

Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes

Mohamad Fajarianditya Nugroho¹, Setyoningsih Wibowo²

¹Teknik Informatika, Fakultas Ilmu Komputer, Universitas AKI

Jl. Imam Bonjol 15-17 Semarang 50139, Telp. (024) 3552555 Fax : 3552111

²Prodi Informatika, Fakultas Teknik dan Informatika, Universitas PGRI Semarang

Jl. Sidadadi Timur No. 24, Dr. Cipto, Semarang 50125, Telp. (024) 8316377 Fax : 8448217

mohamad.fajarianditya@unaki.ac.id

Abstrak

Data mining dalam dunia pendidikan dikenal dengan Educational Data Mining. EDM mengembangkan metode untuk menggali data pendidikan dan menggunakan metode tersebut untuk lebih memahami siswa. EDM dapat membantu pendidik untuk menganalisis cara belajar, mendeteksi mahasiswa yang memerlukan dukungan dan memprediksi kinerja mahasiswa. Perguruan tinggi perlu melakukan prediksi perilaku mahasiswa dan peringatan dini untuk mencegah secara dini kegagalan akademik mahasiswa. Naive Bayes memanfaatkan fungsi seleksi fitur dari Forward Selection untuk pemilihan atribut data dengan karakteristik data itu sendiri, dan meningkatkan ketepatan klasifikasi Naive Bayes. Forward Selection berbasis Naive Bayes lebih akurat dan efektif dalam mengklasifikasikan status kelulusan mahasiswa dengan hasil akurasi 97,14% dan termasuk dalam kategori "excellent classification" dan memperoleh atribut yang berpengaruh yaitu: status pekerjaan dan IPK semester 4.

Kata kunci: Data Mining, Educational Data Mining, Klasifikasi, Naive Bayes, Fitur Seleksi, Forward Selection.

Abstract

Data mining in education is known as Educational Data Mining. EDM develops methods for exploring educational data and using such methods to better understand students. EDM can help educators to analyze how to learn, detect students who need support and predict student performance. Universities need to predict student behavior and early warning to prevent early academic failure of students. Naive Bayes utilizes the feature selection function of Forward Selection for the selection of data attributes with the characteristics of the data itself, and improves the accuracy of Naive Bayes classification. Naive Bayes-based Forward Selection is more accurate and effective in classifying students' graduation status with 97.14% accuracy and is included in the "excellent classification" category and gained influential attributes: status pekerjaan and IPK 4th semester.

Keywords: Data Mining, Educational Data Mining, Classification, Naive Bayes, Feature Selection, Forward Selection.

1. PENDAHULUAN

1.1 Latar Belakang Masalah

Perguruan tinggi merupakan penyelenggara pendidikan akademik bagi mahasiswa [1]. Lima lembaga perguruan tinggi diantaranya adalah universitas, institut, sekolah tinggi, akademi dan politeknik. Perguruan tinggi diharapkan menyelenggarakan pendidikan yang berkualitas bagi mahasiswa sehingga menghasilkan sumber daya manusia yang berilmu, cakap dan kreatif [2]. Semakin bertambah jumlah perguruan tinggi maka semakin meningkat pula jumlah sumber daya

manusia berkualitas yang dihasilkan perguruan tinggi.

Salah satu faktor yang menentukan kualitas perguruan tinggi adalah persentasi kemampuan mahasiswa untuk menyelesaikan studi tepat waktu. Berdasarkan matriks penilaian instrument akreditasi program studi Badan Akreditasi Nasional Perguruan Tinggi [3] bahwa persentase mahasiswa yang lulus tepat waktu merupakan salah satu elemen penilaian akreditasi universitas.

Saat ini, masalah kegagalan studi siswa dan faktor-faktor penyebabnya menjadi topik yang

menarik untuk diteliti [4]. Perguruan tinggi perlu mendeteksi perilaku mahasiswa yang memiliki status “tidak diinginkan” tersebut sehingga dapat diketahui faktor-faktor penyebab kegagalannya. Beberapa penyebab kegagalan mahasiswa diantaranya rendahnya kemampuan akademik, faktor pembiayaan, domisili saat menempuh studi dan faktor lainnya.

