

Analisa Prestasi Siswa Berdasarkan Kedisiplinan, Nilai Hasil Belajar, Sosial Ekonomi dan Aktivitas Organisasi Menggunakan Algoritma Naïve Bayes

Devi Yunita¹, Resti Amalia², dan Perani Rosyani³

^{1,2,3}Teknik Informatika, Fakultas Teknik, Universitas Pamulang, Tangerang Selatan, Indonesia
E-mail: ¹dosen00846@unpam.ac.id; ²dosen00850@unpam.ac.id; ³dosen00837@unpam.ac.id

Abstract

Student achievement is needed to measure the level of knowledge, intelligence, and skills of a person. Data mining is a combination of various fields of science, namely statistics, artificial intelligence, and databases. The purpose of education for students is to be able to change attitudes, perceptions, instill good behavior and instill knowledge and character. The purpose of this study is to be able to find out student achievement during learning as a school evaluation material to improve teaching patterns for students to be more accomplished. Research using the Naïve Bayes Algorithm produces a fairly good value of accuracy of 89%. the results of this study are expected to help to monitor early student academic achievement, in order to improve student achievement and quality.

Keywords: Data Mining, Student Achievement, Naïve Bayes

1. Pendahuluan

Pendidikan merupakan dasar utama untuk mewujudkan poses pembelajaran yang bertujuan agar peserta didik dapat aktif mengembangkan potensi dirinya dalam hal kecerdasan, kepribadian, pemikiran spiritual dalam hal keagamaan, memiliki sifat dan akhlak yang mulia serta kemampuan yang dapat berguna bagi masyarakat dan Negara. hal ini dituangkan dalam undang-undang Negara Republik Indonesia mengenai sistem pendidikan pada tahun 2003 No.20 pasal 1.

Tujuan pendidikan bagi siswa-siswi diantaranya menambah ilmu pengetahuan, konsep-konsep, teori, pendapat serta dapat mengubah sikap, pemahaman dan tingkah laku menjadi lebih baik.

Data mining adalah penggabungan dari berbagai bidang ilmu yaitu ilmu statistik, kecerdasan buatan, dan basisdata. Kata data mining itu sendiri diambil dari sebuah aktivitas penambangan sebuah gunung untuk mengambil hal-hal yang bernilai, dan data mining itu sendiri mengambil informasi yang tersembunyi atau belum diketahui dari sebuah database.

Sederhananya data mining merupakan pengambilan informasi atau data yang penting dari sebuah database. Secara ilmiah data mining ini disebut sebagai *Knowledge Discovery in Databases*.

Ada beberapa langkah dalam data mining yang dapat diterapkan untuk menemukan informasi pada sebuah data.

Klasifikasi merupakan upaya pencarian fungsi untuk melihat perbedaan suatu konsep data guna memeperkirakan kelas dari objek yang belum diketahui. Metode klasifikasi diantaranya adalah Decision Tree, K-Nearest Neighbor, Neural Network dan Naive Bayes.

Kemunculan data mining ini didasari dengan banyaknya permasalahan data dari organisasi atau perusahaan yang memiliki data yang telah menumpuk selama bertahun-tahun. Hubungannya dalam penelitian ini adalah karena adanya penimbunan data di dalam sekolah, data tersebut adalah data kedisiplinan siswa, data aktivitas organisasi dari masing-masing siswa, Ada juga data lain seperti data ekonomi orang tua, prestasi dari siswa tersebut dan lain-lain.

Data-data ini diperoleh oleh pihak sekolah saat proses penerimaan murid baru. Data-data tersebut adalah data hasil perkembangan siswa selama belajar di sekolah dan data ini biasanya di informasikan pihak sekolah saat pembagian raport setiap semester dari sekolah kepada wali murid. Betapa besarnya ukuran data tersebut jika data tersebut tidak diolah dengan baik. Dari hasil permasalahan tersebut apakah data yang didapat oleh pihak sekolah akan dibiarkan saja atau dapat diolah oleh pihak sekolah untuk mengambil informasi yang berguna untuk proses evaluasi prestasi siswa.

2. Metodologi Penelitian

Studi kasus penelitian ini di SMK Al-Hidayah Cinere yang beralamat di jalan Masjid

AI – Mujahidin RT 01/06 No. 3 Meruyung, Kec. Limo Kota Depok – Jawa Barat. Data yang digunakan dalam penelitian ini adalah data primer yang besumber dari nilai hasil belajar siswa, kedisiplinan siswa, aktivitas organisasi dalam tahun pelajaran 2014/2015 sampai 2016/2017 dan data ekonomi keluarga pribadi siswa. Target atau subjek dari penelitian ini adalah siswa SMK kelas XI SMK AI – Hidayah Cinere.

kemudian data tersebut dibagi menjadi 2 variable yaitu variable respond dan variable bebas. Variabel respon membutuhkan data kelulusan, yaitu siswa yang lulus / siswa yang berhasil menyelesaikan masa belajar di SMK AI – Hidayah Cinere. Sedangkan variabel bebas yang digunakan antara lain kedisiplinan, nilai hasil belajar, sosial ekonomi dan aktivitas organisasi.

Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah algoritma yang sederhana dan tiap-tiap atribut bersifat bebas, yang memungkinkan tiap atribut dapat berkontribusi terhadap hasil akhir. salah satu metode algoritma yang termasuk pada teknik klasifikasi adalah algoritma Naïve Bayes . proses klasifikasi yang dilakukan Naïve Bayes menggunakan metode probabilitas dan statistic. Adalah Thomas Bayes yang merupakan ilmuwan yang berasal dari Inggris yang berhasil menemukan *Algoritma Naïve Bayes*. Konsep dasar dari metode algoritma Naïve Bayes adalah dapat melakukan prediksi peluang di masa yang akan datang berdasarkan data pengalaman dari masa lampau, dan teori ini dikenal dengan Teorema Bayesian. Konsep ini digabungkan dengan Naïve yang menerapkan nilai atribut yang bebas.

Algoritma Naïve Bayes menerapkan bahwa ciri tertentu dari suatu kelas tidak berhubungan dengan kelas yang lain.

Persamaan dari *teorema Bayes* adalah:

$$P(H | X) = \frac{P(H | X) \cdot P(X)}{P(X)} \quad (1)$$

Keterangan:

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posterior probability)
- P(H) : Probabilitas hipotesis H (prior probability)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : Probabilitas X

Cross Validation

Adalah langkah yang dilakukan guna mencari *error rate*. Dalam *cross validation* data uji coba dibagi secara random kedalam beberapa bagian dengan nilai pebandingan yang sama, lalu *error rate* tersebut dihitung di tiap-tiap bagian, kemudian dihitung rata-ratanya untuk mendapat nilai *error rate* keseluruhan.

Confusion matrix

Confusion matrix dapat digunakan untuk mengukur nilai keakuratan dari hasil klasifikasi. *Confusion matrix* merupakan aplikasi yang berfungsi untuk menganalisa seberapa tepat *classifier* mampu mengetahui tupel di tiap class yang berbeda. jika terdapat dua class, maka dapat dilakukan pemisahan menjadii tupel positif dan tupel negative. *Classifier* dengan tepat akan memberikan label tepat yang ditujukan kepada tupel positif pada true positif, sedangkan true negative akan diberi label tepat pada tupel positif. False positif merupakan tupel negatif yang memiliki label belum tepat, dan false negatif merupakan tupel positif yang memiliki label belum tepat, pengertian ini dapat digunakan dalam menganalisa kemampuan dari classifier dan dapat dilihat pada table berikut.

Tabel 1. Model *Confusion matrix*

	C1	C2
C	<i>true positives</i>	<i>false negatives</i>
C	<i>false positives</i>	<i>true negatives</i>

Contoh jika terdapat *Confusion matrix* dengan nilai 2x2 seperti pada contoh diatas, maka rumusan perhitungan untuk mengetahui akurasi tersebut adalah:

$$\text{Akurasi} = \frac{\text{Truepositif} + \text{truenegatif}}{\text{Truepositif} + \text{truenegatif} + \text{falsepositif} + \text{falsenegatif}} \times 100\% \quad (2)$$

3. Hasil Dan Pembahasan

Analisa Penelitian

Data primer pada penelitian ini adalah nilai hasil belajar siswa, kedisiplinan siswa, aktivitas organisasi dalam tahun pelajaran 2014/2015 sampai 2016/2017 serta data ekonomi keluarga siswa, data yang didapat adalah dari 225 siswa.

Data Preprocessing

Tahap awal analisa penelitian adalah *preprocessing*. Dari data siswa, tiap variabel

memiliki beberapa sub variabel, maka sub variabel tersebut dijumlahkan untuk menjadi satu variabel. Selanjutnya dilakukan data preprocessing dengan beberapa teknik yaitu:

- Validasi, bertujuan untuk memeriksa dan menghilangkan nilai yang bersifat ganjil data yang kurang tepat, dan data nilai yang belum lengkap.
- Transformasi dan integrasi, bertujuan untuk mengoptimalkan efisiensi dan ketepatan algoritma. Khusus untuk algoritma neural network, transformasi nilai data menggunakan aplikasi rapid miner.
- Pengurangan ukuran data, fungsinya untuk mendapatkan dataset dengan jumlah yang lebih sedikit, namun tidak mengurangi nilai informasi dari data tersebut.

Setelah dilakukan *preprocessing* data yang didapat sebanyak 225 *record* kemudian direduksi dengan meminimalkan duplikasi menjadi 100 *record* untuk data training.

Teknik Analisis

Analisa Perhitungan Algoritma *Naive Bayes* Teknik perhitungan algoritma *Naive Bayes* menggunakan data uji. Dari data uji diketahui jumlah data ada 100, jumlah data lulus ada 73 dan jumlah data tidak lulus ada 27. Untuk perhitungannya adalah:

$$P(X = Lulus | C_i) = 73 / 100 = 0,73$$

$$P(X = Tidak Lulus | C_i) = 27 / 100 = 0,27$$

Nilai $P(X = Lulus)$ adalah jumlah data lulus dibagi total maka hasilnya adalah 0,73, dan untuk nilai $P(X = Tidak Lulus)$ adalah jumlah tidak lulus dibagi total dan hasilnya adalah 0,27. Selanjutnya adalah perhitungan nilai *probabilitas* prior, yaitu *probabilitas* nilai lulus dan tidak lulus masing-masing variabel terhadap total lulus dan tidak lulus dari keseluruhan data.

