

# ANALISIS PREDIKSI TINGKAT KETIDAKDIPLINAN SISWA MENGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER (STUDI KASUS : SMK NEGERI 1 PACITAN)

Niken Puji Astuti<sup>1)</sup>, Kusri<sup>2)</sup>, M. Rudyanto Arief<sup>3)</sup>

<sup>1),2),3)</sup> Magister Teknik Informatika STMIK AMIKOM Yogyakarta  
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281  
Email : [niken.pacitan@gmail.com](mailto:niken.pacitan@gmail.com)<sup>1)</sup>, [kusri@amikom.ac.id](mailto:kusri@amikom.ac.id)<sup>2)</sup>, [rudy@amikom.ac.id](mailto:rudy@amikom.ac.id)<sup>3)</sup>

## Abstrak

*Kedisiplinan sangat penting dalam proses pendidikan. Sebuah proses pendidikan akan berhasil jika ada penerapan kedisiplinan kepada siswa-siswanya. Kedisiplinan siswa merupakan kepatuhan untuk mengikuti setiap peraturan dan tata tertib sekolah. Menurunnya tingkat ketidaksiplinan siswa di SMK Negeri 1 Pacitan ditandai dengan meningkatnya poin pelanggaran setiap siswa.*

*Penelitian ini bertujuan untuk memprediksi siswa yang berpotensi melakukan ketidaksiplinan dengan menggunakan algoritma Naive Bayes Classifier(NBC). Informasi yang diperoleh nantinya dapat digunakan guru Bimbingan Konseling (BK) untuk memberikan pembinaan lebih awal terhadap siswa – siswa yang berpotensi untuk melakukan ketidaksiplinan.*

*Hasil pengujian pada data siswa berjumlah 895 record dengan kondisi reduksi atribut sebesar 75% diurutkan dari nilai information gain(IG) yang terendah dengan perbandingan data training sebesar 75% dan data testing sebesar 25% menghasilkan akurasi sebesar 79,01%.*

**Kata kunci:** Disiplin, poin pelanggaran, prediksi, naïve bayes classifier, information gain.

## 1. Pendahuluan

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

Kedisiplinan merupakan suatu ketaatan yang sungguh-sungguh dan didukung oleh kesadaran untuk menunaikan tugas dan kewajiban serta sikap dan perilaku sesuai dengan aturan atau tata kelakuan yang semestinya di dalam suatu lingkungan tertentu[2].

Penurunan tingkat kedisiplinan ditandai dengan meningkatnya poin pelanggaran setiap siswa. Di SMK Negeri 1 Pacitan menurunnya tingkat kedisiplinan ditunjukkan dengan banyaknya poin pelanggaran yang

didapat oleh siswa. Data rekap pelanggaran diperoleh dari <http://vsm.smkn1pacitan.sch.id> untuk semester genap tahun 2013/2014 terdapat 3483 kasus pelanggaran dimana bulan Januari tercatat 706 pelanggaran, bulan Februari 1079, bulan Maret 572, bulan April 711 dan bulan Mei 415 kasus pelanggaran. Komposisi jenis pelanggaran secara keseluruhan tercatat alpa sebanyak 49,2%, terlambat sekolah 1-10 menit 27,5%, seragam/atribut tidak lengkap 22%, membolos 0,8%, tidak sopan terhadap guru atau karyawan 0,1%, berambut gondrong(putra) 0,1%, tidak membawa buku pelajaran 0,1%, melompat/memanjat pagar 0,1%, tindik atau tato 0,05% dan berkeliaran pada jam diklat 0,05%.

Kedisiplinan siswa sangat berkaitan dengan karakteristik siswa. Karakteristik adalah mengacu kepada karakter dan gaya hidup seseorang serta nilai-nilai yang berkembang secara teratur sehingga tingkah laku menjadi lebih konsisten dan mudah di perhatikan[3]. Karakteristik siswa terbentuk dari faktor internal dan eksternal. Faktor internal merupakan faktor – faktor yang datangnya dari diri sendiri berupa kecerdasan, minat, kesehatan dan lain – lain. Faktor eksternal merupakan faktor dari luar individu yaitu kondisi ekonomi keluarga, orang tua, suasana rumah[4].

