

Klasifikasi Penerima Beasiswa Menggunakan Metode Naïve Bayes (Studi Kasus SMP Negeri 3 Selomerto)

Ahmad Misbachudin Riyadi^{1*}, Hidayatus Sibyan², Iman Ahmad Ihsanuddin³, Nulngafan⁴,
M. Alif Muwafiq Baihaqi⁵

Program Studi Teknik Informatika, Universitas Sains Al-Qur'an, Wonosobo, Indonesia

Email: ^{1*}Ahmadmisbachudin05@gmail.com, ²hsibyan@unsiq.ac.id, ³ahmadihsan@unsiq.ac.id, ⁴affan@unsiq.ac.id,
⁵aviq.baihaqi@gmail.com

Email Coresponding Author: daffadiouf29@gmail.com

Abstrak- SMP 3 Selomerto memberikan bantuan beasiswa untuk kelangsungan belajar siswanya. Pemberian beasiswa dilakukan agar para siswa dapat tetap melanjutkan sekolah atribut yang digunakan yang digunakan terdiri dari jarak, tanggungan, pekerjaan orang tua, pendapatan orang tua, kelengkapan keluarga dan kelayakan. Dikarenakan belum ada metode untuk menentukan penerima beasiswa maka sering salah sasaran dalam memberikan beasiswa. Oleh karena itu diperlukan klasifikasi penerima beasiswa yang tepat dan akurat. Salah satunya data mining dengan metode deskriptif analitis. Bisa dikatakan penelitian deskriptif analitis mengambil masalah atau memperhatikan masalah- masalah yang ada saat penelitian kemudian diolah untuk mendapatkan sebuah kesimpulan. Berdasarkan hasil analisis dan pembahasan klasifikasi penerima beasiswa di SMP N 3 Selomerto dengan menggunakan algoritma Naïve Bayes, dari 186 data siswa yang terdiri dari 150 data training dan 36 data testing diperoleh hasil akurasi yang cukup tinggi sebesar 91,67% baik secara perhitungan manual maupun menggunakan aplikasi Rapidminer. Dapat disimpulkan bahwa Algoritma Naïve Bayes cukup tepat diterapkan untuk membantu pengklasifikasian penerima beasiswa sehingga proses pemberian beasiswa dapat lebih cepat dan akurat.

Kata Kunci: Beasiswa, Klasifikasi, Data mining, Naïve bayes

Abstract- SMP 3 Selomerto provides scholarship assistance to continue their students' studies. Scholarships are provided so that students can continue their education. The attributes used consist of distance, dependents, parents' jobs, parents' income, family completeness and eligibility. Because there is no method for determining scholarship recipients, there is often a wrong target in providing scholarships. Therefore, precise and accurate classification of scholarship recipients is needed. One of them is data mining with analytical descriptive methods. It could be said that analytical descriptive research takes problems or pays attention to the problems that exist during the research and then processes them to reach a conclusion. Based on the results of the analysis and discussion of the classification of scholarship recipients at SMP N 3 Selomerto using the Naïve Bayes algorithm, from 186 student data consisting of 150 training data and 36 testing data, a fairly high accuracy result of 91.67% was obtained both manually and using Rapidminer application. It can be concluded that the Naïve Bayes algorithm is quite appropriate to apply to help classify scholarship recipients so that the scholarship awarding process can be faster and more accurate.

Keywords: Scholarship, Classification, Data mining, Naïve Bayes

1. PENDAHULUAN

Data yang didapat melalui wawancara langsung kepada guru BK terdapat masalah yang dialami pihak sekolah. Salah satu masalahnya adalah pihak sekolah mengalami kesulitan dalam menentukan penerima beasiswa. Hal ini dikarenakan banyaknya permohonan beasiswa tetapi jumlah beasiswa yang terbatas dan Prosedur pengolahan data yang dilakukan meliputi kegiatan pengumpulan data, pengelompokan, pencocokan data dengan data siswa, perkiraan siswa penerima yang rumit membuat guru BK lama dalam mengerjakan dan salah sasaran dalam memberikan beasiswa. Oleh karena itu diperlukan klasifikasi yang tepat dan akurat, salah satunya menggunakan data mining [1].

