

PENERAPAN ALGORITMA *NAIVE BAYES* UNTUK KLASIFIKASI PENERIMA BEASISWA PRESTASI

Mulyadi

Abstract - Scholarship is supporting mean for students in college education. With the scholarship students can still continue their education until they complete their studies. This is in accordance with Act 1945 section 31 (1) that every citizen has the right to get education. Based on the article, of course education providers must help so that every student can enjoy education by providing scholarships to eligible. In deciding awardees sometime there are some errors such as lack of proper grantee selection process and the length of the grantee must fit the criteria and requirements of awardees. In this research, data mining modeling is using Naive Bayes algorithm to get the rules in selecting scholarship. Furthermore, the application of modeling results obtained Naive Bayes algorithm is used for classification of awardees. The data used are primary data that dataset scholarship at the University of BSI. From the results of model testing are done, the value of high accuracy for the classification of awardees with a value of 100% accuracy and AUC 1. Thus the application of Naive Bayes algorithm can be used as an alternative decision-making in the assessment of awardee.

Intisari - Beasiswa adalah sarana penunjang untuk mahasiswa dalam mengenyam pendidikan perguruan tinggi. Dengan adanya beasiswa mahasiswa tetap bisa melanjutkan pendidikannya hingga selesai masa studinya. Hal ini sesuai dengan Undang-Undang 1945 pasal 31 (1) bahwa tiap-tiap warga negara berhak mendapatkan pengajaran. Berdasarkan pada pasal tersebut, tentu penyelenggara pendidikan harus membantu agar setiap mahasiswa dapat menikmati pendidikan dengan memberikan beasiswa kepada yang berhak menerima. Dalam penentu penerima beasiswa sering sekali terjadi kesalahan seperti penerima beasiswa yang kurang tepat dan lamanya proses seleksi penerima beasiswa yang harus sesuai kriteria dan syarat penerima beasiswa. Dalam penelitian ini dilakukan pemodelan data mining dengan menggunakan algoritma Naive Bayes untuk mendapatkan rule dalam penentuan beasiswa. Selanjutnya hasil penerapan pemodelan algoritma Naive Bayes yang didapat digunakan untuk klasifikasi penerima beasiswa. Data yang digunakan merupakan data primer yaitu dataset beasiswa pada Universitas BSI. Dari hasil pengujian model yang dilakukan, diperoleh nilai akurasi yang tinggi untuk klasifikasi penerima beasiswa dengan nilai akurasi 100% dan AUC 1. Dengan demikian penerapan algoritma Naive Bayes dapat dijadikan alternatif pengambilan keputusan dalam penilaian penerimaan beasiswa.

I. PENDAHULUAN

Peran dunia pendidikan sangatlah erat kaitannya

¹Program Studi Manajemen Informatika, AMIK BSI Tasikmalaya, Jl. Dr.Sukarjo No.28, Tasikmalaya. (Telp.0265-312989;email : mulyadi.myd@bsi.ac.id)

dengan kemajuan suatu bangsa. Hal ini dibuktikan dengan adanya peraturan undang-undang di Indonesia yang menegaskan akan pentingnya pendidikan berdasarkan Undang-Undang Dasar 1945 Pasal 31 (1) bahwa tiap-tiap warga negara berhak mendapatkan pengajaran. Dalam hal tersebut perguruan tinggi memiliki andil dalam proses pengajaran, serta memfasilitasi mahasiswa yang kurang mampu dan berprestasi untuk mendapatkan fasilitas beasiswa yang diselenggarakan diperguruan tinggi.

Kesalahan dalam penentuan calon penerima beasiswa seperti terpilihnya penerima beasiswa yang kurang tepat sering terjadi. Selain itu, proses pengambilan keputusan untuk menentukan calon penerima beasiswa yang dilakukan secara manual membutuhkan waktu lama karena proses seleksi harus sesuai kriteria dan syarat penerima beasiswa (Tabrani, 2014). Hal ini sering menimbulkan kesalahan dalam penentuan calon penerima beasiswa seperti terpilihnya penerima beasiswa yang kurang tepat, maka dari itu diperlukan klasifikasi untuk membantu menentukan keputusan penerima beasiswa prestasi.

