

KAJIAN ALGORITMA NAÏVE BAYES DALAM PEMILIHAN PENERIMAAN BEASISWA TINGKAT SMA

Virgana¹⁾, Ulfa Pauziah²⁾ dan Michael Sonny³⁾

^{1), 2), 3)} Teknik Informatika Universitas Indraprasta PGRI
Jl Nangka No. 58 Tanjung Barat, Jakarta Selatan

Email : pelangi_ulfa@yahoo.com²⁾, michael.sonny04@gmail.com³⁾

Abstrak

Dalam rangka meningkatkan kualitas sumberdaya manusia adalah dengan pendidikan, namun kemampuan pembiayaan pendidikan dari setiap peserta didik sangat bervariasi. Yang perlu mendapat perhatian adalah peserta didik dengan kemampuan terbatas dari segi finansial tetapi mereka berprestasi dalam bidang akademik. Disisi lain pemerintah maupun pemerintah daerah telah menyediakan beasiswa bagi peserta didik yang tidak mampu dari biaya operasional sekolah. Namun demikian pemberian beasiswa kepada peserta didik belum memenuhi sasaran. Karena sering terjadi peserta didik yang mampu dari segi finansial mendapat program beasiswa, sementara ada peserta didik yang kurang mampu tidak mendapatkan beasiswa. Oleh karena itu dalam penelitian ini dilakukan kajian algoritma naive bayes yang diaplikasikan terhadap data peserta didik yang akan menerima beasiswa, dengan bantuan datamining. Penelitian ini bertujuan untuk mengukur tingkat akurasi dari kajian algoritma dalam pemilihan calon penerima beasiswa di tingkat SMA maupun SMK dan menerapkan metode tersebut. Pengujian dilakukan dengan mengukur kinerja algoritma tersebut menggunakan metode pengujian Cross Validation, Confusion Matrix dan Kurva ROC.

Keyword: Naïve Bayes Algoritma , Datamining, Beasiswa

1. Pendahuluan

Pendidikan di Indonesia selalu berhadapan dengan persoalan kemiskinan. Kemiskinan ini kemudian menjadi alasan seseorang mengajukan permohonan beasiswa kepada lembaga pendidikan. Berbagai langkahpun ditempuh, antara lain melalui ketersediaan dana pembebasan biaya sekolah melalui program Surat Keterangan Tidak Mampu (SKTM) yang dikeluarkan dari kantor kecamatan setempat. Adakalanya memang benar sang pelajar tidak mampu secara finansial, tapi tak jarang dari kalangan mampu pun memanfaatkan surat ini. Yang penting bisa dekolah gratis. Jikalau targetnya adalah sekolah gratis tanpa pertanggungjawabkan “beasiswa” yang diterima, seharusnya lembaga

pendidikan mematok beberapa peraturan yang memacu prestasi penerima beasiswa.

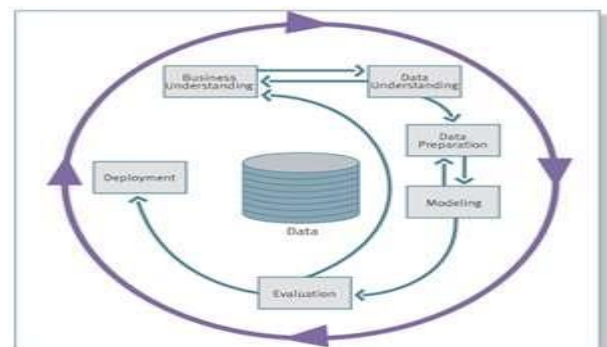
Agar penerima beasiswa ini terus terpacu meningkatkan prestasi akademisnya, tidak keliru apabila lembaga pendidikan tidak serta merta mengabulkan permohonan beasiswa yang diajukan. Harus ada regulasi yang ditetapkan sebagai kategori penerima beasiswa. Dengan kata lain sekalipun seorang siswa berprestasi akademis bagus namun ia berasal dari keluarga mampu, maka ia tidak layak memperoleh beasiswa.