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[5]. Penerapan data mining pada data siswa SMK Negeri 1 Pacitan bertujuan untuk memprediksi siswa yang berpotensi untuk melakukan ketidaksiplinan. Informasi yang diperoleh diharapkan dapat digunakan guru Bimbingan Konseling (BK) untuk memberikan pembinaan lebih awal terhadap siswa – siswa yang berpotensi untuk melakukan ketidaksiplinan.

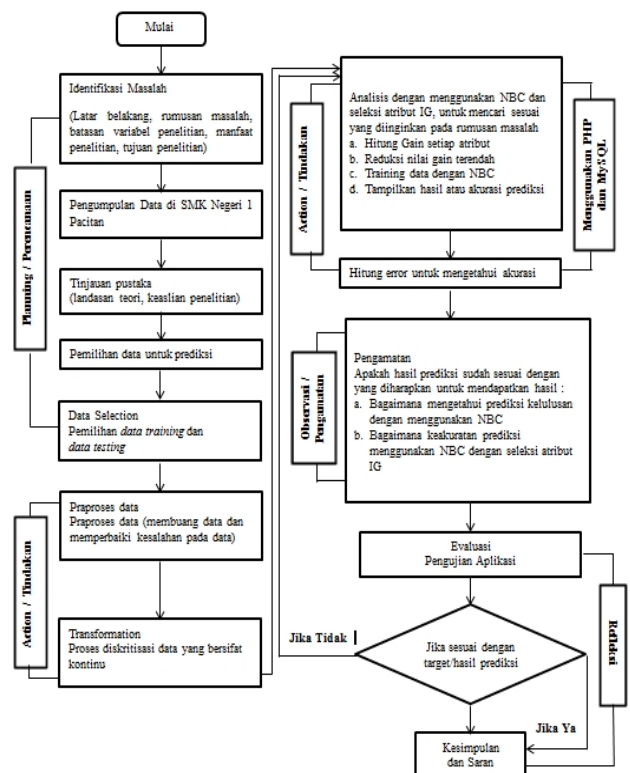
Penelitian yang dilakukan akan membahas tentang prediksi level ketidaksiplinan siswa di SMK Negeri 1 Pacitan dengan menggunakan *Naive Bayes Classifiers* (NBC). Algoritma NBC dikenal sebagai algoritma yang sederhana dalam implementasinya dan memiliki tingkat akurasi yang tinggi[6].

Penelitian Yeffriansjah Salim yang berjudul “Penerapan Algoritma Naive Bayes Untuk Penentuan Status Turn-Over Pegawai” (2012). Penelitian ini bertujuan untuk mengukur akurasi, presisi dari penentuan status turn-over pegawai menggunakan algoritma Naive Bayes dengan pemilihan atribut menggunakan metode PSO[7]. Sedangkan penelitian yang akan dilakukan pemilihan atribut menggunakan metode *information gain*.

Penelitian Sri Kusumadewi, tahun 2009 dengan judul “Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification” bertujuan untuk menentukan status gizi seseorang menggunakan alat ukur antropometri sebagai variabel input. Penelitian ini dilakukan tanpa ada pembobotan atribut[8]. Sedangkan pada penelitian yang akan dilakukan pemilihan atribut menggunakan metode *information gain*.

**Metode Penelitian**

Alur penelitian prediksi tingkat ketidakdisiplinan siswa akan dilakukan dengan langkah – langkah sebagai berikut :



Gambar 1. Diagram Alir Penelitian

**Data Mining**

Data mining merupakan proses menemukan suatu pola atau informasi pada data berukuran besar. Data mining juga merupakan proses dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar[9].

**Klasifikasi**

Klasifikasi merupakan kegiatan menggolongkan, memberikan kelas kepada data berdasarkan atribut. Klasifikasi merupakan proses identifikasi objek kedalam kelas, grup ataupun kategori tertentu berdasarkan prosedur, karakteristik serta definisinya[10].

