

Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)

Syarli

Program Studi Teknik Informatika

Fakultas Ilmu Komputer Universitas AL Asyariah Mandar

Asrul Ashari Muin

Program Studi Sistem Informasi

Fakultas Sains dan Teknologi Universitas Islam Negeri Makassar

Abstract

Timetables system, genetic algorithm, timetable at university Classifier created from a set of data. Bayesian classifier is a statistical classifier for predicting the probability of a particular class membership. This research will try to perform data classification for prediction of new student graduation, Naive Bayes algorithm method used for naïve Bayes classification performance high enough ability to predict future opportunities based on experience or data in the past. Implementation WEKA algorithms in applications that will explore the characteristics of the dataset with superficial attributes Pass options. Evaluation results show classified data correctly (correct classified instances) in accordance with the grouping of choice graduated first choice, second choice and do not pass by the algorithm as much as 93.6288% or as much as 338 data and classified data, but does not match the class predicted (incorrect classified instances) which should be a group of two or Pass options but are included in the group First choice as many as 6.3712% or as much as 23 data. Value Percentage accuracy demonstrated the effectiveness dataset Admissions applied to the methods Naïve Bayes Classification, which reached 94%

Keywords: Naive bayes, klasifikasi, data mining

1 PENDAHULUAN

Klasifikasi adalah salah satu masalah mendasar dan tugas utama dalam data mining (Zhang dkk., 2004), dalam klasifikasi sebuah pengklasifikasi dibuat dari sekumpulan data latih dengan kelas yang telah di tentukan sebelumnya. Performa pengklasifikasi biasanya diukur dengan ketepatan atau tingkat galat (Walpore dkk., 1995).

Pengklasifikasi Bayesian merupakan pengklasifikasi statistik untuk dapat memprediksi probabilitas keanggotaan kelas tertentu (Hang dkk., 2006), menghitung peluang untuk suatu hipotesis, menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal (Hajjaratie, 2011).

Penelitian ini akan menggunakan metode algoritma naïve bayes untuk melakukan prediksi peluang kelulusan mahasiswa baru. Kelulusan yang dimaksud adalah diterimanya seorang mahasiswa pada salah satu program studi di Perguruan Tinggi. Studi kasus dilakukan pada data mahasiswa Universitas Al Asyariah Mandar Sulawesi Barat.

2 TINJAUAN PUSTAKA

2.1. Algoritma Naive Bayes

Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk *machine learning* dan data mining. Performa naïve bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut). Asumsi keidependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keidependenan atribut tersebut dilanggar performa pengklasifikasian naïve bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris.

Naïve 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 sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas. Pada sebuah dataset, setiap baris/dokumen I diasumsikan sebagai vector dari nilai-nilai atribut $\langle x_1, x_2, \dots, x_n \rangle$ dimana tiap nilai-nilai menjadi peninjauan atribut X_i ($i \in [1, n]$). Setiap baris mempunyai label kelas $c_i \in \{c_1, c_2, \dots, c_k\}$ sebagai nilai variabel kelas C , sehingga untuk melakukan klasifikasi dapat dihitung nilai probabilitas $p(C=c_i|X=x_j)$, dikarenakan pada Naïve Bayes diasumsikan setiap atribut saling bebas, maka persamaan yang didapat adalah sebagai berikut :

- Peluang $p(C=c_i|X=x_j)$ menunjukkan peluang bersyarat atribut X_i dengan nilai x_i diberikan kelas c , dimana dalam Naïve Bayes, kelas C bertipe kualitatif sedangkan atribut X_i dapat bertipe kualitatif ataupun kuantitatif.
- Ketika atribut X_i bertipe kuantitatif maka peluang $p(X=x_j|C=c_j)$ akan sangat kecil sehingga membuat persamaan peluang tersebut tidak dapat diandalkan untuk permasalahan atribut bertipe kuantitatif. Maka untuk menangani atribut kuantitatif, ada beberapa pendekatan yang dapat digunakan seperti distribusi normal (Gaussian) :

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

Ataupun kernel density estimation (KDE) :

