PENERAPAN ALGORITMA KLASIFIKASI SEBAGAI PENDUKUNG KEPUTUSAN PEMBERIAN BEASISWA MAHASISWA

Heni Sulistiani, Yohana Tri Utami

Jurusan Informatika, Universitas Teknokrat Indonesia E-mail: henisulistiani@teknokrat.ac.id

Abstrak

Beasiswa merupakan bantuan pemerintah maupun swasta berupa sejumlah uang yang diberikan kepada siswa yang sedang atau yang akan mengikuti pendidikan di sekolah. Beasiswa diberikan dengan harapan dapat menumbuhkan dan meningkatkan semangat mahasiswa untuk berprestasi dilakukan dengan memberikan penghargaan berupa beasiswa tiap semester. Banyaknya calon mahasiswa yang mengajukan beasiswa tersebut dan melebihi kuota yang diberikan mengakibatkan proses penyeleksian penerima memakan waktu yang lama karena penyeleksian harus sesuai dengan kriteria agar penerima beasiswa tepat sasaran. Dalam hal ini penggunaan metode data mining sangatlah tepat untuk menemukan pola di dalam pengolahan datanya. Karena data mining melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data. Classifier Naive Bayes memberikan proses penyeleksian yang cepat dan algoritmanya mudah dimengerti. Dalam beberapa penelitian, pendekatan dengan menggunakan Naive Bayes memiliki kineria yang cukup tinggi untuk mengklasifikasikan data metode Naive Bayes Classifier memiliki keunggulan yaitu kesederhanaan dalam komputasinya. Penelitian ini berfokus pada penerapan algoritma klasifikasi Naive Bayes sebagai pendukung keputusan pemilihan beasiswa Bidikmisi bagi calon mahasiswa untuk klasifikasi pemilihan beasiswa agar mempercepat proses penyeleksian dan tidak terjadi kesalahan dalam penentuan calon penerima beasiswa. Pengujian dilakukan dengan menggunakan teknik pengukuran akurasi dan melihat dari matriks konfusi. Hasil menunjukkan bahwa dengan menggunakan algoritma naive bayes, nilai akurasi mencapai 80%. Kata kunci: Beasiswa, Data Mining, Klasifikasi, Naive Bayes, Seleksi

Pendahuluan

Setiap warga negara Indonesia berhak mendapatkan pendidikan sebagaimana tertuang dalam UUD 1945 pasal 31 ayat 1-2. Namun, tidak semua warga negara Indonesia mendapatkan kesempatan untuk memperoleh pendidikan yang layak dan sesuai yang berkelanjutan [1]. Oleh karena itu bagi peserta didik yang kurang mampu berhak untuk memperoleh biaya pendidikan dan bagi peserta didik yang berprestasi berhak mendapatkan beasiswa [2]. Beasiswa merupakan bantuan pemerintah maupun swasta berupa sejumlah uang yang diberikan kepada siswa yang sedang atau yang akan mengikuti pendidikan di sekolah [3]. Dengan diberikannya beasiswa, diharapkan semangat mahasiswa untuk berprestasi dapat ditingkatkan, dengan begitu mahasiswa berhak mendapatkan penghargaan berupa beasiswa di tiap semester [4].

Salah satu beasiswa yang ada di perguruan tinggi adalah beasiswa bidikmisi. Beasiswa bidikmisi merupakan bantuan biaya pendidikan yang diberikan oleh pemerintah kepada calon mahasiswa yang kekurangan secara ekonomi dan memiliki prestasi bidang akademik maupun non akademik untuk melanjutkan ke perguruan tinggi pada program studi unggulan sampai dengan lulus tepat waktu [5]. Kriteria yang digunakan dalam seleksi penerimaan beasiswa bidikmisi antara lain nama mahasiswa, prestasi, tahun lulus, usia, penghasilan orang tua, pendidikan ayah, pendidikan ibu, pekerjaan ayah, pekerjaan ibu, jumlah tanggungan orang tua.