**Information Gain**

*Information gain* (IG) merupakan suatu pengukuran yang dilakukan untuk melakukan seleksi terhadap atribut-atribut sehingga dapat disimpulkan atribut apa saja yang akan digunakan. Rumus dari entropy adalah sebagai berikut[11] :

$$Entropy (S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots\dots\dots(1)$$

Keterangan :

- S : himpunan kasus
- N : jumlah nilai yang ada pada atribut target (jumlah kelas)
- $p_i$  : jumlah sampel untuk kelas i
- $\log_2$  : logaritma biner

Setelah mendapatkan nilai entropy untuk suatu sampel data, maka selanjutnya adalah mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut sebagai IG. Secara matematis, *information gain* dari suatu atribut A, dituliskan sebagai berikut :

$$IG (S, A) = Entropy (S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} Entropy (S_v) \dots(2)$$

Keterangan :

- A : Atribut
- V : menyatakan suatu nilai yang mungkin untuk atribut A
- |Sv| : jumlah sampel untuk nilai v
- |S| : jumlah sampel seluruh sampel data
- Values(A) : himpunan nilai – nilai yang mungkin untuk atribut A
- Entropy(Sv) : entropy untuk sampel – sampel yang memiliki nilai v

**Naive Bayes Classifier**

*Naive 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 dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*[12]. Teorema *Bayes* memiliki bentuk umum sebagai berikut :

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

- 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 :

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

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

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

**2. Pembahasan**

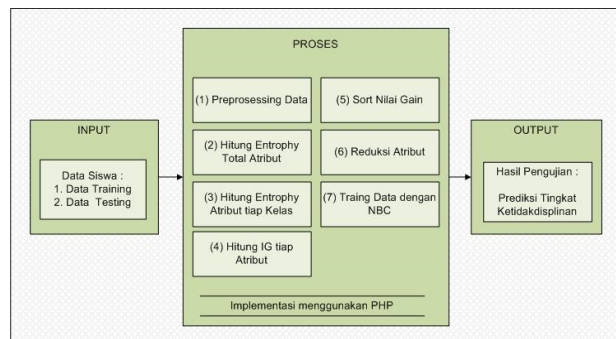
**Analisis Sistem**

Tahap pertama dilakukan preprocessing data, yaitu menyeleksi data dengan membuang atribut – atribut yang tidak relevan kemudian melakukan pembersihan data yaitu membuang atribut – atribut yang *missing value*. Selanjutnya adalah melakukan transformasi data yaitu mendiskritkan atribut yang bernilai kontinu untuk diubah menjadi data kategorikal.

Seleksi atribut merupakan proses untuk mengidentifikasi dan menghilangkan atribut dengan nilai yang tidak relevan. Seleksi atribut menggunakan metode IG bertujuan melakukan pengurutan atribut berdasarkan peringkat (ranking) dimana semakin besar nilai IG dari suatu atribut maka semakin signifikan atribut tersebut untuk tugas prediksi. Atribut yang telah diseleksi diimplementasikan pada algoritma NBC dengan menggunakan bahasa pemrograman PHP.

Data yang digunakan adalah biodata siswa SMK Negeri 1 Pacitan kelas X, XI dan XII tahun pelajaran 2013/2014.

Gambaran umum sistem yang akan dikembangkan untuk prediki potensi ketidakdisiplinan siswa dapat dilihat pada gambar 1 di bawah ini :



**Gambar 2. Blok Diagram Gambaran Umum Sistem**

**Preprocessing**

Preprocessing merupakan tahap awal pengolahan data sebelum masuk pada proses klasifikasi data. Preprocessing bertujuan untuk meningkatkan akurasi data. Tahapan preprocessing data dilakukan secara manual dan dilakukan diluar aplikasi.

