

PENERAPAN DATA MINING METODE NAÏVE BAYES UNTUK PREDIKSI HASIL BELAJAR SISWA SEKOLAH MENENGAH PERTAMA (SMP)

Fathur Rahman¹⁾ dan Muhammad Iqbal Firdaus¹⁾

¹⁾Fakultas Teknologi Informasi Universitas Islam Kalimantan (UNISKA) Banjarmasin
Email: fathur@fti.uniska-bjm.ac.id dan m.iqbal1412@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengetahui tingkat akurasi serta hasil presisi dan Recall dari metode klasifikasi data mining yaitu Naïve Bayes untuk prediksi hasil belajar siswa menengah pertama (SMP). Penentuan hasil belajar dari siswa didik merupakan hal penting dalam dunia pendidikan, Hal ini dianggap serius karena sulitnya menentukan faktor atau variabel yang mempengaruhi hasil belajar siswa. Penerapan data mining yang dapat mengenali dan mengekstrak pola pengetahuan menawarkan solusi untuk masalah pendidikan ini. Ada beberapa model klasifikasi dalam data mining yaitu algoritma ID3 dan C4.5 juga Naïve Bayes yang dapat digunakan untuk memprediksi hasil belajar siswa khususnya siswa menengah pertama. Pada penelitian ini akan digunakan metode klasifikasi naïve bayes untuk memprediksi hasil belajar siswa SMP dengan harapan mendapat akurasi yang lebih baik dari metode sebelumnya yang pernah diuji.

Kata Kunci: *hasil belajar siswa, prediksi, Naïve Bayes*

PENDAHULUAN

Hal penting dalam evaluasi pembelajaran di sekolah adalah hasil dari belajar siswa yang merupakan tujuan semua sekolah dan institusi pendidikan. Untuk mengetahui faktor-faktor internal dan eksternal yang mempengaruhi hasil belajar siswa tersebut merupakan hal penting agar manfaat dari proses belajar dapat tercapai maksimal. Topik penelitian tentang hasil belajar siswa sekarang ini semakin banyak diteliti, Artificial Intelligence dan machine learning mempunyai metode untuk mengenali pola yang mempengaruhi hasil belajar siswa serta dapat membantu memecahkan masalah ini. Hasil belajar siswa bergantung pada banyak faktor seperti personal, sosial, psikologi dan variabel siswa lainnya. Aspek seperti jenis kelamin, tingkat pendidikan orang tua dan keluarga[1]. Selain itu lokasi tempat tinggal, metode pengajaran dan status keluarga juga sangat berpengaruh dengan hasil belajar siswa[2].

Klasifikasi barangkali adalah metode dalam Data Mining yang paling sering digunakan untuk menyelesaikan problem-problem di dunia nyata. Sebagai salah satu yang terpopuler dalam keluarga yang menggunakan teknik '*machine-learning*', *classification* mempelajari pola-pola dari data historis (sekumpulan informasi –seperti ciri-ciri,

variabel-variabel, fitur-fitur – pada berbagai karakteristik item-item yang sudah diberi label sebelumnya) dengan tujuan untuk menempatkan instans (object-object) baru (dengan label yang tak diketahui sebelumnya) ke dalam kelompok atau kelas masing-masing. Contohnya, kita bisa menggunakan *classification* untuk meramalkan apakah cuaca pada hari tertentu akan 'cerah' (sunny), 'hujan' (rainy) atau 'berawan' (cloudy). Beberapa jenis pekerjaan populer dalam menggunakan *classification* adalah 'persetujuan kredit' (misalnya risiko kredit 'baik' atau 'buruk'), lokasi toko (misalnya, 'baik', 'moderate', 'buruk'), target marketing (misalnya, 'kemungkinan menjadi langganan', 'tidak ada harapan sama sekali'), deteksi *fraud* (misalnya, 'ya', 'tidak'), dan telekomunikasi (misalnya, kemungkinan pindah ke perusahaan telpon yang lain, 'yes/no'). Bila yang sedang diprediksikan adalah label kelas (misalnya, 'cerah', 'hujan', atau 'berawan') problem prediksi tersebut disebut dengan '*classification*', sedangkan bila yang diprediksi adalah nilainya yang numerik (misalnya temperature seperti 68 F), problem prediksi tersebut disebut dengan '*regression*' (regresi).

