e-ISSN: 2460-870X | p-ISSN: 2085-1367

Penerapan *Data Mining Classification* Untuk Penentuan Jenis Bantuan Sosial Menggunakan Metode *Naïve Bayes Classifier*

Application of Data Mining Classification to Determine Types of Social Assistance Using the Naïve Bayes Classifier Method

Shinta Siti Sundari*[1], Evi Dewi Sri Mulyani[2] ,Cepi Rahmat Hidayat[3], Dede Syahrul Anwar[4], Teuku Mufizar[5]
[1,2,3,4,5]Universitas Perjuangan Tasikmalaya; Jalan Peta No. 177 Tasikmalaya
Jurusan Teknik Informatika, UNPER, Tasikmalaya
e-mail: *[1]ss.shinta@gmail.com, [2]eviajadech@gmail.com, [3]cepirahmat@unper.ac.id, [4]dedesyahrul@unper.ac.id, [5]teukumufizar@unper.ac.id

ABSTRAK

Program bantuan sosial merupakan program yang diadakan Pemerintah sebagai upaya mengatasi kemiskinan. Desa Mekarjaya adalah salah satu desa yang menjalankan program tersebut. Dalam menjalankan proses bantuan sosial ini terdapat kendala dalam perihal pendataan warganya dikarenakan sering terjadi ketidaksesuaian data penerima yang dikumpulkan oleh masyarakat dengan jenis bantuannya. Untuk mempermudah dalam menentukan jenis bantuan sosial yang sesuai dibutuhkanlah sebuah analisis terhadap data penerima bantuan sosial tersebut. Data yang digunakan sebanyak 1.282 data sampel dengan 15 atribut predictor dan 1 atribut class dengan pembagian 1.025 record data sebagai data training dan 256 record data sebagai data testing. Metode analisis data dalam penelitian ini menggunakan tahapan Data Mining meliputi Data Selection dan Preprocessing, sedangkan metode pengklasifikasiannya menggunakan metode Naïve Bayes Classifier. Pengujian menggunakan Confusion Matrix menghasilkan akurasi sebesar 94,53% dengan perbandingan data training dan data testing 80:20. Dengan dibuatnya model ini, diharapkan pegawai desa dapat menentukan jenis bantuan sosial yang sesuai untuk masyarakat

Kata Kunci : 3-5 Data Mining, Klasifikasi, Naïve Bayes Classifier

ABSTRACT

The social assistance program is a program held by the government as an effort to overcome poverty. Mekarjaya Village is one of the villages running the program. In carrying out this social assistance process, there are obstacles in terms of collecting data on its citizens because there are often discrepancies in the recipient data collected by the community with the type of assistance. To make it easier to determine the appropriate type of social assistance, an analysis of the data on the recipients of the social assistance is needed. Its used 1258 sample data with 15 predictor attributes and 1 class attribute with 1025 record data as data training and 256 record data as data testing. The data analysis method in this research uses Data Mining including Data Selection and Preprocessing, while the classification method uses the Naïve Bayes Classifier. Testing using the Confusion Matrix produces an accuracy of 94.53% with a comparison of training data and testing 80:20. With this model, it is hoped that village officials can determine the type of social assistance that is appropriate for the community..

Keywords : 3-5 Data Mining, Classification, Naïve Bayes Classifier

Info Artikel :
Disubmit:20 Agustus *Direview:19 September* *Diterima :20 Desember*

Copyright © 2022 – CSRID Journal. All rights reserved.

1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan sosial yang masih menjadi tugas pokok utama pemerintah dan termasuk sulit diselesaikan. Kemiskinan adalah suatu keadaan dimana terjadi ketidakmampuan untuk memenuhi kebutuhan pokok hidup seperti sandang, pangan, maupun papan. Kemiskinan disebabkan karena kurangnya alat kebutuhan pokok[1].

