

Sistem Pengelompokan Siswa Berdasarkan Tingkat Kedisiplinan Menggunakan Metode *Naïve Bayes Classifier*

Budi Hartanto¹⁾; Sri Tomo²⁾

^{1,2)}Program Studi S1 Informatika, STMIK Sinar Nusantara

¹⁾budihartanto@sinus.ac.id; ²⁾szrie@sinus.ac.id

ABSTRACT

Discipline is a very important thing in the educational process. Discipline will succeed if it is applied to students correctly. Student discipline is that every student follows every rule and order that has been set by the school. At SMK Muhammadiyah 2 Sukoharjo student discipline. Declining discipline at SMK Muhammadiyah 2 Sukoharjo is marked by the increase in points of violation from students. The purpose of this study was to apply the naive Bayes method in the classification of student discipline levels at SMK Muhammadiyah 2 Sukoharjo. With this information will be obtained that can be used for information on which students need to be given Counseling Guidance to provide direction and guidance to students. The attributes used are cases of fights, not attending apples, not carrying out picket, not entering without explanation, arriving late, noisy in class. Test results with 490 records with a portion of 75% training data and 25% test data. And produces an accuracy of 76%.

Keywords : *Classification, Discipline, Naïve Bayes Classifier.*

I. PENDAHULUAN

Menurunnya tingkat kedisiplinan siswa di sekolah membawa dampak yang sangat penting dalam proses belajar mengajar dan prestasi belajarsiswa di sekolah. Semakin tinggi tingkat kedisiplinan siswa, maka semakin tinggi prestasi belajar. Sebaliknya semakin rendah tingkat kedisiplinan siswa, maka semakin rendah prestasi belajar.[1]

Menurunnya kedisiplinan ditandai dengan meningkatnya poin pelanggaran yang dilakukan oleh setiap siswa. Di SMK Muhammadiyah 2 Sukoharjo menurunnya tingkat kedisiplinan ditunjukkan dengan banyaknya poin ketidaksiplinan yang didapat oleh siswa. Data rekap pelanggaran untuk semester gasal tahun 2019/2020 terdapat 320 kasus pelanggaran. Pihak Sekolah dalam hal ini berusaha untuk mengupayakan akan membuat suatu sistem yang bersifat antisipasi untuk mengatasi menurunnya tingkat kedisiplinan Siswa. Secara teknik hal yang akan dilakukan yaitu memberikan bimbingan konseling secara intensif kepada Siswa yang diperkirakan tidak disiplin.

Masalah yang muncul terkait dengan kedisiplinan apabila Pihak sekolah mendapati bertambahnya kasus pelanggaran yang dilakukan oleh Siswa. Hal tersebut menjadikan mutu atau kualitas dari Sekolah tersebut menjadi turun, dampak negatif yang akan muncul yaitu menurunnya minat masyarakat untuk memilih SMK Muhammadiyah 2

Sukoharjo. Sehingga kuantitas Siswa di Sekolah tersebut akan mengalami penurunan.

Data mining adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar pengambilan keputusan. Manfaat data mining salah satunya untuk melakukan klasifikasi data. Dengan cara mengelompokan data sesuai dengan kelas yang telah ditentukan.[2]

Berdasarkan uraian di atas diperlukan sebuah sistem dengan penerapan algoritma data mining untuk melakukan perkiraan siswa yang dimungkinkan akan berpotensi melakukan pelanggaran di Sekolah. Sistem ini akan diterapkan untuk kelas 10 yang akan naik ke kelas 11 sebagai antisipasi dari sekolah. Sistem tersebut diperlukan karena pihak sekolah dapat melihat informasi mengenai perkiraan siswa yang dimungkinkan akan melakukan pelanggaran atau tidak disiplin. Sehingga dengan adanya hasil informasi tersebut pihak sekolah dapat melakukan tindakan bimbingan penyuluhan yang tepat untuk mengantisipasi siswa yang diperkirakan tidak disiplin.

II. TINJAUAN PUSTAKA

2.1 Sistem Pendukung Keputusan

Sistem Pendukung Keputusan atau dalam Bahasa *inggris Decision Suport System* merupakan suatu system yang dirancang untuk mendukung seluruh tahapan dalam pengambilan keputusan mulai dari identifikasi

masalah, pemilihan data yang terkait dengan kasus, menentukan pendekatan sampai dengan melakukan evaluasi terhadap hasil pengambilan keputusan [1].