Pengklasifikasi data siswa ini memiliki peran penting karena dalam penentuan rekomendasi beasiswa banyak hal yang perlu dipertimbangkan seperti dengan parameter yang digunakan yaitu jarak rumah siswa ke sekolah, tanggungan orang tua, gaji orang tua, pekerjaan orang tua dan status kelengkapan keluarga (orang tua lengkap/ yatim) [2]. Maka pihak sekolah memerlukan cara yang efektif dan efisien dalam waktu dan tenaga agar dapat meningkatkan pelayanan [3].

Tujuan yang ingin dicapai dari klasifikasi penerima bantuan beasiswa di SMP Negeri 3 Selomerto yaitu menerapkan Algoritma *Naïve Bayes* untuk membantu proses klasifikasi penerima bantuan beasiswa di SMP Negeri 3 Selomerto. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya[4]

2. KERANGKA TEORI

Persamaan dari teorema *Bayes* adalah[5]:

$$P(H|X) = \frac{P(X|H).P(H)}{p(x)}$$

Di mana:

X: Data dengan class yang belum diketahui

H: Hipotesis data merupakan suatu *class* spesifik

P(H|X): Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H): Probabilitas hipotesis H (*prior* probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)}$$

Di mana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $(C|F_1 \dots F_n)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= (C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

3. METODE PENELITIAN

Jenis penelitian yang dipakai adalah deskriptif analitis. Penelitian deskriptif analitis merupakan suatu metode yang berguna untuk mendeskripsikan atau memberikan gambaran pada objek yang diteliti melalui data atau sampel yang sudah dikumpulkan sebagaimana adanya tanpa melakukan analisis dan membuat kesimpulan yang berlaku untuk umum. Bisa dikatakan penelitian deskriptif analitis mengambil masalah atau memutuskan perhatian kepada masalah-masalah yang ada saat penelitian dilakukan dan kemudian diolah untuk mendapatkan kesimpulan [6].

Pengumpulan data yang digunakan dalam penelitian pada SMP Negeri 3 Selomerto ini adalah sebagai berikut : [7]

a. *Studi Pustaka/ Literatur*

Penulis melakukan pencarian, pembelajaran dari berbagai macam dokumen yang menunjang pengerjaan Tugas Akhir ini khususnya materi yang berkaitan dengan penerima beasiswa, beserta cara penentuan penerima beasiswa menggunakan metode *Naive Bayes* kedalam klasifikasi penerima beasiswa.

b. *Observasi*

Pengamatan dan penginderaan dengan mendatangi langsung ke lokasi penelitian yaitu di SMP N 3 Selomerto.

c. *Wawancara (Interview)*

Melakukan tatap muka dan tanya jawab langsung dengan *staff* guru SMP N 3 Selomerto terkait data yang akan digunakan sebagai bahan penelitian.

d. *Dokumentasi*

Melakukan dokumentasi terhadap data yang akan digunakan untuk objek penelitian di SMP N 3 Selomerto.

Dalam Penelitian ini penulis menggunakan 2 (dua) jenis data yaitu data *Training* dan data *Testing*. Data *training* adalah himpunan data yang digunakan untuk melatih atau membangun model. Kemudian data testing adalah data yang akan dipakai untuk menguji dan mengetahui performa model yang didapatkan pada tahap *testing*. Data *training* yang digunakan pada penelitian ini sejumlah 150 data siswa yang terdiri dari siswa penerima beasiswa pada tahun 2020-2021 dan data testing yang digunakan sejumlah 36 yang didapat dari penerima tahun 2022.

4. HASIL

Pengolahan data mining yang dilakukan dalam penelitian ini adalah mengikuti tahap dalam *Knowledge Discovery In Database (KDD)*, untuk menghasilkan informasi dengan urutan yang sudah ditentukan [8].

4.1 Data Selection

Pada proses ini adalah proses pemilihan data dari data-data yang harus dilakukan sebelum menuju tahap penggalian informasi dalam KDD dimulai.

Semua atribut yang ada pada data seleksi beasiswa diantaranya Nama, Jenis Kelamin, Jarak rumah ke sekolah (J), Penghasilan orang tua (PEN), Tanggungan orang tua (T), kelengkapan keluarga (Yatim/ Piatu) (Y/P), Pekerjaan orang tua (PE) dan Kelayakan (KL).