Pengujian Model

Model yang terbentuk diuji keakuratannya dengan memasukan data dari hasil dari data training.

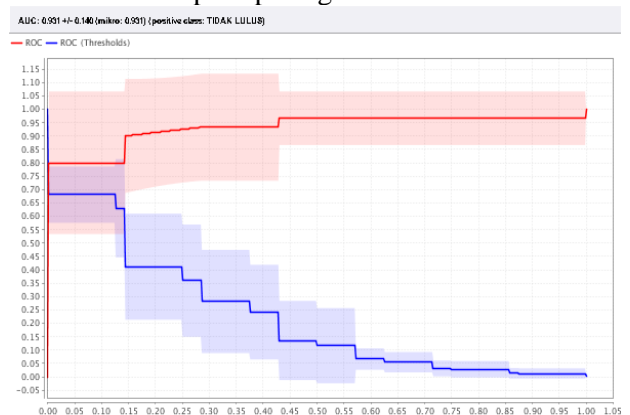
Sesuai dengan data uji, pada Tabel 2. Terdapat 100 data, 66 diklasifikasikan Lulus sesuai prediksi, lalu 4 data diperkirakan lulus tetapi hasilnya tidak lulus, 23 data tidak lulus sesuai perkiraan, dan 7 data diperkirakan tidak lulus namun dinyatakan lulus.

Tabel 2. Model Confusion Matrix untuk Metode *Naive Bayes*

Accurasi: 89.00 % +/- 13 (mikro 89%)			
	True LULUS	True TIDAK LULUS	Class precision
Prediksi LULUS	66	4	94,29%
Prediksi TIDAK LULUS	7	23	76,67%
Class recall	90,41%	85,19%	

Kurva ROC

Hasil perhitungan digambarkan dengan kurva ROC seperti pada grafik di bawah ini.



Gambar 1. Kurva ROC Algoritma *Naive Bayes*

Dari grafik kurva ROC tersebut menunjukkan hasil data perhitungan algoritma *Naive Bayes*. dalam gambar menunjukkan nilai false positive rate yang ditunjukkan dengan garis vertical, memiliki rentan nilai 0 – 1, dan nilai true positive rate yang ditunjukkan dengan nilai horizontal, memiliki rentan nilai 0 – 1,1. Berdasarkan hasil kinerja pada gambar diatas, menunjukkan kurva berwarna merah dan warna biru memiliki kinerja yang hampir sama, karena garis kurva berwarna merah dan warna biru sama-sama menjauh dari garis bujur (0,0). *Area Under Curve* (AUC) atau area dibawah kurva bertujuan untuk membandingkan nilai kurva merah dan biru dalam bentuk angka.

4. Kesimpulan

Kesimpulan yang didapat berdasarkan hasil penelitian dalam menganalisa data kelulusan siswa yaitu:

- Berdasarkan hasil perhitungan metode klasifikasi data mining *Naive Bayes* menunjukkan bahwa nilai akurasi metode

- Naïve Bayes* cukup baik yakni sebesar 89%.
2. Berdasarkan nilai AUC, perhitungan metode *Naïve Bayes* masuk dalam kategori *klasifikasi* yang sangat baik yakni sebesar 0,931.

Daftar Pustaka

- Kusrini, Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*, Penerbit Andi, Yogyakarta
- Eka Sabna, Muhandi, (2016), "*Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Berdasarkan Dosen, Motivasi, Kedisiplinan, Ekonomi, dan Hasil Belajar*", Pekanbaru: STMIK Hang Tuah Pekanbaru
- Sofi Defiyanti, (2014), "*perbandingan: prediksi prestasi belajar mahasiswa menggunakan teknik data mining*", Karawang: Fakultas Ilmu Komputer Universitas Singaperbangsa
- Linda Setiawati, (2015), "*Faktor-Faktor Yang Mempengaruhi Prestasi Belajar Praktik Kejuruan Siswa Smk Program Studi Keahlian Teknik Komputer Dan Informatika*", NTB: SMK Negeri 2 Sumbawa
- I Kadek Juni Arta, Gede Indrawan dan Gede Rasben Dantes, (2016), "*Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Stmik Denpasar Menggunakan Metode Technique For Others Reference By Similarity To Ideal Solution*", Bali: Universitas Pendidikan Ganesha
- Ni Komang Sri Julyantari dan I Ketut Dedy Suryawan, (2014), "*Data Mining Prestasi Akademik Dengan Naive Bayes Berdasarkan Atribut Importance (AI)*", Bali: STMIK STIKOM
- Imam Syaifuddin dan Reddy Alexandro Harianto, (2016), "*analisa perbandingan teknik-teknik data mining untuk prediksi prestasi akademik siswa*", Surabaya: Sekolah Tinggi Teknik Surabaya