**a. Selection Data**

Seluruh atribut *dataset* akan diseleksi untuk mendapatkan atribut – atribut yang relevan. Data asli siswa berjumlah 920 *record* dan memiliki 24 atribut. Atribut-atribut data siswa yang digunakan dalam penelitian ini selengkapnya terdiri dari : NIS, Nama, Jenis kelamin, Tempat lahir, Tanggal lahir, Kelas, Jurusan, Jalur masuk, Agama, Alamat, Nama ayah, Pekerjaan ayah, Pendidikan ayah, Nama ibu, Pekerjaan ibu, Pendidikan ibu, Penghasilan, Jumlah UN, Rata-rata UN, No HP, Asal Sekolah, Tahun lulus, No ijasah dan Poin pelanggaran.

Tidak semua atribut dimasukan dalam *dataset* akan digunakan dalam proses data mining karena hanya berperan sebagai referensi identifikasi, seperti NIS, Nama, No HP, Nama Ayah dan Nama Ibu. Beberapa atribut lain perlu diubah bentuknya, seperti “Alamat Rumah” diubah menjadi atribut “Kecamatan” dan “Kota”, sedangkan “Tanggal Lahir” digunakan untuk menentukan “Usia” saat masuk SMK.

Selanjutnya pada tahap ini atribut – atribut yang tidak relevan akan dibuang. Atribut yang awalnya berjumlah 24 atribut menjadi 16 atribut, yang terdiri dari 15 atribut prediktor dan 1 atribut target, dimana atribut target menjadi kelas output level ketidakdisiplinan siswa. Kelas output level ketidakdisiplinan siswa dikelompokkan menjadi tiga kelas yaitu :

**Tabel 1. Kelas Output Level Ketidakdisiplinan**

No	Kelas	Keterangan
1.	Hijau	Potensi level ketidakdisiplinan rendah
2.	Kuning	Potensi level ketidakdisiplinan sedang
3.	Merah	Potensi level ketidakdisiplinan tinggi

**b. Cleaning Data**

Selanjutnya dilakukan proses pembersihan data dari *missing value* dan redundansi data. Data disebut *missing value* jika pada *record* tertentu pada salah satu atribut tidak mempunyai nilai. *Record* yang *missing value* akan dihapus. Pada *dataset* yang sama jika *record* berisi nilai yang sama, maka *record* tersebut juga akan dihapus. Setelah dilakukan seleksi dan pembersihan data, total seluruh data yang digunakan awalnya berjumlah 920 *record* menjadi 895 *record*.

**c. Transformation Data**

Untuk atribut yang bernilai kontinu perlu didiskritisasi. Berikut diskritisasi atribut bernilai kontinu dapat dilihat pada tabel 2 :

**Tabel 2. Diskritisasi Atribut Bernilai Kontinu**

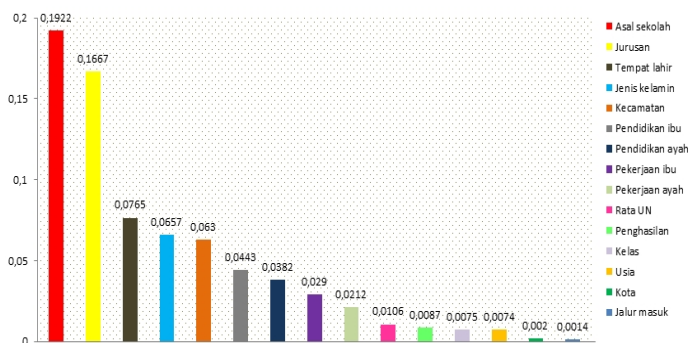
Nama Atribut	Interval	Keterangan
Penghasilan	$\leq 1.000.000$	Kurang
	$>1.000.000$ dan $\leq 2.000.000$	Cukup
	$>2.000.000$ dan $\leq 3.000.000$	Menengah
	$\geq 4.000.000$	Tinggi
Usia	$\leq 15$	1
	$\geq 16$ dan $\leq 17$	2
	$\geq 18$	3
Rata – rata UN	$<4,50$	Rendah
	$\geq 4,50$ dan $<6,00$	Cukup Rendah
	$\geq 6,00$ dan $<7,50$	Cukup
	$\geq 7,50$ dan $<9,00$	Cukup Tinggi
Poin	$\leq 30$	Rendah
	$>31$ dan $\leq 60$	Sedang
	$\geq 61$	Tinggi

