

PENGAPLIKASIAN DATA MINING PENENTUAN PENERIMAAN BANTUAN SOSIAL MASYARAKAT DENGAN PROSEDUR ALGORITMA NAIVE BAYES

by Nurfalah Nurfalah

Submission date: 02-Oct-2024 11:44AM (UTC+0700)

Submission ID: 2472382927

File name: TeknikInformatika_DataMining_Nurfalah_1.pdf (1.09M)

Word count: 3933

Character count: 22187

PENGAPLIKASIAN DATA MINING PENENTUAN PENERIMAAN BANTUAN SOSIAL MASYARAKAT DENGAN PROSEDUR ALGORITMA NAIVE BAYES

Nurfalah*, Rouli Doharma MS²

^{1,2} STMIK Widuri

² *nurfalah23412021@kampuswiduri.ac.id¹, rouli.doharma@kampuswiduri.ac.id²

Alamat: Jl. Palmerah Barat No.353, RT.3/RW.5, Grogol Utara, Kby. Lama, Kota Jakarta Selatan, DKI Jakarta 11480.

Korespondensi penulis: nurfalah23412021@kampuswiduri.ac.id

Abstract. Social assistance is an important aspect of government and non-government programs that can help on a large scale for the community so that the impact is to lighten life in the short term, but social assistance has several criteria such as income, social conditions, family status and the impact of the economic situation. . Knowing the criteria for social assistance is done by applying data mining to social assistance using the Naive Bayes algorithm procedure which produces accuracy calculations from 100 testing data, obtained good values, namely accuracy of 95.00%, precision of 92.31%, and recall of 97.95%.

Keywords Social Assistance, Data Mining, Naive Bayes

Abstrak. Bantuan sosial merupakan aspek penting dalam program pemerintah ataupun non pemerintah yang dapat membantu dalam jumlah skala besar bagi masyarakat sehingga dampak yang ditimbulkan meringankan kehidupan jangka waktu pendek, namun bantuan sosial mempunyai beberapa kriteria seperti, pendapatan, kondisi sosial, status keluarga dan dampak dari situasi ekonomi. Mengetahui kriteria dari bantuan sosial dilakukan dengan cara pengaplikasian data mining pada bantuan sosial menggunakan prosedur algoritma naive bayes yang menghasilkan perhitungan akurasi dari 100 data testing, diperoleh nilai yang baik yaitu akurasi sebesar 95,00%, presisi sebesar 92,31%, dan recall sebesar 97,95%..

Kata kunci: Bantuan Sosial, Data Mining, Naive Bayes

1. LATAR BELAKANG

Bantuan sosial dimasyarakat mencakup program atau kegiatan yang diselenggarakan oleh pemerintah atau lembaga non-pemerintah untuk membantu individu atau kelompok yang membutuhkan. Kemudian bantuan sosial sering kali ditujukan kepada keluarga miskin, lansia, anak-anak, penyandang disabilitas, atau korban bencana alam. selain itu diimplementasikan melalui berbagai mekanisme seperti subsidi, beasiswa, atau layanan komunitas (Alfiah, 2021; Asri et al., 2022; Fitriani, 2020; Purnama et al., 2020).

Kriteria penerimaan bantuan sosial meliputi evaluasi berbagai aspek seperti pendapatan, kondisi sosial, status keluarga dan dampak dari situasi ekonomi, menurut Peraturan Gubernur No. 49 tentang Jaring Pengaman Sosial, kriteria penerima bantuan pada individu yang memiliki pendapatan rendah dalam Keluarga, Pekerja yang

menghadapi pemutusan hubungan kerja, dan pedagang yang mengalami penurunan pendapatan serta orang tua yang tidak lagi bekerja seperti lansia dan janda (Damuri et al., 2021; Nurriqfi Fakhri Fikrillah et al., 2023; Oktha Pratiwi et al., 2022; Putri et al., 2021).

Pelaksanaan program pemberian bantuan sosial yang diinisiasi oleh pemerintah provinsi DKI Jakarta menghadapi tantangan seperti distribusi yang tidak merata maupun tidak sesuai dengan sasaran yang seharusnya menerima sehingga perlu adanya evaluasi yang menjangkau seluruh aspek elemen dilingkungan masyarakat.

Adapun yang menjadi tujuan dengan memproses penerapan prosedur algoritma *naive bayes* dalam kelayakan penerimaan bantuan sosial bagi masyarakat menggunakan pengaplikasian *data mining*, algoritma *naive bayes* ialah algoritma klasifikasi yang menggunakan teorema Bayes dengan asumsi independensi fitur. menghitung probabilitas kelas untuk setiap instans dari probabilitas fitur yang teramati (Alfitri et al., 2023; Azlil Huriah & Dienwati Nuris, 2023; Hayati, 2023; Sundari et al., 2024).

2. KAJIAN TEORITIS

Data mining merupakan proses mengekstraksi informasi atau pola-pola berharga dari kumpulan data yang sangat besar menggunakan teknik statistik, matematika, dan kecerdasan buatan. Tujuan utama dari data mining adalah menemukan hubungan tersembunyi, tren, dan pola yang tidak langsung terlihat oleh pengamatan manual. Proses ini melibatkan beberapa tahapan seperti pengumpulan data, pembersihan data, transformasi data, serta analisis menggunakan algoritma tertentu. Metode yang umum digunakan dalam data mining termasuk klasifikasi, klusterisasi, regresi, dan asosiasi. Aplikasi data mining sangat luas, mencakup berbagai bidang seperti bisnis, kesehatan, keuangan, dan pemasaran. Dalam dunia bisnis, data mining membantu perusahaan memahami perilaku konsumen, memperkirakan tren pasar, dan meningkatkan pengambilan keputusan berdasarkan data (Asri et al., 2022; Putri et al., 2021).

