

METODE KLASIFIKASI DENGAN ALGORITMA NAÏVE BAYES UNTUK REKOMENDASI PENJURUSAN SMA TERANG BANGSA

Andrew Yova Kencana¹, Setia Astuti²

^{1,2}Jurusan Teknik Ilmu Komputer, Universitas Dian Nuswantoro, Semarang

Jl. Imam Bonjol 205-207 Semarang 50131

e-mail: andrewyova@rocketmail.com¹, setia.astuti@dsn.dinus.ac.id²

Abstrak

SMA Terang Bangsa adalah salah satu sekolah yang sedang menerapkan pendidikan kurikulum 2013. Di mana kurikulum 2013 ini adalah penjurusan yang dilakukan di kelas X. Berbeda dengan tahun lalu di mana penjurusan di lakukan di kelas XI menggunakan faktor nilai-nilai SMA maka di kurikulum baru ini menggunakan faktor nilai SMP mereka. Penelitian ini bertujuan untuk membuat aplikasi data mining klasifikasi menggunakan algoritma Naive Bayes. Algoritma Naive Bayes ini adalah algoritma yang menggunakan teknik probabilitas sederhana berdasarkan penerapan aturan bayes. Atribut yang di gunakan dalam penelitian ini adalah NIS (No Induk Siswa), nilai rata-rata pelajaran IPA, nilai rata-rata Matematika, nilai rata-rata Bahasa Inggris, dan minat. Hasil dari penelitian ini adalah terimplementasinya aplikasi data mining sebagai klasifikasi penjurusan Siswa SMA Terang Bangsa.

Kata kunci : Data Mining, Algoritma Naïve Bayes, Penjurusan SMA, Kurikulum 2013.

Abstract

SMA Terang Bangsa is one of the schools that are implementing the education curriculum in 2013. Where the 2013 curriculum is majors that is done in class X. Unlike the last year where the majors in class XI using Senior High School grades, so in this new curriculum they using junior high school grades. This research aims to make the application of data mining classification using Naive Bayes algorithm. Naive Bayes algorithm is an algorithm that uses simple probability techniques based on the application of Bayes rule. Attributes used in this study are NIS (Student ID), the average value of science courses, the average value of Mathematics, the average value of English, and majors. The results of this study is data mining application as majors classification of SMA Terang Bangsa student.

Keywords : Data Mining, Naïve Bayes Algorithm, High School Majors, Curriculum 2013.

1. PENDAHULUAN

Teknologi saat ini sudah sangat berkembang dengan cepat dalam tahun-tahun ini. Di jaman ini banyak teknologi di gunakan untuk membantu pekerjaan manusia dalam hal pencatatan barang dagangan, membantu membuat keputusan dalam suatu masalah, dan sebagainya. Salah satunya adalah membuat keputusan dalam penjurusan anak kelas SMA ke jurusan IPA maupun IPS. Sekarang anak-anak SMP saat ingin melanjutkan bidang studinya

ke SMA kelas X sudah dalam penjurusan IPA atau IPS. Di karenakan adanya kurikulum tahun 2013 yang baru dalam pendidikan SMA tahun ini. Berbeda dengan tahun 2006 di mana saat penjurusan di lakukan di kelas X ke kelas XI tahun ini kelas X sudah dalam penjurusan IPA maupun IPS.

SMA Terang Bangsa adalah Sekolah SMA yang sedang melaksanakan kurikulum pendidikan tahun 2013 ini. Dengan adanya penjurusan ini siswa di harapkan mampu menguatkan bakat,

minat dan kemampuan yang dia miliki agar siswa mampu berprestasi di bidang yang mereka sukai ini. Di SMA Terang Bangsa ini memiliki 2 jurusan yaitu IPA dan IPS. Di Kelas X jurusan IPA ada 3 kelas sedangkan jurusan IPS ada 6 kelas. Saat ini di SMA Terang Bangsa saat melakukan jurusan mereka masih menggunakan data yang ada misalnya angket minat jurusan. Di angket minat jurusan terdapat pertanyaan minat yang diinginkan siswa, dan nilai-nilai rapor smp mereka yang terdiri dari nilai matematika, IPA, dan Bahasa Inggris . Dari situlah mereka menarik kesimpulan untuk mengelompokkan siswa ke dalam IPA maupun IPS.