4.2 Transaformation

Tahap ini merupakan tahap merubah data yang telah dipilih, sehingga data tersebut sesuai dalam proses data mining. Proses transformasi dalam KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

Tabel 1. Tabel Pembagian Interval Jarak

Jarak	Kode
< 1 km	1
1-2 km	2
3-4 km	3
> 5 km	4

Tabel 2. Transformasi Penghasilan Orang Tua

Penghasilan	Kode
< Rp 1.000.000	1
Rp 1.001.000 – Rp 1.500.000	2
Rp 1.501.000 – Rp 2.000.000	3
Rp 2.001.000 – Rp 3.000.000	4
Rp 3.000.000	5

Tabel 3. Transformasi Pekerjaan Orang Tua

Pekerjaan	Kode
Buruh	1
Petani	2
Pedagang	3
Wiraswasta	4

Tabel 4. Transformasi Pekerjaan Orang Tua

Kelengkapan keluarga	Kode
Lengkap	1
Piatu	2
Yatim	3

Tabel 5. Data Hasil Transformasi

NO	J	PEN	T	Y/P	PE	KL
1	2	2	2	1	1	L
2	2	3	3	1	3	TL
3	2	3	1	1	4	TL
4	3	3	2	1	2	TL
5	2	1	1	3	1	L
6	2	2	2	1	2	TL
7	1	2	1	1	2	TL
8	2	2	3	1	2	TL
9	2	1	3	1	1	L
10	2	3	2	1	3	TL
11	4	2	2	1	4	TL
12	1	3	1	1	4	TL

13	3	2	3	1	1	L
14	2	2	2	1	4	TL
15	2	2	2	1	2	TL
16	2	2	3	1	1	L
17	3	2	1	1	2	TL
18	2	2	1	1	1	L
19	2	2	2	1	2	TL
20	3	5	2	3	1	L
21	2	1	2	1	1	L
...
150	2	2	3	1	4	L

4.3 Perhitungan Naïve Bayes

a. Menghitung class probabilities

Langkah selanjutnya menghitung class probabilities. Dataset Kelayakan memiliki 2 kelas masalah. Sehingga dapat diketahui probabilitas (P) masing masing kelas dengan cara membagi nilai frekuensi atau jumlah data untuk masing masing kelas masalah, dibagi dengan jumlah total nilai atau frekuensi data pada kelas tersebut. Seperti pada perhitungan berikut:

$$P_{layak} = \frac{jumlah_{layak}}{jumlah_{layak} + jumlah_{tidak layak}}$$

$$P_{layak} = \frac{71}{71 + 79}$$

$$P_{layak} = \mathbf{0,4733}$$

$$P_{tidak layak} = \frac{jumlah_{tidak layak}}{jumlah_{layak} + jumlah_{tidak layak}}$$

$$P_{tidak layak} = \frac{79}{71 + 79}$$

$$P_{tidak layak} = \mathbf{0,5267}$$

Class probabilities ini adalah nilai peluang atau probabilitas masing – masing kelas masalah dari tujuan yang diharapkan atau diprediksikan. Sehingga dari perhitungan data training yang ada, diperoleh nilai peluang untuk kelayakan = layak adalah 0,4733 dan kelayakan = tidak layak adalah 0,5267.

Dari perhitungan di atas bisa dirangkum sebagai berikut :

Tabel 6. Probabilitas Kelas

Kelayakan		probabilitas	
Layak	Tidak layak	Layak	Tidak layak
71	79	0,4733	0,5267

b. Menghitung conditional probabilities

Langkah selanjutnya adalah menghitung conditional probabilities, yaitu probabilitas setiap nilai input terhadap nilai class kelayakan.

Tabel 7. Probabilitas Jarak

Jarak	Kelayakan		Probabilitas	
	Layak	Tidak layak	Layak	Tidak layak
1	17	27	0,2394	0,3418
2	38	45	0,5352	0,5696
3	14	6	0,1972	0,0759
4	2	1	0,0282	0,0127
Jumlah	71	79	1	1

Tabel 8. Probabilitas Tanggungan

T	Kelayakan		Probabilitas	
	Layak	Tidak layak	Layak	Tidak layak
1	5	21	0,0704	0,2658
2	26	35	0,3662	0,4430
3	19	20	0,2676	0,2532
4	21	3	0,2958	0,0380
Jumlah	71	79	1	1