Dalam proses penyeleksian mahasiswa yang berhak mendapatkan beasiswa bidik misi membutuhkan waktu yang lama, hal tersebut dikarenakan banyakanya calon mahasiswa yang mengajukan dan kuota penerimaan yang terbatas. Sedangkan penerima beasiswa bidikmisi harus tepat sasaran sesuai dengan kriteria yang berlaku [6]. Dalam kondisi lapangan sering muncul 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. Klasifikasi merupakan salah satu teknik yang digunakan secara luas di berbagai bidang, termasuk data pertambangan, yang tujuannya adalah untuk mengklasifikasikan set besar objek ke dalam kelas yang telah ditetapkan, dijelaskan oleh satu set atribut, menggunakan metode pembelajaran *supervised* [7]. Analisis klasifikasi mampu membuat deskripsi yang akurat atau membangun model yang akurat untuk setiap kategori [8].

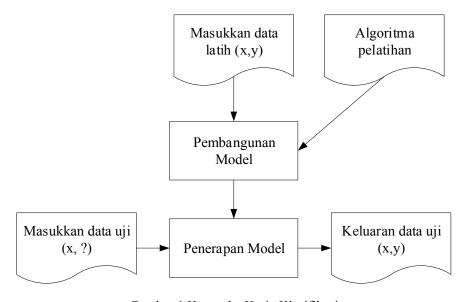
Beberapa metode klasifikasi yang digunakan untuk seleksi penerima beasiswa antara lain *Decission Tree, k-Nearest Neighbor* dan *Naïve Bayes*. Metode *k-Nearest Neighbor* pernah diusulkan untuk melakukan klasifikasi beasiswa PPA (Peningkatan Prestasi Akademik) dan BBM (Bantuan Belajar Mahasiswa) [9]. Namun, metode *k-Nearest Neighbor* memiliki kelemahan yaitu ketika proses prediksi membutuhkan waktu yang lama. Hal ini dikarenakan algoritma tersebut hanya menyimpan sebagian atau seluruh data latih dan model harus membaca kembali semua data latihnya untuk dapat memberikan keluaran label kelas dengan benar pada data uji yang diberikan. Salah satu metode klasifikasi yang algoritmanya mudah dimengerti dan proses prediksinya cepat adalah *Naïve Bayes* [10]. Dalam beberapa penelitian, pendekatan dengan menggunakan *Naive Bayes* memiliki kinerja yang cukup tinggi untuk mengklasifikasikan data metode *Naive Bayes* memiliki keunggulan yaitu kesederhanaan dalam komputasinya.

Penelitian ini membahas mengenai pengujian kegunaan dan keakuratan dari penerapan metode *naïve bayes* sebagai pendukung keputusan pemberian beasiswa bidikmisi bagi mahasiswa pada universitas XYZ. Diharapkan dengan penerapan metode tersebut mampu meminimalisir kesalahan dalam penentuan pemberian beasiswa bidikmisi, sehingga penerima beasiswa bidikmisi dapat tepat sasaran sesuai dengan kriteria-kriteria yang telah ditentukan.

Studi Pustaka

A. Klasifikasi

Klasifikasi merupakan salah satu metode dalam penambangan data (data mining) yang digunakan dalam menganalisis sekumpulan data penting. Metode-metode dalam klasifikasi dapat secara otomatis memprediksi kelas dari data lain yang belum diklasifikasikan. Dua masalah penelitian utama yang berkaitan dengan hasil klasifikasi adalah evaluasi kesalahan klasifikasi dan kekuatan prediksi. Klasifikasi dan prediksi adalah dua bentuk analisis data yang dapat digunakan untuk menggambarkan ekstrak model yang penting pada kelas data atau untuk memprediksi tren data masa depan [11]. Kerangka kerja klasifikasi ditunjukkan pada gambar 1. Pada gambar tersebut, disediakan sejumlah data latih (x, y) untuk digunakan sebagai data membangun model, kemudian menggunakan model tersebut untuk memprediksi kelas dari data uji (x, ?) sehingga data uji (x, ?) diketahui kelas y yang seharusnya. Kerangka kerja yang ditunjukkan pada gambar 1 meliputi dua langkah proses yaitu induksi dan deduksi. Induksi merupakan suatu langkah untuk membangun klasifikasi dari data latih yang diberikan atau disebut juga dengan proses pelatihan, sedangkan deduksi merupakan suatu langkah untuk menerapkan model tersebut pada data uji sehingga data uji dapat diketahui kelas yang sesungguhnya atau disebut juga dengan proses prediksi.



