

Klasifikasi Proses Penjurusan Siswa Tingkat SMA Menggunakan Data Mining

Fata Nidaul Khasanah^{1,*}

¹ Teknik Informatika; STMIK Bina Insani; Jl. Siliwangi No 6 Rawa Panjang Bekasi Timur 17114 Indonesia. Telp. (021) 824 36 886 / (021) 824 36 996. Fax. (021) 824 009 24; e-mail: fatanidaul@gmail.com

* Korespondensi: e-mail: fatanidaul@gmail.com

Diterima: 19 Oktober 2016; Review: 25 Oktober 2016; Disetujui: 2 November 2016

Cara sitasi: Khasanah NF. 2016. Klasifikasi Proses Penjurusan Siswa Tingkat SMA Menggunakan Data Mining. *Informatics for Educators and Professionals*. 1(1): 65 – 69.

Abstrak: Pengarahan jurusan sejak dini dapat memudahkan siswa memilih bidang ilmu yang akan ditekuninya di akademik atau universitas yang tentunya akan mengarah pula kepada karirnya kelak. Kurikulum yang digunakan saat ini penjurusan dilakukan di awal kelas X. Penjurusan siswa SMA merupakan upaya untuk mengarahkan siswa berdasarkan kemampuan akademik dan minat siswa. Proses penjurusan melibatkan beberapa attribute yang digunakan sebagai bahan pertimbangan dimana setiap satuan pendidikan dapat menambah attribute penjurusan sesuai dengan karakteristik dan kebutuhan setiap satuan pendidikan. Perkembangan teknologi saat ini, proses penjurusan siswa tingkat SMA dapat diatasi dengan teknik pengelompokan data didapat dari hasil data mining. Penelitian ini menggunakan metode klasifikasi untuk proses penentuan jurusan siswa tingkat SMA. Penelitian ini bertujuan untuk membandingkan hasil yang diperoleh dari empat algoritma, yaitu J48, Naïve Bayes, OneR dan ZeroR. Algoritma Naive Bayes merupakan algoritma yang mempunyai nilai akurasi tertinggi yaitu 96.74%, selanjutnya adalah algoritma J48 dengan 93.48%, algoritma OneR 90.22%, dan algoritma ZeroR merupakan algoritma yang mempunyai nilai akurasi terendah yaitu 59.78%.

Kata kunci: J48, Klasifikasi, Naïve Bayes, OneR, ZeroR

Abstract: Briefing the Department early on can ease the students choose the fields that will be practiced in the academic or University which certainly will lead to his career in the future. The curriculum used in addition currently done at the beginning of class x. Addition high school students is an attempt to direct the students based on academic ability and interest of students. The process of addition involves some attribute that is used as a material consideration where each unit of education can add attribute in addition according to the characteristics and needs of each educational unit. The current technological developments, HIGH SCHOOL level students in addition process can be overcome by the technique of grouping data obtained from the results of data mining. This research uses a method of classification for students majoring in the process of determining the level of the high school. This study aims to compare the results obtained from four different algorithms, namely the J48, Naïve Bayes, OneR and ZeroR. The algorithm is the algorithm of Bayes Naive had the highest accuracy value i.e. 96.74%, next is J48 algorithm with 93.48% 90.22% OneR, algorithm, and the algorithm is an algorithm which ZeroR has the lowest i.e. 59.78% accuracy.

Keywords: J48, Classification, Naïve Bayes, OneR, ZeroR

1. Pendahuluan

Pemilihan jurusan siswa Sekolah Menengah Atas (SMA) ditentukan berdasarkan kemampuan akademik serta faktor keminatan siswa. Keminatan siswa dianggap penting dalam proses penjurusan hal tersebut dikarenakan siswa akan merasa senang ketika mempelajari ilmu tersebut. Selain itu penjurusan yang disesuaikan terhadap minat akan berpengaruh secara

positif terhadap prestasi akademik. Dengan demikian penjurusan bukan masalah kecerdasan tetapi masalah minat dan bakat siswa.