Menurut beberapa penelitian dengan memakai algoritma decision tree, random forests, neural

network dan support vector machines. data untuk memprediksikan berhasil dan gagal nya siswa dalam belajar didapatkan nilai akurasi 59,4% dengan algoritma CHAID dan 60,5% dengan algoritma CART, yang mana kedua algoritma ini merupakan jenis metode pohon keputusan (decision tree). Yavev et al. (2011) menggunakan tingkat kehadiran siswa, hasil ujian, seminar, tugas dan praktek untuk memprediksi hasil belajar siswa diakhir semester dengan menggunakan algoritma ID3, CART dan C4.5 yang menghasilkan akurasi 52,08%, 56,25% dan 45,83%.

Pendekatan ini menggunakan teori probabilitas untuk membuat model-model klasifikasi berdasarkan kejadian-kejadian di masa lalu yang bisa untuk menempatkan suatu instans baru ke dalam kelas (atau kategori) yang paling mungkin.

Dengan demikian sebenarnya banyak algoritma yang dapat digunakan untuk klasifikasi dan prediksi hasil belajar siswa dan pada penelitian ini akan digunakan metode klasifikasi naïve bayes dengan harapan mendapatkan akurasi yang lebih tinggi dari metode sebelumnya untuk mengklasifikasikan hasil belajar siswa.

METODE PENELITIAN

Metode yang digunakan adalah eksperimen dengan menggunakan software data mining Rapidminer dan tahapan sebagai berikut:

1. Pengumpulan Data

Data didapatkan dari repository dataset terbuka (<http://archive.ics.uci.edu>) yang merupakan data sekunder yang didapatkan dari dua sekolah. Data berasal dari umpan balik siswa yang dinilai dari beberapa faktor internal dan eksternal seperti faktor demografi (pendidikan orang tua dan kemampuan ekonomi keluarga) selain itu juga data dari sekolah (tingkat kehadiran) yang mana faktor-faktor itu berpengaruh pada hasil belajar siswa.

2. Pengolahan Data Awal

Pada penelitian ini data yang digunakan mempunyai 263 record dan terdiri dari 12 atribut. Faktor seperti pendapatan keluarga dan faktor konsumsi alkohol dihilangkan karena bersifat pribadi dan menjadikan data tidak konsisten yang mana hampir 100% siswa masih tinggal bersama keluarga serta memiliki komputer pribadi di rumah.

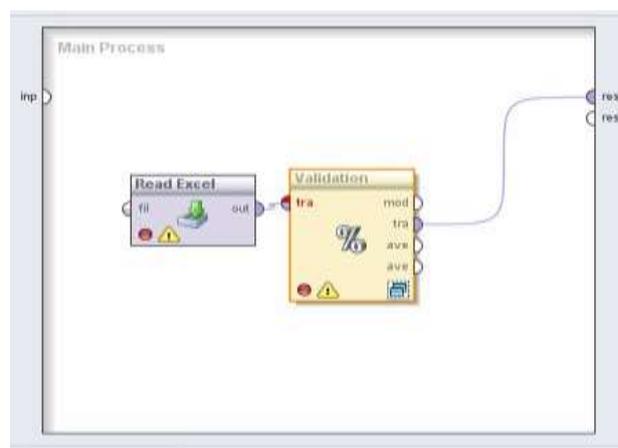
Beberapa nilai diklasifikasikan kembali agar hasil dari setiap algoritma menjadi lebih optimal. Atribut yang tersisa bisa dilihat pada Tabel 1.

Tabel 1. Atribut yang Dinilai

Atribut	Nilai
Jenis Kelamin	Laki-Laki, Perempuan
Anggota Keluarga	>3, <=3
Status Orang Tua	Tinggalbersama, Pisah
Pendidikan Ibu	Tidak, SD, SMP, SMA, >=S1
Pendidikan Ayah	Tidak, SD, SMP, SMA, >=S1
Kerja Ibu	Tidak, PNS, Guru, Medis, Lainnya
Kerja Ayah	Tidak, PNS, Guru, Medis, Lainnya
Pengasuh	Ibu, Ayah, Lainnya
Jarak Rumah	Sangat Dekat, Dekat, Sedang, Jauh
Waktu Belajar	Jarang, Biasa, Rajin, Sangat Rajin
Dukungan Keluarga	Ya, Tidak
Ikut Ekstrakurikuler	Ya, Tidak
Internet di Rumah	Ya, Tidak
Hasil	Rendah, Biasa, Baik

3. Metode Yang Diusulkan

Penelitian ini dilakukan dengan menggunakan software data mining Rapidminer. Percobaan akan menggunakan metode klasifikasi yaitu algoritma Naïve Bayes.



