

KLASIFIKASI PENJURUSAN SISWA MA AL-QODIRI MENGGUNAKAN METODE NAIVE BAYES

Viki firdausi¹, Triawan Adi Cahyanto²

Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Jember

e-mail: vikifirdausi04@gmail.com, triawanac@unmuhjember.ac.id

ABSTRAK

MA Al-qodiri Jember 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 dilakukan 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 digunakan dalam penelitian ini adalah nama siswa, nilai rapor IPA, nilai rapor IPS, nilai rapor Matematika, nilai psikotest dan minat. Hasil dari penelitian ini adalah. Dari 304 data yang di training dan di testing mendapat akurasi terbaik sebesar 81,26% dan cocok untuk menentukan rekomendasi pada siswa di MA Al-qodiri Jember.

Kata Kunci : Naïve Bayes, Klasifikasi, Penjurusan, Kurikulum 2013.

KLASIFIKASI PENJURUSAN SISWA MA AL-QODIRI MENGGUNAKAN METODE NAIVE BAYES

Viki firdausi¹, Triawan Adi Cahyanto²

Program Studi Informatika Fakultas Teknik Universitas Muhammadiyah Jember

e-mail: vikifirdausi04@gmail.com, triawanac@unmuhjember.ac.id

ABSTRACT

MA Al-qodiri Jember is one of the schools that is implementing 2013 curriculum education. Where this 2013 curriculum is a course conducted in class X. Different from last year where majors were conducted in class XI using factors of high school grades then in the curriculum this new uses their junior high school grades. This study aims to make data mining classification applications using the Naive Bayes algorithm. This Naive Bayes algorithm is an algorithm that uses simple probability techniques based on the application of Bayes rules. Attributes used in this study are student names, science report cards, IPS report cards, Mathematics report cards, psychotest scores and interests. The results of this study are. Of the 304 data that were trained and tested the best accuracy was 81.26% and was suitable for determining recommendations for students at MA Al-qodiri Jember.

Keywords : Naïve Bayes, Classification of Majoring, 2013 Curriculum.

A. PENDAHULUAN

Penjurusan siswa adalah suatu proses pengambilan keputusan dalam memilih bidang keahlian studi berdasarkan kemampuan potensi diri dan peluang yang ada. Secara formal pemilihan jurusan merupakan ketentuan yang sudah ditetapkan oleh pemerintah melalui kurikulum 2013 yang berlaku saat ini. Untuk menentukan penjurusan siswa dilakukan pada awal masuk, yaitu pada kelas X Sekolah Menengah Atas. Penerapan kurikulum 2013 ini bertujuan untuk penyesuaian program pendidikan pada satuan pendidikan dengan kondisi dan kekhasan potensi yang ada di daerah siswa (Kementrian Pendidikan dan Kebudayaan, 2013).

MA Al-qodiri Jember adalah Sekolah Madrasah Aliyah 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 MA Al-qodiri Jember ini memiliki 2 penjurusan yaitu IPA dan IPS. Di Kelas X penjurusan IPA ada 3 kelas sedangkan penjurusan IPS ada 4 kelas. Saat ini di MA Al-qodiri Jember saat melakukan penjurusan mereka masih menggunakan data yang ada misalnya angket minat penjurusan. Di angket minat penjurusan terdapat pertanyaan minat yang di inginkan siswa,

dan nilai-nilai raport smp mereka yang terdiri dari nilai matematika, IPA, IPS dan Psikotest. Dari situlah mereka menarik kesimpulan untuk mengelompokkan siswa ke dalam IPA maupun IPS. Saat ini MA Al-qodiri Jember sudah menerapkan pekerjaan yang melibatkan komputer seperti mengatur absensi , Sistem keuangan , Daftar Siswa, dan lain sebagainya. Program yang mereka gunakan adalah kebanyakan MS-Office. Di sini mereka menunjukan bahwa MA Al-qodiri juga sedang meningkatkan pekerjaan mereka dengan menggunakan.

Penelitian ini menggunakan metode *Naïve Bayes* karena berdasarkan penelitian sebelumnya yang telah dilakukan oleh Maghriza Fakri Adilla dan T. Sutojo, “penelitian ini menghasilkan prediksi penjurusan siswa menggunakan matlab diperoleh akurasi sebesar 83,8798 % dengan error rate sebesar 16,1202%”. Dengan algoritma *Naïve Bayes* tersebut diharapkan dapat menjadi alat yang membantu dalam menentukan klasifikasi penjurusan siswa di MA Al-qodiri Jember.