Pengarahannya sejak dini dapat memudahkan siswa memilih bidang ilmu yang akan ditekuninya di akademik atau universitas yang tentunya akan mengarah pula kepada karirnya kelak. Kurikulum yang digunakan saat ini penjurusan dilakukan di awal kelas X. Penjurusan siswa SMA merupakan salah satu upaya untuk mengarahkan siswa berdasarkan kemampuan akademik dan minat siswa. Proses penjurusan melibatkan beberapa attribute yang digunakan sebagai bahan pertimbangan, yaitu nilai akademik siswa, hasil psikotes dan angket keinginan (minat) siswa. Satuan pendidikan dapat menambah attribute penjurusan sesuai dengan karakteristik dan kebutuhan setiap satuan pendidikan.

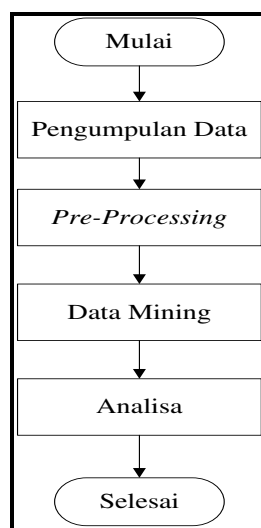
Sistem penjurusan yang selama ini digunakan oleh guru masih dilakukan secara manual. Guru harus menyeleksi satu persatu dalam menentukan jurusan untuk setiap siswa berdasarkan nilai akademik siswa, nilai *placement test* siswa dan angket keinginan (minat) siswa, sehingga menyebabkan proses penentuan jurusan siswa tersebut memakan waktu yang cukup lama untuk mendapatkan hasil penjurusan.

Seiring dengan perkembangan teknologi saat ini, proses penjurusan siswa tingkat SMA dapat diatasi dengan teknik pengelompokan data didapat dari hasil data mining. Data mining merupakan disiplin yang mempelajari metode untuk mengekstrak pengetahuan atau pola dari suatu data, sehingga data mining juga sering disebut *Knowledge Discovery in Database* (KDD). Data mining dapat digunakan untuk mengelompokkan data, memprediksi, mengestimasi dan menentukan kaidah asosiasi dalam suatu data yang ada. Karena adanya data dalam jumlah yang besar diperlukan adanya data mining untuk menghasilkan informasi dan *knowledge* yang berguna. Informasi dan pengetahuan yang didapatkan dapat digunakan untuk mengetahui pola dalam suatu data yang banyak, dan besarnya kebutuhan untuk mengubah data tersebut menjadi informasi yang berguna.

Penelitian ini menggunakan metode klasifikasi untuk penentuan jurusan siswa tingkat SMA. Metode klasifikasi mempunyai beberapa algoritma. Dan dalam penelitian kali ini peneliti menggunakan empat algoritma dan kemudian membandingkan hasil dari keempat algoritma tersebut. Algoritma dalam klasifikasi yang digunakan yaitu algoritma J48, Naïve Bayes, OneR dan algoritma ZeroR.

2. Metode Penelitian

Metode penelitian yang diunakan dalam penelitian ini dimulai dengan proses mencari dan menyusun data, dilanjutkan dengan melakukan studi pustaka yang nantinya mendapatkan kesimpulan dari penelitian ini. Metode yang diusulkan dalam proses penentuan jurusan siswa tingkat SMA adalah metode klasifikasi dengan empat algoritma yaitu algoritma J48, Naïve Bayes, OneR dan algoritma ZeroR.



Gambar 1. Diagram Alir Penelitian

Tahap pengumpulan data peneliti melakukan proses mencari dan menyusun data, sebagai sampel data peneliti menggunakan 40 data siswa SMA pada kelas X. Pada tahap *pre-processing* dengan menjadikan data menjadi berkualitas, sehingga data dapat dilanjutkan pada tahap data mining. Tahapan selanjutnya adalah proses penentuan penjurusan siswa dengan data mining menggunakan empat algoritma yang ada pada metode klasifikasi. Algoritma dalam klasifikasi yang digunakan yaitu algoritma J48, Naïve Bayes, OneR dan algoritma ZeroR. Tahap terakhir yaitu melakukan analisa dengan membandingkan hasil dari keempat algoritma yang memberikan nilai akurasi terbaik.