Konsep penelitian terhadap bantuan sosial menggunakan prosedur algoritma *naive bayes* sebagai klasifikasi karena kesederhanaan dan efisiensinya dalam memproses data berukuran besar. Algoritma ini bekerja dengan menghitung probabilitas posterior dari setiap kelas berdasarkan fitur yang ada, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi. Meskipun asumsi independensi antar fitur jarang

sepenuhnya terpenuhi di dunia nyata, Naive Bayes tetap menunjukkan kinerja yang baik dalam berbagai aplikasi, seperti pengenalan teks, deteksi spam, dan diagnosis penyakit (Damuri et al., 2021; Oktha Pratiwi et al., 2022)..

3. METODE PENELITIAN

3.1 Penentuan Knowledge Discovery in Database (KDD)

Pada tahapan penelitian ini terdapat data mentah warga untuk sampel menerima bantuan sosial dari 500 orang sebagai contoh dan nilai keterangan atribut dalam proses knowledge discovery in database (KDD), data mentah maupun nilai keterangan atribut akan diproses lebih lanjut untuk tahapan – tahapan algoritma naive bayes untuk data mining. berikut data mentah warga yang dapat dijabarkan sesuai dengan klasifikasi knowledge discovery in database dibawah ini

N o	NAMA KRT	Jenis Kelamin	PROV INSI	R W	R T	Namases uaiNIK	Pekerj aan	Jumla h Tangg ungan	Penghasil an	Status Rumah	Keterangan
1	Hadi Saputra Rahman	Laki-Laki	DKI Jakarta	3	3	Hadi Saputra Rahman	Karyaw an	2	3.100.000	Kontrak	Layak
2	Maya Ayu Safitri	Peremp uan	DKI Jakarta	3	3	Maya Ayu Safitri	Karyaw an	1	4.200.000	Rumah Sendiri	Tidak Layak
3	Dian Pratama Hidayat	Peremp uan	DKI Jakarta	3	3	Dian Pratama Hidayat	Karyaw an	1	2.500.000	Rumah Sendiri	Tidak Layak
4	Rina Wahyu Utami	Peremp uan	DKI Jakarta	3	3	Rina Wahyu Utami	Karyaw an	2	1.500.000	Rumah Sendiri	Tidak Layak
5	Ryan Budi Pramono	Laki-Laki	DKI Jakarta	3	3	Ryan Budi Pramono	Karyaw an	3	2.500.000	Kontrak	Tidak Layak

Table 1 Keterangan Nilai Atribut

Kode	Nama Atribut	Nilai	Keterangan
A1	nama	Nama Penerima	Dijadikan Sebagai ID
A2	Jenis Kelamin	L	Laki-Laki
	Jenis Kelamin	P	Perempuan
A3	Pekerjaan	Karyawan	
	Pekerjaan	Ibu Rumah Tangga	
	Pekerjaan	Pedagang	
	Pekerjaan	Pekerja Lepas	

Kode	Nama Atribut	Nilai	Keterangan
A4	Penghasilan	K1	1.000.000 - 2.500.000
	Penghasilan	K2	2.500.000 - 4.000.000
	Penghasilan	K3	4.000.000 - UMR DKI Jakarta
	Penghasilan	Tidak Ada	Tidak Ada
A5	Status Rumah	Kontrak	
	Status Rumah	Rumah Sendiri	
A6	Tanggungan	1	Tanggungan, dimaksud jumlah anak dalam satu keluarga
	Tanggungan	2	
	Tanggungan	3	
A7	RT	3	Jumlah terdiri dari 9 RT yaitu RT 03-011
	RT	4	
	RT	5	
	RT	6	
	RT	7	
	RT	8	
	RT	9	
	RT	10	
	RT	11	

17

4. HASIL DAN PEMBAHASAN

4.1 Data Transformasi dan Cleaning

Data sampel yang sudah ditentukan maka tahapan selanjutnya transformasi sekaligus *cleaning* di *knowledge discovery in database* dari *microsoft excel*. Berikut ini transformasi serta *cleaning* yang dapat dilakukan antara lain:

Table 2 Contoh Data Transformasi dan *Cleaning*

No	Nama	Jenis Kelamin	Pekerjaan	Penghasilan	Status Rumah	Jumlah Tanggungan	RT	Keterangan
1	Hadi Saputra Rahman	L	Karyawan	K2	Kontrak	2	3	Layak
2	Maya Ayu Safitri	P	Karyawan	K3	Rumah Sendiri	1	3	Tidak Layak
3	Rina Wahyu Utami	P	Karyawan	K3	Rumah Sendiri	2	3	Tidak Layak
4	Ryan Budi Pramono	L	Karyawan	K3	Kontrak	3	3	Tidak Layak
5	Wulan Ayu Sari Utami	P	Ibu Rumah Tangga	Tidak Ada	Kontrak	3	3	Layak

Setelah mengetahui data transformasi dan *cleaning* yang sudah dilakukan, data ini dilakukan *import data set* melalui aplikasi *rapid miner* yang dapat dilihat pada gambar dibawah ini.