Saat ini SMA Terang Bangsa sudah menerapkan pekerjaan yang melibatkan komputer seperti mengatur absensi , Sistem keuangan , Daftar Siswa, dan lain sebagainya. Program yang mereka gunakan adalah kebanyakan MS-Office. Pada sistem Absensi di SMA Terang Bangsa mereka sudah menggunakan program dengan bahasa PHP. Di sini mereka menunjukkan bahwa SMA Terang Bangsa juga sedang meningkatkan pekerjaan mereka dengan menggunakan bantuan komputer. Dalam Penelitian ini Peneliti menggunakan program dengan bahasa JAVA. Data yang ada berupa data angket nilai SMP mereka dan Minat mereka.

Metode yang digunakan untuk penelitian ini adalah metode Naïve Bayes. Metode ini adalah teknik metode data mining dalam klasifikasi sebuah data yang berlabel. Data yang ada di SMA Terang Bangsa cocok di gunakan untuk metode ini karena datanya berlabel. Data dari Variabel tersebut bersifat bebas, dan tidak saling berkaitan sehingga bila dalam

penggunaan program Konvensional (if ... then ...) tidak mungkin di lakukan sehingga data ini cocok di gunakan dengan metode Naïve Bayes. Maka dari permasalahan yang telah di jelaskan di atas maka penulis ingin meneliti dan menyelesaikan permasalahan dengan menggunakan Teknik Data Mining yaitu klasifikasi dengan judul “Metode Klasifikasi Dengan Algoritma Naïve Bayes untuk rekomendasi Jurusan SMA Terang Bangsa”.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Teknik *data mining* digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna[1].

Data mining, sering juga di sebut *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar[2].

2.2 Klasifikasi

Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Metode-metode classification yang lain adalah Bayesian, Neural network, genetic algorithm, fuzzy, case based reasoning, dan k-nearest neighbor [1].

2.3 Algoritma Naïve Bayes

Algoritma Naïve Bayes Classifier merupakan algoritma yang di gunakan

untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. [3][4][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 di ketahui

H= Hipotesis data X merupakan suatu class spesifik

P(H|X) = probabilitas hipotesis h berdasar 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.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

Data yang di dapat di SMA Terang bangsa sebanyak 205. Sebelum di proses ke dalam aplikasi maka di olah dulu di excel ke bentuk yang lebih mudah untuk dijadikan sebagai formula Naïve Bayes seperti gambar berikut :

Tabel 1 : Tabel Data Siap Olah

No	NIS	Rata-Rata Mat	Rata-rata IPA	Rata-rata Binggris	Minat	Kelas
20	2015.10.1470	Middle	Middle	Low	IPS	IPS
21	2015.10.1471	High	Middle	Middle	IPS	IPS
22	2015.10.1472	Middle	Middle	Middle	IPA	IPA
23	2015.10.1474	High	Middle	High	-	IPA
24	2015.10.1475	Middle	Middle	Middle	IPS	IPS
25	2015.10.1476	Middle	Middle	Middle	IPS	IPS
26	2015.10.1477	Middle	Middle	Middle	IPS	IPS
27	2015.10.1478	Middle	Middle	Middle	-	IPS
28	2015.10.1479	High	High	High	-	IPS
29	2015.10.1480	Middle	High	Middle	IPS	IPA
30	2015.10.1481	Middle	Middle	Middle	-	IPS
31	2015.10.1482	Middle	Middle	High	-	IPS
32	2015.10.1483	Middle	Middle	Middle	-	IPS
33	2015.10.1484	Middle	Middle	Middle	-	IPS
34	2015.10.1486	Low	Middle	Middle	IPA	IPA

Ketentuan untuk mengubah nilai nominalnya adalah sebagai berikut :

1. Di katakan High bila nilai di atas 84 (X > 84)
2. Di katakan Middle bila nilai di antara/ sama dengan 60 sampai di bawah atau sama dengan 84. (60 <= X <= 84)
3. Di katakan low bila nilai di bawah 60 (X < 60)

3.2 Hasil Implementasi

Setelah 205 Data sudah di konversikan maka mendapatkan Formula Sebagai berikut :