**Seleksi Atribut**

Dengan menggunakan keseluruhan data yaitu 100% sebagai data training sekaligus sebagai data testing, akan dilakukan perhitungan nilai IG setiap atribut. Nilai IG untuk setiap atribut dapat dilihat pada tabel 3 :

**Tabel 3. Nilai Gain Setiap Atribut Untuk 100% Data Training sekaligus Data Testing**

Atribut	Nama Atribut	Nilai IG
1.	Asal sekolah	0,1922
2.	Jurusan	0,1667
3.	Tempat lahir	0,0765
4.	Jenis kelamin	0,0657
5.	Kecamatan	0,063
6.	Pendidikan ibu	0,0443
7.	Pendidikan ayah	0,0382
8.	Pekerjaan ibu	0,029
9.	Pekerjaan ayah	0,0212
10.	Rata – rata UN	0,0106
11.	Penghasilan	0,0087
12.	Kelas	0,0075
13.	Usia	0,0074
14.	Kota	0,002
15.	Jalur masuk	0,0014



**Gambar 3. Nilai Gain Setiap Atribut Untuk 100% Data Training Sekaligus Data Testing**

Dari tabel 3 dan gambar 2 dapat dilihat nilai IG untuk setiap atribut. Nilai IG yang tertinggi adalah asal sekolah sebesar 0,1922 dan yang terendah adalah jalur masuk sebesar 0,0014. Setelah nilai IG untuk setiap atribut diketahui, selanjutnya akan dilakukan uji coba skenario untuk seleksi atribut dengan jumlah reduksi atribut berbeda. Skenario percobaan nilai reduksi atribut dapat dilihat pada tabel 4 :

**Tabel 4. Skenario Percobaan Nilai Reduksi Atribut**

No	Nilai reduksi atribut (%)	Jumlah atribut yang direduksi	Jumlah atribut untuk proses prediksi	Daftar atribut yang dipakai
1.	0%	0	15	Atribut 1 - Atribut 15
2.	25% (tertinggi)	4	11	Atribut 5 - Atribut 15
3.	25% (terendah)	4	11	Atribut 1 - Atribut 11
4.	50% (tertinggi)	8	7	Atribut 9 - Atribut 15
5.	50% (terendah)	8	7	Atribut 1 - Atribut 7
6.	75% (tertinggi)	11	4	Atribut 12 - Atribut 15
7.	75% (terendah)	11	4	Atribut 1 - Atribut 4

Percobaan pertama, tidak ada atribut yang direduksi. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 1 sampai atribut 15 yaitu Asal sekolah, Jurusan, Tempat lahir, Jenis kelamin, Kecamatan, Pendidikan ibu, Pendidikan ayah, Pekerjaan ibu, Pekerjaan ayah, Rata – rata UN, Penghasilan, Kelas, Usia, Kota dan Jalur masuk. Seluruh atribut digunakan untuk proses prediksi.

Percobaan kedua, jumlah atribut yang direduksi berjumlah 4 diurutkan dari nilai IG tertinggi. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 5 sampai dengan atribut 15 yaitu Kecamatan, Pendidikan ibu, Pendidikan ayah, Pekerjaan ibu, Pekerjaan ayah, Rata – rata UN, Penghasilan, Kelas, Usia, Kota dan Jalur masuk. Jumlah seluruh atribut yang digunakan untuk proses prediksi berjumlah 11.

Percobaan ketiga, jumlah atribut yang direduksi berjumlah 4 diurutkan dari nilai IG terendah. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 1 sampai atribut 11 yaitu Asal sekolah, Jurusan, Tempat lahir, Jenis kelamin, Kecamatan, Pendidikan ibu, Pendidikan ayah, Pekerjaan ibu, Pekerjaan ayah, Rata – rata UN dan Penghasilan. Jumlah seluruh atribut yang digunakan untuk proses prediksi berjumlah 11.