Select the cells to import.

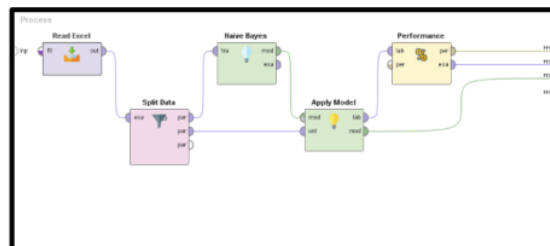
Sheet: Data Cleaning Cell range: A:H Select All Define head: 1

	A	B	C	D	E	F	G	H
1	Nama	Jenis Kela...	Pekerjaan	Penghasilan	Status Rum...	Jumlah Ta...	RT	Keterangan
2	Hadi Saput...	L	Karyawan	K2	Kontrak	2.000	3.000	Layak
3	Maya Ayu S...	P	Karyawan	K3	Rumah Sen...	1.000	3.000	Tidak Layak
4	Dian Prata...	P	Karyawan	K3	Rumah Sen...	1.000	3.000	Tidak Layak
5	Rina Wahy...	P	Karyawan	K3	Rumah Sen...	2.000	3.000	Tidak Layak
6	Ryan Budi...	L	Karyawan	K3	Kontrak	3.000	3.000	Tidak Layak
7	Rima Ayu D...	P	Ibu Rumah ...	Tidak Ada	Kontrak	1.000	3.000	Layak
8	Wulan Ayu ...	P	Ibu Rumah ...	Tidak Ada	Kontrak	3.000	3.000	Layak
9	Ratna Purn...	P	Ibu Rumah ...	Tidak Ada	Kontrak	3.000	3.000	Layak
10	Made Suda...	L	Pedagang	K2	Kontrak	2.000	3.000	Layak

Gambar 1 Import Data Set Aplikasi Rapid Miner

4.2 Proses Operator dan Split Data Rapid Miner

Tahapan proses operator dan *split data* melalui *rapid miner*, dilakukan awal tahapan operator terlebih dahulu, berikut operator dalam *rapid miner* antara lain:



Gambar 2 Operator Rapid Miner

Sedangkan split data dapat diperoleh dari operator pada *rapid miner*, oleh karena itu *split data* tersebut diproses dalam gambar sebagai hasil split data.

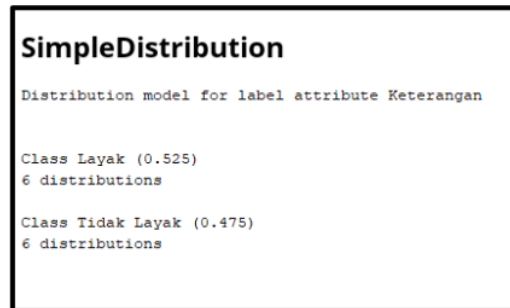
ExampleSet (Split Data)	Data Training
Filter (400 / 400 examples):	
ExampleSet (Split Data)	Data Testing
Filter (100 / 100 examples):	

Gambar 3 Split Data Rapid Miner

Gambar *split data* diatas dapat dijelaskan sebagai berikut, untuk data training dengan *rasio* 0.8 atau sebanyak 400 data dan untuk data *testing* dengan *rasio* 0.2 atau sebanyak 100 data.

4.3 Hasil Prediksi Metode Naive Bayes

Setelah tahapan operator dan *split* data, memperoleh hasil prediksi menggunakan metode *naive bayes* pada pengaplikasian *rapid miner*, berikut hasil prediksi metode *naive bayes* antara lain:



Gambar 4 Hasil Prediksi Metode Naive Bayes

Ilustrasi di atas, dapat dijelaskan bahwa probabilitas kelas “Layak” memiliki skor 0.525 atau setara dengan 52,5%, sedangkan probabilitas kelas “Tidak Layak” memiliki skor 0.475 atau setara dengan 47,5%. Informasi ini berasal dari data yang diperoleh dari label “Keterangan”.

4.4 Perhitungan Manual Metode Naive Bayes

Berdasarkan hasil prediksi metode *naive bayes* yang dilakukan menggunakan *rapid miner* sebelumnya, dapat diperhitungkan manual metode *naive bayes* antara lain:

Table 3 Data Jumlah Keterangan Kriteria

	Keterangan	Jumlah Kasus	Hasil (Layak)	Hasil (Tidak Layak)
TOTAL		500	262	238
Jenis Kelamin	L	384	184	200
	P	116	78	38
Pekerjaan	Karyawan	276	70	206
	Ibu Rumah Tangga	49	48	1
	Pedagang	84	75	9
	Pekerja Lepas	91	69	22
Penghasilan	K1	130	130	0
	K2	142	84	58
	K3	180	0	180
	Tidak Ada	48	48	0
Status Rumah	Kontrak	446	258	188
	Rumah Sendiri	54	4	50

	Keterangan	Jumlah Kasus	Hasil (Layak)	Hasil (Tidak Layak)
Tanggungan	1	154	56	98
	2	165	95	70
	3	181	111	70
RT	3	30	14	16
	4	38	21	17
	5	53	41	12
	6	50	24	26
	7	65	27	38
	8	68	30	38
	9	78	52	26
	10	56	32	24
	11	62	21	41