B. KAJIAN PUSTAKA

Sebelum melakukan penelitian penulis terlebih dahulu melakukan tinjauan pustaka dari penelitian lain dan penelitian tentang prediksi penjurusan sekolah menengah atas ini bukanlah penelitian yang pertama kalinya. Sebelumnya sudah ada penelitian yang membahas mengenai klasifikasi penjurusan sekolah menengah atas menggunakan metode algoritma *Naive*

Bayes. Berikut adalah beberapa penelitian yang terkait dengan masalah tersebut.

Penelitian yang pertama adalah penelitian yang dilakukan oleh Claudia Clarentina Ciptohartono yang berjudul “Algoritma Klasifikasi *Naive Bayes* untuk Menilai Kelayakan Kredit”(Claudia Clarentia Ciptohartono,2014).

Metode yang digunakan pada penelitian ini adalah Algoritma *Naive Bayes*. Data yang digunakan adalah data dari nasabah perusahaan BCA Finance Jakarta tahun 2013. Tool yang digunakan untuk implementasi sistem adalah matlab. Penelitian ini menghasilkan bukti bahwa algoritma *Naive Bayes* bisa diterapkan dalam pemberian kelayakan kredit pada BCA Finance Jakarta. Tahap pengolahan data awal dapat menghasilkan akurasi yang tinggi dan menghasilkan akurasi akhir yang *Excellent*. Dalam penelitian kelayakan kredit ini pada data awal dengan melakukan pre-processing mendapatkan akurasi sebesar 85.57%, tetapi jika data awal dilakukan pengolahan dan dengan melakukan pre-processing dapat menghasilkan akurasi sebesar 92.53%. Jadi Algoritma *Naive Bayes* akan lebih unggul dalam pemberian penilaian kelayakan kredit jika dilakukan proses pengolahan data awal, meskipun algoritma *Naive Bayes* adalah algoritma yang sanggup menangani data yang hilang.

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 menentukan pola yang baru dan berguna (J. Han and M. Kamber,2006).

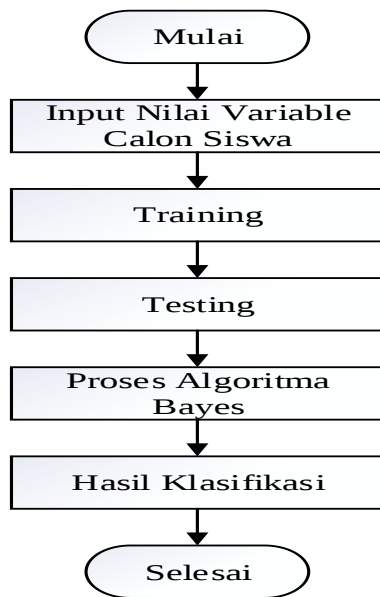
Data mining, sering juga disebut knowledge discovery in database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini biasa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Sehingga istilah pattern recognition sekarang jarang digunakan karena termasuk bagian dari data mining (B. Santosa,2007).

2.2. Klasifikasi

Klasifikasi adalah sebuah proses untuk menemukan sebuah model yang menjelaskan dan membedakan konsep atau kelas data dengan tujuan memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui (J. Han and M. Kamber,2006).

C. METODE PENELITIAN

3.1 Tahap-tahap Kegiatan Penelitian



Gambar 3.1 Langkah-langkah kegiatan penelitian

Dari table di atas yang sudah di ambil dari sekian banyak dataset, maka akan di proses dengan metode naive bayes:

1. Membaca data training
2. Hitung jumlah class/label

Tabel 3.3 Data kemunculan setiap kelas/label

No	atribute	jumlah data	IPA	IPS	P(IPA)	P(IPS)
1	Total	304	165	139	0,542763158	0,457236842
	Nilai IPA					
	High	98	2	96	0,020408163	0,979591837
	Middle	206	137	69	0,665048544	0,334951456
	Low	0	0	0	0	0
2	Nilai IPS					
	High	58	16	42	0,275862069	0,724137931
	Middle	246	123	123	0,5	0,5
	Low	0	0	0	0	0
3	Nilai MTK					
	High	64	21	43	0,328125	0,671875
	Middle	239	117	122	0,489539749	0,510460251
	Low	1	1	0	1	0
4	Psikotest					
	High	54	10	44	0,185185185	0,814814815
	Middle	228	121	107	0,530701754	0,469298246
	Low	22	8	14	0,363636364	0,636363636

