ANALISIS KELULUSAN MAHASISWA MENGGUNAKAN ALGORITMA NAÏVE BAYES

JONTINUS MANULLANG[1)], JONAS FRANKY R PANGGABEAN[2)]

[1)]Komputerisasi Akuntansi, AMIK Medicom, Medan
Jl. Darat No. 74, Medan 20153, Sumatera Utara, Indonesia

Jhoe6590@gmail.com

[2)]Manajemen Informatika, AMIK Medicom, Medan
Jl. Darat No. 74, Medan 20153, Sumatera Utara, Indonesia

jonasfrankypanggabean@gmail.com

ABSTRAK

Dari data yang dikumpulkan diputuskan Penentuan waktu, algoritma genetika,jadwal di university classifier mengklasifikasi statistik untuk memetakan kemungkinan keanggotaan kelas tertentu merupakan fungsi pengklasifikasian bayesian penelitian ini akan berusaha mengemukakan data klasifikasi prediksi kelulusan mahasiswa baru menggunakan metode algoritma Naive Bayes Kemampuan yang baik untuk memperkirakan kesempatan di hari yang akan datang berdasarkan peristiwa atau data di waktu yang terdahulu Dari proses penilaian yang telah dikerjakan didapat data yang terklasifikasi dengan baik berdasarkan kumpulan pilihan yang telah lulus terlebih dahulu pada pilihan pertama, pilihan kedua dan tidak lolos algoritma berjumlah 93.6288% atau berjumlah 338 data dan data rahasia, tetapi tidak cocok dengan kelas yang diperkirakan (instance salah diklasifikasikan) yang seharusnya merupakan grup dari dua atau opsi Lulus tetapi disertakan pada kelompok Pilihan I sebanyak 6,3712% atau berjumlah 23 data. Nilai Persentase ketepatan menampilkan efektivitas dataset Penerimaan diterapkan dengan metode Naïve Bayes Classification,yang mencapai 94%

Kata kunci: Naive bayes, klasifikasi, data mining

ABSTRACT

From the data collected, it is decided that the timing, genetic algorithm, schedule at the university classifier to classify statistics to map the possibility of certain class membership is a bayesian classification function. future days based on events or data in the past. From the assessment process that has been carried out, data is obtained that is well classified based on a collection of choices that have passed first in the first choice, the second choice and did not pass the algorithm totaling 93,6288% or totaling 338 data and confidential data, but does not match the estimated class (instance misclassified) which should be a group of two or the Pass option but is included in the Option I group as much as 6.3712% or a total of 23 data. The accuracy percentage value displays the effectiveness of the Acceptance dataset applied by the Naïve Bayes Classification method, which reaches 94%

Keywords: *Naive bayes, classification, data mining*

1. Pendahuluan

Klasifikasi merupakan salah satu masalah dasar dan tugas dari data mining (Zhangdkk., 2004), dalam . klasifikasi dibuat dari sekumpulan data dengan kelas yang di tentukan. Performa pengklasifikasi diukur dengan ketepatan atau tingkat galat (Walporedkk., 1995). Pengklasifikasi Bayesian memprediksi probabilitas keanggotaan kelas tertentu (Hangdkk., 2006). Masing-masing kelompok atribut yang ada, dan menentukan , kelas mana yang paling optimal (Hajjaratie, 2011).

Penelitian ini menggunakan metode algoritma naïve bayes untuk memprediksi peluang kelulusan mahasiswa baru. Dengan maksud adalah diterimanya seorang mahasiswa pada salah satu program studi di Perguruan Tinggi. Studi kasus dilakukan pada data kampus AMIK Medicom.

2. Tinjauan Pustaka

2.1. Algoritma Naive Bayes

Pada machine learning dan data mining, Naïve Bayes adalah algoritma yang pembelajaran induktif paling efektif dan efisien. Asumsi keidependenan atribut pada performa naïve bayes yang kompetitif dalam klasifikasi. Asumsi keidependenan sebenarnya jarang terjadi, namun asumsi keidependenan atribut tersebut dilanggar performa pengklasifikasian naïve bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian.

Metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes Naïve Bayes yang merupakan pengklasifikasian dengan memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumny. Setiap baris/dokumen I diasumsikan sebagai vector dari nilai atribut dimana tiap nilai-nilai menjadi peninjauan atribut Xi (i€[1,n])). Setiap baris mempunyai label kelas ci € {c1,c2,…,ck} sebagai nilai variabel kelas C, sehingga untuk melakukan klasifikasi dapat dihitung nilai probabilitas p(C=ci|X=xj) , dikarenakan pada Naïve Bayes diasumsikan setiap atribut saling bebas, maka persamaan yang didapat adalah sebagai berikut :

- Peluang p(C=ci|X=xj) menunjukkan peluang bersyarat atribut Xi dengan nilai xi diberikan kelas c,
- Ketika atribut Xi bertipe kuantitatif maka peluang p(X=xi|C=cj) akan sangat kecil sehingga membuat persamaan peluang tersebut bertipe kuantitatif. ada beberapa pendekatan yang dapat digunakan seperti distribusi normal (Gaussian) :

$$\hat{f} = N(X_i;\mu_c,\sigma_c) = \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(X_j-\mu_c)^2}{2\sigma_c^2}}, \qquad (1)$$

Ataupun kernel density estimation (KDE) :

$$\hat{f} = \frac{1}{n_c}\sum_j N(X_i;\mu_{ij},\sigma_c), \qquad (2)$$

Terdapat cara kerja yang beda untuk mengerjakan atribut kuantitatif yaitu diskritisasi daripada dua cara distribusi tersebut. atribut numerik X diubah menjadi atribut nominal X* terjadi saat pre processing data yang merupakan proses diskritisasi. Asumsi Pendekatan distribusi [Dougherty] tidak terlalu baik dari hasil klasifikasi naïve bayes ketika atribut numeric didiskritisasi. Nilai-nilai numerik akan dikelompokkan ke nilai nominal dalam bentuk interval yang tetap memperhatikan tingkatan dari tiap-tiap nilai numerik yang dikelompokkan. penggambaran perhitungan Naive Bayesnya seperti berikut:

Tabel 1. Pehitungan Naives Bayes

	Interval 1 (i1) Interval 2 (i2)
Kelas 1 (c1)	Rumus Naive Bayes : $p(I=i_j\|C=c_i) = \frac{p(I=ij)\,p(C=ci\|I=ij)}{p(C=ci\|)}$
Kelas 2 (c2)	ket : p(I=ij|C=ci) : peluang interval i ke-j untuk kelas ci p(C=ci|I=ij) : peluang kelas ci pada interval i ke-j p(I=ij) : peluang sebuah interval ke-j pada semua interval yang terbentuk p(C=ci) : peluang sebuah kelas ke-i untuk semua kelas yang ada di dataset

Berikut adalah algoritma dari metode Naïve Bayes :

a. hitung jumlah kelas
b. hitung jumlah kasus dan class class yang sama
c. kalikan semua hasil variable TEPAT & TERLAMBAT
d. Bandingkan hasil class TEPAT & TERLAMBAT

Kekuatan dan kelemahan Naïve Bayesian: Kekuatan:

a. Mudah diimplementasi.
b. Memberikan hasil yang baik untuk banyak kasus.

Kelemahan:

mengasumsi antar fitur tidak terkait (independent) dalam realita, keterkaitan itu ada Keterkaitan tersebut tidak dapat dimodelkan oleh Naïve Bayesian Classifier.

3. Metode Penelitian

3.1. Data

Pada proses pengklasifikasian ini menggunakan 2 dataset yang diolah yaitu Pertama : Data Penerimaan Mahasiswa Baru yang memiliki 10 atribut dan 363 record yang terdiri nama, tempat tanggal lahir, asal sekolah, tahun lulus, nama orang tua, nomor ujian, nilai tes tertulis, nilai rata-rata, ket. Lulus, lulus pilihan. Kedua, Data hasil Ujian Masuk Perguruan tinggi yang terdiri dari 5 atribut dan 362 record yang terdiri dari Nomor Ujian, Nilai tes tulisan, rata-rata dan keterangan Lulus.

Tabel 2. Data Hasil Ujian

No Ujian	Tes Tertulis		Total	Keterangan
	Bahasa Indonesia dan Inggris	Pengatahuan Umum dan Matematika		
2001	30	30	60	lulus
2002	21	23	44	lulus
2003	34	13	47	lulus
2004	35	34	69	lulus
2005	15	17	32	lulus
2006	18	20	38	lulus
2007	20	25	45	lulus
2008	12	30	42	lulus
2009	20	15	35	lulus
2010	30	15	45	lulus