Langkah selanjutnya adalah melakukan perhitungan dari 500 dataset dengan menggunakan rumus probabilitas metode *Naïve Bayes* yang diberikan di bawah ini:

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

Menghitung jumlah kelas berdasarkan klasifikasi yang sudah terbentuk pada label Keterangan :

1. $P(\text{Class Keterangan} = \text{"Layak"} \mid \text{Total Keterangan}) = \frac{262}{500} = 0,524$
2. $P(\text{Class Keterangan} = \text{"Tidak Layak"} \mid \text{Keterangan}) = \frac{238}{500} = 0,476$

Perhitungan pada atribut Jenis Kelamin terdapat 2 nilai yaitu "L" untuk Laki-Laki dan "P" untuk Perempuan, berdasarkan jumlah data yang ada pada label Keterangan Layak atau Tidak Layak. Dibawah ini perhitungan probabilitas pada atribut Jenis Kelamin dengan nilai "L", untuk perempuan sebagai berikut:

1. $P(\text{Jenis Kelamin} = \text{"L"} \mid \text{Class Keterangan} = \text{"Layak"})$
 $= \frac{184}{262} = 0,702290$
2. $P(\text{Jenis Kelamin} = \text{"L"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$
 $= \frac{200}{238} = 0,840336$

Perhitungan probabilitas pada atribut Jenis Kelamin dengan nilai "P", untuk Laki-Laki sebagai berikut:

3. $P(\text{Jenis Kelamin} = \text{"P"} \mid \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{78}{262} = 0,297709$$

$$4. P(\text{Jenis Kelamin} = "P" \mid \text{Class Keterangan} = "Tidak Layak")$$

$$= \frac{38}{238} = 0,159663$$

Perhitungan pada atribut Pekerjaan terdapat 4 class yaitu "Karyawan", "Ibu Rumah Tangga", "Pedagang" dan "Pekerja Lepas", berdasarkan jumlah data yang ada pada label Keterangan Layak atau Tidak Layak. Dibawah ini perhitungan probabilitas pada atribut Pekerjaan dengan nilai "Karyawan", sebagai berikut:

$$1. P(\text{Pekerjaan} = "Karyawan" \mid \text{Class Keterangan} = "Layak")$$

$$= \frac{70}{262} = 0,267175$$

$$2. P(\text{Pekerjaan} = "Karyawan" \mid \text{Class Keterangan} = "Tidak Layak")$$

$$= \frac{206}{238} = 0,865546$$

Perhitungan probabilitas pada atribut Pekerjaan dengan nilai "Ibu Rumah Tangga", sebagai berikut:

$$3. P(\text{Pekerjaan} = "Ibu Rumah Tangga" \mid \text{Class Keterangan} = "Layak")$$

$$= \frac{48}{262} = 0,183206$$

$$4. P(\text{Pekerjaan} = "Ibu Rumah Tangga" \mid \text{Class Keterangan} = "Tidak Layak")$$

$$= \frac{1}{238} = 0,004201$$

Perhitungan probabilitas pada atribut Pekerjaan dengan nilai "Pedagang", sebagai berikut:

$$5. P(\text{Pekerjaan} = "Pedagang" \mid \text{Class Keterangan} = "Layak")$$

$$= \frac{75}{262} = 0,286259$$

$$6. P(\text{Pekerjaan} = "Pedagang" \mid \text{Class Keterangan} = "Tidak Layak")$$

$$= \frac{9}{238} = 0,037851$$

Perhitungan probabilitas pada atribut Pekerjaan dengan nilai "Pekerja Lepas", sebagai berikut:

$$7. P(\text{Pekerjaan} = "Pekerja Lepas" \mid \text{Class Keterangan} = "Layak")$$

$$= \frac{69}{262} = 0,263358$$

$$8. P(\text{Pekerjaan} = "Pekerja Lepas" \mid \text{Class Keterangan} = "Tidak Layak")$$

$$= \frac{22}{238} = 0,092436$$

Perhitungan pada atribut Penghasilan terdapat 4 nilai yang dikategorikan sebagai berikut “K1”, ”K2”, ”K3” dan “Tidak Ada”, berdasarkan jumlah data yang ada pada label Keterangan Layak atau Tidak Layak. Dibawah ini perhitungan probabilitas pada atribut Penghasilan dengan nilai “K1”, sebagai berikut:

$$1. P(\text{Penghasilan} = \text{"K1"} \mid \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{130}{262} = 0,496183$$

$$2. P(\text{Penghasilan} = \text{"K1"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{0}{238} = 0$$

Perhitungan probabilitas pada atribut Penghasilan dengan nilai “K2”, sebagai berikut:

$$3. P(\text{Penghasilan} = \text{"K2"} \mid \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{84}{262} = 0,320610$$

$$P(\text{Penghasilan} = \text{"K2"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{58}{238} = 0,243697$$

Perhitungan probabilitas pada atribut Penghasilan dengan nilai “K3”, sebagai berikut:

$$4. P(\text{Penghasilan} = \text{"K3"} \mid \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{0}{262} = 0$$

$$5. P(\text{Penghasilan} = \text{"K3"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{180}{238} = 0,756302$$