Untuk probabilitas Kelas IPA
Untuk probabilitas Kelas IPS

$$P(\text{IPA}) = 83$$

$$P(\text{IPS}) = 122$$

$$P(\text{Matematika} | \text{Mid}) = 53$$

$$P(\text{Matematika} | \text{Mid}) = 111$$

$$P(\text{Matematika} | \text{High}) = 30$$

$$P(\text{Matematika} | \text{High}) = 5$$

$$P(\text{Matematika} | \text{low}) = 0$$

$$P(\text{Matematika} | \text{low}) = 6$$

$$P(\text{IPA} | \text{Mid}) = 44$$

$$P(\text{IPA} | \text{Mid}) = 112$$

$$P(\text{IPA} | \text{High}) = 39$$

$$P(\text{IPA} | \text{High}) = 10$$

$$P(\text{IPA} | \text{low}) = 0$$

$$P(\text{IPA} | \text{low}) = 0$$

- P(B Inggris | Mid) = 45
- P(B Inggris | Mid) = 106
- P(B Inggris | High) = 38
- P(B Inggris | High) = 15
- P(B Inggris | low) = 0
- P(B Inggris | low) = 1
- P(Minat | -) = 25
- P(Minat | -) = 71
- P(Minat | IPA) = 51
- P(Minat | IPA) = 19
- P(Minat | IPS) = 7
- P(Minat | IPS) = 32

Sehingga membentuk Probabilitas sebagai berikut :

1. Untuk Probabilitas Kelas IPA = $P_{(IPA)} = 83/205 = 0.404878049$

Tabel 2: Tabel Probabilitas IPAClass

IPA			
	High	Middle	Low
Average Mat	0.3614457 83	0.6385542 17	0
Average IPA	0.4698795 18	0.5301204 82	0
Average Binggris	0.4578313 25	0.5301204 82	0
	-	IPA	IPS
Minat	0.3012048 19	0.6144578 31	0.0843373 49

2. Untuk Probabilitas Kelas IPS = $P_{(IPS)} = 122/205 = 0.595121951$

Tabel 3: Tabel Probabilitas IPS Class

IPS			
	High	Middle	Low
Average Mat	0.0409836 07	0.9098360 66	0.0491803 28
Average IPA	0.0819672 13	0.9180327 87	0

Average Binggris	0.1229508 2	0.8688524 59	0.0081967 21
	-	IPA	IPS
Minat	0.5819672 13	0.1557377 05	0.2622950 82

Ketika Mendapat probabilitas IPA dan IPS bernilai 0 maka dapat di klasifikasi sebagai kelas IPS. Di dapat hasil 81 Percobaan yang menyatakan Siswa mendapat kelas IPA terdapat 18 percobaan dan 63 percobaan menghasilkan Nilai IPA.

Tabel 4: Percobaan 1 sampai 20

Percobaan ke	Rata-Rata Mat	Rata-rata IPA	Rata-rata Binggris	Minat	Kelas
1	High	High	High	-	IPA
2	High	High	High	IPA	IPA
3	High	High	High	IPA	IPA
4	High	High	Middle	-	IPA
5	High	High	Middle	IPA	IPA
6	High	High	Middle	IPS	IPA
7	High	High	Low	-	IPS
8	High	High	Low	IPA	IPS
9	High	High	Low	IPS	IPS
10	High	Middle	High	-	IPA
11	High	Middle	High	IPA	IPA
12	High	Middle	High	IPS	IPA
13	High	Middle	Middle	-	IPA
14	High	Middle	Middle	IPA	IPA
15	High	Middle	Middle	IPS	IPA
16	High	Middle	Low	-	IPS
17	High	Middle	Low	IPA	IPS
18	High	Middle	Low	IPS	IPS
19	High	Low	High	-	IPS
20	High	Low	High	IPA	IPS