3.2. Praproses

Pada awal proses KDD pembersihan data merupakan hal yang harus dilakukan. Proses pembersihan data merangkum beberapa hal seperti mempelajari data yang berubah-ubah (missing value), hilangnya beberapa data dan data yang berulang-ulang(rerudant data). Ketentuan yang harus dipenuhi terlebih dahulu untuk proses data mining yang akan menghasilkan dataset yang bersih dan siap digunakan pada tahap mining data adalah semua atribut pada dua kumpulan data (tabel) harus dibersihkan terlebih dahulu. Jika terdapat salah satu atribut nilai yang hilang maka record yang dikehendaki tersebut akan dihapus karena record tersebut dinilai kehilangan data, Jika ada record dalam dataset yang sama berisi nilai yang sama maka record yang dikehendaki tersebut akan dihapus karena akan memberikan informasi yang tidak memiliki arti jika tetap dipertahankan. Pada proses ini bukan hanya membersihkan data yang mengandung missing value saja akan tetapi terhadap data yang tidak konsisten juga dilakukan. Data pada penelitian ini sudah menggunakan data yang konsisten. tidak ada data yang dicleaning karena dua kelompok data (tabel) diambil seluruhnya, maka jumlah atribut dan record pada kelompok data (tabel) adalah tetap. Pada tahap ini data sudah bersih dan siap untuk diproses pada tahap selanjutnya yaitu integrasi data. Data hasil cleaning kami sajikan pada tabel 3.

3.3. Implementasi perhitungan pada Naïve Bayes Clasification

Hasil data cleaning telah rampung maka selanjutnya, di implementasikan dengan formula Naïve Bayes Classifier. cara kerja dari proses perhitungan Naïve Bayes Classifier diawali dengan mengabil data sample atau contoh data.

Tabel 3 Data tes mahasiswa

NO	JURUSAN SEKOLAH	PILIHAN 1	PILIHAN 2	NILAI RATA	KET	PILIHAN
1	IPA	MI	KA	60	lulus	MI
2	IPA	MI	TK	44	lulus	MI
3	IPS	KA	MI	47	lulus	KA
4	TKJ	TK	MI	69	lulus	TK

5	TKJ	TK	MI	32	lulus	TK
6	TKJ	TK	MI	38	lulus	TK
7	IPS	KA	TKJ	45	lulus	KA
8	IPS	KA	MI	42	lulus	KA
9	IPS	MI	KA	35	lulus	MI
10	IPA	MI	TKJ	45	lulus	MI

dari tabel diatas dapat dihitung menggunakan formula Naïve Bayes Clasification, dengan cara :
masalah klasifikasi, yang dihitung adalah P(H|X), yaitu peluang bahwa hipotesa benar (valid) untuk data sample X yang diamati:
Dimana :
X adalah data sampel dengan kelas (label) yang tidak diketahui
H merupakan data dengan klas (label) C.
P(H) adalah peluang dari hipotesa H P(X) data sampel yang diamati
P(X|H) adalah peluang data sampel X, diasumsikan hipotesa benar (valid).
Jadi Rumusnya adalah

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

Sehingga,
P(Pil1) = 10/15 = 0.67
P(Pil2) = 5/15 = 0.34
P(Prodi=IPA|Pil1) = 3/10 = 0.3
P(Prodi=IPA|Pil2) = 0/3 = 0.00
P(Pilihan Pertama=Sistem Informasi|Pil1) = 7/10 = 0.7
P(Pelican Pertama=Sistem Informasi|Pil2) = 3/3 = 1.00
P(Pilihan Kedua=Ilmu Pemerintahan|Pil1) = 7/10 = 0.7
P(Pilihan Kedua=Ilmu Pemerintahan|Pil2) = 3/3 = 1.00
P(Nilai Rata=55.83|Pil1) = 1/10 = 0.10
P(Nilai Rata=55.83|Pil1) = 0/3 = 0.00

Kemudian,
P(Pil1) P(IPA|Pil1) P(Sistem Informasi|Pil1)
P(Ilmu Pemerintahan|Pil1) P(55.83|Pil1)
= 0.67 x 0.30 x 0.70 x 0.30 x 0.10
= 0.004221 P(Pil2)

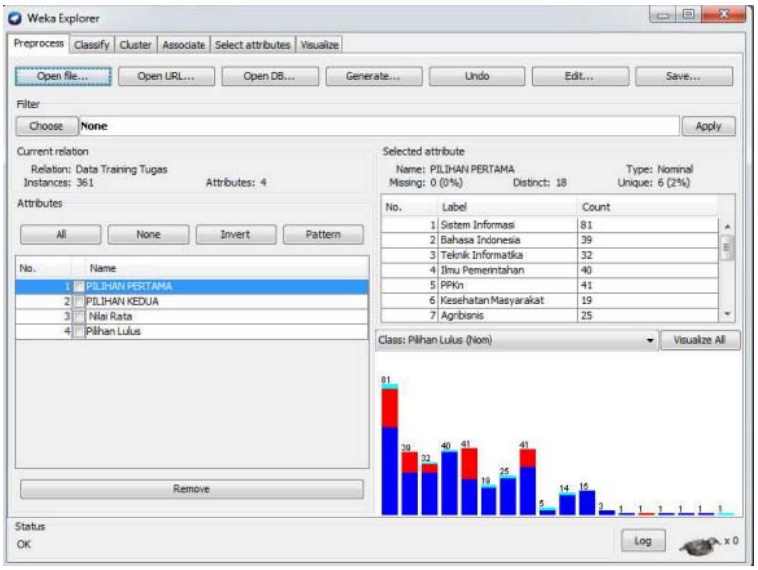
P(IPA|Pil2) P(Sistem Informasi|Pil2)
P(Ilmu Pemerintahan|Pil2) P(55.83|Pil2)
= 0.34 x 0 x 0.1 x 1 x 0
= 0