Dalam hal ini penggunaan metode *data mining* sangatlah tepat untuk menemukan pola di dalam pengolahan datanya. Karena *data mining* adalah melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data (Witten, Frank, dan Hall, 2011).

Diantara beberapa metode *data mining*, *Naive Bayes* dapat digunakan untuk mengklasifikasikan sekumpulan data dengan mengidentifikasi dari pola data yang diinput. *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar (Kusrini dan Luthfi, 2009). Dalam beberapa penelitian, pendekatan dengan menggunakan *Naive Bayes* memiliki kinerja yang cukup tinggi untuk mengklasifikasikan data (Hamzah, 2012).

Pada penelitian ini, akan dilakukan pengujian metode *Naive Bayes* dengan melihat nilai akurasi yang diperoleh. Sehingga dengan diketahui hasil yang didapat, dapat terlihat pola untuk menentukan siapa yang berhak menerima beasiswa prestasi.

II. KAJIAN LITERATUR

A. Beasiswa

Dalam Kamus Besar Bahasa Indonesia (2008:153) Beasiswa adalah tunjangan yang diberikan kepada pelajar atau mahasiswa sebagai bantuan biaya belajar.

Pada penyelenggara pendidikan khususnya tingkat perguruan tinggi pada universitas banyak sekali program beasiswa yang ditawarkan kepada mahasiswa yang berprestasi maupun yang tidak mampu. Hal ini didasari dari Undang – Undang Dasar 1945 Pasal 31 (1) bahwa tiap-tiap warga negara berhak mendapatkan pengajaran. Berdasarkan pasal tersebut, tentu penyelenggara pendidikan sudah sepatutnya memberikan pelayanan dan kemudahan bagi mahasiswanya dalam menikmati pendidikan yang bermutu. Untuk itu, mahasiswa yang berprestasi atau tidak mampu sebagai bagian dari terselenggaranya pendidikan berhak mendapatkan beasiswa dari penyelenggara pendidikan sebagai bentuk tambahan penghasilan atau bantuan.

Pada penerima beasiswa BBM (Bantuan Belajar Mahasiswa) dengan menggunakan salah satu algoritma yang terdapat pada *data mining* mengelompokkan beasiswa menjadi tiga kelompok dalam pengklasifikasian yaitu mahasiswa yang berhak menerima beasiswa, mahasiswa yang di pertimbangkan menerima dan mahasiswa yang tidak berhak menerima beasiswa (Rohmawati, Defiyanti dan Jajuli, 2015). Dalam penelitian lain penentu penerima beasiswa, nilai siswa juga menjadi acuan untuk pengambilan keputusan ketika untuk masuk pendidikan lanjutan ke universitas favorit. Data nilai siswa perlu dikelompokkan untuk membedakan nilai yang baik dan buruk dengan jangkauan kelompok nilai tertentu. Hasil pengelompokan nilai ini dapat digunakan untuk membuat suatu kebijakan sekolah untuk memberikan beasiswa (Muzakir, 2014). Sementara itu, dalam menentukan penerima beasiswa simpati proses yang paling banyak menemui kesulitan adalah dalam membandingkan mahasiswa yang satu dengan mahasiswa yang lain sesuai kriteria yang ada, karena jumlah mahasiswa yang mengajukan permohonan semakin meningkat. Dengan semakin banyak pemohon beasiswa dan beragam kriteria/pertimbangan yang ada, maka tim dosen pembimbing akademik menemui kesulitan dalam menentukan penerima beasiswa secara tepat (Tjandra, 2014).

Prosedur penerimaan Beasiswa Prestasi Universitas BSI Bandung:

1. Rangkaing 1-10 di sekolah
 - a. Fotocopy rapot sampai kelas 12 semester 1.
 - b. Mendapat surat rekomendasi dari kepala sekolah.
2. Tidak diperkenankan merokok dan mematuhi tata tertib kampus.
3. Mendaftar secara *online* di laman <http://beasiswa.bsi.ac.id> lalu klik *Form* Beasiswa *Online* atau secara *Offline* dengan mengirim ke-Universitas BSI Bandung.
4. Seleksi berkas beasiswa prestasi.
5. Mengikuti Tes Potensi Akademik

B. *Data Mining*

Menurut Turban, dkk dalam Kusri dan Luthfi (2009:3) “*Data mining* adalah proses yang menggunakan teknik statistik, matematik, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar”.