Database perguruan tinggi menyimpan data akademik, administrasi dan biodata mahasiswa. Data tersebut apabila digali dengan tepat maka dapat diketahui pola atau pengetahuan untuk mengambil keputusan [5]. Serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining [6]. Data mining memecahkan masalah dengan menganalisis data yang telah ada dalam database.

Data mining dalam dunia pendidikan dikenal dengan Educational Data Mining [7]. EDM mengembangkan metode untuk menggali data pendidikan dan menggunakan metode tersebut untuk lebih memahami siswa. EDM dapat membantu pendidik untuk menganalisis cara belajar, mendeteksi mahasiswa yang memerlukan dukungan dan memprediksi kinerja mahasiswa.

Perguruan tinggi perlu melakukan prediksi perilaku mahasiswa untuk mencegah secara dini kegagalan akademik mahasiswa. Penelitian yang dilakukan oleh Kotsiantis, Pierrakeas dan Pintelas [8] menyebutkan bahwa sangat penting bagi dosen untuk mendeteksi mahasiswa yang cenderung drop out sebelum mereka memasuki pertengahan masa studi.

Berdasarkan berlimpahnya data mahasiswa dan data jumlah kelulusan mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data mahasiswa sehingga berguna bagi pihak universitas [9]. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (knowledge Discovery), misalnya informasi mengenai pengklasifikasian data mahasiswa berdasarkan profil dan data akademik. Pengetahuan baru tersebut dapat membantu pihak universitas untuk melakukan klasifikasi mengenai tingkat kelulusan mahasiswa guna menentukan strategi untuk meningkatkan kelulusan pada tahun-tahun berikutnya.

Berdasarkan data yang diperoleh dari IASol yaitu sebuah sistem informasi akademik

Universitas AKI [10] yang didalamnya terdapat beberapa informasi akademik diantaranya Nomor Induk Mahasiswa (NIM), nama, jurusan, umur, jenis kelamin, daerah asal, status pernikahan, status pekerjaan, kelompok atau jenis beasiswa, indeks prestasi dari tiap semester, IPK, jumlah sks yang ditempuh dan jenis konsentrasi jalur peminatan. Berdasar data yang ada, total jumlah kelulusan mahasiswa fakultas ilmu komputer pada tahun 2011 sampai dengan tahun ajaran 2016 yang jumlah kelulusannya fluktuatif dijadikan sebagai dasar acuan dilakukannya proses klasifikasi. Jika hasil klasifikasi kelulusan mahasiswa menunjukkan tingkat peningkatan maupun penurunan, maka hasil klasifikasi tersebut dapat dijadikan sebagai salah satu bahan evaluasi dalam menentukan kebijakan Fakultas Ilmu Komputer UNAKI dengan menggunakan teknik Data Mining.

Pengkategorian atribut yang berpengaruh pada klasifikasi kelulusan mahasiswa masih cukup rumit dan cenderung subyektif. Masih sangat sulit untuk mendefinisikan permasalahan tersebut. Tetapi ada beberapa ciri khusus yang dapat digunakan untuk pengkategorian kelulusan berdasar atribut pada mahasiswa.

Beberapa algoritma klasifikasi data mining telah digunakan untuk memprediksi perilaku mahasiswa yang berpotensi lulus tidak tepat waktu diantaranya decision tree, neural network, naïve bayes, instance-based learning, logistic regression dan support vector machine. Penelitian yang dilakukan oleh Gerben W. Dekker [11] menyebutkan bahwa monitoring dan dukungan terhadap mahasiswa di tahun pertama sangat penting dilakukan. Mahasiswa jurusan teknik elektro Universitas Eindhoven yang berhenti studi pada tahun pertama mencapai hingga 40%. Kurikulum yang sulit dianggap sebagai salah satu penyebab tingginya jumlah mahasiswa drop out. Selain itu, nilai, prestasi, kepribadian, latar belakang sosial mempunyai peran dalam kesuksesan akademik mahasiswa. Dekker menggunakan algoritma Decision tree, Bayesian classifiers, logistic models, rule-based learner dan random forest.

Berdasar penelitian sebelumnya algoritma klasifikasi Data Mining tentang penentuan status kelulusan salah satunya menggunakan Naïve bayes. Pada penelitian kali ini selain mendapatkan nilai akurasi yang baik juga bertujuan mendapatkan model atribut yang

berpengaruh dengan cara menerapkan Feature Selection.