Tabel 5: Percobaan 21 sampai 40

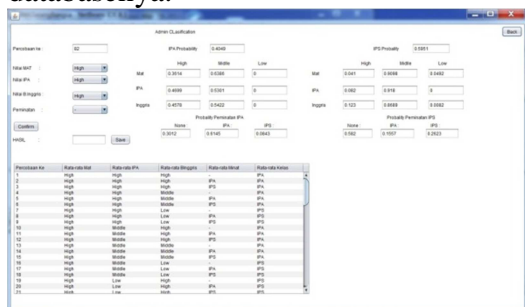
Percobaan ke	Rata-Rata Mat	Rata-rata IPA	Rata-rata Binggris	Minat	Kelas
21	High	Low	High	IPS	IPS
22	High	Low	Middle	-	IPS
23	High	Low	Middle	IPA	IPS
24	High	Low	Middle	IPS	IPS
25	High	Low	Low	-	IPS
26	High	Low	Low	IPA	IPS
27	High	Low	Low	IPS	IPS
28	Middle	High	High	-	IPA
29	Middle	High	High	IPA	IPA
30	Middle	High	High	IPS	IPA
31	Middle	High	Middle	-	IPA
32	Middle	High	Middle	IPA	IPA
33	Middle	High	Middle	IPS	IPS
34	Middle	High	Low	-	IPS
35	Middle	High	Low	IPA	IPS
36	Middle	High	Low	IPS	IPS
37	Middle	Middle	High	-	IPS
38	Middle	Middle	High	IPA	IPA
39	Middle	Middle	High	IPS	IPS
40	Middle	Middle	Middle	-	IPS

Contoh Perhitungan manual :
 siswa NIS 2016.10.0002 dengan atribut Nilai average :
 Matematika : Middle
 IPA : High

B.Ingggris : Middle
 Minat : IPS
 Maka siswa tersebut masuk ke kelas ?
 Untuk Probabilitas IPA :
 $P(\text{IPA}) = 0.638554217 * 0.469879518 * 0.530120482 * 0.084337349$
 $P(\text{IPA}) = 0.01341463$
 Untuk Probabilitas IPS
 $P(\text{IPS}) = 0.909836066 * 0.081967213 * 0.868852459 * 0.262295082$
 $P(\text{IPS}) = 0.01699571$
 $P(\text{IPA}) =$
 $P(\text{IPA}) / (P(\text{IPA}) + P(\text{IPS}))$
 $P(\text{IPA}) = 0.01341463 / (0.01341463 + 0.01699571)$
 $P(\text{IPA}) = 0.4411206$
 $P(\text{IPS}) =$
 $P(\text{IPS}) / (P(\text{IPA}) + P(\text{IPS}))$
 $P(\text{IPS}) = 0.01699571 / (0.01341463 + 0.01699571)$
 $P(\text{IPS}) = 0.5588793$
 Karna P(IPS) lebih besar dari P(IPA) maka kelasnya masuk ke kelas IPS

3.3 Simulasi Aplikasi

Aplikasi di buat dengan menggunakan Java Netbeans dan XAMPP sebagai databasenyaa.



Gambar 1. Simulasi Program

4. KESIMPULAN

Kesimpulan dari penelitian di atas adalah dapat mengklasifikasi Siswa SMA Terang Bangsa dengan teknik klasifikasi data mining menggunakan metode naïve bayes yang di imlementasi dengan bahasa pemrograman java.

5. SARAN

Saran untuk peneliti selanjutnya agar dapat memperbaiki kelemahan penelitian ini, di antaranya :

1. Perlu adanya penggunaan smoothing correction agar menghindari probabilitas 0 (nol).
2. Data yang digunakan berupa nominal, bisa di ubah ke bentuk numeric.
3. Peneliti menggunakan metode Naïve Bayes classification, di harapkan peneliti selanjutnya bisa menggunakan metode-metode lainnya seperti id3, c.4.5 , dan sebagainya.

DAFTAR PUSTAKA

- [1] J. Han and M. Kamber, *Data Mining : Concept and Techniques Second Edition*. San Francisco: Morgan Kaufmann Publishers, 2006.
- [2] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk keperluan Bisnis*, 1st ed. Yogyakarta: Graha Ilmu, 2007.
- [3] K Bambang, E. Syahril, and S.S. Opim, "Dunia Teknologi Informasi," *Klasifikasi Konten Berita Dengan Metode Text Mining*, vol. 1, pp. 14-19, 2012.
- [4] Y. S. NUGROHO, "Data Mining Menggunakan Algoritma Naive

Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro.," Dian Nuswantoro Fakultas Ilmu Komputer Skripsi, 2014.

- [5] O. KRISTANTO, "PENERAPAN ALGORITMA KLASIFIKASI DATA MINING ID3 UNTUK MENENTUKAN PENJURUSAN SISWA SMAN6 SEMARANG," Dian Nuswantoro Fakultas Ilmu Komputer Skripsi, 2014.