Menurut Hand, Mannila, dan Smyth dalam Widodo, Handayanto dan Herlawati (2013:2) *Data mining* adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

Data mining adalah Melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data (Witten, Frank, dan Hall, 2011).

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan menurut Larose dalam Kusri dan Luthfi (2009:10) sebagai berikut:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas mengumpulkan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.
2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik. Pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan dihasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya. Contoh lain, estimasi nilai indeks prestasi kumulatif mahasiswa program pascasarjana dengan melihat nilai indeks prestasi mahasiswa tersebut pada saat mengikuti program sarjana.
3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah:

 - a. Prediksi harga beras dalam tiga bulan yang akan datang.

- b. Prediksi presentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, diantaranya pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- b. Memperkirakan apakah pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

5. Penglusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yang lainnya dan memiliki tidak kemiripan dengan *record-record* dalam kluster lain.

Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain kan bernilai minimal.

Contoh pengklusteran dalam bisnis dan penelitian:

- a. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari suatu produk bagi perusahaan yang tidak memiliki dana pemasaran yang besar.
- b. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku finansial dalam baik dan mencurigakan.
- c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisa keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

- a. meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respons yang positif terhadap penawaran *upgrade* layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli secara bersamaan.

C. Naive Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *Class*. (kusrini dan luthfi, 2009:189).

Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Klasifikasi Bayesian didasarkan pada teorema *Bayes*, diambil dari nama seorang ahli matematika yang juga menteri *Prebysterian* Inggris, Thomas Bayes (1702-1761) (Bramer, 2013)

Bayesian classification didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan tinggi saat diaplikasikan kedalam *database* dengan data yang besar.

Naive Bayes adalah penyederhanaan metode *bayes*. Teorema *bayes* disederhanakan menjadi persamaan berikut:

$$P(H|X) = P(X|H) P(X) \dots \dots \dots (1)$$

- X : data dengan class yang belum diketahui
- H : hipotesis data X merupakan *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

Hastuti 2012 dalam penelitiannya komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif dengan algoritma yang digunakan adalah *Logistic Regression*, *Decision Tree*, *Naive Bayes* dan *Neural Network*. Hasil penelitian tersebut menunjukkan *Decision Tree* memiliki nilai *accuracy* tertinggi 95,29% Sedangkan pada uji ROC curve menunjukkan bahwa *Neural Network* dan *Naive Bayes* mencapai nilai AUC yang terbaik yaitu 0,976.

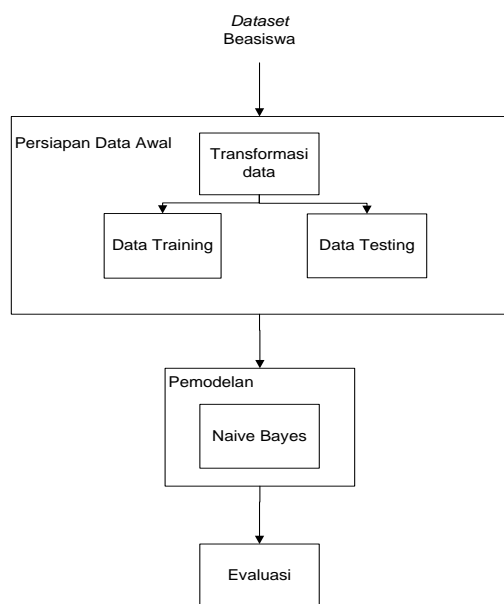
Tabrani 2014 dalam penelitiannya kajian penerepan algoritma untuk klasifikasi penerima beasiswa. Algoritma yang digunakan dalam penelitian tersebut menggunakan algoritma C4.5 dan *Naive Bayes*. Hasil penelitiannya algoritma C.45 memiliki tingkat akurasi yang paling tinggi dengan tingkat akurasi 87.74%.