Feature Selection adalah salah satu cara untuk menentukan atribut yang paling berpengaruh di dalam dataset. Feature Selection berperan memilih subset yang tepat dari set fitur asli, karena tidak semua fitur/atribut relevan dengan masalah [12]. Bahkan beberapa dari fitur atau atribut tersebut mengganggu dan dapat mengurangi akurasi. Noisy Features atau fitur yang tidak terpakai tersebut harus dihapus untuk meningkatkan akurasi. Selain itu dengan fitur atau atribut yang banyak akan memperlambat proses komputasi.

Wrapper Feature Selection terdiri dari Forward Selection, Backward Elimination dan Stepwise Selection. Forward Selection dan Stepwise Selection memiliki hasil yang lebih memuaskan dibandingkan dengan proses Backward Elimination. Forward Selection juga memerlukan waktu komputasi yang relatif lebih pendek dibandingkan dengan Backward Elimination maupun dengan Stepwise Selection.

2. TINJAUAN PUSTAKA

2.1 Tinjauan Studi

Penelitian terkait yang tentang klasifikasi yang menggunakan algoritma Naive Bayes dan Forward Selection:

L. Ladha dan T .Deepa [13] dalam penelitiannya yang berjudul Feature Selection Methods And Algorithms. Bayesian classifier adalah statistik, algoritma klasifikasi Naive Bayes didasarkan pada aturan Bayes dan mengasumsikan bahwa kelas yang diberikan adalah fitur yang independen. Secara teori pengklasifikasian Bayesian memiliki tingkat kesalahan minimal dibandingkan dengan algoritma lainnya. Tetapi hal ini tidak selalu terjadi dalam prakteknya, karena asumsi yang disebutkan sebelumnya. Meski begitu Feature Selection pada Naive Bayes classifier menunjukkan akurasi dan kecepatan tinggi bila diterapkan pada database yang besar.

Mark A. Hall dan Geoffrey Holmes [14] dalam penelitiannya menyajikan perbandingan patokan metode Feature Selection dari beberapa algoritma klasifikasi. Dan menyimpulkan bahwa metode Forward Selection sangat cocok untuk Naive Bayes.

2.2 Landasan Teori

Data mining dalam dunia pendidikan dikenal dengan Educational Data Mining [15]. EDM muncul terkait dengan pengembangan metode

untuk menggali data pendidikan untuk lebih memahami perilaku mahasiswa. Dengan memahami perilaku mahasiswa maka dapat diprediksi mahasiswa yang berpotensi gagal dalam akademik.

Dalam penelitian ini akan dijelaskan tentang teknik klasifikasi algoritma data mining. Algoritma yang akan dianalisis adalah: Naive Bayes atau Bayes classifier dengan Algoritma fitur seleksi yaitu Wrapper Feature Selection khususnya Forward Selection.

Data Mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar. Data mining memiliki hubungan dari bidang ilmu seperti artificial intelligent, machine learning, statistik dan database [6].

2.2.1 Teknik Klasifikasi Data Mining

Menurut Witten [6], serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data disebut dengan data mining. Data mining memecahkan masalah dengan menganalisis data yang telah ada dalam database. Proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui [16].

2.2.1.1 Naive Bayes

Naive Bayes merupakan salah satu penerapan teorema Bayes. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output [16]. Bayes merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class [17]. Bayes memiliki akurasi dan kecepatan yang sangat tinggi saat diaplikasi ke dalam database dengan data yang besar.

Naive Bayes adalah metode yang baik karena mudah dibuat, tidak membutuhkan skema estimasi parameter perulangan yang rumit, ini berarti bisa diaplikasikan untuk dataset berukuran besar [16]. Mudah diintegrasikan sehingga pengguna yang tidak punya keahlian dalam bidang teknologi klasifikasi pun bisa mengerti. Naive Bayes merupakan algoritma yang dapat meminimalkan tingkat kesalahan dibandingkan dengan semua pengklasifikasi lainnya. Namun, dalam praktek ini tidak selalu terjadi, karena untuk ketidakakuratan dalam asumsi yang dibuat untuk penggunaannya class yang tidak utuh dan kurangnya data

probabilitas yang tersedia.