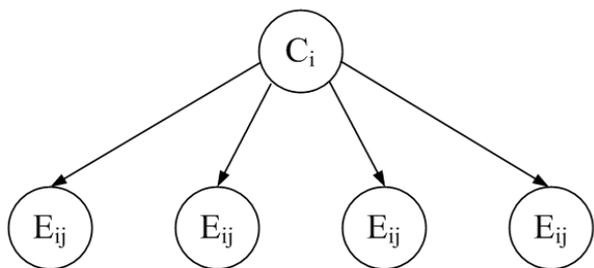
$$\hat{f} = \frac{1}{n_c} \sum_j N(X_i; \mu_{ij}, \sigma_c) \quad (2)$$

Selain dua pendekatan distribusi tersebut, ada mekanisme lain untuk menangani atribut kuantitatif (numerik) yaitu Diskritisasi. Proses diskritisasi sendiri terjadi saat proses

persiapan data atau saat data preprocessing, dimana atribut numerik X diubah menjadi atribut nominal X*. Performansi klasifikasi Naive Bayes akan lebih baik ketika atribut numerik didiskritisasi daripada diasumsikan dengan pendekatan distribusi seperti di atas [Dougherty]. Nilai-nilai numerik akan dipetakan ke nilai nominal dalam bentuk interval yang tetap memperhatikan kelas dari tiap-tiap nilai numerik yang dipetakan, penggambaran perhitungan Naive Bayesnya seperti berikut:

Tabel 1. Perhitungan Naive Bayes

	Interval 1 (i ₁)	Interval 2 (i ₂)
Kelas 1 (c ₁)	Rumus Naive Bayes nya menjadi :	
	$p(I=i_j C=c_j) = \frac{p(I=i_j) p(C=c_i I=i_j)}{p(C=c_i)}$	
	ket :	
Kelas 2 (c ₂)	p(I=i _j C=c _i) : peluang interval i ke-j untuk kelas c _i p(C=c _i I=i _j) : peluang kelas c _i pada interval i ke-j p(I=i _j) : peluang sebuah interval ke-j pada semua interval yang terbentuk p(C=c _i) : peluang sebuah kelas ke-i untuk semua kelas yang ada di dataset	



Gambar 1. Ilustrasi Naive Bayes

Berikut adalah algoritma dari metode Naive Bayes :

- Menghitung jumlah kelas
- Menghitung jumlah kasus yang sama dengan class yang sama
- kalikan semua hasil variable TEPAT & TERLAMBAT
- Bandingkan hasil class TEPAT & TERLAMBAT

Kekuatan dan kelemahan Naive Bayesian:

Kekuatan:

- Mudah diimplementasi.
- Memberikan hasil yang baik untuk banyak kasus.

Kelemahan:

- Harus mengasumsi bahwa antar fitur tidak terkait (independent) Dalam realita, keterkaitan itu ada Keterkaitan tersebut tidak dapat dimodelkan oleh Naive Bayesian Classifier.

Dalam pembahasan ini terdapat 2 dataset yang digunakan yaitu Pertama : Data Penerimaan Mahasiswa Baru yang memiliki 15 atribut dan 363 record yang terdiri nama, tempat tanggal lahir, asal sekolah, tahun lulus, nama orang tua, nomor ujian, nilai tes tertulis, nilai tes wawancara, nilai rata-rata, ket. Lulus, lulus pilihan dan sebagainya. Kedua, Data hasil Ujian Masuk Perguruan tinggi yang terdiri dari 5 atribut dan 362 record yang terdiri dari Nomor Ujian, Nilai tes tulisan, nilai tes wawancara, rata-rata dan keterangan Lulus.