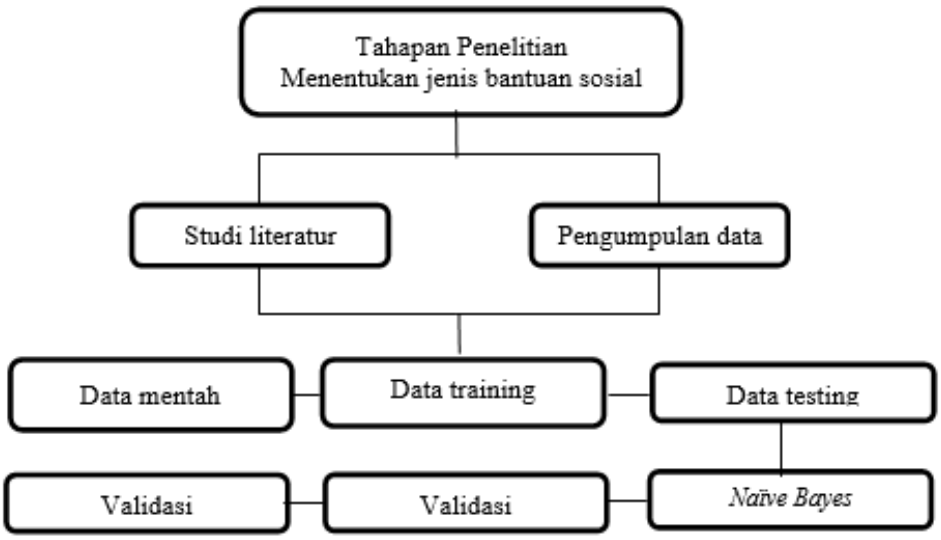
Dalam upaya mengatasi kemiskinan, banyak program yang telah dilakukan pemerintah, tetapi bantuan tersebut tidak sampai kepada masyarakat sesuai harapan. Salah satu penyebabnya karena dalam penentuan status keluarga miskin sebagai penerima bantuan belum optimal, sehingga tidak sesuai dengan jenis bantuan yang diterima[2].

Kementerian Sosial (Kemensos) mengelola program bantuan sosial yang terdiri dari bantuan sosial reguler dan bantuan sosial khusus. Bantuan sosial reguler diantaranya adalah Program Keluarga Harapan (PKH) dan Bantuan Pangan Non Tunai (BPNT) atau disebut juga Kartu Sembako. Sedangkan bantuan sosial khusus yaitu Bantuan Sosial Tunai (BST)[3]. Program bantuan sosial lainnya adalah Bantuan Langsung Tunai (BLT) Desa yang diarahkan dari Dana Desa[4]. Menteri Desa PDTT, Abdul Halim Iskandar atau yang biasa disapa Gus Halim, mengatakan bahwa pada tahun 2022 penggunaan Dana Desa diatur sebesar 40% untuk program Bantuan Langsung Tunai (BLT) Desa[5].

Dalam menjalankan proses bantuan sosial ini, Desa Mekarjaya mengalami kendala dalam perihal pendataan warganya dikarenakan sering terjadi ketidaksesuaian data penerima yang dikumpulkan oleh masyarakat dengan jenis bantuannya. Oleh karena itu, penulis akan mengklasifikasi jenis bantuan yang sesuai berdasarkan pendataan ajuan penerima bantuan sosial.
Penulis dalam penelitian ini akan menggunakan metode *Naïve Bayes* karena dalam penelitian sebelumnya *Naïve Bayes* mampu untuk menyelesaikan permasalahan dalam pendataan untuk penerimaan bantuan PKH, dengan jumlah persentase data training dan data testingnya 80:20. Tingkat akurasi yang dihasilkan sebesar 82,14%[6]. Penelitian lain mampu memprediksi penerima bantuan sembako dengan pendataan yang lebih valid terkait kelayakan dari penerima bantuan tersebut sehingga pemberian bantuan tersebut tepat sasaran. Dengan jumlah persentase data training dan data testing 80:20, dihasilkan tingkat akurasi sebesar 86%[7].
Berdasarkan permasalahan diatas, teknik klasifikasi *Naïve Bayes* diharapkan dapat mengklasifikasi jenis bantuan sosial yang sesuai untuk masyarakat

2. METODE PENELITIAN

Tahapan penelitian dalam klasifikasi jenis bantuan sosial dapat digambarkan pada bagan berikut ini:



Gambar 1. Tahapan penelitian dalam klasifikasi jenis bantuan sosial

 e-ISSN: 2460-870X | p-ISSN: 2085-1367

A. Tahapan Review

Algoritma *Naïve Bayes* memiliki fungsi untuk menemukan pengetahuan atau pola kesamaan karakteristik dalam kelompok atau kelas tertentu[7]. Algoritma *Naïve Bayes* merupakan suatu algoritma yang dikemukakan oleh *Thomas Bayes* seorang ilmuwan Inggris yang dapat memprediksi peluang masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal dengan *Teorema Bayes* dan merupakan algoritma yang terdapat pada teknik klasifikasi probabilitas dan statistik[8].

Metode *Naïve Bayes* hanya membutuhkan jumlah yang kecil dari data yang diperlukan untuk proses pengklasifikasian merupakan suatu keuntungan penggunaan algoritma *Naïve Bayes*[18]. *Naïve Bayes* memiliki akurasi dan kecepatan yang tinggi saat diterapkan pada basis data dengan jumlah yang besar[9]. *Naïve Bayes* memiliki kelemahan yang mengakibatkan nilai akurasi klasifikasinya menjadi rendah yaitu sangat sensitif pada fitur yang terlalu banyak[10].

1. Teorema Bayes

Teorema Bayes memiliki bentuk umum sebagai berikut:

$$P(H \mid X) = \frac{P(X \mid H)(P \mid H)}{P(X)} \qquad \text{.. (1)}$$

Keterangan:
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[8]

B. Confusion Matrix

Confusion Matrix merupakan matriks yang berisi informasi tentang kelas aktual dan prediksi dari hasil proses klasifikasi yang dilakukan oleh sistem klasifikasi[11]. *Confusion Matrix* digunakan untuk mendapatkan tingkat akurasi klasifikasi terhadap kelas dengan algoritma yang digunakan[12].

Tabel 1. Confusion Matrix

Confusions Matrix		**Nilai Prediksi**	
		True	**False**
Nilai Aktual	**True**	TP *(True Positive)* *Correct result*	FN *(False Negative)* *Unexpected result*
	False	FP *(False Positive)* *Missing result*	TN *(True Negative)* *Correct absence of result*

Confusion Matrix menghasilkan nilai sebagai berikut:
1. *Accuracy* yaitu proporsi kasus yang diidentifikasi benar dari semua kasus.
2. *Precision* yaitu proporsi dari kasus dengan hasil positif yang benar.
3. *Recall* yaitu proporsi kasus positif yang diidentifikasi dengan benar[11].

Berikut formulasi pada pembentukan model klasifikasi untuk menghitung nilai *accuracy*, *presicion*, dan *recall*[13]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad (2)$$

$$Pr\,ecision = \frac{TP}{TP + FP} \times 100\% \quad (1)$$

$$Re\,call = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

3. HASIL DAN PEMBAHASAN

Metode yang digunakan dalam penentuan jenis bantuan sosial ini adalah metode *Naïve Bayes* dengan teknik klasifikasi. Klasifikasi *Naïve Bayes* diterapkan untuk menentukan hasil klasifikasi pada *data training* yng sudah ada dengan menggunakan rumus-rumus *Naïve Bayes*. Pada klasifikasi *Naïve Bayes* terdapat beberapa tahapan yaitu:

1. Menghitung jumlah kelas atau label.
2. Menghitung jumlah kasus pada kelas.
3. Mengalikan semua variabel.
4. Membandingkan hasil per kelas.

Analisis data digunakan untuk memperoleh data yang sesuai dengan kebutuhan. Analisis data dilakukan melalui data penerima bantuan sosial di Desa Mekarjaya dengan 15 variabel, data diolah dan diuji dalam pengujian pada algoritma *Naïve Bayes*.
Dengan pengujian tersebut dapat diperoleh nilai akurasi, *presicion*, dan *recall* sehingga dapat diimplementasikan dengan baik pada sebuah sistem yang dapat digunakan untuk menentukan jenis bantuan sosial yang sesuai.