Berikut teorema bayes:

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

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data x merupakan suatu 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 berdasar kondisi pada hipotesis H

$P(X)$: Probabilitas dari X

Pengklasifikasi Bayesian juga berguna dalam membenaran teoritis untuk pengklasifikasi lain yang tidak secara eksplisit menggunakan teorema Bayes [17].

2.2.2 Feature Selection

Feature Selection atau seleksi fitur adalah sebuah proses yang biasa digunakan pada Machine Learning dimana sekumpulan dari fitur yang dimiliki oleh data digunakan untuk pembelajaran algoritma. Feature selection menurut Oded Maimon [12] telah menjadi bidang penelitian aktif dalam pengenalan pola, statistik, dan Data Mining.

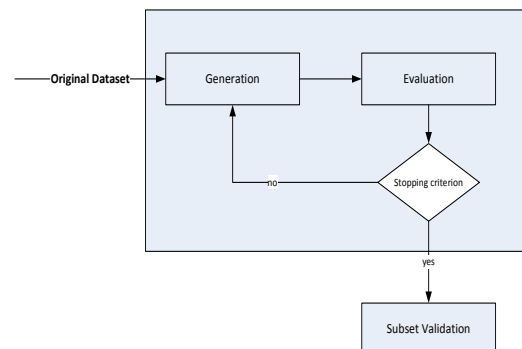
Seleksi fitur adalah salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi karena jika dataset berisi sejumlah fitur, dimensi dataset akan menjadi besar hal ini membuat rendahnya nilai akurasi klasifikasi. Masalah dalam seleksi fitur adalah pengurangan dimensi, dimana awalnya semua atribut diperlukan untuk memperoleh akurasi yang maksimal.

Empat alasan utama untuk melakukan pengurangan dimensi menurut Maimon [12]:

1. Decreasing the learning cost atau penurunan biaya pembelajaran.
2. Increasing the learning performance atau meningkatkan kinerja pembelajaran.
3. Reducing irrelevant dimensions atau mengurangi dimensi yang tidak relevan.
4. Reducing redundant dimensions atau mengurangi dimensi yang berlebihan.

Ide utama dari Feature Selection adalah memilih subset dari fitur yang ada tanpa transformasi karena tidak semua fitur/atribut relevan dengan masalah. Bahkan beberapa dari fitur atau atribut tersebut mengganggu dan mengurangi akurasi. Noisy Features atau fitur

yang tidak terpakai tersebut harus dihapus untuk meningkatkan akurasi. Selain itu dengan fitur atau atribut yang sangat banyak akan memperlambat proses komputasi. Berikut gambar tahapan Feature Selection.



Gambar 2.1 Feature Selection

2.2.2.1 Forward Selection

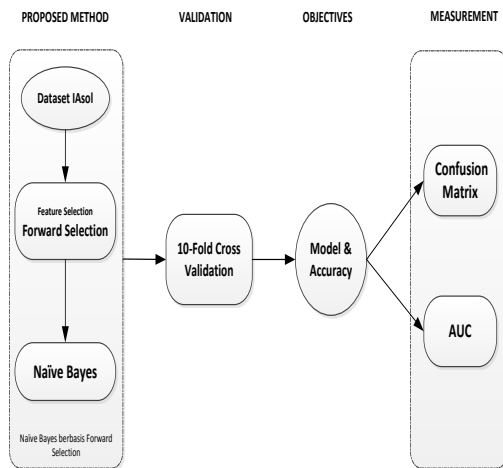
Metode Sequential Forward Selection atau metode seleksi maju adalah algoritma pencarian paling sederhana. Forward Selection didasarkan pada model Regresi Linear. Forward Selection adalah salah satu teknik untuk mereduksi dimensi dataset dengan menghilangkan atribut-atribut yang tidak relevan atau redundan [17]. Metode Forward Selection adalah pemodelan dimulai dari nol peubah (empty model), kemudian satu persatu peubah dimasukan sampai kriteria tertentu dipenuhi.

2.2.3 Evaluasi dan Validasi Hasil Klasifikasi Data Mining

Evaluasi klasifikasi didasarkan pada pengujian pada obyek benar dan salah [14]. Dalam penelitian ini menggunakan metode Confusion Matrix dan AUC untuk mengukur hasil proses klasifikasi.