Tabel 2. Data Hasil Ujian Masuk Perguruan Tinggi

NO. UJIAN	Tes Tertulis		Tes Wawancara	Total	Keterangan Lulus
	Bahasa Indonesia & Inggris	Pengetahuan Umum & Matematika dasar			
0001	27.50	25.00	95.00	49.17	LULUS
0002	27.50	38.33	90.00	51.94	LULUS
0003	25.00	30.00	90.00	48.33	LULUS
0004	17.50	25.00	95.00	45.83	LULUS
0005	27.50	43.33	85.00	51.94	LULUS
0006	20.00	28.33	90.00	46.11	LULUS
0007	30.00	20.00	85.00	45.00	LULUS
0008	15.00	23.33	90.00	42.78	LULUS
0009	32.50	28.33	80.00	46.94	LULUS
0010	32.50	36.67	90.00	53.06	LULUS
0011	27.50	26.67	95.00	49.72	LULUS
0012	15.00	36.67	90.00	47.22	LULUS
0013	22.50	26.67	95.00	48.06	LULUS
0014	27.50	43.33	85.00	51.94	LULUS
0015	22.50	21.67	90.00	44.72	LULUS
0016	20.00	18.33	80.00	39.44	LULUS
0017	17.50	23.33	90.00	43.61	LULUS

3.2. Praproses

Tahap pembersihan data merupakan awal dari proses KDD. Proses pembersihan data mencakup antara lain memeriksa data yang tidak konsisten, data dengan *missing value* dan *redundant* data. Seluruh atribut pada dua kelompok data (tabel) dibersihkan karena hal tersebut merupakan syarat awal untuk proses data mining yang akan menghasilkan dataset yang bersih dan siap digunakan pada tahap mining data. Dikatakan *missing value* jika pada salah satu atribut nilai *record* tersebut hilang maka *record* yang dimaksud akan dihapus, karena *record* tersebut dinilai kehilangan data atau *missing value*. Apabila dalam dataset yang sama terdapat lebih dari satu *record* yang berisi nilai yang sama, maka *record* yang dimaksud juga harus dihapus karena tidak akan memberi informasi yang berarti jika dipertahankan. Tahap ini tidak hanya membersihkan data yang mengandung *missing value* saja akan tetapi terhadap data yang tidak konsisten juga dilakukan. Data pada penelitian ini merupakan data yang sudah konsisten. Karena dua kelompok data (tabel) diambil seluruhnya tidak ada data yang *dicleaning*, maka jumlah atribut dan record pada kelompok data (tabel) adalah tetap. Pada tahap ini data sudah bersih dan siap untuk digunakan pada tahap selanjutnya yaitu integrasi data. Data hasil *cleaning* kami sajikan pada tabel 3.

3.3. Implementasi perhitungan pada Naive Bayes Clasification

Setelah data hasil *cleaning* rampung maka selanjutnya, data tersebut akan di implementasikan dengan formula *Naive Bayes Classifier*. Adapun cara kerja dari proses perhitungan *Naive Bayes Classifier* yaitu sebagai berikut : Tahapan diawali dengan mengambil data sample atau contoh data seperti pada table 4 data tes mahasiswa.