Tabel siap di olah

N0	Nama	Nilai Rapot IPA	Nilai Rapot IPS	Nilai Rapot Mtk	Nilai Psikotest	Minat	Kelas
1	Shelfi Hartining	Middle	High	High	High	IPA	IPS
2	Abdul Aziz	High	High	High	Middle	IPA	IPS
3	Abdul Karim	Middle	High	High	Low	IPS	IPS
4	A. Jailani	Middle	Middle	Middle	High	IPA	IPS
5	Jamiatul B	High	High	High	Middle	IPA	IPA
6	Khoirul R	High	High	High	Low	IPA	IPA
7	Imam Nawawi	Middle	Middle	Middle	Middle	IPA	IPS
8	Imam Sholihin	Middle	Middle	Middle	Middle	IPA	IPS
9	Ismail Malik M	Middle	High	High	High	IPA	IPS
10	Rudi	High	Middle	High	Middle	IPS	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 70 sampai di bawah atau sama dengan 84. ($70 \leq X \leq 84$)
3. Di katakan low bila nilai di bawah 70 ($X < 70$)

Tabel 3.4 Data kemunculan pada atribut total

Atribut	Jumlah Kasus (S)	P(IPA)	P(IPS)
Total	304	165	139

Terdapat 2 class dari data training tersebut, yaitu:

- a. $P(IPA) \Rightarrow$ Jurusan = IPA $\Rightarrow$ 165 record
- b. $P(IPS) \Rightarrow$ Jurusan = IPS $\Rightarrow$ 139 record

Total = 304 record

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Maka :

$$P(IPA) = \frac{165}{304} = 0,5427631579$$

$$P(IPS) = \frac{139}{304} = 0,4572368421$$

Tabel 3.5 Klasifikasi perhitungan *naive bayes*

No	Nama	Nilai Rapor IPA			Nilai Rapor IPS			Nilai Rapor Matematika			Nilai Palanet			BATA X			Kelas	K1
		Value	P(IPA)	P(IPS)	Value	P(IPA)	P(IPS)	Value	P(IPA)	P(IPS)	Value	P(IPA)	P(IPS)	Value	P(IPA)	P(IPS)		
1	SHIEL F I	High	0	1	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
2	ARORA A Z	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
3	ADOL K A R	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
4	A J A L A N	High	0	1	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
5	J A H R A T U L B	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
6	K O R D I L A N	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
7	J e m a n M a n a n i	High	0	1	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
8	J e m a n S t r o t h e n	High	0	1	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
9	J e m a n M a n a n i	High	0	1	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
10	R I D I	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
11	T I D I A D I U S	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
12	D e n a R a h a y a	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
13	I n a N u r F	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
14	M u h a m m a d N	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
15	H u s t a m a t i n	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
16	R i z q u n A d i a	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
17	F a t m a A d i	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
18	F a t m a A n e s	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
19	F a t m a A n e s	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	
20	H A L I K A T U L	High	0,84615385	0,153846	High	0,6	0,4	High	0,636363636	0,363636364	High	0,63	0,37	0	0,545454545	0,454545455	IPA	

D. HASIL DAN PEMBAHASAN

4.1 Hasil implementasi

Setelah 304 data sudah di konversi maka dapat formula sebagai berikut :

Untuk (P)Kelas IPA

Untuk (P)Kelas IPS

$$P(\text{IPA}) = 165$$

$$P(\text{IPS}) = 139$$

$$P(\text{IPA} | \text{High}) = 2$$

$$P(\text{IPA} | \text{High}) = 96$$

$$P(\text{IPA} | \text{Mid}) = 137$$

$$P(\text{IPA} | \text{Mid}) = 69$$

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

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

$$P(\text{IPS} | \text{High}) = 16$$

$$P(\text{IPS} | \text{High}) = 42$$

$$P(\text{IPS} | \text{Mid}) = 123$$

$$P(\text{IPS} | \text{Mid}) = 123$$

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

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

$$P(\text{MTK} | \text{High}) = 43$$

$$P(\text{MTK} | \text{Mid}) = 117$$

$$P(\text{MTK} | \text{Mid}) = 122$$

$$P(\text{MTK} | \text{low}) = 1$$

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

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

$$P(\text{Psikotest} | \text{High}) = 44$$

$$P(\text{Psikotest} | \text{Mid}) = 121$$

$$P(\text{Psikotest} | \text{Mid}) = 107$$

$$P(\text{Psikotest} | \text{low}) = 8$$

$$P(\text{Psikotest} | \text{low}) = 14$$

$$P(\text{Minat} | \text{IPA}) = 165$$

$$P(\text{Minat} | \text{IPS}) = 139$$

Sehingga membentuk Probabilitas sebagai berikut :

$$\text{Untuk (P) Kelas IPA} = 165/304 =$$

$$0.5427631579$$

$$\text{Untuk (P) Kelas IPS} = 139/304 =$$

$$0.4572368421$$

Berikut ini adalah gambar kemunculan setiap kelas pada sistem klasifikasi penjurusan.