2.3 Kerangka Pemikiran

Kerangka pemikiran pada penelitian ini berdasarkan klasifikasi data mining dari dataset IAsol UNAKI. Algoritma klasifikasi Naïve Bayes diketahui bisa menangani masalah dataset yang besar, Sedangkan proses fitur seleksi yaitu Forward Selection digunakan untuk menentukan atribut yang paling berpengaruh dan dapat membantu meningkatkan hasil akurasi klasifikasi Naïve Bayes.

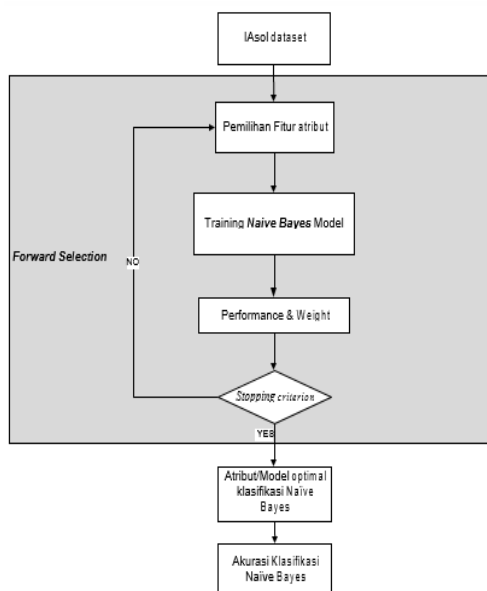


Gambar 2.2 Kerangka pemikiran

3. METODE PENELITIAN

3.1 Desain Penelitian

Dalam konteks penelitian, metode yang dilakukan mengacu kepada pemecahan masalah yang meliputi mengumpulkan data, merumuskan hipotesis atau proposisi, pengujian hipotesis, menafsirkan hasil, dan kesimpulan [17]. Menurut Dawson [18] ada empat metode penelitian yang umum digunakan yaitu: Action Research, Experiment, Case Study, dan Survey. Pada Penelitian kali ini menggunakan metode penelitian eksperimen, dengan metode sebagai berikut:



Gambar 3.1 Tahapan proposed method
Tahap ini akan membahas metode yang akan digunakan untuk penelitian. Berikut ini adalah tahap yang akan dilakukan dalam penelitian. Tahapan dilakukan mengikuti langkah-langkah metode Forward Selection dengan

algoritma Naive Bayes yaitu:

Dataset dari IAsol UNAKI diseleksi fitur menggunakan Forward Selection, Metode Forward Selection adalah pemodelan dimulai dari nol peubah (empty model).

Pemilihan fitur seleksi forward selection diuji menggunakan training atau metode Naive Bayes.

Dari training Naive Bayes yang diujikan mendapatkan hasil dan pembobotan.

Apabila proses tersebut lolos maka akan mendapatkan suatu atribut/model yang optimal dari klasifikasi Naive Bayes.

Sedangkan bila proses tersebut berhenti pada stopping criterion maka proses tersebut diulang dari awal (pemilihan fitur seleksi forward selection) sampai mendapatkan atribut/model optimal.

Setelah mendapatkan atribut/model yang optimal pada klasifikasi Naive Bayes maka akan muncul hasil akurasi dari klasifikasi Naive Bayes yang sudah di fitur seleksi.

Tahap ini akan membahas metode yang akan digunakan untuk penelitian nanti. Berikut ini adalah tahap yang akan dilakukan dalam penelitian. Seleksi fitur digunakan sebagai input untuk proses klasifikasi. Seleksi fitur dilakukan dengan mengambil sebagian variabel pada seluruh atribut yang ada pada data untuk dijadikan atribut penentu dalam melakukan pemberian keputusan. Dataset diseleksi fitur menggunakan Forward Selection, proses selanjutnya adalah melakukan klasifikasi menggunakan algoritma Naive Bayes, hasil proses klasifikasi di evaluasi dengan menggunakan Confusion Matrix dan AUC untuk mengukur performan atau tingkat akurasi.

3.2 Pengujian Model/Metode

Pada tahapan ini menjelaskan tentang teknik pengujian yang digunakan. Tahap modeling untuk mengklasifikasikan status kelulusan dengan menggunakan dua metode yaitu algoritma Naive Bayes dan Forward Selection-Naive Bayes. Proses eksperimen dan pengujian model menggunakan dataset IAsol [6] yaitu dataset kelulusan mahasiswa yang memiliki 2 class atau 2 kategori kelulusan yaitu tepat dan terlambat. Memiliki 210 record dan 22 attribute. Metode eksperimen dan pengujian ini mengikuti cara pengklasifikasian menggunakan RapidMiner.