Perhitungan probabilitas pada atribut Penghasilan dengan nilai “Tidak Ada”, sebagai berikut:

$$6. P(\text{Penghasilan} = \text{"Tidak Ada"} \mid \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{48}{262} = 0,183206$$

$$7. P(\text{Penghasilan} = \text{"Tidak Ada"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{0}{238} = 0$$

Perhitungan pada atribut Status Rumah terdapat 2 nilai yaitu “Kontrak” dan “Rumah Sendiri”, berdasarkan jumlah data yang ada pada label Keterangan Layak atau Tidak Layak. Dibawah ini perhitungan probabilitas pada atribut Status Rumah dengan nilai “Kontrak”, sebagai berikut:

$$1. P(\text{Status Rumah} = \text{"Kontrak"} \mid \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{258}{262} = 0,984732$$

2. $P(\text{Status Rumah} = \text{"Kontrak"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{188}{238} = 0,789915$$

Perhitungan probabilitas pada atribut Status Rumah dengan nilai "Rumah Sendiri", sebagai berikut:

3. $P(\text{Status Rumah} = \text{"Rumah Sendiri"} \mid \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{4}{262} = 0,015267$$

4. $P(\text{Status Rumah} = \text{"Rumah Sendiri"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{50}{238} = 0,210084$$

Perhitungan pada atribut Jumlah Tanggungan terdapat 3 nilai yaitu "1", "2" dan "3", berdasarkan jumlah data yang ada pada label Keterangan Layak atau Tidak Layak. Dibawah ini perhitungan probabilitas pada atribut Jumlah Tanggungan dengan nilai "1", sebagai berikut:

1. $P(\text{Jumlah Tanggungan} = \text{"1"} \mid \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{56}{262} = 0,213740$$

2. $P(\text{Jumlah Tanggungan} = \text{"1"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{98}{238} = 0,411764$$

Perhitungan probabilitas pada atribut Jumlah Tanggungan dengan nilai "2", sebagai berikut:

3. $P(\text{Jumlah Tanggungan} = \text{"2"} \mid \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{95}{262} = 0,362595$$

4. $P(\text{Jumlah Tanggungan} = \text{"2"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{70}{238} = 0,294117$$

Perhitungan probabilitas pada atribut Jumlah Tanggungan dengan nilai "3", sebagai berikut:

5. $P(\text{Jumlah Tanggungan} = \text{"3"} \mid \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{111}{262} = 0,423664$$

6. $P(\text{Jumlah Tanggungan} = \text{"3"} \mid \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{70}{238} = 0,294117$$

Perhitungan pada atribut RT terdapat 9 nilai yaitu “3”, “4”, “5”, “6”, “7”, “8” dan “9”, “10” dan “11”, berdasarkan jumlah data yang ada pada label Keterangan Layak atau Tidak Layak. Dibawah ini perhitungan probabilitas pada atribut RT dengan nilai “1”, sebagai berikut:

1. $P(RT = "3" | \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{14}{262} = 0,053435$$

2. $P(RT = "3" | \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{16}{238} = 0,067226$$

Perhitungan probabilitas pada atribut RT dengan nilai “4”, sebagai berikut:

3. $P(RT = "4" | \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{21}{262} = 0,080152$$

4. $P(RT = "4" | \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{17}{238} = 0,071428$$

Perhitungan probabilitas pada atribut RT dengan nilai “5”, sebagai berikut:

5. $P(RT = "5" | \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{41}{262} = 0,156488$$

6. $P(RT = "5" | \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{12}{238} = 0,050420$$

Perhitungan probabilitas pada atribut RT dengan nilai “6”, sebagai berikut:

7. $P(RT = "6" | \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{24}{262} = 0,091603$$

8. $P(RT = "6" | \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{26}{238} = 0,109243$$

Perhitungan probabilitas pada atribut RT dengan nilai “7”, sebagai berikut:

9. $P(RT = "7" | \text{Class Keterangan} = \text{"Layak"})$

$$= \frac{27}{262} = 0,103053$$

10. $P(RT = "7" | \text{Class Keterangan} = \text{"Tidak Layak"})$

$$= \frac{38}{238} = 0,159663$$

Perhitungan probabilitas pada atribut RT dengan nilai “8”, sebagai berikut:

$$11. P(RT = "8" | \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{30}{262} = 0,114503$$

$$12. P(RT = "8" | \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{38}{238} = 0,159663$$

Perhitungan probabilitas pada atribut RT dengan nilai "9", sebagai berikut:

$$13. P(RT = "9" | \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{52}{262} = 0,198473$$

$$14. P(RT = "9" | \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{26}{238} = 0,109243$$

Perhitungan probabilitas pada atribut RT dengan nilai "10", sebagai berikut:

$$15. P(RT = "10" | \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{32}{262} = 0,122137$$

$$16. P(RT = "10" | \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{24}{238} = 0,100840$$

Perhitungan probabilitas pada atribut RT dengan nilai "11", sebagai berikut:

$$17. P(RT = "11" | \text{Class Keterangan} = \text{"Layak"})$$

$$= \frac{21}{262} = 0,080152$$

$$18. P(RT = "11" | \text{Class Keterangan} = \text{"Tidak Layak"})$$

$$= \frac{41}{238} = 0,172268$$

Hasil perhitungan diatas, agar mempermudah dalam proses analisis dibuat sebuah tabel probabilitas dari setiap atribut sebagai berikut:

Table 4 Nilai *Probabilitas* Atribut

Kode	Atribut	Parameter(Value)	Tidak Layak	Layak
A1	Jenis Kelamin	L	0,702290	0,840336
	Jenis Kelamin	P	0,297709	0,159663
A2	Pekerjaan	Karyawan	0,267175	0,865546
	Pekerjaan	Ibu Rumah Tangga	0,183206	0,004201
	Pekerjaan	Pedagang	0,286259	0,037851
	Pekerjaan	Pekerja Lepas	0,263358	0,092436

Kode	Atribut	Parameter(Value)	Tidak Layak	Layak
A3	Penghasilan	K1	0,496183	0
	Penghasilan	K2	0,320610	0,243697
	Penghasilan	K3	0	0,756302
	Penghasilan	Tidak Ada	0,183206	0
A4	Status Rumah	Kontrak	0,984732	0,789915
	Status Rumah	Rumah Sendiri	0,015267	0,210084
A5	Jumlah Tanggungan	1	0,213740	0,411764
	Jumlah Tanggungan	2	0,362595	0,294117
	Jumlah Tanggungan	3	0,423664	0,294117
A6	RT	3	0,053435	0,067226
	RT	4	0,080152	0,071428
	RT	5	0,156488	0,050420
	RT	6	0,091603	0,109243
	RT	7	0,103053	0,159663
	RT	8	0,114503	0,159663
	RT	9	0,198473	0,109243
	RT	10	0,122137	0,100840
	RT	11	0,080152	0,172268

Hasil dari setiap atribut telah didapatkan, maka dilanjutkan dengan perkalian dari seluruh atribut yang tersedia antara lain “Jenis Kelamin”, “Pekerjaan”, “Penghasilan”, “Status Rumah”, “Jumlah Tanggungan”, dan “RT” .

1. Perkalian dari semua atribut dengan *class* “Layak”:

$$P(X \mid \text{Class Keterangan} = \text{“Layak”}) = 0,702290 * 0,297709 * 0,267175 * 0,183206 * 0,286259 * 0,263358 * 0,496183 * 0,320610 * 0,183206 * 0,984732 * 0,015267 * 0,213740 * 0,362595 * 0,423664 * 0,053435 * 0,080152 * 0,156488 * 0,091603 * 0,103053 * 0,114503 * 0,198473 * 0,122137 * 0,080152$$

$$= 1,56236E-17$$

2. Perkalian dari semua atribut dengan class “Tidak Layak” :

$$\begin{aligned} P(X | \text{Class Keterangan} = \text{“Tidak Layak”}) &= 0,840336 * 0,159663 * 0,865546 * \\ &0,004201 * 0,037851 * 0,092436 * 0,243697 * 0,756302 * 0,789915 * 0,210084 * \\ &0,411764 * 0,294117 * \\ &0,294117 * 0,067226 * 0,071428 * 0,050420 * 0,109243 * 0,159663 * 0,159663 * \\ &0,109243 * 0,100840 * 0,172268 \\ &= 2,37940E-18 \end{aligned}$$

Kemudian perkalian dari semua atribut dengan class “Layak” dan “Tidak Layak”, kemudian dilakukan perhitungan perkalian pada probabilitas sebelumnya dengan semua atribut Class dari Keterangan = “Layak” :

$$\begin{aligned} P(\text{Class Keterangan} = \text{“Layak”} | \text{Total Keterangan}) * P(X | \text{Class Keterangan} = \text{“Layak”}) \\ &= 0,524 * 1,56236E-17 \\ &= 8,18676E-18 \end{aligned}$$

Setelah itu perkalian dari semua atribut dengan class “Layak” dan “Tidak Layak”, kemudian dilakukan perhitungan perkalian pada probabilitas sebelumnya dengan semua atribut Class dari Keterangan = “Tidak Layak” :

$$\begin{aligned} P(\text{Class Keterangan} = \text{“Tidak Layak”} | \text{Total Keterangan}) * P(X | \text{Class Keterangan} = \text{“Tidak Layak”}) \\ &= 0,476 * 2,37940E-18 \\ &= 1,13260E-18 \end{aligned}$$

Perhitungan diatas sudah didapatkan, langkah terakhir ialah membandingkan kedua hasil tersebut :

$$\begin{aligned} P(\text{Class Keterangan} = \text{“Layak”} | \text{Total Keterangan}) * P(X | \text{Class Keterangan} = \text{“Layak”}) \\ > P(\text{Class Keterangan} = \text{“Tidak Layak”} | \text{Total Keterangan}) * P(X | \text{Class Keterangan} = \text{“Tidak Layak”}) \\ &= 8,18676E-18 > 1,13260E-18 \end{aligned}$$

Lalu perhitungan dari perkalian probabilitas sebelumnya dengan semua atribut “Layak” dan “Tidak Layak”²¹ menunjukkan bahwa nilai Class Keterangan “Layak” lebih besar dibandingkan nilai Class dari Keterangan “Tidak Layak”.