2.2 Naïve Bayes Classifier

Algoritma *Naïve Bayes Classifier* (NBC) merupakan algoritma yang memanfaatkan teori probabilitas, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Klasifikasi-klasifikasi *Bayes* adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. Untuk klasifikasi *Bayes* sederhana yang lebih dikenal sebagai *naïve Bayesian Classifier* dapat diasumsikan bahwa efek dari suatu nilai atribut sebuah kelas yang diberikan adalah bebas dari atribut-atribut lain. Asumsi ini disebut *class conditional independence* yang dibuat untuk memudahkan perhitungan-perhitungan pengertian ini dianggap "naive", dalam bahasa lebih sederhana *naïve* itu mengasumsikan bahwa kemunculan suatu term kata dalam suatu kalimat tidak dipengaruhi kemungkinan kata-kata yang lain dalam kalimat padahal dalam kenyataannya bahwa kemungkinan kata dalam kalimat sangat dipengaruhi kemungkinan keberadaan kata-kata yang dalam kalimat [4].

NBC merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi[5].

Teorema *Bayes* memiliki bentuk umum sebagai berikut :

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

Dalam hal ini:

- X : data dengan class yang belum diketahui
- H : hipotesis data X merupakan suatu *class* spesifik
- P(H|X) : probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
- P(H) : probabilitas hipotesis H (*prior probability*)
- P(X|H) : probabilitas X berdasarkan kondisi pada hipotesis H
- P(X) : probabilitas dari X

Hasil dari rumus diatas untuk mencari probabilitas yang paling tinggi dengan membandingkan seluruh data yang ada yaitu P(H|X).

Teorema *Bayesian* dengan asumsi bahwa setiap variabel bersifat bebas (*independence*). Sehingga dengan asumsi seperti itu menyimpulkan bahwa setiap atribut tidak saling terkait (*conditionally independence*) maka

$$P(X|Ci) = \prod_{k=1}^n P(X_k|Ci) \quad (2)$$

Setelah diperoleh hasil dari seluruh data pada setiap class, maka hasil akhirnya dapat menggunakan rumus:

$$P(X|Ci) = \arg \max P(X_i|Ci) * P(Ci) \quad (3)$$

Teori-teori Metode *Naïve Bayes* dan Metode *Konfusi Matriks*.

2.3 Konfusi Matrik

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 3 keluaran, yaitu: recall, precision dan accuracy

1. *Recall* adalah proporsi kasus positif yang diidentifikasi dengan benar.
2. *Precision* adalah proporsi kasus dengan hasil positif yang benar
3. *Accuracy* adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus.

III. METODE PENELITIAN

Untuk memperoleh data yang tepat dan akurat guna kesempurnaan sistem yang akan dibuat sehingga membentuk alur yang sistematis diperlukan beberapa metode penelitian.

3.1 Metode Pengumpulan Data

3.1.1 Data Primer

Dikumpulkan dengan dua cara yaitu:

a) Metode Observasi.

Dalam teknik ini penulis melakukan pengumpulan data dengan mengadakan pengamatan secara langsung pada SMK Muhammadiyah 2 Sukoharjo.

b) Metode Dokumentasi

Dalam teknik ini penulis melihat laporan yang telah ada agar data yang diperoleh lebih terperinci dan jelas serta sesuai kebutuhan.

c) Metode Wawancara

Dalam teknik ini penulis mengadakan tanya jawab dengan pihak-pihak yang terkait dengan kedisiplinan siswa, dalam hal ini adalah Guru Bimbingan Konseling dan Wakil Kepala Sekolah Bidang Kesiswaan.