Penelitian kali ini menggunakan empat algoritma dari metode klasifikasi, diantaranya yaitu:

- a. Algoritma J48
Algoritma J48 didasarkan pada algoritma ID3 dikembangkan oleh Ross Quinlan, dengan fitur tambahan untuk mengatasi masalah dari algoritma ID3 yang sulit. Algoritma ini merupakan salah satu *decision tree* yang telah dipelajari di bidang pengenalan pola dan pembelajaran mesin.
- b. Algoritma Naïve Bayes
Klasifikasi Naïve Bayes menggunakan teori Bayesian dengan mengasumsikan tidak ada hubungan antar attribute. Model ini mudah digunakan serta sederhana sehingga dianggap tepat untuk database yang besar. Hasil dari naïve bayes dianggap baik karena mampu melakukan klasifikasi dengan baik [4].
- c. Algoritma OneR
Algoritma OneR merupakan singkatan dari *One Rule*. Algoritma OneR akan membangkitkan sebuah aturan untuk setiap attribute kemudian memilih aturan dengan error paling kecil yang selanjutnya digunakan sebagai *One Rule*. Untuk membuat aturan (*rule*) setiap attribute yang ada maka perlu membuat tabel kemunculan untuk setiap attribute dengan targetnya [9].
- d. Algoritma ZeroR
Algoritma ZeroR merupakan skema belajar paling primitive pada Weka yang memprediksi mayoritas pelatihan data untuk masalah dengan nilai kategoris dan nilai rata-rata kelas untuk masalah prediksi numerik [9].

3. Hasil dan Pembahasan

Data yang digunakan pada penelitian ini berasal dari data siswa baru tahun ajaran 2015/2016. Jumlah data yang digunakan sebanyak 92 data. Atribut yang digunakan dalam proses penjurusan siswa tingkat SMA, yaitu nilai rapor, nilai tes penempatan, nilai Ujian Akhir Nasional (UAN), nilai Ujian Akhir Sekolah (UAS) dan minat siswa. Pada atribut target terdapat dua kelas yang menjadi tujuan yaitu IPA atau IPS.

Data *pre-processing* merupakan salah satu langkah penting di KDD. Tujuan dari langkah ini adalah untuk membuat dataset yang bersih untuk dilanjutkan pada langkah data mining. Untuk mendapatkan data yang bersih atau berkualitas terdapat beberapa teknik yang digunakan yaitu dengan *data cleaning*, *data integration and transformation* dan *data reduction*.

Tabel 1. Tabel Atribut

Atribut	Variabel
Nilai Rapor	X1
Nilai Placement	X2
Nilai UAS	X3
Minat	X4
Nilai UN	X5
Jurusan	Y

Penelitian ini menggunakan empat algoritma klasifikasi untuk mendapatkan hasil akurasi klasifikasi tertinggi. Algoritma itu terdiri dari J48, Naïve Bayes, OneR dan ZeroR. Perhitungan akurasi menggunakan 10 *cross-validation*. *Cross-validation* merupakan metode algoritma statistik dengan membagi data menjadi dua segmen yaitu data latih dan data tes untuk memvalidasi model yang diusulkan.

Tabel 2. Hasil Perbandingan Algoritma

Metode	Correctly Classified Instances	Incorrectly Classified Instances
Naïve Bayes	96.74%	3.26%
J48	93.48%	6.52%
OneR	90.22%	9.78%
ZeroR	59.78%	40.22%

Precision adalah jumlah kelompok dokumen relevan dari total jumlah dokumen yang ditemukan oleh sistem. Presisi juga merupakan cara mengukur tingkat efektivitas sistem informasi. *Recall* merupakan istilah yang berhubungan dengan kemampuan suatu sistem temu balik dalam menemukan dokumen yang relevan. *Recall* berhubungan dengan kemampuan sistem untuk memanggil dokumen yang relevan, sedangkan *precision* berkaitan dengan kemampuan sistem untuk tidak memanggil dokumen yang tidak relevan. *Recall* sebenarnya sulit diukur karena jumlah seluruh dokumen yang relevan dalam database sangat besar. Oleh karena itu *precision* yang biasanya menjadi salah satu ukuran yang digunakan untuk menilai efektivitas suatu sistem temu balik informasi.

