

PREDIKSI PERFORMA MAHASISWA MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Hasnan Afif¹, Kusrini², M. Rudyanto Arief³

Magister Teknik Informatika
Universitas Amikom Yogyakarta

hasnan.afif@gmail.com¹, kusrini@amikom.ac.id², rudy@amikom.ac.id³

Abstrak

Salah satu dalam menentukan performa mahasiswa dapat dilihat dari ketepatannya dalam menyelesaikan masa studinya di universitas. Penelitian ini bertujuan untuk melakukan klasifikasi terhadap data mahasiswa Universitas Amikom Yogyakarta Fakultas Ilmu Komputer angkatan 2012 dan 2013 jenjang S1 dengan memanfaatkan proses data mining dengan menggunakan teknik klasifikasi. Algoritma yang digunakan untuk klasifikasi kelulusan adalah algoritma Naïve Bayes, dimulai dari proses Diagnosing, Action Planning, Action Taking, Evaluating dan Specifying Learning. Penelitian ini juga bertujuan untuk menemukan faktor-faktor yang mempengaruhi prediksi kelulusan mahasiswa melalui data kinerja akademik mahasiswa dan beberapa variabel lainnya. Variabel yang digunakan adalah Jenis Kelamin, umur, asal sekolah, IPS1 (Indeks Prestasi Semester satu) sampai dengan IPS4, SKS, IPK, dan kehadiran. Pengujian data dalam penelitian ini adalah sebanyak 1170 data, 936 data digunakan untuk data training dan 234 data untuk data testing dengan menggunakan 5 k fold cross validation Hasil pengujian menunjukkan bahwa faktor yang paling berpengaruh dalam penentuan klasifikasi kinerja akademik mahasiswa yaitu SKS dengan nilai akurasi 96,15%, umur 79,02%, IPK 77,72%, dan kehadiran 77,67%. Dalam penelitian ini hasil yang dicapai dalam penerapan metode algoritma Naïve Bayes menggunakan seluruh variabel dengan tingkat akurasi sebesar 93,93% kemudian setelah menggunakan variabel terbaik akurasi meningkat menjadi 96,15%.

Kata Kunci: performa mahasiswa, kelulusan mahasiswa, data mining, naïve bayes classifier, variabel, akurasi.

1. Pendahuluan

Kelulusan mahasiswa yang tepat waktu merupakan isu yang penting karena tingkat kelulusan sebagai dasar efektifnya suatu kelembagaan. Jika terjadi penurunan tingkat kelulusan secara signifikan dan terus berkembang maka akan menjadi permasalahan yang serius. Bahkan dapat mempengaruhi akreditasi perguruan tinggi tersebut. Untuk itu pemantauan dan evaluasi secara berkala terhadap kecenderungan tingkat kelulusan mahasiswa diperlukan. Berdasar pada peraturan yang disampaikan dalam buku II standard dan prosedur tentang akreditasi institusi perguruan tinggi oleh BAN-PT (Badan Akreditasi Nasional Perguruan Tinggi) tahun 2011 menyatakan bahwa salah satu aspek penilaian akreditasi adalah mahasiswa dan lulusan (Marwing, 2011). Jadi, tingkat kelulusan dan jumlah mahasiswa akan berpengaruh dalam proses akreditasi yang dilakukan oleh pemerintah.

Jumlah mahasiswa lulus tepat waktu dapat ditingkatkan dengan meningkatkan kualitas pembelajaran dan layanan akademik untuk mahasiswa. Selain itu, jika waktu penyelesaian studi mahasiswa dapat diprediksikan maka penanganan mahasiswa akan lebih efektif (Salmu dan Solichin, 2017). Salah satu teknik melakukan prediksi yang dapat digunakan adalah dengan teknik penggalian data atau data mining terutama algoritma Naïve Bayes Clasifier.

Naive Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naive Bayes didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Naive Bayes terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Prasetyo, 2012). Algoritma naive bayes merupakan algoritma yang memiliki akurasi tertinggi, termasuk good classification dan memiliki waktu pembangunan model yang paling cepat dibandingkan dengan algoritma decision tree dan artificial neural network (Defiyanti, 2014). Naïve Bayes Classifier mengenali setiap atribut pada data set sebagai atribut yang independen, sehingga disebut algoritma yang naïve.