Metode analisis data dalam penelitian ini menggunakan tahapan *Knowledge Discovery in Database* (KDD) meliputi *Data Selection* dan *Preprocessing*. Data yang digunakan adalah data penduduk penerima bantuan sosial di Desa Mekarjaya Kecamatan Sukaraja Kabupaten Tasikmalaya. Data yang diperoleh dari desa masih berupa raw data, kemudian data penduduk penerima bantuan sosial akan digali berdasarkan kriteria jenis kelamin, usia, pekerjan, penghasilan, jumlah anggota rumah tangga (ART), Terdampak Covid-19, ART sakit kronis, ART tunggal lanjut usia (>60 tahun), ART ibu hamil, ART anak usia dini (0-6 tahun), ART anak sekolah (SD, SMP, SMA), ART disabilitas, memiliki KKS, dan status DTKS.

Nilai dari setiap kriteria tersebut akan menjadi sebuah patokan untuk penentuan jenis bantuan sosial yang sesuai. Dengan mengamati persoalan diatas *data mining* dengan metode *Naïve Bayes* sangat tepat digunakan untuk menghasilkan *knowledge* jenis bantuan sosial yang sesuai diantara ribuan jumlah penduduk penerima bantuan sosial.

Data yang akan digunakan berjumlah 1.282 data. Adapun data yang akan dijadikan sebagai *data training* sebanyak 80% dari jumlah data mentah yaitu berjumlah 1025 data, sedangkan untuk *data testing* sebanyak 20% dari jumlah data mentah yaitu berjumlah 256 data.

1. Data Selection

Data yang digunakan dalam penelitian ini adalah data penerima bantuan sosial tahun 2021-2022 yang didapat dari Desa Mekarjaya. Data yang didapat berjumlah 1.282 *record* data dengan pembagian *training* dan *testing* 80:20.. Data tersebut berisi 15 atribut *predictor* yang digunakan yaitu jenis kelamin, usia, pekerjaan, penghasilan, jumlah anggota rumah tangga (ART), ART rentan sakit kronis, bantuan sosial lain, ART tunggal lanjut usia (>60 tahun), ART ibu hamil, ART anak usia dini (0-6 tahun), ART anak sekolah (SD, SMP, SMA), ART disabilitas, memiliki KKS, dan status DTKS, dan 1 atribut *Class*. Penjelasan untuk masing-masing atribut dapat dilihat pada tabel berikut.

Tabel 2. Atribut Data Selection

e-ISSN: 2460-870X | p-ISSN: 2085-1367

No	Atribut	Keterangan	Type
1	Jenis Kelamin	Berisi informasi jenis kelamin	Atribut *predictor*
2	Usia	Berisi informasi usia	Atribut *predictor*
3	Pekerjaan	Berisi informasi pekerjaan	Atribut *predictor*
4	Penghasilan	Berisi informasi jumlah penghasilan	Atribut *predictor*
5	Jumlah Anggota Rumah Tangga (ART)	Berisi informasi jumlah anggota keluarga dalam satu Kartu Keluarga (KK)	Atribut *predictor*
6	Terdampak Covid-19	Berisi informasi keluarga terdampak Covid-19 atau tidak	Atribut *predictor*
7	ART Rentan Sakit Kronis	Berisi informasi anggota keluarga yang sakit kronis	Atribut *predictor*
8	Bantuan Sosial Lain	Berisi informasi keluarga menerima bantuan sosial lain atau tidak	Atribut *predictor*
9	ART Tunggal Lanjut Usia	Berisi informasi anggota rumah tangga tunggal yang sudah lanjut usia >60 tahun	Atribut *predictor*
10	ART Ibu Hamil	Berisi informasi keluarga yang memiliki ibu hamil	Atribut *predictor*
11	ART Anak Usia Dini	Berisi informasi keluarga yang memiliki anak usia dini yaitu usia 0-6 tahun	Atribut *predictor*
12	ART Anak Sekolah	Berisi informasi keluarga yang memiliki anak sekolah jenjang SD, SMP, SMA	Atribut *predictor*
13	ART Disabilitas	Berisi informasi memiliki anggota keluarga disabilitas	Atribut *predictor*
14	Memiliki Kartu Keluarga Sejahtera (KKS)	Berisi informasi kepemilikan Kartu Keluarga Sejahtera (KKS)	Atribut *predictor*
15	Status DTKS	Berisi informasi status terdaftar atau tidak pada DTKS (Data Terpadu Kesejahteraan Sosial)	Atribut *predictor*
16	Jenis Bantuan	Berisi informasi hasil jenis bantuan sosial yang akan diterima, yaitu PKH, BPNT, BST, atau BLT	*Class* atau *Label*