2. Pembahasan

Beasiswa adalah penghasilan bagi yang menerimanya [1]. Dalam ketentuan pasal 4 ayat (1) UU PPh/2000 pengertian penghasilan adalah tambahan kemampuan ekonomis dengan nama dan dalam bentuk apapun yang diterima atau diperoleh dari sumber Indonesia atau luar Indonesia yang dapat digunakan untuk konsumsi atau menambah kekayaan wajib pajak (WP).

Gartner group menyebutkan bahwa data mining adalah proses menelusuri pengetahuan yang baru, pola, dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan tehnik pengenalan pola serta statistik dan teknik matematika[2].

Data Mining adalah sebuah proses, yang mana dalam melakukan prosesnya harus sesuai dengan prosedur dari proses tersebut, yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari keseluruhan proses, preprosesing data, pembentukan model, model evaluasi, dan tahap akhir penyebaran model [2]. Gambar 1 merupakan proses *Data Mining* CRISP-DM sebagai berikut :



Gambar 1 Proses CRISP-DM [2]

Algoritma Naive bayes merupakan salah satu metode pengklasifikasi berpeluang sederhana yang berdasarkan pada penerapan Teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen).

Dua kelompok peneliti, satu oleh Pantel dan Lin, dan yang lain oleh Microsoft Research memperkenalkan metode statistik Bayesian ini pada teknologi anti spam filter. Tetapi yang membuat algoritma Bayesian filtering ini populer adalah pendekatan yang dilakukan oleh Paul Graham

Weka (*Waikato Environment for Knowledge Analysis*) adalah sebuah alat (*tool*) yang merupakan aplikasi *data mining* berbasis *open source (GPL)* yang ditulis dengan Java. *Software* ini mulai dikembangkan sejak tahun 1994 oleh University of Waikato, *New Zealand*. Kelebihan dari Weka yaitu, mudah digunakan, selalu *up to date* dengan algoritma-algoritma baru, Teknik klasifikasi dan algoritma yang digunakan di Weka disebut *classifier*. Weka berbasis *GUI (Graphical Interface User)* dan dapat digunakan untuk mengintegrasikan metode baru yang dibuat sendiri dengan beberapa ketentuan, selain itu Weka tidak saja digunakan untuk akademik namun banyak juga dipakai



untuk memprediksi suatu bisnis perusahaan.

Gambar 2 Aplikasi Weka

Untuk menguji model, pada penelitian ini, digunakan metode *Cross Validation*, *Confusion Matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

1. *Confusion Matrix*

Metode ini hanya menggunakan table matriks seperti pada Tabel 1, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [3]

Tabel 1 Tabel Model Confision Matrix

Correct classification	Classified as	
	+	-
+	true positives	false negatives
-	false positives	true negatives

True Positive adalah jumlah record positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah record negative yang diklasifikasikan sebagai positif, *false negative* adalah jumlah record positif yang diklasifikasikan sebagai negative, *true negative* adalah

jumlah record negative yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *Specifity*, *precision*, dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah record yang positif sedangkan *Specifity*, *precision* adalah perbandingan jumlah TN terhadap jumlah record yang negative. Untuk menghitung digunakan persamaan dibawah ini [4]:

$$Sensitivity = \frac{TP}{P}$$

$$Specifity = \frac{TN}{N}$$

$$Precision = \frac{TP}{TP+FP}$$

$$accuracy = Sensitivity \frac{P}{(P+N)} + Specifity \frac{N}{(P+N)}$$

Keterangan :

TP = Jumlah *true positives*

TN = Jumlah *true negative*

P = Jumlah *record positives*

N = Jumlah *tupel negatives*

FP = Jumlah *false positives*

2. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* sebagai garis vertical[5]. *The area under curve (AUC)* dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC digunakan dengan menggunakan rumus [6]:

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(x_i^r, x_j^r)$$

Dimana :

$$(x, y) = \begin{cases} 1 & y < x \\ \frac{1}{2} & y = x \\ 0 & y > x \end{cases}$$

Keterangan :

K = Jumlah algoritma klasifikasi yang dikomparasi

X = Output positif

Y = Output negatif

dikembangkan sejak tahun 1994 oleh University of Waikato, *New Zealand*.

Metode Penelitian

Jenis penelitian ini adalah penelitian eksperimental merupakan penelitian yang bersifat uji coba, memanipulasi dan mempengaruhi hal-hal yang terkait dengan seluruh variabel atau atribut.