4.5 Evaluasi *Confusion Matrix* Algoritma *Naive Bayes*

Pada aspek terakhir mengenai *confusion matrix* algoritma *naive bayes*, dievaluasi secara keseluruhan untuk mengukur seberapa baik memberikan nilai akurasi, berikut *confusion matrix* algoritma *naive bayes data testing* di *rapid miner* antara lain:

accuracy: 95.00%		Data Testing	
	true Layak	true Tidak Layak	class precision
pred. Layak	48	1	97.99%
pred. Tidak Layak	4	47	92.16%
class recall	92.31%	97.92%	

Gambar 5 Evaluasi *Confusion Matrix* Algoritma *Naive Bayes*

Table 5 Perhitungan *Confusion Matrix*

<i>Correct Classification</i>	True Terlambat	True Tepat	<i>Class Precision</i>
Pred Layak	48 (<i>True Positives</i>)	1 (<i>False Negatives</i>)	
Pred Tidak Layak	4 (<i>False Positives</i>)	47 (<i>True Negatives</i>)	
<i>Class Recall</i>			

Accuracy merupakan Jumlah data yang benar dibandingkan dengan jumlah data seluruhnya (Hadianti et al., 2023). Dapat dihitung dengan rumus:

$$Accuracy = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \times 100\% \dots\dots\dots(2)$$

Rumus diatas jika diterapkan dengan kasus pada tabel *confusion matrix* diatas dapat menghasilkan perhitungan sebagai berikut:

$$Accuracy = \frac{48 + 47}{48 + 47 + 4 + 1} \times 100\% = \frac{95}{100} = 0.95 \times 100\% = \mathbf{95.00\%}$$

Precision merupakan persentase total kelas data positif yang diprediksi dengan benar (Aisyah Fatmawati, 2024). Dapat dihitung dengan rumus dibawah ini:

$$Precision = \frac{Tp}{Tp+Fp} \dots\dots\dots(3)$$

Rumus diatas jika diterapkan dengan kasus pada tabel *confusion matrix* diatas dapat menghasilkan perhitungan sebagai berikut:

$$Precision = \frac{48}{48 + 4} \times 100\% = \frac{48}{52} = 0.9231 \times 100\% = \mathbf{92.31\%}$$

Recall digunakan untuk membuktikan persentase kelas data positif yang telah berhasil diprediksi dengan benar dari keseluruhan data kelas positif (Attamami et al., 2023).

$$Recall = \frac{Tp}{Tp+Fn} \dots\dots\dots(4)$$

Rumus diatas jika diterapkan dengan kasus pada tabel *confusion matrix* diatas dapat menghasilkan perhitungan sebagai berikut:

$$Recall = \frac{48}{48 + 1} \times 100\% = \frac{48}{49} = 0.9795918 \times 100\% = \mathbf{97,95\%}$$

5. KESIMPULAN DAN SARAN

Proses yang sudah dilakukan terhadap algoritma *naive bayes* dapat disimpulkan bahwa, hasil perhitungan akurasi dari 100 data testing, diperoleh nilai yang baik yaitu akurasi sebesar 95,00%, presisi sebesar 92,31%, dan recall sebesar 97,95%.

DAFTAR REFERENSI

- Aisyah Fatmawati, A. I. P. & I. A. (2024). Implementasi Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sosial. *Mahasiswa Teknik Informatika*, 8(1), 745–750.
- Alfiah, N. (2021). Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes. *Respati*, 16(1), 32. <https://doi.org/10.35842/jtir.v16i1.386>
- Alfitri, M. M., Nurahman, N., & ... (2023). Evaluasi Performa Algoritma Naïve Bayes Dalam Mengklasifikasi Penerima Bantuan Pangan Non Tunai. *Jurnal Media ...*, 7, 1433–1445. <https://doi.org/10.30865/mib.v7i3.6151>
- Asri, A., Arifin, A., & Handoko, W. (2022). Implementasi Metode Naive Bayes Untuk Klasifikasi Penerima Program Keluarga Harapan. *Journal of Computer*, 2(1), 21–26.
- Attamami, N., Triayudi, A., & Aldisa, R. T. (2023). Analisis Performa Algoritma Klasifikasi Naive Bayes dan C4.5 untuk Prediksi Penerima Bantuan Jaminan Kesehatan. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(2), 262–269. <https://doi.org/10.35870/jtik.v7i2.756>
- Azlil Huriyah, D., & Dienwati Nuris, N. (2023). Klasifikasi Penerima Bantuan Sosial Ukm Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 360–365. <https://doi.org/10.36040/jati.v7i1.6300>
- Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 219.