Beberapa penelitian telah banyak dilakukan dengan menggunakan teknik data mining untuk menggali berbagai informasi dari sebuah database, seperti penelitian yang dilakukan oleh Mueen dkk (2016), penelitian dilakukan untuk memprediksi dan menganalisis prestasi akademik mahasiswa dengan membandingkan tiga teknik klasifikasi yaitu Algoritma Naïve Bayes, Neural Network dan Decision Tree. Dari penelitian diperoleh hasil bahwa akurasi prediksi menggunakan algoritma Naïve Bayes lebih tinggi atau melebihi dua pengklasifikasi lain. Rakhman (2017) melakukan penelitian tentang prediksi ketepatan kelulusan mahasiswa menggunakan metode *Decision Tree* berbasis *Particle Swarm Optimization*(PSO). Dari penelitian yang dilakukan diperoleh nilai akurasi yang didapat pada model algoritma Decision Tree berbasis PSO adalah 97,67% lebih baik jika dibandingkan dengan algoritma Decision Tree yaitu 96,67%.

Berdasarkan penjelasan diatas penelitian ini akan melakukan uji coba penerapan algoritma Naïve Bayes Classifier untuk mengklasifikasikan kelulusan mahasiswa dengan cara mengevaluasi performa mahasiswa menggunakan beberapa variabel.

2. Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian tindakan (*action research*). penelitian tindakan merupakan prosedur

yang sistematis yang dilakukan peneliti untuk mendapatkan informasi tentang tindakan dan akibat tindakan tersebut dalam rangka untuk memperbaiki kinerja organisasi. Menurut Susman dan Evered (1978) terdapat lima fase dalam penelitian tindakan (*Action Research*) yaitu *Diagnosing*, *Action Planning*, *Action Taking*, *Evaluating* dan *Specifying Learning*.

1. *Diagnosing*: Fase ini merupakan fase awal dari tahapan pemodelan. Fase ini berfokus pada proses identifikasi masalah-masalah yang menjadi dasar penelitian, pengumpulan data awal dan pembelajaran data yang sudah ada. Dalam penelitian ini menggunakan data mahasiswa fakultas ilmu komputer program studi S1 teknik informatika Universitas AMIKOM Yogyakarta tahun 2012 dan 2013.
2. *Action planning*: Fase ini terdiri dari pemilihan data, pembersihan data, dan transformasi data untuk membangun dataset mahasiswa yang akan diterapkan kedalam tahap selanjutnya (*action taking*), dari data mentah awal berupa dataset mahasiswa dan selanjutnya akan melakukan proses data mining. Pada tahap ini merupakan tahap untuk memastikan data mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan.
 - a. *Data selection*: Seluruh atribut dataset akan diseleksi untuk mendapatkan atribut-atribut yang relevan. Data asli yang di dapat peneliti dari jurusan Teknik Informatika Universitas Amikom adalah sebanyak 1259 record dan memiliki 16 atribut. Tidak semua atribut dimasukkan dalam dataset yang akan digunakan dalam proses data mining karena hanya berperan sebagai referensi identifikasi, seperti NIM dan Tahun Kelulusan mahasiswa.
 - b. *Cleaning data*: Selanjutnya dilakukan proses pembersihan data. Pada tahap ini cara yang dilakukan peneliti yaitu dengan menghapus data. Data yang dihapus adalah data yang tidak memiliki atribut lengkap sehingga tidak dapat diproses. Atribut yang datanya tidak lengkap dan tidak bisa digunakan dalam proses pengklasifikasian sehingga total atribut yang akan digunakan dalam penelitian ini adalah berjumlah 11 atribut. Atribut yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.
 - c. *Transformation data*: Untuk mengubah dataset maka digunakan transformasi data sehingga informasi terbaik bisa diambil dengan cara melakukan pengurangan ataupun perubahan tipe data standar menjadi data kategorikal sehingga data siap digunakan untuk klasifikasi teknik data

mining menggunakan algoritma naive bayes classifier. Beberapa atribut lain perlu diubah bentuknya. Adapun atribut-atribut yang perlu diubah bentuknya yaitu sebagai berikut:

- 1) Asal Sekolah: Atribut Asal Sekolah terdiri dari “Negeri” dan “Swasta”.
- 2) Umur: Umur dibagi menjadi 2 kategori yaitu umur kurang dari atau sama dengan 23 tahun (≤ 23) dan umur yang lebih dari 23 tahun (>23).