Mulyadi 2015 dalam penelitiannya komparasi algoritma Naive Bayes dan Suport Vector Mechine untuk klasifikasi penerima beasiswa prestasi. Hasil penelitiannya algoritma Suport Vector Mechine memiliki Tingkat Akurasi yaang paling tinggi dengan tingkat akurasi 97,94% dan AUC 0,998.

Khan dan Peer pada tahun 2013 dalam penelitiannya untuk klasifikasi risiko kredit menunjukkan akurasi dari beberapa metode diantaranya RBT 69%, *OneR Classifier* 64%, *Naive Bayes* 75,5%, ANN 68,5%, C4.5 69,5%, dan *Random Forest* 74,5%. Dilihat dari hasil evaluasi diketahui bahwa *Naive Bayes* memiliki akurasi paling tinggi untuk klasifikasi *German credit dataset*. Namun hal ini mungkin tidak sama untuk semua *dataset*. Umumnya sebuah *classifier* yang digunakan harus disesuaikan dengan berbagai tipe *dataset*(Khan & Peer, 2013).

III. METODE PENELITIAN

Pada bagian ini, akan disajikan gambaran mengenai metodologi penelitian secara keseluruhan yang didasarkan pada CRISP-DM. Berikut adalah tahapan yang dilakukan dalam penelitian ini seperti yang terlihat pada Gambar



Gambar 1. Desain Penelitian

a. Dataset Beasiswa

Pada penelitian ini, dataset beasiswa prestasi terdiri dari 179 informasi data pengajuan aplikasi beasiswa prestasi pada tahun 2015. *Dataset* ini bisa digunakan untuk mengetahui klasifikasi yang berhak menerima beasiswa prestasi. Hasil dari klasifikasi yang digunakan adalah untuk menentukan data beasiswa prestasi yang “lulus” dan “tidak”. Dalam penelitian ini permasalahan yang terjadi adalah untuk memperoleh pola klasifikasi yang lulus dan tidak lulus setelah melewati tahapan penyeleksian beasiswa prestasi. Di dalam *dataset* penelitian ini terdapat 82 aplikasi data yang dinyatakan lulus dan 97 aplikasi data yang dinyatakan tidak lulus.

b. Data Transformation

Atribut yang dilakukan transformasi data diantaranya nilai rata-rata semester 1, semester 2, semester 3, semester 4, semester 5, tes TPA, hasil. Hasil dari transformasi data akan digunakan untuk proses *dataset* untuk algoritma *Naive Bayes*. Selain itu untuk menentukan jumlah kelas interval dalam algoritma *Naive Bayes* dilakukan dengan menggunakan rumus Struges dalam menentukan banyak kelas interval seperti yang terlihat pada rumus:

$$K = 1 + 3,3 \log N \dots\dots\dots (2)$$

K =Banyak kelas.

N =Jumlah data observasi

TABEL 1. HASIL KATEGORIKAL NAIVE BAYES

No	Atribut	Kategorikal
1	Nilai Rata-rata	Nil1 = ... ≤ 74,98
		Nil2 = 74,98> ... ≤ 78,46
		Nil3 = 78,46> ... ≤ 81,95
		Nil4 = 81,95> ... ≤ 85,43
		Nil5 = 85,43> ... ≤ 88,91
		Nil6 = ... > 88,91
2	Semester 1	S11 = ... ≤ 2,5
		S12 = 2,5> ... ≤ 4
		S13 = 4> ... ≤ 5,5
		S14 = 5,5> ... ≤ 7
		S15 = 7 > ... ≤ 8,5
		S16 = ... > 8,5
3	Semester 2	S21 = ... ≤ 2,5
		S22 = 2,5> ... ≤ 4
		S23 = 4> ... ≤ 5,5
		S24 = 5,5> ... ≤ 7
		S25 = 7 > ... ≤ 8,5
		S26 = ... > 8,5
4	Semester 3	S31 = ... ≤ 2,5
		S32 = 2,5> ... ≤ 4
		S33 = 4> ... ≤ 5,5
		S34 = 5,5> ... ≤ 7
		S35 = 7 > ... ≤ 8,5
		S36 = ... > 8,5
5	Semester 4	S41 = ... ≤ 2,5
		S42 = 2,5> ... ≤ 4
		S43 = 4> ... ≤ 5,5
		S44 = 5,5> ... ≤ 7
		S45 = 7 > ... ≤ 8,5
		S46 = ... > 8,5
6	Semester 5	S51 = ... ≤ 2,5
		S52 = 2,5> ... ≤ 4
		S53 = 4> ... ≤ 5,5
		S54 = 5,5> ... ≤ 7
		S55 = 7 > ... ≤ 8,5
		S56 = ... > 8,5
7	Tes TPA	TPA1 = ... ≤ 272