Gambar 1. Pengaturan Proses Utama Rapidminer

Penilaian akan dilakukan dengan melihat akurasi yang dihasilkan, semakin tinggi persentase dari hasil klasifikasi berarti semakin tinggi pula akurasi metode tersebut. Simulasi data akan

menggunakan 10-fold validation, yaitu sebanyak 85% data akan digunakan sebagai data training dan 15% lainnya akan digunakan sebagai data uji.

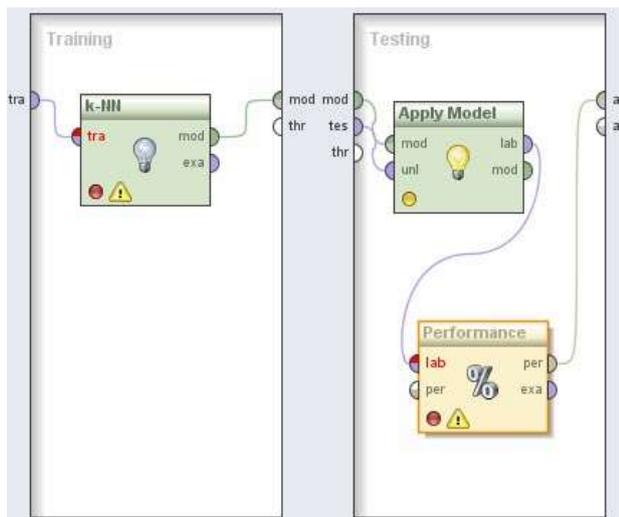
HASIL DAN PEMBAHASAN

Evaluasi penelitian ini akan menggunakan confusion matrix untuk melihat akurasi. Dan untuk mengukur kinerja juga dibandingkan dari precision dan recall masing-masing algoritma. Hasil bisa dilihat pada tabel 2.

Tabel 2 Hasil Pengukuran Akhir

	Naïve Bayes
Akurasi	56,79%
Presisi	62,80%
Recall	72,56%

Dari tabel 2 dapat dilihat algoritma Naïve Bayes mendapatkan nilai akurasi, presisi dan recall sebesar 56,79%; 62,80% dan 72,56%.



KESIMPULAN

Berdasarkan hasil pengujian klasifikasi hasil belajar siswa pada penelitian ini disimpulkan bahwa algoritma Naïve Bayes memiliki kinerja yang baik dari segi akurasi, presisi dan recall walaupun dari jumlah akurasi yang hanya 56,79% masih tergolong hasil yang rendah dibandingkan beberapa model data mining yang lainnya. Hal ini karena banyaknya input tidak terlalu relevan dan algoritma yang diujikan tidak terlalu sensitif dengan input yang tidak terlalu relevan tersebut.

DAFTAR PUSTAKA

Destuardi & Sumpeno, S. 2009. Klasifikasi Emosi untuk Teks Bahasa Indonesia menggunakan Metode *Naive Bayes*. Seminar Nasional Pascasarjana Institut Teknologi Sepuluh Nopember.

Indranandita, A., Susanto, B. & Rachmat, A. 2008. Sistem Klasifikasi dan Pencarian Jurnal dengan Menggunakan Metode Naive Bayes dan Vector Space Model. *Jurnal Informatika*, Volume 4 Nomor 2, pp. 9 – 18.

Nurani, A., Susanto, B. & Proboyekti, U. 2007. Implementasi *naive bayes classifier* pada Program Bantu Penentuan Buku Referensi Matakuliah. *Jurnal Informatika*, Volume 3 Nomor 2, pp. 32 – 36.

Domingos, P., and Pazzani, M. (1997). *On the optimality of the Simple Bayesian Classifier under Zero-One Loss*.

Kusrini, 2006, *Algoritma Data Mining*, Penerbit ANDI, Yogyakarta.

Han, J., and Kamber M. (2000). *Data Mining, Concept and Techniques*. New York, NY : Morgan Kaufmann.

Duda, R.O., and Hart, P.E. (1973). *Pattern classification and scene analysis*. New York, NY: Wiley.