3 METODE PENELITIAN

3.1. Data

A	B	C	D	E	F	G
JURUSAN SEKOLAH	PILIHAN PERTAMA	PILIHAN KEDUA	Nilai Rata	Keterangan Lulus	Pilihan Lulus	Jurusan Lulus
IPA	Sistem Informasi	Ilmu Pemerintahan	55.83	LULUS	PIL 1	Sistem Informasi
IPA	Sistem Informasi	Ilmu Pemerintahan	51.94	LULUS	PIL 1	Sistem Informasi
IPA	Sistem Informasi	Ilmu Komunikasi	50.07	LULUS	PIL 1	Sistem Informasi
TKJ	Sistem Informasi	Teknik Informatika	52.50	LULUS	PIL 1	Sistem Informasi
IPA	Bahasa Indonesia	Teknik Informatika	51.94	LULUS	PIL 1	Bahasa Indonesia
IPA	Teknik Informatika	Peternakan	46.11	LULUS	PIL 2	Peternakan
IPA	Teknik Informatika	Peternakan	45.00	LULUS	PIL 2	Peternakan
IPS	Ilmu Pemerintahan	Sistem Informasi	42.78	LULUS	PIL 1	Ilmu Pemerintahan
KESEHATAN	PPKn	Kesehatan Masyarakat	50.28	LULUS	PIL 1	PPKn
KESEHATAN	PPKn	Teknik Informatika	53.06	LULUS	PIL 1	PPKn
KESEHATAN	PPKn	Teknik Informatika	51.06	LULUS	PIL 1	PPKn
KESEHATAN	Kesehatan Masyarakat	Peternakan	47.22	LULUS	PIL 1	Kesehatan Masyarakat
KESEHATAN	Bahasa Indonesia	Ekonomi Islam	48.06	LULUS	PIL 2	Ekonomi Islam
KESEHATAN	Teknik Informatika	Ilmu Pemerintahan	51.94	LULUS	PIL 1	Teknik Informatika
BAHASA	Ilmu Pemerintahan	Matematika	44.72	LULUS	PIL 1	Ilmu Pemerintahan
IPA	Bahasa Indonesia	Sistem Informasi	50.88	LULUS	PIL 1	Bahasa Indonesia
IPS	PPKn	Ilmu Pemerintahan	43.61	LULUS	PIL 2	Ilmu Pemerintahan
IPA	Agribisnis	Ekonomi Islam	51.39	LULUS	PIL 1	Agribisnis
IPA	Matematika	Sistem Informasi	50.11	LULUS	PIL 1	Matematika
KESEHATAN	Kesehatan Masyarakat	Sistem Informasi	52.22	LULUS	PIL 1	Kesehatan Masyarakat
IPS	Sistem Informasi	Ilmu Pemerintahan	50.28	LULUS	PIL 1	Sistem Informasi
KESEHATAN	Peternakan	Teknik Informatika	19.44	TIDAK LULUS	TIDAK LULUS	TIDAK LULUS
IPS	Teknik Informatika	Ilmu Komunikasi	51.00	LULUS	PIL 1	Teknik Informatika
IPA	PPKn	Kesehatan Masyarakat	51.39	LULUS	PIL 1	PPKn
IPA	PPKn	Kesehatan Masyarakat	53.33	LULUS	PIL 1	PPKn
TKJ	Sistem Informasi	Kesehatan Masyarakat	50.83	LULUS	PIL 2	Kesehatan Masyarakat
IPS	Sistem Informasi	Matematika	50.34	LULUS	PIL 1	Sistem Informasi

Tabel 4 Data tes mahasiswa

JURUSAN SEKOLAH	PILIHAN PERTAMA	PILIHAN KEDUA	Nilai Rata	Keterangan Lulus	Pilihan Lulus
IPA	Sistem Informasi	Ilmu Pemerintahan	55.83	LULUS	PIL 1
BAHASA	Sistem Informasi	Ilmu Pemerintahan	42.78	LULUS	PIL 2
IPS	Sistem Informasi	Ilmu Pemerintahan	50.28	LULUS	PIL 1
TKJ	Sistem Informasi	Teknik Informatika	52.50	LULUS	PIL 1
IPS	Ilmu Pemerintahan	Sistem Informasi	42.78	LULUS	PIL 1
KESEHATAN	Kesehatan Masyarakat	Peternakan	47.22	LULUS	PIL 1
BAHASA	Ilmu Pemerintahan	Matematika	44.72	LULUS	PIL 1
PERKANTORAN	Ilmu Pemerintahan	Ekonomi Islam	47.50	LULUS	PIL 1
IPS	PPKn	Ilmu Pemerintahan	43.61	LULUS	PIL 2
IPS	Bahasa Indonesia	Ilmu Pemerintahan	44.44	LULUS	PIL 2
IPA	Matematika	Sistem Informasi	50.11	LULUS	PIL 1
KESEHATAN	Kesehatan Masyarakat	Sistem Informasi	52.22	LULUS	PIL 1
IPS	Sistem Informasi	Ilmu Pemerintahan	50.28	LULUS	PIL 1