Percobaan keempat, jumlah atribut yang direduksi berjumlah 8 diurutkan dari nilai IG tertinggi. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 9 sampai atribut 15 yaitu Pekerjaan ayah, Rata –

rata UN, Penghasilan, Kelas, Usia, Kota dan Jalur masuk. Jumlah seluruh atribut yang digunakan untuk proses prediksi berjumlah 7.

Percobaan kelima, jumlah atribut yang direduksi berjumlah 8 diurutkan dari nilai IG terendah. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 1 sampai atribut 7 yaitu Asal sekolah, Jurusan, Tempat lahir, Jenis kelamin, Kecamatan, Pendidikan ibu dan Pendidikan ayah. Jumlah seluruh atribut yang digunakan untuk proses prediksi berjumlah 7.

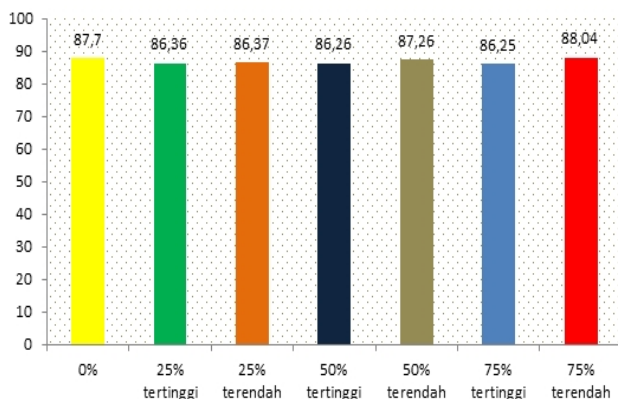
Percobaan keenam, jumlah atribut yang direduksi berjumlah 11 diurutkan dari nilai IG tertinggi. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 12 sampai atribut 15 yaitu Kelas, Usia, Kota dan Jalur masuk. Jumlah seluruh atribut yang digunakan untuk proses prediksi berjumlah 4.

Percobaan ketujuh, jumlah atribut yang direduksi berjumlah 11 diurutkan dari nilai IG terendah. Jumlah atribut yang digunakan untuk proses prediksi adalah atribut 1 sampai atribut 4 yaitu Asal sekolah, Jurusan, Tempat lahir dan Jenis kelamin. Jumlah seluruh atribut yang digunakan untuk proses prediksi berjumlah 4.

Dari hasil skenario uji coba untuk seleksi atribut dengan nilai reduksi atribut yang berbeda dihasilkan akurasi seperti pada tabel 5 :

**Tabel 5.** Hasil Akurasi Skenario Percobaan Nilai Reduksi Atribut

Percobaan	Nilai reduksi atribut (%)	Akurasi (%)
1	0	87,70
2	25% (tertinggi)	86,59
3	25% (terendah)	86,37
4	50% (tertinggi)	86,26
5	50% (terendah)	87,26
6	75% (tertinggi)	86,25
7	75% (terendah)	88,04



**Gambar 4.** Grafik Hasil Akurasi Skenario Percobaan Nilai Reduksi Atribut

Dari tabel 5 dan gambar 3, dapat dilihat pada percobaan pertama dengan tidak ada reduksi atribut 0% menghasilkan akurasi sebesar 87,7%. Percobaan kedua dengan nilai reduksi 25% diurutkan dari nilai IG tertinggi menghasilkan akurasi sebesar 86,36%. Percobaan ketiga dengan nilai reduksi 25% diurutkan dari nilai IG terendah menghasilkan akurasi sebesar 86,37%. Percobaan keempat dengan nilai reduksi 50% diurutkan dari nilai IG tertinggi menghasilkan akurasi sebesar 86,25%. Percobaan kelima dengan nilai reduksi 50% diurutkan dari nilai IG terendah menghasilkan akurasi sebesar 87,26%. Percobaan keenam dengan nilai reduksi 75% diurutkan dari nilai IG tertinggi menghasilkan akurasi sebesar 86,25%. Sedangkan pada percobaan ketujuh dengan nilai reduksi 75% diurutkan dari nilai IG terendah menghasilkan akurasi sebesar 88,04%. Hasil terbaik diperoleh pada percobaan dengan nilai reduksi atribut 75% diurutkan dari nilai IG terendah dengan nilai akurasi mencapai 88,04%.