Gambar 1 Kerangka Kerja Klasifikasi

B. Naive Bayes Classifier

Metode *Naive Bayes Classiffier* merupakan penyederhanaan dari teorema *Bayes*, penemu metode ini adalah seorang ilmuwan Inggris yang bernama Thomas Bayes [10]. Probabilitas bersyarat adalah dasar dari teorema Bayes yang dinyatakan dalam persamaan 1.

$$P(X|Y) = \frac{P(X \cap Y)}{P(Y)} \tag{1}$$

Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y atau dengan bahasa lain P(X|Y) adalah prosentase banyaknya X di dalam Y. Dalam metode *Naive Bayes Classiffier* terdapat dua fase vaitu fase *training* dan *testing*.

1. Data Pelatihan (training)

Jika jenis data bersifat kategorial untuk sebagian data yang telah diketahui kelasnya diproses untuk membentuk model perkiraan. Namun jika jenis data bersifat numerik, maka proses training harus melewati beberapa tahapan yaitu perhitungan nilai *mean*, *variance*, dan *deviasi standart* pada tiap kriteria untuk masing-masing golongan.

2. Data Percobaan (testing)

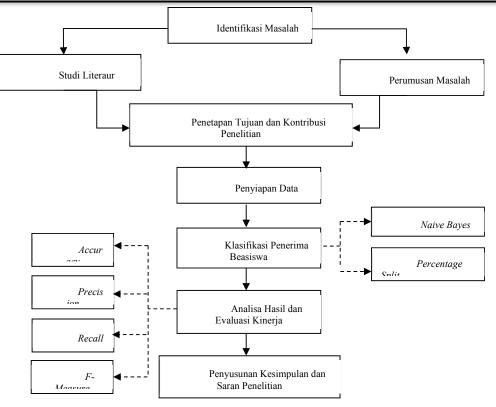
Proses ini merupakan proses perhitungan data yang mengacu pada data *training*. Pada proses ini ada beberapa tahapan, yaitu menghitung peluang kriteria terhadap golongan, menghitung peluang setiap golongan, dan menentukan nilai maksimal pada masing-masing *posterior*.

Metodologi Penelitian

Data yang digunakan dalam penelitian ini merupakan data pengusul dan penerima beasiswa bidikmisi Tahun 2017 di Universitas XYZ yang merupakan salah satu perguruan tinggi swasta di Proponsi Lampung. Terdapat 98 calon mahasiswa yang mengusulkan untuk memperoleh beasiswa bidikmisi, namun kuota yang biberikan hanya sebanyak 35 penerima beasiswa untuk lima program studi yang ada. Atribut dan nilai atribut diperoleh dari data mahasiswa dan nilai tes. Jumlah atribut yang digunakan sebanyak 13, antara lain tahun lulus, usia, jumlah prestasi, tingkat prestasi (sekolah, lokal, regional, nasional atau internasional), jumlah organisasi yang diikuti semasa sekolah, pekerjaan ayah, pekerjaan ibu, penghasilan ayah, penghasilan ibu, nilai tes 1, nilai tes 2, nilai keseluruhan dan hasil (rekomendasi atau tidak direkomendasi). Alur pikir penelitian terdiri dari beberapa tahapan yang dapat dilihat pada gambar 2.