Sehingga,

$$\begin{aligned}
 P(\text{Pil1}) &= 10/13 = 0.77 \\
 P(\text{Pil2}) &= 3/13 = 0.23 \\
 P(\text{Prodi=IPA}|\text{Pil1}) &= 2/10 = 0.20 \\
 P(\text{Prodi=IPA}|\text{Pil2}) &= 0/3 = 0.00 \\
 P(\text{Pilihan Pertama}=\text{Sistem Informasi}|\text{Pil1}) &= 4/10 = 0.40 \\
 P(\text{Pilihan Pertama}=\text{Sistem Informasi}|\text{Pil2}) &= 1/3 = 0.33 \\
 P(\text{Pilihan Kedua}=\text{Ilmu Pemerintahan}|\text{Pil1}) &= 3/10 = 0.30 \\
 P(\text{Pilihan Kedua}=\text{Ilmu Pemerintahan}|\text{Pil2}) &= 3/3 = 1.00 \\
 P(\text{Nilai Rata}=55.83|\text{Pil1}) &= 1/10 = 0.10 \\
 P(\text{Nilai Rata}=55.83|\text{Pil2}) &= 0/3 = 0.00
 \end{aligned}$$

Kemudian dari tabel diatas dapat dihitung dengan menggunakan formula Naïve Bayes Clasification, adapun cara kerjanya sebagai berikut :

- Untuk 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 hipotesa bahwa X adalah data dengan klas (label) C.
P(H) adalah peluang dari hipotesa H
P(X) adalah peluang data sampel yang diamati
 $P(X|H)$ adalah peluang data sampel X, bila diasumsikan bahwa hipotesa benar (valid).
Jadi Rumusnya adalah

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

Kemudian,

$$\begin{aligned}
 &P(\text{Pil1}) P(\text{IPA}|\text{Pil1}) P(\text{Sistem Informasi}|\text{Pil1}) \\
 &P(\text{Ilmu Pemerintahan}|\text{Pil1}) P(55.83|\text{Pil1}) \\
 &= 0.77 \times 0.20 \times 0.40 \times 0.30 \times 0.10 \\
 &= 0.00203
 \end{aligned}$$

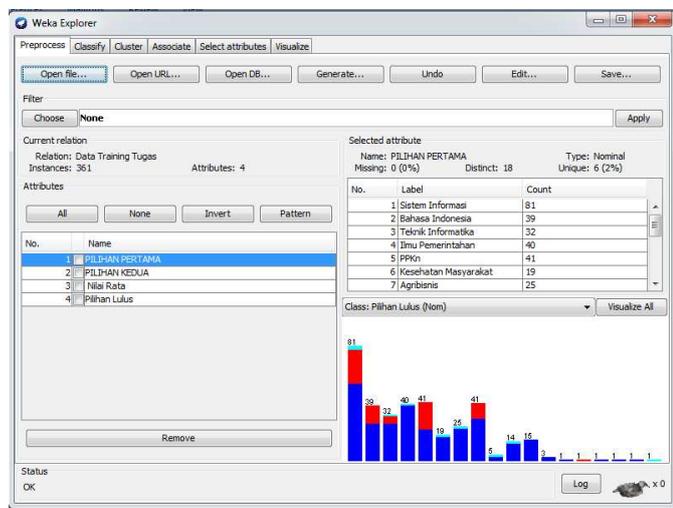
$$\begin{aligned}
 &P(\text{Pil2}) P(\text{IPA}|\text{Pil2}) P(\text{Sistem Informasi}|\text{Pil2}) \\
 &P(\text{Ilmu Pemerintahan}|\text{Pil2}) P(55.83|\text{Pil2}) \\
 &= 0.23 \times 0 \times 0.33 \times 1 \times 0 \\
 &= 0
 \end{aligned}$$

Jadi Keputusan Pilihan Lulus adalah Pil 1 (Pertama)

4. Hasil

4.1. Implementasi pada Aplikasi WEKA

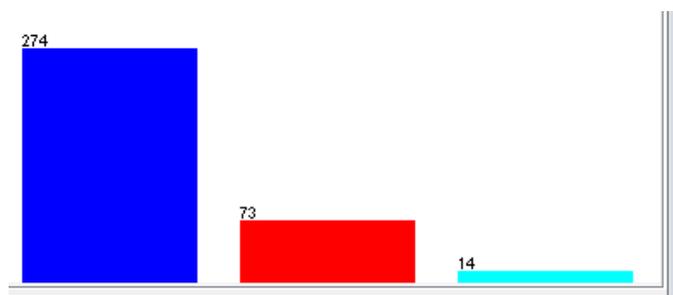
Weka akan menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokkan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata. Adapun tahapan praproses pada weka adalah sebagai berikut:



Gambar 2. Praproses Weka Tools.