Tabel 9. Probabilitas Pekerjaan

Pekerjaan	Kelayakan		Probabilitas	
	Layak	Tidak layak	Layak	Tidak layak
1	47	13	0,6619	0,1645
2	12	27	0,1690	0,3418
3	3	8	0,0423	0,1013
4	9	31	0,1268	0,3924
Jumlah	71	79	1	1

Tabel 10. Probabilitas Pendapatan

Pendapatan	Kelayakan		Probabilitas	
	Layak	Tidak layak	Layak	Tidak layak
1	31	1	0,4211	0,0127
2	39	37	0,5493	0,4684
3	1	25	0,0263	0,3164
4	0	12	0,0132	0,1519
5	0	4	0,0132	0,0506
Jumlah	71	79	1	1

Tabel 11. Probabilitas Kelengkapan Keluarga

Kelengkapan keluarga	Kelayakan		Probabilitas	
	Layak	Tidak layak	Layak	Tidak layak
1	62	78	0,8732	0,9634
2	2	1	0,0282	0,0244
3	7	0	0,0986	0,0122
Jumlah	71	79	1	1

c. Pengujian *Naïve Bayes*

Setelah menghitung probabilitas setiap atribut selanjutnya dilakukan pengujian pada data testing yang berjumlah 36 data

Tabel 12. Tabel Data *Testing*

NO	J	PEN	T	Y/P	PE	KL
1	3	3	1	1	1	TL
2	3	3	3	1	1	TL
3	3	3	3	1	3	TL
4	3	1	3	1	1	L
5	3	2	2	1	1	L
6	1	2	1	1	1	TL
...
36	1	3	1	1	4	TL

Untuk mengujinya kita masukkan data *training* ke rumusan *Naïve bayes* dimana kriteria diambil dari data kemudian kita hitung setiap kriteria tersebut dengan probabilitas sebelumnya. Berikut rumus yang digunakan:

$$P(LAYAK | X) = (p_{jarak(layak)} * p_{pendapatan(layak)} * p_{tanggungan(layak)} * p_{KL(layak)} * p_{pekerjaan(layak)}) * P(kelayakan=layak)$$

$$P(TIDAK LAYAK | X) = (p_{jarak(tidak layak)} * p_{pendapatan(tidak layak)} * p_{tanggungan(tidak layak)} * p_{KL(tidak layak)} * p_{pekerjaan(tidak layak)}) * P(kelayakan=tidak layak)$$

Berikut pengujian datanya:

Jarak = 3, pendapatan = 3, tanggungan = 1, kelengkapan keluarga = 1, dan pekerjaan = 1

$$P(LAYAK | X) = (0,1973 * 0,0268 * 0,0704 * 0,8732 * 0,6619) * 0,4733$$

$$P(LAYAK | X) = \mathbf{0,000099879}$$

$$P(tidak layak | X) = (0,0759 * 0,3164 * 0,2658 * 0,9873 * 0,1646) * 0,5267$$

$$P(tidak layak | X) = \mathbf{0,00054636}$$

Dari perhitungan tersebut diperoleh nilai probabilitas $P(TIDAK LAYAK | X)$ adalah tertinggi yaitu sebesar **0,00054636** sehingga prediksinya adalah kelayakan **TIDAK LAYAK**. Berikut adalah tabel hasil *confusion matrix* dari hasil testing 36 data di atas.

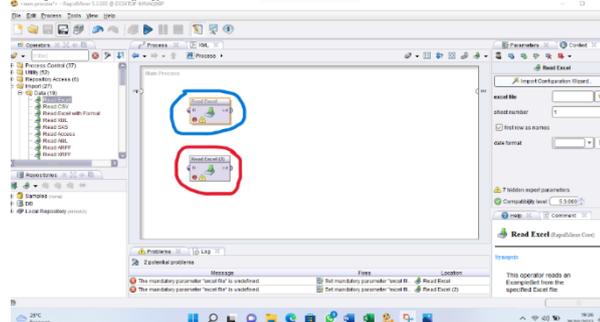
Tabel 13. Confusion Matrix

f_{ij}		Kelas Hasil Prediksi (j)	
		Layak	Tidak Layak
Kelas asli (i)	Layak	11	2
	Tidak layak	1	22