Tabel 1. Atribut mahasiswa Universitas Amikom Yogyakarta 2012 dan 2013

Kode Atribut	Nama Atribut
JK	Jenis Kelamin
AS	Asal Sekolah
Umur	Umur mahasiswa saat lulus
IPS1	Indeks prestasi semester satu
IPS2	Indeks prestasi semester dua
IPS3	Indeks prestasi semester tiga
IPS4	Indeks prestasi semester empat
IPK	Indeks prestasi kumulatif
SKS	Jumlah SKS yang diambil
Kehadiran	Persentase Kehadiran mahasiswa
Status Mahasiswa	Lulus tepat waktu atau terlambat (atribut target)

- 3) IPS 1, IPS 2, IPS 3, IPS 4 dan IPK dibagi menjadi 3 (tiga) kategori, yang pertama yaitu “Good” untuk range IP 3.50 – 4.00, kemudian “Average” untuk range IP 3.00 – 3.49, dan “Poor” untuk nilai IP dibawah 3.00 (<3.00).
- 4) Jumlah SKS: Jumlah SKS dibagi menjadi 2 atribut yaitu mahasiswa yang mengambil lebih dari atau sama dengan 144 SKS (≥ 144 SKS) dan mahasiswa yang mengambil kurang dari 144 SKS (< 144 SKS).
- 5) Persentase Kehadiran: Persentase Kehadiran dibagi menjadi 3 kategori yaitu “Good” untuk persentase kehadiran mahasiswa yang lebih dari atau sama dengan 80%, kemudian “Average” untuk range persentase kehadiran mahasiswa dari 60% - 79,99 %, dan “Poor” untuk persentase kehadiran mahasiswa yang kurang dari 60% ($<60\%$).

Selanjutnya pada tahap ini atribut – atribut yang tidak relevan dan atribut-atribut yang tidak lengkap tidak diproses pada implementasi naive bayes classifier, sehingga data yang layak digunakan untuk proses implementasi naive bayes classifier adalah sebanyak 1170 data.

3. Action taking: Fase ini merupakan tahap implementasi tindakan. Fase ini secara langsung melibatkan teknik data mining yaitu dengan melakukan

pemilihan teknik data mining dan menentukan algoritma yang akan digunakan. Model yang digunakan pada penelitian ini adalah klasifikasi dan algoritma yang digunakan adalah Naïve Bayes Classifier yang akan dibahas pada sub judul selanjutnya yaitu pembahasan dan hasil.

4. Evaluating: Pada fase ini bertujuan untuk mengevaluasi akurasi yang didapat pada fase sebelumnya untuk menilai sejauh mana hasil implementasi tindakan (action taking) memenuhi tujuan data mining yang telah ditentukan pada tahap diagnosing. Pada tahap evaluation dilakukan pengujian terhadap model dengan menggunakan *k-fold cross validation* dan *confusion matrix* untuk mendapatkan informasi model yang akurat untuk memprediksi tingkat kelulusan.
5. Specifying learning: Specifying Learning merupakan tahapan akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir yang berisi mengenai pengetahuan yang diperoleh atau pengenalan pola pada data dalam proses data mining serta implementasi dari prediksi performa mahasiswa menggunakan algoritma Naive Bayes.