TPA2 = 272 > ... ≤ 344
TPA3 = 344 > ... ≤ 416
TPA4 = 416 > ... ≤ 488
TPA5 = 488 > ... ≤ 560
TPA6 = ... > 560

Sumber: Hasil pengolahan data (2016)

c. *Data Training* dan *Data Testing*

Pada penelitian ini, menggunakan *dataset private* dari data beasiswa prestasi Universitas BSI Bandung tahun 2015 dengan banyak *record* 179. Untuk menguji model, data akan dibagi menjadi dua, yaitu *data training* dan *data testing*. *Data training* digunakan untuk pengembangan model, sedangkan *data testing* digunakan untuk pengujian model. Dari data 179 dilakukan pembagian 90%, 80%, 70%, 60% untuk *data training* dan 10%, 20%, 30%, 40%, *data testing*, dengan sebaran dijelaskan pada Tabel 2.

TABEL 2. SEBARAN DATA TRAINING DAN DATA TESTING

Pembagian Data	Data training	Data testing
90% - 10%	161	18
80% - 20%	143	36
70% - 30%	125	54
60% - 40%	107	72

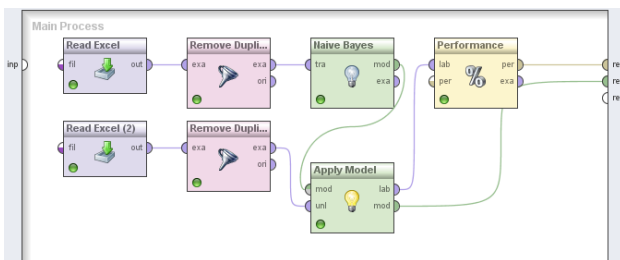
Sumber: Hasil pengolahan data (2016)

d. *Pemodelan*

Dalam penelitian ini akan dilakukan sebuah percobaan *data mining* dengan menggunakan algoritma *Naive Bayes*. *Pemodelan* yang dilakukan dalam penelitian ini menggunakan *software Rapidminer*.

e. *Evaluasi*

Proses evaluasi yang akan dilakukan menggunakan metode ROC (*Receiver Operating Characteristic*) untuk mencari AUC (*Area Under the ROC Curve*) tertinggi dengan menggunakan *software Rapidminer* untuk mengetahui kualitas dan efektivitas dari masing-masing model yang diterapkan dalam penelitian ini.



Sumber: pemodelan data (2016)

Gambar 2. Evaluasi pemodelan menggunakan *data training*

IV. HASIL DAN PEMBAHASAN

Dari hasil pengujian masing model yang sudah dilakukan terhadap *data testing*, dapat diketahui bahwa model klasifikasi mempunyai nilai diagnosa AUC yang

menunjukkan baik pada klasifikasi *Naive Bayes* dengan menggunakan 7 set atribut. Dilihat pada tingkat akurasi yang dihasilkan, model klasifikasi dengan menggunakan algoritma *Naive Bayes* dengan 7 set atribut mempunyai akurasi yang tinggi dengan nilai akurasi 100% dengan kategori klasifikasi *Excellent Classification* seperti yang dapat dilihat pada Tabel 3.