3.3 Evaluasi Dan Validasi Hasil

Pada tahap ini akan dibahas tentang hasil evaluasi dari eksperimen yang telah dilakukan.

Model yang terbentuk akan diuji dengan menggunakan Confusion Matrix untuk mengetahui tingkat akurasi. Confusion Matrix akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negative yang benar, dan prediksi negative yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan.

Pengujian juga diukur dengan menggunakan AUC, semakin tinggi nilai AUC, maka semakin baik pula model klasifikasi yang terbentuk

4. HASIL DAN PEMBAHASAN

4.1 Hasil

Pada penelitian ini menguji keakuratan klasifikasi kelulusan mahasiswa fakultas ilmu komputer UNAKI dengan menggunakan algoritma Naïve Bayes, setelah itu Naïve Bayes dengan Forward Selection sebagai fitur seleksi.

4.1.1 Evaluasi Naïve bayes

Hasil klasifikasi dengan metode Naïve Bayes diperoleh hasil nilai akurasi sebesar 90,95%.

accuracy: 90.95% +/- 4.40% (mikro: 90.95%)

	true Tepat	true Tertambat	class precision
pred. Tepat	140	2	98.59%
pred. Tertambat	17	51	75.00%
class recall	88.17%	98.23%	

Gambar 4.1 hasil akurasi metode naïve bayes

Evaluasi dan validasi menggunakan aturan AUC mendapatkan nilai 0,984



Gambar 4.2 hasil evaluasi nilai AUC

4.1.2 Evaluasi Forward Selection dengan Naïve Bayes

Forward Selection berdasarkan pada model Wrapper Feature Selection yaitu metode yang mengadopsi dari algoritma pembelajaran yang utama. Pada penelitian kali ini berdasar dari algoritma Naïve Bayes untuk mengevaluasi kinerja dari subset fitur yang optimal sesuai dengan kriteria akhir berdasarkan perhitungan metode X-Validation.

accuracy: 97.14% +/- 4.06% (mikro: 97.14%)

	true Tepat	true Tertambat	class precision
pred. Tepat	157	6	96.32%
pred. Tertambat	0	47	100.00%
class recall	100.00%	88.59%	

Gambar 4.3 hasil akurasi forward selection naïve bayes



Gambar 4.4 nilai AUC forward selection naïve bayes

attribute	weight
Nim	0
Nama	0
Jurusan	0
Jenis Kelamin	0
Umur	0
Daerah Asal	0
Status Pernikahan	0
Status Pekerjaan	1
Kelompok	0
IPS 1	0
IPS 2	0
IPS 3	0
IPS 4	1
IPS 5	0
IPS 6	0
IPS 7	0
IPS 8	0
IPK	0
SKS	0
Konsentrasi	0

Gambar 4.5 atribut yang berpengaruh

4.2 Pembahasan

Tujuan dari penelitian ini adalah mendapatkan model fitur/atribut parameter yang relevan. Eksperimen algoritma Naïve Bayes memperoleh hasil akurasi sebesar 90,95% dengan dengan nilai evaluasi AUC 0,984.

Eksperimen berikutnya algoritma Naïve Bayes dengan Forward Selection sebagai fitur seleksi memperoleh hasil akurasi sebesar 97,14% dengan dengan nilai evaluasi AUC 0,981.

Berdasarkan pemodelan seleksi atribut dengan menggunakan Forward Selection-Naïve Bayes diperoleh model atribut yang berpengaruh pada klasifikasi kelulusan mahasiswa, yaitu atribut *status pekerjaan* dan *IPK semester 4*.

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Metode Forward Selection dapat mereduksi dimensi dataset yang besar dan dapat membantu meningkatkan hasil akurasi klasifikasi Naïve Bayes.

Dalam hal ini Naive Bayes memanfaatkan fungsi seleksi fitur dari Forward Selection untuk pemilihan atribut data dengan karakteristik data itu sendiri, dan meningkatkan ketepatan klasifikasi Naïve Bayes.