3.1.2 Data Sekunder

Diperoleh dengan cara mengumpulkan data dengan membaca buku (studi pustaka) yang berkaitan permasalahan yang dijadikan objek penelitian

3.2 Metode Pengembangan Sistem

Metodologi untuk mengembangkan sistem ini adalah Daur Hidup Pengembangan Sistem (*System Development Life Cycle*) yang digunakan untuk menggambarkan tahapan utama dan langkah pengembangan sistem. Tahapan tersebut meliputi:

1. Tahap Analisa Sistem
Untuk mengembangkan sistem ini dibutuhkan perangkat lunak (*software*) dan perangkat keras (*hardware*). Perangkat lunak yang digunakan terdiri dari teks editor, scripting language dan database. Sedangkan perangkat keras berupa komputer untuk menjalankan sistem.
2. Tahap Desain Sistem
Pemodelan pembangunan sistem menggunakan UML (*Unified Modeling Language*). Perancangan interface sistem menggunakan Bahasa Web Programming PHP yang dibuat dengan upaya pemakai mudah mengerti (*user friendly*).
3. Tahap Implementasi Sistem
Implementasi sistem dilakukan pada tahap ini, ditampilkan dengan komputer dengan kriteria program dapat digunakan dengan mudah dan dipahami oleh user. Pada tahap ini perlu ditegaskan mengenai pemakaian program pada calon user.
4. Tahap Pengujian
Pengujian dilakukan dengan hanya menjalankan unit atau modul kemudian diamati apakah hasil dari unit tersebut sesuai dengan yang diharapkan

IV. HASIL DAN PEMBAHASAN

Dalam laporan penelitian ini penulis hanya melampirkan contoh pembersihan data pada periode 2014 sampai 2019, dengan rincian sebagai berikut:

- Data keseluruhan periode tahun 4 tahun = 510 records.
- Data yang tidak lengkap = 20 records

Setelah dilakukan seleksi dan pembersihan data, total seluruh data yang digunakan awalnya berjumlah 510 records menjadi 490 records. Rincian jumlah masing – masing kelas tertera pada Tabel 1.

Tabel 1. Rincian jumlah data untuk setiap kelas output

No	Kelas Output	Jumlah
1.	Disiplin	360
2.	Tidak Disiplin	130

4.1 Seleksi Data Atribut

Data latih merupakan suatu data yang akan digunakan untuk melakukan penelusuran pola dalam data mining. Dimana proporsi data latih dalam suatu konsep data mining adalah 75% dari data secara keseluruhan. Sehingga proporsi data Uji sebesar 25% dari sisa data yang telah digunakan sebagai data uji. Berdasarkan data yang telah dinyatakan lengkap untuk melakukan proses prediksi, maka komposisi data latih dan data uji adalah seperti pada Tabel 2.

Tabel 2. Komposisi Data Latih dan Data Uji

Jenis Data	Prosentase	Jumlah
Data Latih	75%	360
Data Uji	25%	130

4.2 Prediksi Dengan Naïve Bayes

Pada tahap ini akan mengelola data latih dan data uji kedalam metode Naïve Bayes. Langkah awal yang akan dilakukan pada tahap ini adalah mengumpulkan data training dan data testing. Dalam laporan penelitian ini Penulis hanya menampilkan data testing yang akan digunakan untuk melakukan pengujian.

Setelah Penelusi menentukan data latih dan data uji. Maka selanjutnya Penulis akan menentukan nilai tepat dan nilai tidak menggunakan metode Naïve Bayes. Dengan mengacu pada langkah perhitungan tersebut. Akan menghasilkan nilai disiplin dan nilai tidak disiplin. Dengan menggunakan langkah parhitungan tersebut, maka akan diperoleh hasil data seperti terlihat pada tabel lampiran 3 hasil prediksi.

Dari hasil data prediksi diketahui bahwa data prediksi yang sesuai dengan data asli sebanyak 117. Sehingga data prediksi yang tidak sesuai dengan data asli sebanyak 13. Dengan demikian proses analisa menggunakan metode naïve bayes telah selesai. Langkah selanjutnya yaitu menguji kinerja sistem. Tujuan dari menguji kinerja sistem yaitu untuk mengetahui kinerja sistem prediksi yang telah dibangun oleh peneliti.

4.3 Pengujian Dengan Konfusi Matriks

Pada tahap ini Peneliti akan melakukan pengujian terhadap sistem yang telah dibangun. Pada tahap ini Peneliti akan menguji sistem menggunakan metode pengujian *Confusion Matrix*. Adapun rincian pengujian ditunjukkan pada Tabel 3.