Berdasarkan gambar 2 dapat dilihat tahapan yang dilakukan dalam penelitian ini yaitu dengan melakukan identifikasi masalah terhadap pemilihan penerima beasiswa bidik misi di Perguruan Tinggi pada tahapan awal. Dari tahapan identifikasi masalah, tahapan selanjutnya yaitu melakukan studi literatur dan perumusan masalah yang nantinya akan ditetapkan tujuan dan kontribusi penelitian. Tahapan selanjutnya melakukan penyiapan data, dengan megambil data *history* pengusul dan penerima beasiswa bidik misi untuk dilakukan klasifikasi menggunakan metode *Naive Bayes* dan divalidasi dengan metode *percentage split*. Data yang akan digunakan sebagai *data training* sebanyak 90% dan *data testing* sebanyak 10% dalam melakukan pengklasifikasian. Sebelum dilakukan proses pengolahan, data akan dilakukan normalisasi terlebih dahulu untuk beberapa atribut yang bernilai kategorikal.

Penerapan metode *Naive Bayes Classifier* akan diklasifikasikan ke dalam dua kelas yaitu kelas rekomendasi dan kelas tidak direkomendasi untuk mendapatkan beasiswa. Model prediksi diolah dengan menggunakan WEKA. Analisis hasil dan evaluasi kinerja klasifikasi menggunakan pengukuran *accuracy*, *precision*, *recall* dan *f-measure*. Selanjutnya melakukan penyusunan kesimpulan untuk menjelaskan kesesuaian antara rencana penelitian dengan hasil penelitian serta melakukan penyusunan saran penelitian guna pengembangan lebih lanjut.



Gambar 2 Alur pikir penelitian

Hasil dan Pembahasan

Dataset yang terkumpul, akan dilakukan normalisasi terlebih dahulu. Hal ini dilakukan untuk menghapus beberapa atribut yang tidak digunakan seperti atribut nama mahasiswa, tempat tanggal lahir, alamat dan program studi. Gambar 3 merupakan contoh dari data sebelum dilakukan proses normalisasi.

Prestasi	Organisasi	Pekerjaan Ayah	Pekerjaan Ibu	Penghasilan Aya	h Penghasilan Ibu	Alamat	Jurusan	Nilai Tes l	Nilai Tes 2
- 1	Pramuka	Buruh	Ibu rumah tangga	Rp 2.500.000	-	Л. Urip Sumoharjo Lk.II Kel. Gunung Sula- Kec. Wayhalim Bandarlampung	Informatika 18	60	
Juara 1 Roket Air Kontes Robot	Pramuka		- 1	Rp 500.000		Dusun Wonorejo Desa Pesawaran Indah Kec. Way	Informatika	50	93
Juara 3 Mading Koprasi Unila	Robotik	Buruh Tani) -				
2015	Mading Sekolah					Ratai Kab Pesawaran	24-02-02-02-02-02-02-02-02-02-02-02-02-02-		
Berprestasi dikelas	-	Supir	Ibu rumah tangga	Rp 2.000.000		Perum PU Blok B No.11 RT/RW 001/002 Desa Burni Raya Kec. Abung Selatan lampung utara	Informatika	0	122
Lulusan terbaik	-	Buruh	Ibu rumah tangga	Rp 1.500.000		Л. Soekarno Hatta Km. 15 Kec. Panjang kel Serengsem Bandarlampung	Informatika	54	52
	-	Buruh	Ibu rumah tangga	Rp 1.000.000	-	Л. P. Senopati No. 30 Koorpri Jaya, Sukarame Bandarlampung	Informatika	25	75
Juara II Karate Nasional di Bandung		Tidak Bekerja	Pedagang	*	Rp 750.000	Л. Soekarno Hatta LK. 1 RT.001 Kel. Way Gubak	Informatika	44	108
Juara III Debat Konstitusi Unila	1					Kei. Way Gubak			
Juara 4 Pocari Sweet 2016	Futsal	Buruh	Ibu rumah tangga	Rp 2.000.000	Kampung Karang Jaya, Rt. 14 - LK.1 Kel. Karang Maritim kec panjang bandarlampung		Informatika	68	87
Juara 3 Futsal Walikota Cup						panjang cancariampung	Informatika	E AAII	Idow

Gambar 3 Contoh Data Sebelum Normalisasi

Sebelum data diproses lebih lanjut, data perlu dilakukan normalisasi agar data dapat terbaca pada perangkat lunak Weka dan atribut yang tidak terpakai dihapus agar tidak mempengaruhi nilai akurasi model klasifikasi. Data yang dikumplkan masih berbentuk file dengan ekstensi .xlsx (excel workbook), sedangkan untuk terbaca di aplikasi Weka data harus berbentuk ekstensi .csv (comma delimited). Gambar 4 menunjukkan hasil normalisasi dari dataset yang akan digunakan.