TABEL 3. HASIL PEMODELAN NAIVE BAYES

Dataset	P(Lulus)	P(Tidak)
90% - 10%	0.390	0.610
80% - 20%	0.656	0.344
70% - 30%	0.345	0.655
60% - 40%	0.621	0.379

Sumber: Hasil pengolahan data (2016)

TABEL 4. HASIL PENGUJIAN

Pembagian Data	<i>Naive Bayes</i>	
	Accuracy	AUC
<i>Data training</i> 90%- <i>data testing</i> 10%	100%	1
<i>Data training</i> 80%- <i>data testing</i> 20%	100%	1
<i>Data training</i> 70%- <i>data testing</i> 30%	100%	1
<i>Data training</i> 60%- <i>data testing</i> 40%	100%	1

Sumber: Hasil pengolahan data (2016)

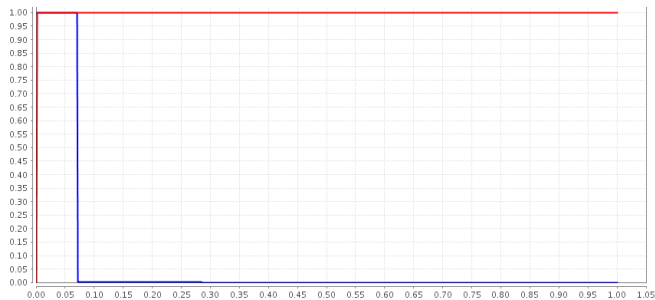
Dilihat pada Tabel 4., model klasifikasi penerapan algoritma *Naive Bayes* dengan 7 set atribut memiliki tingkat akurasi yang tinggi. Oleh karena itu, penggunaan algoritma *Naive Bayes* dengan 7 atribut dapat dijadikan sebagai alternatif untuk klasifikasi penerima beasiswa prestasi karena memiliki akurasi yang tinggi yaitu sebesar 100% dengan kategori klasifikasi *Excellent Classification*.

TABEL 5. CONFUSION MATRIX ALGORITMA NAIVE BAYES 90% DATA TRAINING DAN 10% DATA TESTING

Prediksi	Aktual Lulus	Aktual Tidak
Lulus	14	0
Tidak	0	3

Sumber: Hasil pengolahan data (2016)

Dari *confusion matrix* pada Tabel 5. dapat diukur tingkat akurasinya dari klasifikasi sebagai berikut:
 TP = 14, FP = 0, TN = 3, FN = 0
 Akurasi = ((TP+TN) / (TP+FP+TN+FN)) x 100%
 Akurasi = ((14+3) / (14+0+3+0)) x 100%
 Akurasi = 100%



Sumber: Hasil pengolahan data (2016)

Gambar 3. Kurva ROC Pemodelan *Naive Bayes* dengan perbandingan dataset 90%-10%

TABEL 6. *CONFUSION MATRIX* ALGORITMA *NAIVE BAYES* 80% *DATA TRAINING* DAN 20% *DATA TESTING*

Prediksi	Aktual Lulus	Aktual Tidak
Lulus	27	0
Tidak	0	7

Sumber: Hasil pengolahan data (2016)

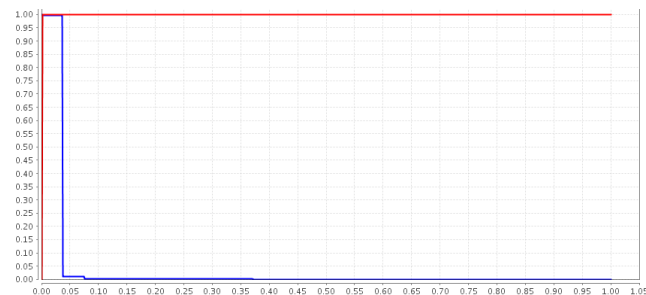
Dari *confusion matrix* pada Tabel 6. dapat diukur tingkat akurasi dari klasifikasi sebagai berikut:

$$TP = 27, FP = 0, TN = 7, FN = 0$$

$$\text{Akurasi} = ((TP+TN) / (TP+FP+TN+FN)) \times 100\%$$

$$\text{Akurasi} = ((27+7) / (27+0+7+0)) \times 100\%$$

$$\text{Akurasi} = 100\%$$



Sumber: Hasil pengolahan data (2016)