Dataset diproses dengan menggunakan teknik *classifier Naive Bayes Updateable* dengan luaran Pilihan Lulus. Jenis test yang digunakan adalah training set karena luaran menemukan pola baru dalam data, yang menghubungkan pola data yang sudah ada dengan data yang baru. *Classifier* panel memungkinkan untuk mengkonfigurasi dan menjalankan salah satu dari pengklasifikasian yang dipilih untuk diterapkan pada dataset yang ada.

Proses klasifikasi dipengaruhi oleh atribut-atribut terpilih yang mendukung untuk menentukan kelompok Lulus Pilihan Pertama dan Pilihan Kedua. Hasil dari klasifikasi divisualisasikan dengan diagram batang dengan luaran class Pilihan Lulus yang terdiri dari 13 kategori pilihan lulus. 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.



Gambar 3. Diagram Hasil Klasifikasi Pilihan Lulusan

Pada weka classifier juga dapat dilihat rincian kelompok Pilihan Lulus pada masing-masing Prodi dan didapatkan hasil bahwa class Pilihan Lulus Pilihan Pertama lebih tinggi dibandingkan class Pilihan Lulus Pilihan Kedua serta Pilihan Lulus Tidak Lulus seperti pada gambar dibawah ini

Attribute	Class		
	PIL 1 (0.76)	PIL 2 TIDAK LULUS (0.2)	(0.04)
PILIHAN PERTAMA			
Sistem Informasi	55.0	25.0	4.0
Bahasa Indonesia	27.0	14.0	1.0
Teknik Informatika	27.0	6.0	2.0
Ilmu Pemerintahan	40.0	1.0	2.0
PPKn	23.0	20.0	1.0
Kesehatan Masyarakat	18.0	1.0	3.0
Agribisnis	24.0	1.0	3.0
Matematika	31.0	12.0	1.0
Peternakan	4.0	1.0	3.0
Agroteknologi	13.0	1.0	3.0
Ekonomi Islam	16.0	1.0	1.0
Ilmu Komunikasi	4.0	1.0	1.0
PPKn	2.0	1.0	1.0
Sistem Informasi	1.0	2.0	1.0
Imu Pemerintahan	2.0	1.0	1.0
Sistem Infomasi	2.0	1.0	1.0

Gambar 4. Rincian Kelompok Pada pilihan Pertama

4.2. Evaluasi

Pada hasil evaluasi menunjukkan data yang diklasifikasikan secara benar (*correct classified instances*) sesuai dengan pengelompokkan pilihan lulus pilihan pertama, Pilihan kedua dan tidak lulus oleh algoritma sebanyak 93.6288 % atau sebanyak 338 data dan data yang diklasifikasikan namun tidak sesuai dengan class yang diprediksi (*incorrect classified instances*) yang seharusnya merupakan kelompok Pilihan Kedua atau tidak Lulus tetapi dimasukkan ke dalam kelompok Pilihan Pertama yaitu sebanyak 6.3712 % atau sebanyak 23 data.

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      338      93.6288 %
Incorrectly Classified Instances    23       6.3712 %
    
```

Gambar 5. Hasil Evaluasi

Setelah dilakukan pengolahan data training maka diperoleh akurasi pada model tersebut. Akurasi pada model dihitung dengan menggunakan *confusion matrix*. Berikut ini dijabarkan *confusion matrix* dengan metode *naive bayes classifier*. Huruf a,b dan c pada tabel berturut-turut menunjukkan class Pil 1, Pil 2 dan Tidak Lulus.