Berdasarkan perhitungan manual di atas, dimana dari 36 data testing yang diuj terdapat 33 data kelayakan yang berhasil diklasifikasi dengan benar dan sebanyak 3 data yang diklasifikasi dengan salah. Maka presentase untuk *correct classified instance* adalah 91,6667% dan untuk *incorrect classifield instance* 8,3333%

Pengujian menggunakan aplikasi Rapidminer

Pengujian di *software Rapidminer*, sebelum data *training* dan data *testing* disimpan dalam bentuk *.exe* atau *excel* agar bisa di import ke aplikasi *Rapidminer*. Kemudian klik menu "Import data" dan pilih yang *read excel* yang sesuai dengan format data pada penelitian, dan masukkan 2 *read excel*. *Read excel* (biru) untuk data *Training* dan yang *read excel* 2 (merah) untuk data *Testing*. Klik *inport configuration wizard* untuk memasukkan data *Training* dan *Testing*.



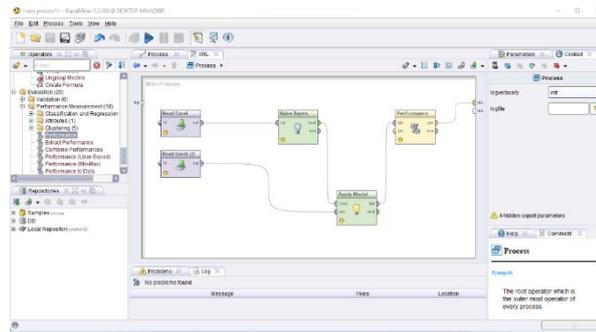
Gambar 1. Import Data Training dan Testing (sumber: Penulis)

Masukkan model *Naïve Bayes*, *Apply model* dan *Performance* dalam proses dan hubungkan seperti pada gambar berikut

Multiclass Classification Performance			
accuracy: 91.67%			
	true TIDAK LAYAK	true LAYAK	class precision
pred. TIDAK LAYAK	22	2	91.67%
pred. LAYAK	1	11	91.67%
class recall	95.65%	84.82%	

Gambar 2. Proses Pengolahan Data Pada Rapidminer.

Klik tombol *Run* untuk pengujian dan akan terlihat hasil dari pengujiannya.



Gambar 3. Hasil Pengujian *Rapidminer*

Pada hasil pengujian dengan aplikasi *Rapidminer* diperoleh hasil *Accuracy* sebesar 91,67%. Hasil tersebut sama dengan perhitungan manual sebelumnya.

5. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan klasifikasi penerima beasiswa di SMP N 3 Selomerto dengan menggunakan algoritma *Naïve Bayes*, dari 186 data siswa yang terdiri dari 150 data *training* dan 36 data *testing* diperoleh hasil akurasi yang cukup tinggi sebesar 91,67% baik secara perhitungan manual maupun menggunakan aplikasi *Rapidminer*. Dapat disimpulkan bahwa Algoritma *Naïve Bayes* cukup tepat diterapkan untuk membantu pengklasifikasian penerima beasiswa sehingga proses pemberian beasiswa dapat lebih cepat dan akurat.

DAFTAR PUSTAKA

- [1] Reksoatmojo, W. (2018). *Analisis dan perancangan sistem basis data*. Yogyakarta: ANDI.
- [2] Prasetyo. (2012). *Data Mining Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: ANDI Yogyakarta.
- [3] Ridwan, M. S. (2013). *Penerapan data mining untuk evaluasi kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier*. Jurnal EECIS Vol. 7, No. 1 hal. 59-64, juni 2013.
- [4] Bustami. (2013). *Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*, *TECHSI : Jurnal Penelitian Teknik Informatika*, Vol. 3, hal. 127-146.
- [5] Patil, Tina , S. S. Sherekar, (2013). *Performance Analysis of Naive Bayes and j48 Classification Algoritm for Data Classification*, *International journal of computer science and applications*, Vol. 6, hal 256-261 No. 2 April 2013.
- [6] Susanto. (2002). *pengertian data menurut ahli*. Retrieved from creator media: <https://creatormedia.my.id/pengertian-data-menurut-para-ahli-jurnal-terbaru/>
- [7] Amirullah, Widayat. (2002). *metode penelitian pemasaran*. Malang : CV. Cahaya Press
- [8] Larose, D. T. (2005). *Discovering Knowledge in data: An Introduction to Data Mining*. Retrieved from A Jhon Willey & Sons, Inc., Publication