Gambar 4. Kurva ROC Pemodelan *Naive Bayes* dengan perbandingan dataset 80%-20%

TABEL 7. *CONFUSION MATRIX* ALGORITMA *NAIVE BAYES* 70% *DATA TRAINING* DAN 30% *DATA TESTING*

Prediksi	Aktual Lulus	Aktual Tidak
Lulus	35	0
Tidak	0	17

Sumber: Hasil pengolahan data (2016)

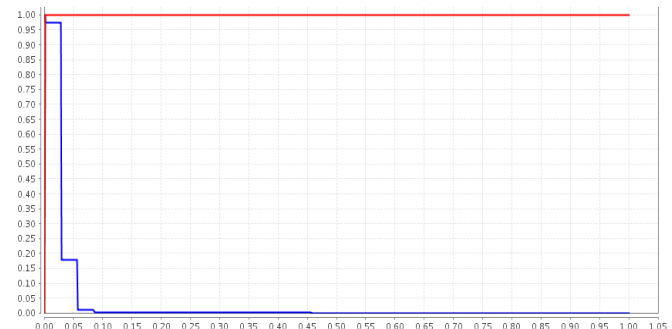
Dari *confusion matrix* pada Tabel 7. dapat diukur tingkat akurasi dari klasifikasi sebagai berikut:

$$TP = 35, FP = 0, TN = 17, FN = 0$$

$$\text{Akurasi} = ((TP+TN) / (TP+FP+TN+FN)) \times 100\%$$

$$\text{Akurasi} = ((35+17) / (35+0+17+0)) \times 100\%$$

$$\text{Akurasi} = 100\%$$



Sumber: Hasil pengolahan data (2016)

Gambar 5. Kurva ROC Pemodelan *Naive Bayes* dengan perbandingan dataset 70%-30%

TABEL 8. *CONFUSION MATRIX* ALGORITMA *NAIVE BAYES* 60% *DATA TRAINING* DAN 40% *DATA TESTING*

Prediksi	Aktual Lulus	Aktual Tidak
Lulus	38	0
Tidak	0	32

Sumber: Hasil pengolahan data (2016)

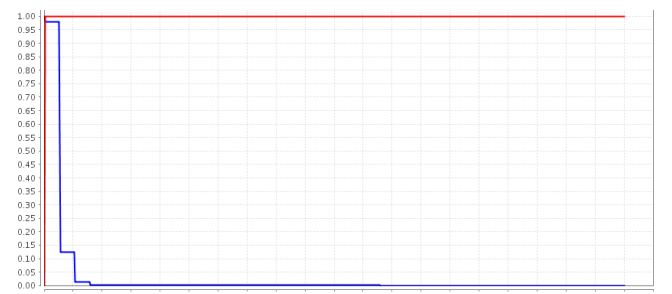
Dari *confusion matrix* pada Tabel 8. dapat diukur tingkat akurasi dari klasifikasi sebagai berikut:

$$TP = 38, FP = 0, TN = 32, FN = 0$$

$$\text{Akurasi} = ((TP+TN) / (TP+FP+TN+FN)) \times 100\%$$

$$\text{Akurasi} = ((38+32) / (38+0+32+0)) \times 100\%$$

$$\text{Akurasi} = 100\%$$



Sumber: Hasil pengolahan data (2016)

Gambar 6. Kurva ROC Pemodelan *Naive Bayes* dengan perbandingan dataset 60%-40%

V. KESIMPULAN

Beasiswa prestasi diharapkan dapat membantu mahasiswa yang berprestasi maupun kurang mampu untuk membantu perekonomian bagi penerimanya. Disamping itu bagi lembaga penyelenggara beasiswa prestasi, merupakan salah satu langkah untuk turut membantu mencerdaskan anak bangsa dan tentu memiliki keuntungan tersendiri bagi lembaga penyelenggara beasiswa prestasi. Berdasarkan kondisi tersebut, diperlukan adanya aspek kontrol untuk memberikan beasiswa prestasi kepada calon mahasiswa yang menerimanya agar tepat sasaran dan tidak berhenti pendidikannya selama masa studi berlangsung.