Pengumpulan data dilakukan untuk mendapatkan informasi dan data yang diperoleh yang berhubungan dengan penelitian ini. Untuk mengumpulkan data dan

informasi tersebut, dilakukan metode pengumpulan data sebagai berikut:

1. Pengumpulan data primer

Metode ini digunakan untuk mendapatkan data primer, yaitu dengan wawancara langsung dengan pihak sekolah.

2. Pengumpulan data sekunder

Data sekunder dikumpulkan dengan mengamati data, membaca, mempelajari dan mengutip dari buku literatur, serta sumber-sumber yang berhubungan erat dengan penelitian ini.

Hasil Penelitian

Penelitian ini bertujuan untuk menerapkan dan mendapat nilai akurasi dari algoritma naïve bayes dalam pemilihan penerima beasiswa.

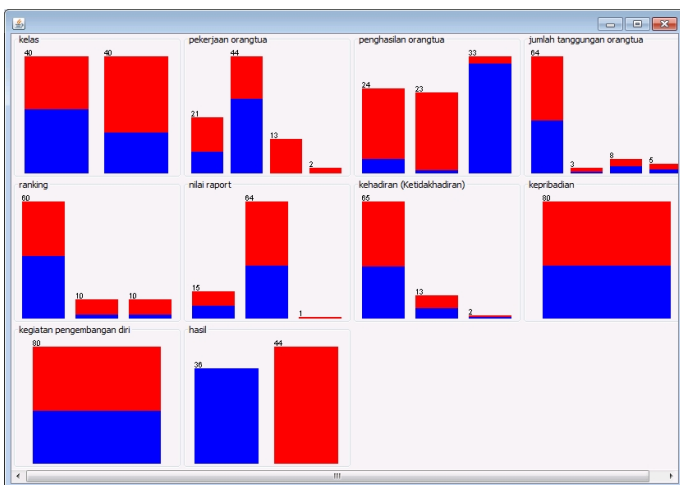
Dalam penerapan ada data yang diolah, dimana dalam mengolah data dibantu dengan aplikasi weka.

Adapun data yang dipakai adalah data training yang berjumlah 80 data:

Gambar 3 data training

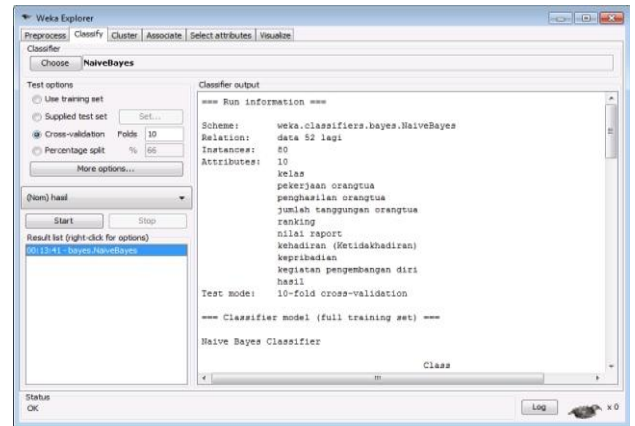
1. Dihitung dengan metode algoritma naïve bayes dengan bantuan aplikasi weka

Didapat diagram batang keseluruhan atributnya:

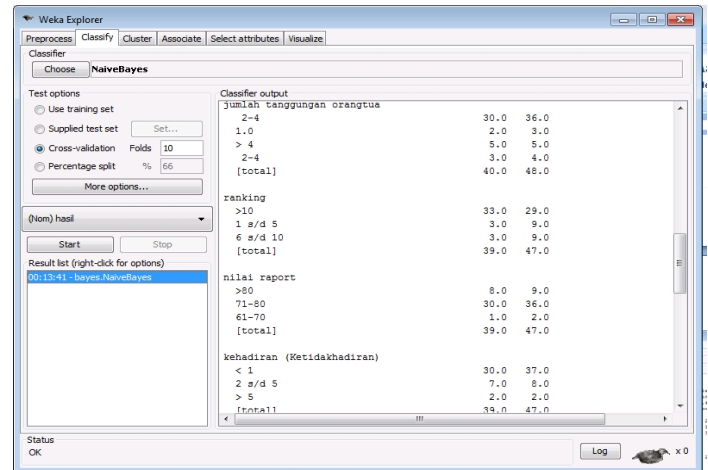


Gambar 4 diagram batang

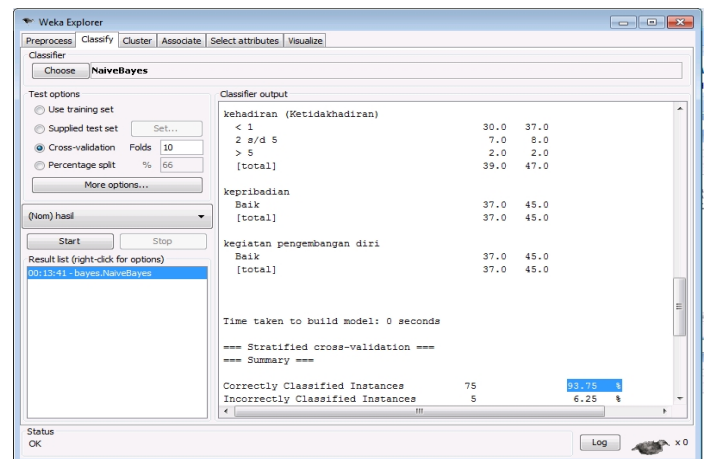
2. Selanjutnya data diuji dengan Cross Validation untuk mendapatkan nilai akurasi dari metode algoritma naïve bayes.



Gambar 5 perhitungan data dengan algoritma naïve bayes



Gambar 6 perhitungan algoritma naïve bayes



Gambar 7 nilai akurasi algoritma Naïve bayes.

Dari hasil perhitungan algoritma naïve bayes diatas dengan menggunakan bantuan tools weka didapat hasil bahwa 93,75 % algoritma naïve bayes dapat membantu dalam pengambilan keputusan pada pemilihan penerima beasiswa disekolah, sedangkan 6,25% tidak dapat membantu untuk pengambilan keputusan.

3. Evaluasi dan Validasi

Model yang telah dibentuk diuji tingkat akurasi dengan memasukan data uji yang berasal dari data *training*. Karena data yang didapat dalam penelitian ini setelah proses *preprocessing* hanya 80 data maka digunakan metode *cross validation* untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode *naïve bayes* sebesar 93,75 %, Selain itu dalam penelitian ini diuji juga dengan menggunakan *confution matrix* dan kurva ROC

a. Confution matrix

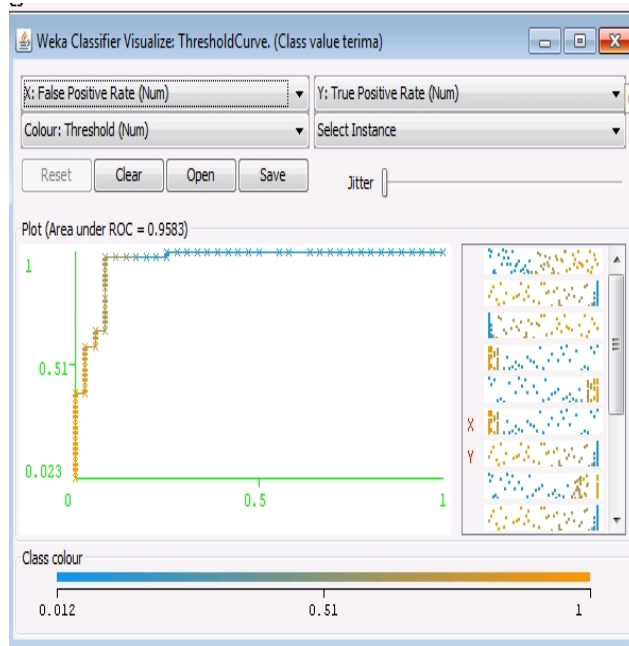
Tabel 5.1 adalah perhitungan berdasarkan data *training*, diketahui dari 80 data, 43 diklasifikasikan terima, lalu 32 data diprediksi tidak.