Tabel 3. Rumus Pengujian Confusion Matrix

Confusion Matrix		Data Sebenarnya	
		TRUE	FALSE
Data Prediksi	TRUE	TP (True Positive Correct Result)	FP (False Positive Unexpected Result)
	FALSE	FN (False Negative Missing Result)	TN (True Negative Correct Absence Of Result)

Nilai *Precision* = $TP / (TP+FP)$

Nilai *Recall* = $TP / (TP+FN)$

Nilai *Accuracy* = $TP + TN / (TP+TN+FP+FN)$

Mendefinisikan sebagai berikut

Precision = Merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem

Recall = Merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Accuracy = Sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual

Tabel 4. Data Hasil Prediksi

Confusion Matrix		Data Sebenarnya	
		TRUE	FALSE
Data Prediksi	TRUE	117	13
	FALSE	130	360

Sehingga didapatkan hasil :

Nilai *Precision* = $117 / (117+13) = 0,99$

Nilai *Recall* = $117 / (117+130) = 0,47$

Nilai *Accuracy* = $(117+360) / (117+360+130+13) = 0,766$

Apabila dikonversi kedalam bentuk persentase adalah sebagai berikut:

Nilai *Precision* = $0,88 \times 100 = 88\%$

Nilai *Recall* = $0,46 \times 100 = 46\%$

Nilai *Accuracy* = $0,766 \times 100 = 76\%$

4.4 Contoh Studi Kasus

Pada tahap ini akan diberikan contoh perhitungan data baru untuk diklasifikasi kedalam metode Naïve Bayes. Setelah Penulis menentukan data latih dan data uji. Maka selanjutnya Penulis akan menentukan nilai disiplin dan nilai tidak disiplin menggunakan metode Naïve Bayes.

Bila terdapat data baru yang belum memiliki kelas atau keterangan adalah:

- a Kasus Perkelahian = Tidak
(Berkelahi dengan teman) Pernah
- b Tidak Ikut Apel = Sering
(tidak mengikuti upacara)
- c Tidak Melaksanakan = Tidak
Piket (Tidak melakukan piket sesuai jadwal yang ditentukan) Pernah
- d Tidak Masuk Tanpa = Tidak
Keterangan (Membolos atau tidak masuk tanpa surat ijin) Pernah
- e Datang Terlambat = Tidak
(Datang ke sekolah terlambat) Pernah
- f Gaduh Di Kelas = Tidak
(Membuat gaduh sehingga kelas menjadi ramai) Pernah

Maka untuk menentukan data baru tersebut masuk kelas yang mana akan dilakukan perhitungan sebagai berikut :

a. Menghitung P(Ci)

$$P(\text{Keterangan}=\text{'Disiplin'}) = \frac{315}{360} = 0.875$$

$$P(\text{Keterangan}=\text{'Tidak Disiplin'}) = \frac{45}{360} = 0.125$$

b. Menghitung P(X|Ci) untuk setiap kelas i

$$P(\text{Kasus Perkelahian} = \text{"Tidak Pernah"} | \text{keterangan}=\text{Disiplin}) = \frac{290}{300} = 0.966$$

$$P(\text{Kasus Perkelahian} = \text{"Tidak Pernah"} | \text{keterangan}=\text{"Tidak Disiplin"}) = \frac{10}{300} = 0.03$$

$$P(\text{Tidak Ikut Apel} = \text{"Sering"} | \text{keterangan}=\text{"Disiplin"}) = \frac{75}{90} = 0.833$$

$$P(\text{Tidak Ikut Apel} = \text{"Sering"} | \text{keterangan} = \text{"Tidak Disiplin"}) = \frac{15}{90} = 0.166$$

$$P(\text{Tidak Melaksanakan Piket} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Disiplin"}) = \frac{300}{315} = 0.952$$

$$P(\text{Tidak Melaksanakan Piket} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Tidak Disiplin"}) = \frac{15}{315} = 0.047$$

$$P(\text{Tidak Masuk Tanpa Keterangan} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Disiplin"}) = \frac{315}{325} = 0.969$$

$$P(\text{Tidak Masuk Tanpa Keterangan} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Tidak Disiplin"}) = 0.030 \frac{10}{325}$$

$$P(\text{Datang Terlambat} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Disiplin"}) = \frac{270}{300} = 0.9$$

$$P(\text{Datang Terlambat} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Tidak Disiplin"}) = \frac{30}{300} = 0.1$$

$$P(\text{Gaduh di Kelas} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Disiplin"}) = \frac{310}{320} = 0.967$$

$$P(\text{Gaduh di Kelas} = \text{"Tidak Pernah"} \mid \text{keterangan} = \text{"Tidak Disiplin"}) = \frac{10}{320} = 0.031$$