Dalam penelitian ini dilakukan pengujian model dengan menggunakan algoritma *Naive Bayes* dengan menggunakan *dataset* beasiswa prestasi untuk analisis beasiswa yang lulus dan tidak. Model yang dihasilkan diuji untuk mendapatkan nilai akurasi dan AUC dari algoritma yang diterapkan. Dari hasil pengujian model yang dilakukan, diperoleh nilai akurasi tinggi untuk klasifikasi penerima beasiswa yaitu *Naive Bayes* dengan nilai akurasi 100% dan AUC 1. Untuk itu, algoritma *Naive Bayes* dapat dijadikan alternatif dalam klasifikasi karena memiliki akurasi yang tinggi dalam penelitian ini.

Dengan demikian, dari hasil pengujian model yang dilakukan dalam penelitian ini dapat disimpulkan bahwa penerapan algoritma *Naive Bayes* dapat dijadikan rekomendasi dan alternatif dalam menentukan penerima beasiswa prestasi yang selanjutnya bisa dilakukan ke proses berikutnya pada tahapan penerimaan beasiswa prestasi.

REFERENSI

- [1] Alfisahrin, S. N. (2014). *Komparasi Algoritma C4.5, Naive Bayes dan Neural Network Untuk Memprediksi Penyakit Jantung*. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- [2] Bramer, M. (2013). *Principle of Data Mining Second Edition*. London: Springer.
- [3] Hamzah, A. (2012). Klasifikasi Teks Dengan *Naive Bayes Classifier* (NBC) Untuk Pengelompokan Teks Berita dan Abstract Akademis. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III*, (hal. B-269). Yogyakarta.
- [4] Han, J., & Kamber, M. (2006). *Data Mining Concepts and Techniques Second Edition*. San Francisco: Diane Cerra.
- [5] Hastuti, K. (2012). Analisa Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non-Aktif. *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan*. Semarang.
- [6] Kamus Besar Bahasa Indonesia (2008). Jakarta: PT. Gramedia Pustaka Utama.
- [7] Khan, S., & Peer, M. (2013). Evaluation of Knowledge Extraction Using Various Classification Data Mining Techniques. *IJARCSSE*, 251.
- [8] Kusri., & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [9] Mulyadi. (2015). *Komparasi Algoritma Naive Bayes dan Support Vector Machine* untuk klasifikasi penerima beasiswa prestasi. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- [10] Muzakir, A. (2014). Analisa dan Pemanfaatan Algoritma *K-Means Clustering* Pada Data Nilai Siswa

Sebagai Penentu Penerima Beasiswa. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014*. Yogyakarta.

- [11] Rohmawati, N., Defiyanti, S., & Jajuli, M. (2015). Implementasi Algoritma *K-Means* Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jurnal Ilmiah Teknologi Informasi Terapan*.
- [12] Tabrani, M. (2014). *Kajian Penerapan Algoritma C4.5 dan Naive Bayes untuk klasifikasi Penerima Beasiswa Kopertis*. Jakarta: Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- [12] Tjandra, E. (2014). Sistem Pendukung Keputusan Untuk Menentukan Penerima Beasiswa Simpati di Universitas "X" Dengan Metode Multi_Attribute Global Inference Of Quality (MAGIQ). *Seminar Nasional Teknologi Informasi dan Multimedia*. Surabaya
- [13] Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Cornwall: John Wiley & Sons, Ltd.
- [14] Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). *Penerapan Data Mining Dengan Matlab*. Bandung: Rekayasa Sains.
- [15] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques 2nd Edition*. USA: Elsevier.



Mulyadi. Tahun 2015 lulus dari Program Strata Dua (S2) Program Pasca Sarjana Ilmu Komputer STMIK Nusa Mandiri Jakarta. Saat ini aktif sebagai Dosen di AMIK BSI Tasikmalaya