```

=== Confusion Matrix ===

  a  b  c  <-- classified as
271  2  1 | a = PIL 1
 16 57  0 | b = PIL 2
  3  1 10 | c = TIDAK LULUS
    
```

Gambar 6. Confusiin Matrix

Pengolahan ini menggunakan data sebanyak 361 record. Berdasarkan hasil confusion matrix terlihat bahwa 271 record pada class a diprediksi tepat sebagai class a dan sebanyak 3 record diprediksikan tidak tepat sebagai kelompok data class Pilihan 1 , karena record tersebut diprediksi sebagai class Pilihan Kedua dan Tidak Lulus. Kemudian dari 57 record pada class b diprediksi tepat sebagai class b dan sebanyak 16 record diprediksikan tidak tepat sebagai kelompok data class Pilihan 2 karena record tersebut diprediksi sebagai class Pilihan

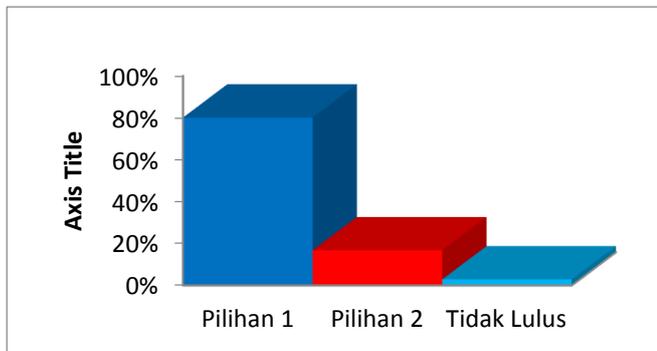
Pertama dan Tidak Lulus. Kemudian 10 record pada class c diprediksi tepat sebagai class c dan sebanyak 4 record diprediksikan tidak tepat sebagai kelompok data class Pilihan 1 dan 1 record karena record tersebut diprediksi sebagai class Pilihan Pertama dan Pilihan Kedua.

Dalam menentukan persentase akurasi dari data yang diolah maka formula yang digunakan adalah sebagai berikut :

$$\text{Presentase Akurasi} = \frac{\text{Banyaknya Prediksi yang benar}}{\text{Total banyaknya prediksi}} \times 100\%$$

$$\begin{aligned} \text{Presentase Akurasi} &= \frac{271 + 57 + 10}{271 + 2 + 1 + 16 + 57 + 0 + 3 + 1 + 10} \times 100\% \\ &= 94\% \end{aligned}$$

Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naïve Bayes Clasification* yang mencapai 94 % . Untuk lebih mempermudah pemahaman dalam menganalisa hasil klasifikasi dataset maka dilampirkan vizualisasi hasil tersebut dalam bentuk grafik seperti pada gambar 7



Gambar 7. Grafik hasil klasifikasi berdasarkan presentase keakuratan

5. Kesimpulan

Naïve Bayes dapat melakukan pengklasifikasian dengan metode probabilitas dan statistik, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Nilai Presentase keakuratan menunjukkan keefektifan dataset Penerimaan Mahasiswa Baru yang diterapkan ke dalam metode *Naïve Bayes Clasification*. Impelementasi *Naïve bayes* menggunakan aplikasi WEKA dapat menelusuri karakteristik atribut dari dataset dengan luaran Pilihan Lulus. Pengelompokkan Pilihan Lulus dilakukan berdasarkan atribut terpilih yaitu Prodi, Pilihan Pertama, Pilihan Kedua dan Nilai Rata-rata.

DAFTAR PUSTAKA

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.

BIOGRAPHY OF AUTHORS



Syarli. Menerima gelar sarjana S.Kom Teknik Informatika di Universitas Al Asyariah Mandar, Sulawesi Barat tahun 2010 dan Saat ini tengah menyelesaikan program pascasarjana Pada Magister Sistem Informasi Universitas Diponegoro, Semarang. Dia adalah seorang dosen pada Fakultas Ilmu Komputer Universitas Al Asyariah Mandar, Polewali Mandar, Sulawesi Barat, Indonesia. Dia juga aktif pada kegiatan-kegiatan sosial dan pendampingan. Research interests termasuk Artificial intelingence.



Asrul Azhari Muin. Menerima gelar sarjana S.kom pada Universitas Satria Makassar tahun 2011 dan gelar M.Kom Pada Program Pascasarjana pada Magister Sistem Informasi Universitas Diponegoro Semarang. Dia adalah seorang dosen pada Fakultas Sains dan Teknologi Universitas Islam Negeri Makassar. Aktif pada kegiatan-kegiatan perusahaan.