Forward Selection berbasis Naive Bayes lebih akurat dan efektif dalam mengklasifikasikan status kelulusan mahasiswa dengan hasil akurasi 97,14% dan termasuk dalam kategori "excellent classification". Dengan memperoleh atribut yang berpengaruh yaitu: *status pekerjaan* dan *IPK semester 4*.

5.2 Saran

Metode Forward Selection berbasis Naive Bayes terbukti akurat dalam klasifikasi status kelulusan mahasiswa dari dataset yang bersifat class imbalance dengan dimensi data yang besar, tetapi dalam penelitian ini terdapat beberapa saran dalam pengembangannya antara lain prosedur ini tidak selalu mengarahkan ke model pemilihan atribut yang terbaik. Forward Selection berbasis Naive Bayes hanya mempertimbangkan sebuah subset kecil dari semua model-model yang mungkin, sehingga resiko melewatkan atau kehilangan model terbaik akan bertambah, seiring dengan penambahan jumlah variabel bebas.

Membantu administrasi perguruan tinggi untuk memberikan peringatan dini dan pembimbingan awal bagi mahasiswa yang kemungkinan tidak lulus tepat waktu dan

membantu perguruan tinggi dalam membuat kebijakan untuk bisa meningkatkan kelulusan mahasiswa

Penelitian ini dapat dikembangkan dengan metode klasifikasi data mining lainnya, penggunaan metode fitur seleksi atau metode optimasi lainnya.

DAFTAR PUSTAKA

- [1] "Peraturan Pemerintah Republik Indonesia Nomor 66 tahun 2010 tentang Perubahan atas Peraturan Pemerintah Nomor 17 tahun 2010 tentang Pengelolaan dan Penyelenggaraan Pendidikan.,".
- [2] "Undang Undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional,".
- [3] BAN-PT, "Buku VI Matriks Penilaian Instrumen Akreditasi Program Studi Badan Akreditasi Nasional Perguruan Tinggi," 2008.
- [4] C. Marquez-Vera, C. Romero, and S. Ventura, "Predicting School Failure Using Data Mining," *Department of Computer Science*, 2011.
- [5] Alaa El-Halees, "Department of Computer Science," *Mining Students Data to Analyze Learning Behaviour: A Case Study*, 2009.
- [6] Ian H. Witten, Frank Eibe, and Mark A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Asma Stephan and Burlington, Eds. United States of America: Morgan Kaufmann, 2011.
- [7] Ryan S.J.D. Baker and Kalinana Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," 2009.
- [8] S.B. Kotsiantis, C.J. Pierrakeas, and P.E. Pintelas, "Preventing Student Dropout in Distance Learning Using Machine Learning Techniques," *Educational Software Development Laboratory, Department of Mathematics University of Patra, Greece*, 2003.
- [9] Johan Oscar Ong, "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing

President University," *Jurnal Ilmiah Teknik Industri*, Juni 2013.

- [10] <http://www.unaki.ac.id/>.
- [11] Gerben W. Dekker, "Predicting Students Drop Out: A Case Study," *Department of Electrical Engineering Eindhoven University of Technology*, 2009.
- [12] Oded Maimon and Lior Rokach, *Data Mining and Knowledge Discovery Handbook*, 2nd ed.: Springer, 2010.
- [13] L. Ladha and T. Deepa, "Feature Selection Methods And Algorithms," *International Journal on Computer Science and Engineering (IJCSE)*, vol. 3, p. 5, May 2011.
- [14] Mark A. Hall and Geoffrey Holmes, "Benchmarking Attribute Selection Techniques for Discrete Class Data Mining," *IEEE Transactions On Knowledge And Data Engineering*, vol. 15, p. 3, May/June 2003.
- [15] Sotiris Kotsiantis, "Educational Data Mining: A Case Study for Predicting Dropout-Prone Students," *Int. J. of Knowledge Engineering and Soft Data Paradigms*, vol. X, 2010.
- [16] Xindong Wu and Vipin Kumar, *The top ten Algorithms in Data Mining.*: Taylor & Francis Group, LLC, 2009.
- [17] Jiawei Han, *Data Mining Concept And Technique*, 2nd ed., Asma Stephan, Ed. Champaign, United States of America: Multiscience Press, 2007.
- [18] Jacob Cohen, "A Coefficient Of Agreement For Nominal Scale ,".