b. Menghitung $P(X|Ci)$ untuk setiap kelas

$$P(X| \text{level} = \text{'Disiplin'}) = 0.966 \times 0.833 \times 0.952 \times 0.969 \times 0.9 \times 0.967 = 0.646028$$

$$P(X| \text{level} = \text{'Tidak Disiplin'}) = 0.03 \times 0.166 \times 0.047 \times 0.030 \times 0.1 \times 0.031 = 0.000000021$$

c. Menghitung $P(X|Ci) \cdot P(Ci)$:

$$P(X| \text{keterangan} = \text{"Disiplin"}) \cdot P(\text{level} = \text{"Disiplin"}) = 0.875 \times 0.646028 = 0.565275$$

$$P(X| \text{keterangan} = \text{"Tidak Disiplin"}) \cdot P(\text{level} = \text{"Tidak Disiplin"}) = 0.125 \times 0.000000021 = 0.000000027$$

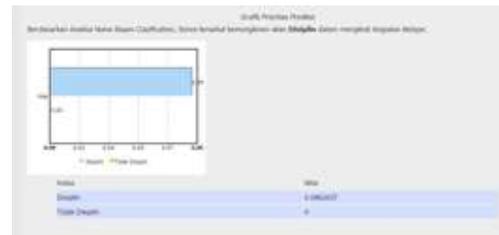
Hasil dari perhitungan dengan menggunakan algoritma NBC diatas didapatkan untuk X masuk dalam class keterangan = $\frac{15}{90}$ 'Disiplin'

4.5 Prediksi Baru

Prediksi baru digunakan untuk memprediksi data prediksi baru terlihat seperti Gambar 1.

Gambar 1. Input Data Prediksi

Hasil yang didapatkan terlihat seperti Gambar 2.



Gambar 2. Hasil Prediksi

V. PENUTUP

1. Terciptanya aplikasi yang dapat mengklasifikasikan kedisiplinan siswa yang mempermudah pihak bimbingan konseling memberi bimbingan kepada siswa yang diprediksi tidak disiplin. Keluaran berupa daftar prediksi siswa yang disiplin dan tidak disiplin.
2. Aplikasi yang dibangun sudah layak digunakan karena nilai *Accuracy* sebesar 76.94%.

DAFTAR PUSTAKA

[1] Anita Dewi S, Muhammad M, and Sri Hartati, "Sistem Pendukung Keputusan Perankingan Calon Siswa Baru Jalur Undangan Menggunakan Simple Additive Weighting (Studi Kasus: SMK Bumi Nusantara Wonosobo)," Seminar Nasional Teknologi Informasi dan Multimedia 2017, STMIK AMIKOM Yogyakarta, 4-7 Februari 2017 ISSN : 2302-3805, 2017.

[2] Hanifa, Muhamad Muslihudin, Sri Hartati. Sistem Pendukung Keputusan Menentukan Besar Gaji Untuk Guru Honorar Di Kabupaten Pesawaran Menggunakan Metode Fuzzy SAW, Jurnal Teknologi, Vol. 09 No.01 Juni 2016, hal. 83-88. IST Akprind. Yogyakarta, 2016.

- [3] Lena Magdalena, Abdul Rachman, Aplikasi Pendaftaran Siswa Baru Dengan Sistem Seleksi Menggunakan Metode *Simple Additive Weighting* (SAW) Pada Smk Miftahul Huda Ciwaringin. JURNAL DIGIT, Vol. 7 No.1 Mei 2017, pp. 38~49 ISSN: 2088-589, 2017
- [4] Muhamad Muslihudin & A. Wulan Arumita, Pembuatan Model Penilaian Proses Belajar Mengajar Perguruan Tinggi Menggunakan Fuzzy *Simple Additive Weighting* (SAW)(Studi: STMIK Pringsewu), 2016.
- [5] Wulandari, Ahmad Mustofa, Ponidi, Muhamad Muslihudin, Firza Adi Firdiansah, *Decision Support System* Pemetaan Lahan Pertanian Yang Berkualitas Untuk Meningkatkan Hasil Produksi Padi Menggunakan Metode *Simple Additive Weighting* (SAW), SEMNASTEKNOMEDIA, AMIKOM Yogyakarta, 2016.