Tabel 3. Nilai *Precision*

Metode	IPA	IPS
Naïve Bayes	0.947	0.981
J48	0.878	0.98
OneR	0.868	0.926
ZeroR	0	0.598

Tabel 4. Nilai *Recall*

Metode	IPA	IPS
Naïve Bayes	0.973	0.96
J48	0.973	0.909
OneR	0.892	0.909
ZeroR	0	1

Berdasarkan tabel 2 dapat disimpulkan bahwa metode klasifikasi dengan algoritma naïve bayes merupakan algoritma yang mempunyai nilai akurasi tertinggi 96.74%. Sedangkan algoritma ZeroR mempunyai nilai akurasi terendah 59.78%.

4. Kesimpulan

Klasifikasi proses penjurusan siswa tingkat SMA yang selama ini digunakan oleh guru masih dilakukan secara manual. Guru harus menyeleksi satu persatu dalam menentukan proses penjurusan untuk setiap siswa. Hal tersebut menyebabkan proses penjurusan siswa memakan waktu yang cukup lama untuk mendapatkan hasil penjurusan. Untuk mengatasi permasalahan tersebut penelitian ini menggunakan data mining untuk membantu klasifikasi proses penjurusan siswa tingkat SMA. Sebelum mengimplementasikan ke sistem, perlu dilakukan penelitian untuk mengetahui algoritma klasifikasi yang mempunyai nilai akurasi tertinggi.

Penelitian ini mengusulkan empat algoritma untuk klasifikasi proses penjurusan siswa tingkat SMA. Algoritma yang digunakan yaitu J48, Naïve Bayes, OneR dan algoritma ZeroR.

Pada penelitian ini diperoleh hasil bahwa algoritma Naive Bayes merupakan algoritma yang mempunyai nilai akurasi tertinggi yaitu 96.74%, selanjutnya adalah algoritma J48 dengan 93.48%, algoritma OneR 90.22%, dan algoritma ZeroR merupakan algoritma yang mempunyai nilai akurasi terendah yaitu 59.78%.

Usulan penelitian selanjutnya, dapat dilakukan dengan memasukkan algoritma terbaik ke sistem sehingga dapat digunakan untuk membantu guru untuk melakukan proses penjurusan siswa tingkat SMA. Sistem tersebut nantinya juga dapat membantu siswa untuk memberikan informasi awal mengenai proses penjurusan.

Referensi

- Santosa B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta *Graha Ilmu*.
- Departemen Pendidikan Nasional. 2006. *Panduan Penyusunan Laporan Hasil Belajar Peserta Didik Sekolah Menengah Atas (SMA)*. Jakarta.
- Departemen Pendidikan Nasional. 2004. *Panduan Penilaian Penjurusan Kenaikan Kelas dan Pindah Sekolah*. Jakarta.
- Khasanah NF. 2015. Fuzzy MADM for Major Selection at High School. *Proceedings of 2nd Int Conference on Information Technology, Computer and Electrical Engineering*. 44 - 48.
- Kerdegari H, Samsudin K, Ramli AR, and Mokaram S. 2012. Evaluation of fall detection classification approaches. *ICIAS 2012 - 2012 4th Int. Conf. Intell. Adv. Syst. A Conf. World Eng. Sci. Technol. Congr. - Conf. Proc.*, vol. 1, pp. 131–136, 2012.
- Witten IH and Frank E. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Han J, Kamber M, and Pei J. 2011. *Data mining: concepts and techniques: concepts and techniques*. Elsevier.
- Kusrini. *Algoritma Data Mining*, 2009th ed. Yogyakarta: Penerbit Andi.
- Winanta SG, Yetli O. 2013. *Implementasi Metode Bayesian Dalam Penjurusan Di Sma Bruderan Purworejo Studi Kasus: Sma Bruderan Purworejo*.
- Kristanto O. 2013. *Penerapan Algoritma Klasifikasi Data Mining ID3 Untuk Menentukan Penjurusan Siswa SMAN 6 Semarang*.