Jadi Keputusan Pilihan Lulus adalah Pil 1 (Pertama)

4. Hasil

4.1. Implementasi pada Aplikasi WEKA

Luaran pilihan lulus akan ditelusuri dengan weka dengan menggunakan atribut dari dataset. Atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata dikerjakan berdasarkan pengelompokan pilihan lulus. Adapun tingkatan praproses pada weka adalah sebagai berikut:



Gambar 1. Praproses Weka Tools

Luaran pilihan lulus pada dataset diproses dengan menggunakan teknik classifier naïve bayes updateable. luaran menemukan struktur yang tidak ada sebelumnya dalam data, yang menghubungkan struktur data yang sudah ada dengan data yang belum ada sebelumnya merupakan training set dari jenis test yang dilakukan. Pada dataset yang tersedia Classifier panel memungkinkan untuk membentuk dan menjalankan pilihan pengklasifikasian yang dikehendaki.

Alam menetapkan kelompok lulus pilihan pertama dan pilihan kedua proses klasifikasi dipengaruhi oleh atribut-atribut yang ditetapkan. luaran class Pilihan Lulus yang terdiri dari 15 kategori pilihan lulus dari hasil klasifikasi dapat dilihat dengan diagram batang. Hasil dari diagram akan menampilkan class Pilihan pertama yang lebih tinggi yaitu 274 data dan Pilihan Kedua sebanyak 73 data serta yang tidak Lulus sebanyak 14 data.

5. Kesimpulan

Memperkiraan peluang dimasa yang akan datang dapat diklasifikasikan dengan naïve bayes dengan metode probabilitas dan statistik bersumber dari hal maupun pengalaman yang terjadi sebelumnya. Penerimaan mahasiswa baru yang diterapkan kedalam metode naïve bayes classification menunjukkan keakuratan nilai presentase berdasarkan dataset yang digunakan. Dalam penelusuran karakteristik atribut dari dataset dapat menggunakan aplikasi weka yang diimplementasikan dengan naïve bayes berdasarkan dataset dengan luaran pilihan lulus.Pengelompokkan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata.

Daftar Pustaka

Fithri, D. L., & Darmanto, E. (2014). SISTEM PENDUKUNG KEPUTUSAN UNTUK MEMPREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES. Prosiding SNATIF, ISBN: 978-602-1180-04-4.

Samponu, Y. B., & Kusrini. (2017). OPTIMASI ALGORITMA NAIVE BAYES MENGGUNAKAN METODE CROSS VALIDATION UNTUK MENINGKATKAN AKURASI PREDIKSI TINGKAT KELULUSAN TEPAT WAKTU. Jurnal ELTIKOM, Vol. 1 No. 2, ISSN 2598-3245 (Print).

Syarli, & Muin, A. A. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). Jurnal Ilmiah Ilmu Komputer, Vol. 2, No. 1 (P) ISSN 2442-4512.

Testiana, G. (2018). Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah. JUSIFO (Jurnal Sistem Informasi), Vol. 02, No.1 ISSN: 2460-0921.

Han J, Kember M., 2006, Data Mining: Concepts and Techniques, Elsevier Inc. All rights reserved

Hadjaratie, L., 2011, Jaringan Saraf Tiruan Untuk Prediksi Tingkat Kelulusan Mahasiswa Diploma Program Studi Manajemen Informatika Universitas Negeri Gorontalo, Thesis, Postgraduated IPB, Bogor

Rennie J, Shih L, Teevan J, and Karger D. Tackling 2003, The Poor Assumptions of Naive Bayes Classifiers. In Proceedings of the Twentieth International Conference on Machine Learning (ICML).

Walpole, E. R., Myers, R. H, 1995, Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuan, Edisi ke-4. Bandung, ITB.

Zhang, Harry, 2004, The Optimality of Naive Bayes. FLAIRS conference.