Tahun Lulus, Usia, Jumlah Prestasi, Tingkat Prestasi, Jumlah Organisasi, Pekerjaan Ayah , Pe 2017,17,0,0,1,7,7,Rp2.500.000,0,18,60,156,Tidak Direkomendasi 2017,19,2,2,2,7,7,Rp500.000,0,50,93,286,Rekomendasi 2017,18,1,1,0,6,7,Rp2.000.000,0,0,122,244,Tidak Direkomendasi 2017.18.1.1.0.7.7.Rp1.500.000.0.54.52.212.Rekomendasi 2017,19,0,0,0,7,7,Rp1.000.000,0,25,75,200,Rekomendasi 2017.19.2.5.0.8.2.0.Rp750.000.44.108.304.Tidak Direkomendasi 2016, 19, 2, 2, 1, 7, 7, Rp2.000.000, 0, 68, 87, 310, Rekomendasi 2017.19.0.0.1.7.7.Rp2.500.000.0.36.65.202.Tidak Direkomendasi 2017,18,2,3,1,2,6,Rp1.000.000,Rp1.000.000,94,143,474,Rekomendasi 2017.18.2.3.2.2.7.Rp1.000.000.Rp1.000.000.42.100.284.Tidak Direkomendasi 2017,18,2,2,0,6,7,Rp1.000.000,0,94,120,428,Tidak Direkomendasi 2017.19.0.0.0.7.7.Rp2.000.000.0.7.83.180.Rekomendasi 2017,17,0,0,0,7,2,Rp600.000,Rp750.000,81,131,424,Tidak Direkomendasi 2017,18,0,0,1,6,7,Rp1.500.000,0,0,0,Tidak Direkomendasi 2017,19,2,2,1,7,7,Rp1.091.250,0,76,121,394,Rekomendasi 2016,19,0,0,0,4,7,Rp1,500,000,0,53,78,262,Tidak Direkomendasi 2017,19,2,4,0,1,7,Rp1.250.000,0,0,0,Tidak Direkomendasi

2017,18,0,0,0,5,7,Rp1.000.000,0,30,70,200,Tidak Direkomendasi 2017,19,0,0,0,1,7,Rp1.500.000,0,0,0,Tidak Direkomendasi 2017,18,0,0,0,7,7,Rp750.000,0,80,100,360,Rekomendasi 2016,19,1,2,0,7,7,Rp2.500.000,0,51,98,298,Tidak Direkomendasi 2017,18,0,0,0,8,6,0,Rp1.000.000,0,0,0,Tidak Direkomendasi 2017,19,0,0,0,5,4,Rp750.000,0,0,0,Tidak Direkomendasi 2017,18,0,0,0,6,7,Rp1.000.000,0,55,79,268,Tidak Direkomendasi

Gambar 4 Data Sudah Dinormalisasi

Pengujian pemodelan metode naïve bayes dilakukan dengan menggunakan perangkat lunak Weka. Beberapa aspek pengujian yang dilakukan berdasarkan kebutuhan penggalian data yaitu dengan mengukur nilai akurasi, *precision*, *recall* dan *f-measure*. Data divalidasi dengan menggunakan teknik *percentage split* yaitu dengan membagi data untuk *training* dan *testing* sebanyak 90% dan 10%. Data yang digunakan dalam peneilitian ini berisi 98 *record* dengan 13 atribut. Tabel 2 menunjukkan hasil pengujian data menggunakan model naïve bayes, diperoleh hasil sebagai berikut:

Tabel 2 Hasil Pengujian

1 40 01 2 114011 1 01184 1 411					
Pengukuran	Nilai				
Akurasi	80%				
Presicion	80%				
Recall	80%				
f-measure	80%				

Berdasarkan tabel 2 dapat diketahui bahwa nilai keakuratan dari model naïve bayes sebesar 80%. Hal ini berarti bahwa dari 20% data yang digunakan untuk testing menghasilkan 8 (80%) data yang dapat diklasifikasikan dengan benar dan terdapat 2 (20%) data yang salah diklasifikasikan. Tabel 3 merupakan tabel matrik konfusi dari data *testing*.