3. Pembahasan dan Hasil

3.1 Implementasi Menggunakan RapidMiner

Implementasi pada penelitian ini menggunakan RapidMiner Studio versi 9.1.000 win64. Langkah pertama proses klasifikasi yang dilakukan yaitu dengan penentuan dataset mahasiswa yang disimpan dalam format excel seperti yang terlihat pada Gambar 1.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Npm	JK	Asal Sek	umur	IP Sem1	IP Sem2	IP Sem3	IP Sem4	IPK	JumlahSk	Persentas	StatusMhs
2	13.11.674	L	Negri	<=23	good	good	good	good	good	≥144	good	TEPAT
3	13.11.674	L	Swasta	<=23	good	good	average	average	average	≥144	good	TEPAT
4	13.11.674	L	Swasta	<=23	average	average	average	average	average	≥144	good	TEPAT
5	13.11.674	P	Negri	<=23	good	average	good	average	good	≥144	good	TEPAT
6	13.11.674	L	Swasta	<=23	average	average	average	poor	poor	<144	Poor	Terlambat
7	13.11.674	L	Swasta	<=23	average	average	average	average	average	≥144	good	TEPAT
8	13.11.674	L	Swasta	<=23	good	good	good	average	good	≥144	average	TEPAT
9	13.11.675	L	Swasta	>23	average	average	good	average	average	<144	good	Terlambat
10	13.11.675	L	Swasta	>23	average	average	average	average	average	<144	average	Terlambat
11	13.11.675	L	Negri	<=23	good	average	good	good	good	≥144	good	TEPAT
12	13.11.675	P	Swasta	<=23	average	average	average	good	average	≥144	good	TEPAT
13	13.11.675	L	Swasta	>23	average	average	average	poor	average	<144	average	Terlambat
14	13.11.675	L	Swasta	<=23	average	average	good	average	average	≥144	good	TEPAT
15	13.11.675	L	Swasta	>23	average	average	good	average	average	<144	average	Terlambat
16	13.11.676	L	Swasta	>23	good	average	average	average	average	≥144	good	TEPAT
17	13.11.676	P	Negri	<=23	good	good	average	average	average	≥144	good	TEPAT
18	13.11.676	L	Negri	>23	average	average	average	average	average	≥144	average	TEPAT
19	13.11.676	L	Swasta	<=23	good	average	average	average	average	>144	good	TEPAT

Gambar 1. Dataset yang disimpan dalam format excel

3.2 Uji Skenario Kombinasi Atribut

Pada tahap ini dilakukan Uji Skenario Kombinasi atribut, mulai dari atribut JK (jenis kelamin) sampai dengan atribut SKS. Masing – masing atribut tersebut dikombinasikan antara satu dengan yang lainnya dengan cara seperti yang disajikan dalam Tabel 2.

Tabel 2. Kombinasi atribut data mahasiswa

Kombinas JK	Kombinas as	Kombinas Umur	Kombinas ips1	Kombinas ips2	Kombinas ips3	Kombinas ips4	Kombinas lpk	Kombinas Sks
jk – as	-	-	-	-	-	-	-	-
jk–umur	as –umur	-	-	-	-	-	-	-
jk – ips1	as – ips1	umur-ips1	-	-	-	-	-	-
jk – ips2	as – ips2	umur-ips2	ips1 - ips2	-	-	-	-	-
jk – ips3	as – ips3	umur-ips3	ips1 – ips3	ips2–ips3	-	-	-	-
jk – ips4	as – ips4	umur-ips4	ips1 – ips4	ips2–ips4	ips3–ips4	-	-	-
jk – ipk	as – ipk	umur- ipk	ips1 – ipk	ips2–ipk	ips3 –ipk	ips4–ipk	-	-
jk – sks	as – sks	umur-sks	ips1 – sks	ips2–sks	ips3–sks	ips4–sks	ipk – sks	-
jk – kehadiran	as – kehadiran	umur – kehadiran	ips1 – kehadiran	ips2–kehadiran	ips3 – kehadiran	ips4–kehadiran	ipk – kehadiran	sks – kehadiran

3.3 Hasil Klasifikasi Kombinasi Variabel

Hasil klasifikasi menggunakan kombinasi variabel disajikan pada Tabel 3.

Tabel 3. Hasil klasifikasi kombinasi variabel

Kombinasi	JK	AS	Umur	IPS1	IPS2	IPS3	IPS4	IPK	SKS
AS	65,98%	-	-	-	-	-	-	-	-
Umur	77,35%	79,92%	-	-	-	-	-	-	-
IPS1	70,09%	68,46%	75,64%	-	-	-	-	-	-
IPS2	67,95%	70,09%	73,44%	69,32%	-	-	-	-	-
IPS3	70,17%	71,71%	75,47%	71,45%	68,38%	-	-	-	-
IPS4	69,06%	71,37%	75,64%	72,48%	69,23%	72,31%	-	-	-
IPK	75,90%	75,73%	79,06%	75,04%	73,25%	74,45%	73,34%	-	-
SKS	96,15%	96,15%	96,15%	96,15%	96,15%	96,15%	96,15%	96,15%	-
Kehadiran	75,85%	74,36%	78,55%	74,59%	74,62%	74,10%	74,20%	76,58%	96,15%

Berdasarkan hasil klasifikasi kombinasi tiap-tiap atribut yang telah dilakukan, maka didapatlah variabel dengan akurasi tertinggi atau disebut juga

sebagai variabel terbaik yaitu variabel SKS, umur, IPK dan kehadiran. Variabel-variabel terbaik ini juga akan digunakan untuk menghitung kemungkinan tepat waktu atau terlambat dalam menentukan prediksi performa atau kelulusan mahasiswa serta juga akan dibandingkan dengan hasil klasifikasi menggunakan seluruh atribut. Adapun akurasi rata-rata untuk kombinasi variabel terbaik disajikan dalam Tabel 4.