**Prediksi dengan NBC**

Setelah diketahui reduksi atribut 75% diurutkan dari nilai IG terendah menghasilkan akurasi yang terbaik. Selanjutnya dilakukan skenario percobaan untuk pengujian dengan data training dan data testing yang berbeda.

a. Pengujian pertama

Pembagian jumlah data training 75% yaitu 671 record dan data testing 25% yaitu 224 record. Hasil pengujian didapatkan jumlah prediksi benar keputusan dengan NBC berjumlah 177 record.

Akurasi pengujian pertama

$$= \frac{177}{224} \times 100 = 79,01\%$$

b. Pengujian kedua

Pembagian jumlah data training 50% yaitu 448 record dan data testing 50% yaitu 447 record. Hasil pengujian didapatkan jumlah prediksi benar keputusan dengan NBC berjumlah 333 record.

Akurasi pengujian kedua

$$= \frac{333}{447} \times 100 = 68,38\%$$

c. Pengujian ketiga

Pembagian jumlah data training 25% yaitu 224 record dan data testing 75% yaitu 671 record. Hasil pengujian didapatkan jumlah prediksi benar keputusan dengan NBC berjumlah 469 record.

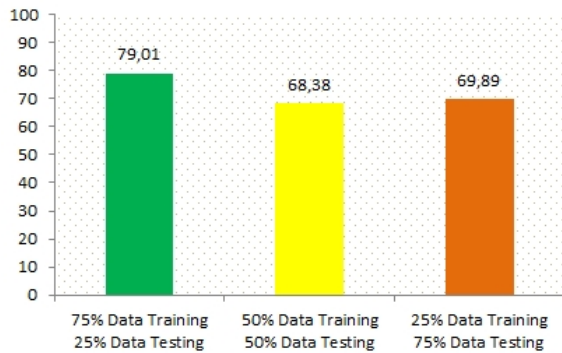
Akurasi pengujian ketiga

$$= \frac{469}{671} \times 100 = 69,89\%$$

Hasil akurasi dari pengujian komposisi data training dan data testing yang dilakukan dapat dilihat pada tabel 6 :

**Tabel 6. Hasil Akurasi Pengujian Pertama Sampai Pengujian Ketiga**

Pengujian	Data Training	Data Testing	Akurasi
1	75%	25%	79,01%
2	50%	50%	68,38%
3	25%	75%	69,89%



**Gambar 5. Grafik Hasil Akurasi Pengujian Pertama sampai Pengujian Ketiga**

Dari tabel 6 dan gambar 4, percobaan pertama dengan pembagian data training 75% dan data testing 25% menghasilkan akurasi sebesar 79,01%. Percobaan kedua dengan pembagian data training 50% dan data testing 25% menghasilkan akurasi sebesar 68,38%. Sedangkan pada percobaan ketiga dengan pembagian data training 25% dan data testing 75% menghasilkan akurasi sebesar 69,89%.

Hasil terbaik dengan menggunakan NBC menghasilkan akurasi terbaik yaitu sebesar 79,01% dengan reduksi atribut 75% diurutkan dari nilai IG terendah dengan pembagian jumlah data training 75% dan jumlah data testing 25%. Dengan tingkat akurasi yang dihasilkan sebesar 79,01%, algoritma NBC menunjukkan performa yang baik untuk memprediksi potensi ketidaksiplinan siswa.