Berdasarkan hasil pengujian menggunakan metode *naive bayes* didapat akurasi terbaik sebesar 81,26% . Dengan jumlah true positif (tp) sebanyak 175 record, false positif (fp) sebanyak 51 record. Jumlah true negatif (tn) sebanyak 69 record dan false negatif (fn) sebanyak 9 record. Untuk lebih jelasnya akan di jelaskan tabel dan perhitungan di bawah ini :

Tabel 4.1 akurasi perkelompok

uji	testing	training	Akurasi
1	60	244	81,26 %
2	100	204	70,58 %
3	140	164	53,94 %
4	180	124	40,78 %
5	220	84	26,25 %
rata-rata akurasi			54,6 %

Tabel di atas merupakan pembagian kelompok data testing dan data training tiap kelompok mempunyai akurasi masing-masing, dari no 1 sampai dengan 5.

percobaan di atas akan di ambil akurasi yang paling baik untuk lebih rincinya perhitungan perkelompok ada di halaman lampiran 1. Perhitungan mencari akurasi di atas sebagai berikut :

$$\text{Rumus} = \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$1. \text{Accuracy} = \frac{175+69}{175+69+51+9} = \frac{244}{304} 81,26\%$$

$$2. \text{Accuracy} = \frac{153+51}{153+51+98+2} = \frac{240}{304} 70,58\%$$

$$3. \text{Accuracy} = \frac{133+31}{133+31+72+68} = \frac{164}{304} = 53,94\%$$

$$4. \text{Accuracy} = \frac{93+31}{93+31+153+27} = \frac{124}{304} = 40,78\%$$

$$5. \text{Accuracy} = \frac{59+25}{59+25+154+66} = \frac{84}{320} = 26,25\%$$

KESIMPULAN DAN SARAN

5. 1. Kesimpulan

Klasifikasi Penjurusan Siswa Madrasah Aliyah. Berdasarkan hasil pengujian menggunakan metode *naive bayes* didapat akurasi terbaik sebesar 81,26%. Dengan jumlah true positif (tp) sebanyak 175 record, false positif (fp) sebanyak 51 record. Jumlah true negatif (tn) sebanyak 69 record dan false negatif (fn) sebanyak 9 record.

1. Sistem klasifikasi ini cocok di gunakan untuk mengklasifikasikan penjurusan siswa di MA Al-qodiri Jember.
2. Hasil dari penerapan metode *naive bayes* sebanyak 304 data yang di training dan di testing mendapat akurasi terbaik sebesar 81,26%.

5. 2. Saran

Saran yang diperlukan untuk pengembangan program lebih lanjut antara lain :

1. Untuk peneliti berikutnya di harapkan menggunakan metode klasifikasi yang lain untuk di analisis antara 2 metode tersebut .
2. Menggunakan data yang lebih bervariasi seperti menambah nilai bahasa inggris dan bahasa indonesia.

DAFTAR PUSTAKA

Abi Sukrisna, "*Penerapan Kurikulum 2013 dan Penjurusan Siswa*," SMA 1 Kajen, Pekalongan, Interview 2015.

Arif Jananto, "*Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Study Mahasiswa*," 2013.

B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk keperluan Bisnis, 1st ed.* Yogyakarta: Graha Ilmu, 2007.

Claudia Clarentia Ciptohartono, "*Algoritma Klasifikasi Naive Bayes Untuk Menilai Kelayakan Kredit*," 2014.

Faid Ari Prastya, "*Penerapan Algoritma C4.5 Untuk Prediksi Jurusan Siswa SMAN 3 Rembang*," 2015.

J. Han and M. Kamber, *Data Mining : Concept and Techniques Second Edition.* San Francisco: Morgan Kaufmann Publishers, 2006.

K Bambang, E. Syahril, and S.S. Opim, "*Dunia Teknologi Informasi*," *Klasifikasi Konten Berita Dengan Metode Text Mining*, vol. 1, pp. 14-19, 2012.

Kementrian Pendidikan dan Kebudayaan, Salinan – *Permendikbud Nomor 81A*

Tahun 2013 Tentang Implementasi Kurikulum. Jakarta, Indonesia, 2013.

O. KRISTANTO, *"Penerapan Algoritma Klasifikasi Data Mining Id3 Untuk Menentukan Penjurusan Siswa Sman6 Semarang,"* Dian Nuswantoro Fakultas Ilmu Komputer Skripsi, 2014.

Y. S. NUGROHO, *"Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro,"*Dian Nuswantoro Fakultas Ilmu Komputer Skripsi,2014.

Yuda Septian Nugroho, *"Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian,"*2014