Tabel 4. Nilai akurasi rata-rata variabel terbaik

SKS		Umur		IPK		Kehadiran	
sks-jk	96,15%	umur-jk	77,35%	IPK-jk	75,90%	kehadran-jk	75,85%
sks-AS	96,15%	umur-AS	79,92%	ipk-AS	75,73%	kehadran -AS	74,36%
sks-umur	96,15%	umur-IPS1	75,64%	IPK-umur	79,06%	kehadran -umur	78,55%
sks-IPS1	96,15%	umur-IPS2	73,44%	IPK-ips1	75,04%	kehadran -IPS1	74,59%
sks-IPS2	96,15%	umur-IPS3	75,47%	IPK-ips2	73,25%	kehadran -IPS2	74,62%
sks-IPS3	96,15%	umur-IPS4	75,64%	IPK-ips3	74,45%	kehadran -IPS3	74,10%
sks-IPS4	96,15%	umur-IPK	79,06%	IPK-ips4	73,34%	kehadran -IPS4	74,20%
sks-IPK	96,15%	umur-SKS	96,15%	IPK-sks	96,15%	kehadran -IPK	76,58%
sks-kehadran	96,15%	umur-kehadran	78,55%	IPK-kehadran	76,58%	kehadran -sks	96,15%
Rata-rata	96,15%	Rata-rata	79,02%	Rata-rata	77,72%	Rata-rata	77,67%

3.4 Hasil Klasifikasi Menggunakan Seluruh Variabel

Hasil klasifikasi dari seluruh variabel dengan menggunakan 5 *k-fold cross validation* dan *confusion matrix* pada prediksi performa mahasiswa disajikan pada Tabel 5.

Tabel 5. Hasil klasifikasi menggunakan seluruh variabel

5 K fold validation	Precision	Recall	Accuracy
1	93.59%	98.65%	94.87%
2	95.21%	95.78%	93.59%
3	97.40%	96.77%	96.15%
4	89.50%	98.18%	90.60%
5	95.61%	93.16%	94.44%
Average	94.26%	96.51%	93.93%

Berdasarkan tabel 4. dapat diketahui bahwa penerapan metode algoritma Naive Bayes Classifier dengan menggunakan parameter jenis kelamin, asal

sekolah, umur, IPS1, IPS2, IPS3, IPS4, IPK, SKS, kehadiran dan status mahasiswa menghasilkan tingkat akurasi sebesar 93,93%. Berdasarkan hasil akurasi tersebut dapat dikatakan bahwa algoritma Naive Bayes Classifier merupakan teknik data mining dengan tingkat akurasi yang tinggi. Sehingga dengan menggunakan metode Naïve Bayes dan dengan atribut akademik mahasiswa yaitu indeks prestasi dari semester 1 sampai dengan semester 4 serta atribut lainnya dapat digunakan untuk menghitung probabilitas dengan kemungkinan tepat waktu atau terlambat dalam menentukan prediksi performa ataupun kelulusan mahasiswa. Adapun hasil prediksi kelulusan mahasiswa dengan menggunakan seluruh atribut disajikan dalam tabel 6 sebagai berikut.

Tabel 6. Hasil prediksi kelulusan mahasiswa

5k fold cross validation	Lulus Tepat Waktu		Terlambat	
	Status Mahasiswa	Prediki mahasiswa	Status Mahasiswa	Prediki mahasiswa
1	156	148	78	86
2	167	166	67	68
3	154	155	80	79
4	181	165	53	69
5	114	117	120	117
Total	772	751	398	419

Berdasarkan Tabel 6. hasil prediksi kelulusan mahasiswa menggunakan metode algoritma Naive Bayes dengan 5 k-fold cross validation menghasilkan akurasi sebesar 93,93% dimana status mahasiswa tepat waktu sebanyak 772 dan prediksi mahasiswa tepat waktu sebanyak 751. Sedangkan untuk status mahasiswa terlambat sebanyak 398 dan prediksi mahasiswa terlambat sebanyak 419 mahasiswa

3.5 Hasil Implementasi Naive Bayes Menggunakan Variabel Terbaik

Setelah dilakukan uji coba klasifikasi kombinasi dari masing-masing variabel sehingga didapatkan variabel dengan akurasi tertinggi atau variabel terbaik yang akan dilakukan penghitungan dengan menggunakan Naive Bayes Classifier. Berikut hasil klasifikasi menggunakan variabel terbaik pada prediksi performa mahasiswa disajikan pada Tabel 7.

