak"}) P(\text{heregistrasi}=\text{"tidak"})$$

$$0.00512 \times 0,28 = 0.001$$

Dari data class yang baru dengan maka dapat diketahui jika mahasiswa tersebut akan melakukan heregistrasi

4. Kesimpulan

STMIK widya pratama melakukan pendaftaran mahasiswa baru untuk tiap tahun akademik. Pada tahun akademik 2011/2012 jumlah pendaftara 913, dan yang melakukan heregistrasi 658, yang tidak melakukan heregistrasi 255. Jika kemungkinan pengunduran diri diketahui lebih dini maka pihak manajemen dapat mengantisipasi. Bayesian Clasification merupakan salah satu algoritma yang ada di data mining untuk mengklasifikasikan. Bayesian classification dapat digunakan untuk memprediksi heregistrasi mahasiswa baru, dengan atribut seperti: kota asal, program studi, status daftar, gelombang. Dari data yang ingin deketahui kelasnya dengan variable kota :pemalang, program studi TI, status daftar: bebas tes, gelombang:3. Maka dapat diprediksi bahwa calon mahasiswa tersebut melakukan hergistrasi. Dari data yang diperoleh di bagian BAAK terdapat data pendaftaran dan heregistrasi sebagai berikut:

Tabel. 7. Data pendaftaran dan hergistrasi tahun akademik 2011/2012

no_daftar	kota_kec	prodi	st_daftar	gelombang	registrasi
11-040-0723	PEMALANG	240	BEBAS TEST	3	Ya
11-040-0747	PEMALANG	240	BEBAS TEST	3	Ya
11-040-0762	PEMALANG	240	BEBAS TEST	3	Tidak
11-040-0783	PEMALANG	240	BEBAS TEST	3	Ya
11-040-0807	PEMALANG	240	BEBAS TEST	3	Ya
11-010-0817	PEMALANG	240	BEBAS TEST	3	Ya
11-040-0869	PEMALANG	240	BEBAS TEST	3	Ya
11-040-0891	PEMALANG	240	BEBAS TEST	3	Ya
11-040-0897	PEMALANG	240	BEBAS TEST	3	Tidak

Terdapat 9 pendaftar dengan *class* yang telah diketahui. Dari table diatas 7 mahasiswa melakukan heregistrasi dan 2 siswa tidak melakukan heregistrasi, keakuratan Bayesian classification adalah 78 % yang dibuktikan dengan data diatas.

7. Daftar Pustaka

- Kusrini and Emha Taufiq Lutfhi, 2009. *Algoritma Data Mining*. Andi Offset. Yogyakarta.
 Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introducing to Data Mining*.: John Willey & Sons, Inc,
 Turban, E, dkk 2005. *Decicion Support Systems and Intelligent Systems*. Yogyakarta: Andi Offcest