- <https://doi.org/10.30865/jurikom.v8i6.3655>
- Fitriani, E. (2020). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*, 9(1), 103. <https://doi.org/10.32520/stmsi.v9i1.596>
- Hadianti, I., Astuti, R., Suprpti, T., Informatika, T., Informasi, S., Dinding, J., Penerangan, S., Minum, S. A., Bakar, B., Penghasilan, J., Terakhir, P., Rumah, K., Sosial, B., Golat, D., & Kelayakan, K. (2023). Penerapan Algoritma Naive Bayes Untuk Menentukan Kelayakan Penerima Bantuan Sosial Di Desa Golat. *Mahasiswa Teknik Informatika*, 7(6), 3616–3620.
- Hayati, A. S. & U. (2023). Implementasi Algoritma Naïve Bayes Untuk Prediksi Penerima Bantuan Sosial. *Mahasiswa Teknik Informatika*, 7(1), 347–352.
- Nurriq Fakhri Fikrillah, H., Hudawiguna, S., Juliane, C., & Studi Rekayasa Sistem Informasi, P. (2023). Klasifikasi Penerima Bansos Menggunakan Algoritma Naive Bayes. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 10(1), 683–695. <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/3624>
- Oktha Pratiwi, N. W., Widya Utami, N., & Gede Juliana Eka Putra, I. (2022). Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 Di Desa Keramas, Gianyar, Bali. *Jurnal Informatika, Teknologi Dan Sains*, 4(3), 101–107. <https://doi.org/10.51401/jinteks.v4i3.1667>
- Purnama, A. I., Aziz, A., & Wiguna, A. S. (2020). Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan Pkh Desa Wae Jare Menggunakan Metode Naïve Bayes. *Kurawal - Jurnal Teknologi, Informasi Dan Industri*, 3(2), 173–180. <https://doi.org/10.33479/kurawal.v3i2.348>
- Putri, H., Purnamasari, A. I., Dikananda, A. R., Nurdiawan, O., & Anwar, S. (2021). Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN. *Building of Informatics, Technology and Science (BITS)*, 3(3), 331–337. <https://doi.org/10.47065/bits.v3i3.1093>
- Sundari, S. S., Dewi, E., Mulyani, S., Hidayat, C. R., Anwar, D. S., Mufizar, T., Tasikmalaya, U. P., & Informatika, J. T. (2024). Penerapan Data Mining Classification Untuk Penentuan Jenis Bantuan Sosial Menggunakan Metode Naïve Bayes Classifier. 16(1).

PENGAPLIKASIAN DATA MINING PENENTUAN PENERIMAAN BANTUAN SOSIAL MASYARAKAT DENGAN PROSEDUR ALGORITMA NAIVE BAYES

ORIGINALITY REPORT

10%
SIMILARITY INDEX

7%
INTERNET SOURCES

4%
PUBLICATIONS

1%
STUDENT PAPERS

PRIMARY SOURCES

1 ejournal.nusamandiri.ac.id 1%
Internet Source

2 journal.aptii.or.id 1%
Internet Source

3 Putri Pratiwi, Dima Dwifa, Anita Desiani, Ali Amran, Bambang Suprihatin. "Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Algoritma Naïve Bayes dan Iterative Dichotomizer 3 (ID3)", Electrician : Jurnal Rekayasa dan Teknologi Elektro, 2024 1%
Publication

4 conference.binadarma.ac.id <1%
Internet Source

5 repository.upy.ac.id <1%
Internet Source

6 text-id.123dok.com <1%
Internet Source

7

Internet Source

<1 %

8

Andra Setiawan, Ryan Randy Suryono.
"Analisis Sentimen Ibu Kota Nusantara
menggunakan Algoritma Support Vector
Machine dan Naïve Bayes", Edumatic: Jurnal
Pendidikan Informatika, 2024

Publication

<1 %

9

Makmul Ikhsan Siahaan, Elly Rosmaini. "Use
of Classification and Regression Tree (CART)
Method for Classification of Labor Force
Participation Levels in Medan City in 2019",
FARABI: Jurnal Matematika dan Pendidikan
Matematika, 2022

Publication

<1 %

10

Submitted to Sriwijaya University

Student Paper

<1 %

11

www.diprozvyazok.kiev.ua

Internet Source

<1 %

12

www.jurnal.polgan.ac.id

Internet Source

<1 %

13

jurnal.darmajaya.ac.id

Internet Source

<1 %

14

jurnal.poltekba.ac.id

Internet Source

<1 %

ppid.bandungkab.go.id

15

Internet Source

<1 %

16

repository.umsu.ac.id

Internet Source

<1 %

17

www.researchgate.net

Internet Source

<1 %

18

123dok.com

Internet Source

<1 %

19

Anggi Pradita, Rasiban. "Implementasi Data Mining dengan Metode Regresi Linear untuk Prediksi Hasil Penjualan di PT Awitama Cyndo Wahana", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2024

Publication

<1 %

20

M Khairul Anam, Rahmiati Rahmiati, Dinda Paradila, Mardainis Mardainis, Machdalena Machdalena. "Application of Naïve Bayes Algorithm for Non-Cash Food Assistance Recipients in Kampar Regency", Sinkron, 2023

Publication

<1 %

21

e-journal.hamzanwadi.ac.id

Internet Source

<1 %

22

ejournal.stmikgici.ac.id

Internet Source

<1 %

23

ejurnal.bangunharapanbangsa.com

Internet Source

<1 %

24

journal.unj.ac.id

Internet Source

<1 %

25

teknologiterkini.org

Internet Source

<1 %

26

Gomgom Yosua Balutaro Sihotang.
"Implementasi Data Mining Dalam
Menentukan Korelasi Antara Pelanggaran
Sikap Siswa Di Lingkungan Sekolah Lanjutan
Advent Purwodadi Menggunakan Algoritma
Association Rule", TeIKa, 2021

Publication

<1 %

27

Sofian Wira Hadi, Muhammad Fahmi Julianto,
Syaifur Rahmatullah, Windu Gata. "ANALISA
CLUSTER APLIKASI PADA APP STORE DENGAN
MENGUNAKAN METODE K-MEANS",
Bianglala Informatika, 2020

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On