Tabel 7. Hasil klasifikasi menggunakan variabel terbaik

5 K fold validation	Precision	Recall	Accuracy
1	98,72%	98,72%	98,29%
2	98,20%	95,35%	95,30%

3	100,00%	95,65%	97,01%
4	95,58%	97,74%	94,87%
5	96,49%	94,02%	95,30%
Average	97,80%	96,30%	96,15%

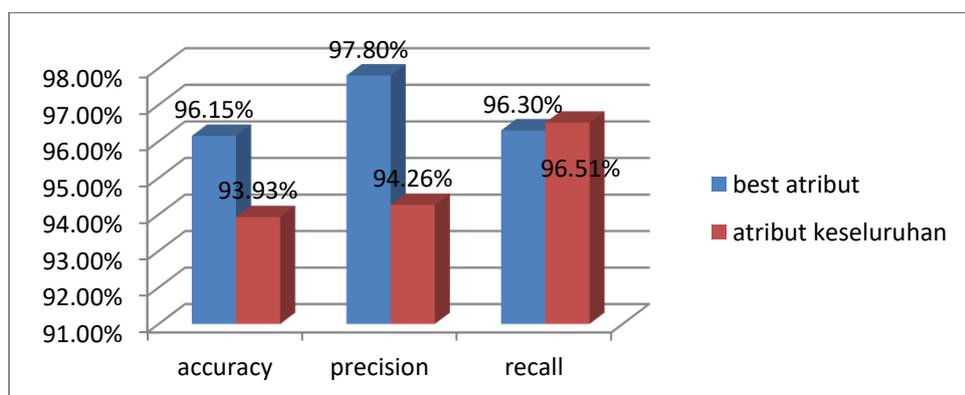
Tabel 7 menunjukkan hasil klasifikasi menggunakan empat variabel terbaik yaitu SKS, umur, IPK dan kehadiran. Dari hasil akurasi tersebut telah dibuktikan bahwa tingkat akurasi ke empat variabel tersebut lebih tinggi jika dibandingkan dengan nilai akurasi dari seluruh variabel. Adapun untuk hasil prediksi performa mahasiswa dengan menggunakan variabel terbaik dalam menentukan kemungkinan lulus tepat waktu atau terlambat disajikan dalam Tabel 8.

Tabel 8. Hasil prediksi kelulusan mahasiswa dengan variabel terbaik

5 k fold cross validation	tepat		terlambat	
	Status mahasiswa	Prediksi mahasiswa	Status mahasiswa	Prediksi mahasiswa
1	156	156	78	78
2	167	172	67	62
3	154	161	80	73
4	181	177	53	57
5	114	117	120	117
Total	772	783	398	387

3.6 Perbandingan Hasil Klasifikasi

Pada tahapan ini dilakukan perbandingan hasil klasifikasi akurasi antara hasil klasifikasi seluruh variabel dengan hasil klasifikasi variabel terbaik. Pada Gambar 2 disajikan grafik perbandingan hasil klasifikasi akurasi antara atribut terbaik dan seluruh atribut.



Gambar 2. grafik perbandingan hasil klasifikasi akurasi

Dari Gambar 2 diketahui nilai akurasi tertinggi 96,15% dari kombinasi variabel terbaik yaitu SKS, IPK, kehadiran dan umur. Sedangkan penggunaan seluruh atribut menghasilkan nilai akurasi 93,93%. Dari hasil ini dapat disimpulkan bahwa penggunaan empat atribut terbaik menghasilkan performa klasifikasi yang lebih baik dibandingkan menggunakan seluruh atribut yang ada.