Tabel 2 Model *Confusion Matrix* untuk Metode Naïve Bayes

```

=== Confusion Matrix ===
  a  b  <-- classified as
32  4  |  a = tidak
 1 43  |  b = terima
    
```

a. Kurva ROC



Gambar 8 kurva ROC

Kurva ROC pada gambar 5.15 diatas mengekspresikan *confusion matrix* Garis X adalah *false positives* dan garis Y *true positives*. Sebesar 0.9583.

3. Impilkasi penelitian

Hasil dari penelitian yang telah dilakukan ini diharapkan dapat memberikan inspirasi dan masukan bagi pihak Sekolah di DKI Jakarta untuk dapat memanfaatkan metode algoritma naïve bayes yang dapat membantu dalam pengambilan keputusan pada pemilihan penerima beasiswa.

Implikasi dari temuan penelitian ini mencakup pada dua bagian aspek, yaitu manajerial dan sistem.

1. Aspek Manajerial

Dari hasil evaluasi ternyata Algoritma Naïve Bayes terbukti dapat dipakai dalam pemilihan penerima beasiswa. Dimana metode klasifikasi ini diterapkan pada data siswa yang menerima beasiswa. Dengan hasil ini, algoritma Naïve Bayes merupakan algoritma yang dapat membantu dan dapat diterapkan untuk memberikan pemecahan untuk permasalahan pemilihan penerima beasiswa pada sekolah.

2. Aspek sistem

a. Penelitian semacam ini dapat dikembangkan pada unit bisnis serupa atau yang lain. Penelitian ini dapat dikembangkan dengan algoritma yang lain misalkan saja dengan metode *Support Vector Machinne (SVM)*.

b. Untuk mendukung pengambilan keputusan dan pengembangan system informasi model ini dapat diterapkan pada sekolah menggunakan *software Weka*.

3. Kesimpulan

A. Kesimpulan

Dapat Berdasarkan penelitian yang dilakukan dapat diambil kesimpulan sebagai berikut:

1. Adanya bentuk dan fungsi aplikasi yang dapat dipakai untuk kalangan sekolah menengah atas, yang bisa digunakan untuk menghitung layak atau tidaknya seorang siswa tersebut mendapatkan beasiswa.
2. Mempermudah bagi instansi sekolah untuk pemilihan beasiswa yang layak bagi siswa-siswi terutama kalangan tidak mampu di sekolah tersebut.

B. Saran

Saran dari perkembangan penelitian ini adalah

1. dikembangkan di berbagai Sekolah Menengah Atas atau Kejuruan.
2. Penelitian ini dapat dikembangkan di tahun yang berikutnya, agar lebih banyak obyek sekolah yang bisa mengetahui aplikasi system keputusan ini.

Daftar Pustaka

- [1] Anneahira (2012). Tujuan Beasiswa Dalam <http://www.anneahira.com/beasiswa.htm>(diakses pada tanggal 3 Juni 2012).
- [2] Larose, Daniel. T. (2005).*Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [3] Asda Bramer, Max. (2007). Principles of Data Mining London : Springer
- [4] Han, J.,&Kamber, M. (2006). Data MiningConcept and Tehniques. San Fransisco: Morgan Kauffman.
- [5] Vercellis,Carlo.(2009), “Business Intelegent: Data Mining and Optimization for Decision Making”, Southern Gate, Chichester, west Sussex: John Willey & Sons, Ltd, 2009.
- [6] Liao. (2007). Recent Advances in Data Miningof Enterprise Data: Algorithms and Application Singapore :WorldScientific Publishing

Biodata Penulis

H. Virgana, memperoleh gelar Sarjana Komputer (S.Pd), Jurusan Pendidikan Matematika IKIP Jakarta, lulus tahun 1980. Memperoleh gelar Magister of Art (MA) University of Virginia, Spec. Education, lulus tahun 1988. Memperoleh gelar Doktorat (DR) IKIP Jakarta, lulus tahun 2010. Saat ini menjadi Dosen di Universitas Indraprasta PGRI Jakarta

Ulfa Pauziah, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Indraprasta PGRI Jakarta, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, lulus tahun 2012. Saat ini menjadi Dosen di Universitas Indraprasta PGRI Jakarta

Michael Sonny, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Indraprasta PGRI Jakarta, lulus tahun 2009. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, lulus tahun 2011. Saat ini menjadi Dosen di Universitas Indraprasta PGRI Jakarta