Tabel 3 Matrik Konfusi

Hasil Testing	Rekomendasi	Tidak Direkomendasi
Benar/True	3	5
Salah/False	1	1

Tabel 3 dapat diartikan bahwa dari 10 data testing dapat dikelompokkan menjadi dua kelas yaitu "rekomendasi" dan "tidak direkomendasi". Dimana terdapat empat data yang masuk kedalam kategori kelas "rekomendasi" dan enam data masuk kedalam kategori kelas "tidak direkomendasi". Dari kelas "rekomendasi" terdapat tiga data yang diklasifikasikan dengan benar dan satu data diklasifikasikan dengan salah. Sedangkan dari kelas "tidak direkomendasi" terdapat lima data yang dapat diklasifikasikan dengan benar dan satu data diklasifikasikan dengan salah.

Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan, kesimpulan yang dapat diperoleh dalam penelitian ini adalah dengan menerapkan metode naïve bayes sebagai pendukung keputusan dalam pemberian beasiswa bidikmisi mahasiswa di Universiatas XYZ mampu memberikan nilai akurasi sebesar 80%. Pengunaan data sebanyak 10% untuk testing terdapat delapan data yang benar untuk diklasifikasikan.

Daftar Pustaka

- [1] N. Hijriana, P. Studi, T. Informatika, and U. Islam, "Penerapan Metode Decision Tree Algoritma C4 . 5 Untuk Seleksi," pp. 9–13, 2017.
- [2] J. Augusto, D. Guterres, C. Algoritma, and P. Keputusan, "Kelayakan Algoritma C45 Sebagai Pendukung Beasiswa," pp. 142–147, 2015.
- [3] N. Zuwida, M. Jabar, and A. Arizal, "Tinjauan Pemanfataan Pemberian Beasiswa Bantuan Khusus Murid (BKM) Pada Siswa SMK Negeri 1 Pariaman," vol. 2, no. 3, 2014.
- [4] M. A. Rahman, P. P. S. Iain, R. Intan, and B. Lampung, "Algoritma C45 Untuk Menentukan Mahasiswa Penerima Beasiswa (Studi Kasus: PPS IAIN Raden Intan Bandar Lampung)," *J. TIM Darmajaya*, vol. 01, no. 02, pp. 118–128, 2015.
- [5] Direktorat Jenderal Pembelajaran dan Kemahasiswaan, "Pedoman Bantuan Biaya Pendidikan Bidikmisi Tahun 2016," 2016.
- [6] M. Tabrani, "Klasifikasi Penerima Beasiswa Kopertis Dengan Menggunakan Algoritma C.45," *J. Pilar Nusa Mandiri*, vol. XII, no. 1, pp. 72–80, 2016.
- [7] D. AL-Nabi and S. Ahmed, "Survey on Classification Algorithms for Data Mining:(Comparison and Evaluation)," *Comput. Eng. Intell. Syst.*, vol. 1719, no. 8, pp. 18–25, 2013.
- [8] M.-S. Chen, J. Han, and P. S. Yu, "Paper Data Mining, An Overview From Database Perspective." pp. 866–883, 1996.
- [9] Sumarlin, "Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM," *J. Sist. Inf. Bisnis*, vol. 01, pp. 52–62, 2015.
- [10] R. Dewi, Yunita, and N. Indrawati, "Rancang Bangun Sistem Pendukung Keputusan Penerimaan Beasiswa Menggunakan Meotde Naive Bayes Classiffier," pp. 79–84.
- [11] B. R. Patel and K. K. Rana, "A Survey on Decision Tree Algorithm For Classification," *Ijedr*, vol. 2, no. 1, pp. 1–5, 2014.