Jika hasil prediksi antara seluruh variabel dibandingkan dengan variabel terbaik, maka dapat disimpulkan bahwa hasil prediksi menggunakan variabel terbaik lebih akurat dibandingkan dengan hasil prediksi menggunakan seluruh variabel. Hasil prediksi menggunakan seluruh variabel adalah 751 mahasiswa yang tepat waktu dari status mahasiswa tepat waktu sebanyak 772 mahasiswa. Hasil tersebut menunjukkan bahwa terdapat selisih 21 angka antara status mahasiswa lulus tepat waktu dan prediksi mahasiswa lulus tepat waktu. Sedangkan, hasil prediksi mahasiswa tepat waktu menggunakan variabel terbaik dengan status mahasiswa sebanyak 772 mahasiswa adalah 783 mahasiswa. Hasil tersebut menunjukkan bahwa selisih angka antara status mahasiswa lulus tepat waktu dan prediksi mahasiswa lulus tepat waktu hanya 11 mahasiswa.

4. Penutup

Prediksi performa mahasiswa dengan menggunakan algoritma Naïve Bayes Classifier untuk menghitung probabilitas dengan kemungkinan mahasiswa lulus tepat waktu atau terlambat dengan menggunakan variabel jenis kelamin, umur, asal sekolah, IPS1, IPS2, IPS3, IPS4, IPK, SKS dan persentase kehadiran, menghasilkan akurasi yang tinggi yaitu 93,93%. Diharapkan dalam penelitian selanjutnya dapat dibandingkan dengan menggunakan teknik-teknik data mining atau memanfaatkan metode klasifikasi lainnya seperti metode decision tree, C.4.5, metode k-nearest neighbor dan lain-lain sehingga dengan menggunakan banyak metode dapat diketahui kelebihan masing-masing metode dan metode mana yang menghasilkan nilai akurasi yang lebih baik. Selain variabel yang digunakan oleh penulis dalam melakukan penelitian, akan lebih baik jika menambahkan atribut lain seperti konsentrasi yang diambil mahasiswa pada program studi yang ditempuh, apakah mahasiswa tersebut mendapatkan beasiswa atau tidak dan jenis beasiswa apa yang didapatkan mahasiswa tersebut serta atribut non-akademik seperti hobi, suku, jumlah organisasi yang diikuti, mengikuti kegiatan lab atau tidak, penghasilan orang tua, penghasilan pribadi dan atribut lain yang dapat mempengaruhi ketepatan waktu lulus mahasiswa dan proses belajar mahasiswa selama mengikuti perkuliahan baik faktor internal maupun eksternal dalam

menyelesaikan studinya. Contoh faktor eksternal mahasiswa yaitu misalnya faktor kurikulum dan dosen. Apakah kedua faktor tersebut dapat mempengaruhi mahasiswa dalam menyelesaikan masa studinya bisa tepat waktu atau tidak tepat waktu.

Daftar Pustaka

- Mueen, A., Zafar, B. dan Manzoor, U., 2016, Modeling and Predicting Students' Academic Performance Using Data Mining Techniques, I.J. Modern Education and Computer Science, DOI: 10.5815/ijmeecs.2016
- Defiyanti, S., 2014. Perbandingan: Prediksi Prestasi Belajar Mahasiswa Menggunakan Teknik Data Mining (Study Kasus Fasilkom Unsika), *Konferensi Nasional Sistem Informasi*, 177-120
- Marwing, A., 2011, Undang-Undang Sistem Pendidikan Nasional (Analisis terhadap UU SISDIKNAS Nomor 20 Tahun 2003), Ulul Albab, Volume 13, Nomor 1, Juni 2011
- Prasetyo, Eko, 2012, Data Mining : Konsep dan Aplikasi menggunakan MATLAB, 1st ed, Andi, Yogyakarta, Indonesia
- Rakhman, A., 2017, Prediksi Ketepatan Kelulusan Mahasiswa Menggunakan Metode Decision Tree Berbasis Particle Swarm Optimization (PSO), *Smart Comp*, ISSN:2089-676X, Volume 6 No 1 Januari 2017
- Salmu, S. dan Solichin, A., 2017, Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes: Studi Kasus UIN Syarif Hidayatullah Jakarta, *Prosiding Seminar Nasional Multidisiplin Ilmu Universitas Budi Luhur*, ISSN : 2087 - 0930, Jakarta, April 2017
- Susman, G.I. and Evered, R.D., 1978. "An Assessment of the Scientific Merits of Action Research". *Administrative Science Quarterly*, (23) 1978, 582 - 603