### 3. Kesimpulan

Dari hasil penelitian didapatkan kesimpulan sebagai berikut :

1. Pada penelitian ini dengan menggunakan algoritma NBC yang digunakan di SMK Negeri 1 Pacitan sudah baik dilihat dari data yang diperoleh yang mempunyai kesesuaian 79,01%.
2. Atribut utama yang menjadi penentu potensi ketidaksiplinan adalah atribut asal sekolah. Hal ini dapat dilihat berdasarkan percobaan dengan menggunakan 100% data training dan juga sebagai data testing didapatkan hasil perhitungan nilai *information gain* tertinggi adalah atribut asal sekolah sebesar 0,1922%.
3. Prediksi ketidaksiplinan dengan menggunakan algoritma *Naive Bayes Classifier* menghasilkan akurasi terbaik yaitu sebesar 79,01% didapatkan dengan kondisi :

- a. Reduksi atribut sebesar 75% diurutkan dari nilai *information gain* yang terendah
- b. Perbandingan data training dan data testing sebesar 75% untuk *data training* dan 25% data testing.

### Daftar Pustaka

- [1] T.L. Widiatuti, "Hubungan Antara Kedisiplinan Dengan Prestasi Belajar Siswa SMA Santo Bernadus Pekalongan", Universitas Katolik Soegijapranata, Semarang, 2008.
- [2] Suratman, *Pembinaan Mental, Fisik dan Disiplin*, LAN, Jakarta, 1999.
- [3] M.U. Usman, *Menjadi Guru Profesional*, Bandung, Remaja Rosdakarya, 1989.
- [4] Tim Pengembangan Ilmu Pendidikan FIP-UPI, *Ilmu dan Aplikasi Pendidikan*, Jakarta, Grasindo, 2007.
- [5] Jr.R. McLeod, dan G.P. Schell, *Management Information System. 10th ed*, Pearson Education, Inc, 2007.
- [6] A.N. Fadillah, dkk, "Island Shape Detector Menggunakan Chain Code Dengan Metode Pengklasifikasian Naive Bayes". Universitas Brawijaya, Malang, 2014.
- [7] Y. Salim, "Penerapan Algoritma Naive Bayes Untuk Penentuan Status Turn-Over Pegawai". Media Sains, Volume 4 Nomor 2, Oktober 2012, ISSN 2085-3548, 2012.
- [8] S. Kusumadewi, "Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification". CommIT, Vol.3 No.1 Mei 2009, hlm 6-11, 2009.
- [9] Turban, Efraim, et al, *Decision Support Systems and Intelligent Systems 7th ed*, New Jersey, Pearson Education, 2005.
- [10] Pang-Ning Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*, Pearson Education, Inc., Boston, 2005.
- [11] N. Rifqi, "Analisis Dan Implementasi Klasifikasi Data Mining Menggunakan Jaringan Syaraf Tiruan Dan Evolution Strategis". Konferensi Nasional Sistem dan Informatika 2011, Bali, November 12, 2011.
- [12] Kusriani dan E.T. Lutfhi, *Algoritma Data Mining*, Andi Offset, Surabaya, 2009

### Biodata Penulis

**Niken Puji Astuti**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika Universitas Kristen Satya Wacana Salatiga, lulus tahun 2008. Saat ini menempuh studi untuk memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika STMIK AMIKOM Yogyakarta. Saat ini menjadi pengajar di SMK Negeri 1 Pacitan.

**Kusrini**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2002. Memperoleh gelar Magister Ilmu Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2006. Memperoleh gelar Doktor (Dr) Program Doktor Ilmu Komputer Universitas Gajah Mada Yogyakarta. Saat ini menjadi Dosen tetap di STMIK AMIKOM Yogyakarta.

**M. Rudyanto Arief**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Ilmu Komputer Universitas Islam Indonesia Yogyakarta, lulus tahun 2001. Memperoleh gelar Magister Ilmu Komputer (M.T) Program Pasca Sarjana Magister Teknik Elektro Universitas Gajah Mada Yogyakarta, lulus tahun 2005. Saat ini menjadi Dosen tetap di STMIK AMIKOM Yogyakarta.