2. Preprocessing

Pada tahapan *Preprocessing* dilakukan proses pembersihan data (*data cleaning*). Jumlah data yang ada sejumlah 1.282 data. Berdasarkan hasil verifikasi kualitas data yang telah dilakukan diperoleh hasil yaitu tidak terdapat *Missing Value* pada seluruh atribut data, sehingga tidak akan dilakukan penghapusan data. Dengan demikian data tetap berjumlah 1.282 data

3. Spliting Data

a. Data Training

Persentase data *training* yang digunakan sebanyak 80% atau setara dengan 1025 *record* data yang nantinya akan digunakan sebagai pemodelan dalam metode *Naïve Bayes Classifier*, model ini akan digunakan sebagai sebagai prediksi kelas data baru yang belum pernah ada. Berikut contoh sampel data *training*.

	JENIS KELAMIN	USIA	PEKERJAAN	PENGHASILAN	JUMLAH ART	TERDAMPAK COVID-19	ART SAKIT KRONIS	ART TUNGGAL LANSIA	ART IBU HAMIL	ART ANAK USIA DINI	ART ANAK SEKOLAH	ART DISABILITAS	MEMILIKI KKS	BANSOS LAIN	STATUS DTKS	JENIS BANTUAN
3	PEREMPUAN	51-60	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
4	PEREMPUAN	<=40	2	SEDANG	2	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
5	PEREMPUAN	51-60	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
6	PEREMPUAN	41-50	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	YA	YA	PKH
7	PEREMPUAN	41-50	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
8	PEREMPUAN	41-50	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	TIDAK	YA	PKH
9	PEREMPUAN	>60	2	RENDAH	2	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
10	PEREMPUAN	41-50	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
11	PEREMPUAN	41-50	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
12	PEREMPUAN	<=40	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
13	PEREMPUAN	>60	2	RENDAH	2	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	YA	YA	PKH
14	PEREMPUAN	41-50	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	TIDAK	YA	PKH
15	PEREMPUAN	41-50	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH

Gambar 2. Data Training

b. Data Testing

Persentase data *testing* yang digunakan sebanyak 20% atau setara dengan 256 *record* data. Data *testing* atau data uji digunakan untuk menguji sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Berikut contoh sampel data *testing*.

	JENIS KELAMIN	USIA	PEKERJAAN	PENGHASILAN	JUMLAH ART	TERDAMPAK COVID-19	ART SAKIT KRONIS	ART TUNGGAL LANSIA	ART IBU HAMIL	ART ANAK USIA DINI	ART ANAK SEKOLAH	ART DISABILITAS	MEMILIKI KKS	BANSOS LAIN	STATUS DTKS	JENIS BANTUAN
3	PEREMPUAN	41-50	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
4	PEREMPUAN	<=40	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
5	PEREMPUAN	>60	2	RENDAH	3	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
6	PEREMPUAN	>60	2	RENDAH	2	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
7	PEREMPUAN	51-60	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	YA	YA	PKH
8	PEREMPUAN	41-50	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	YA	YA	PKH
9	PEREMPUAN	>60	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
10	PEREMPUAN	51-60	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
11	PEREMPUAN	<=40	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
12	PEREMPUAN	<=40	2	RENDAH	3	TIDAK	TIDAK	TIDAK	YA	YA	YA	TIDAK	TIDAK	TIDAK	YA	PKH
13	PEREMPUAN	51-60	2	RENDAH	2	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	TIDAK	YA	PKH
14	PEREMPUAN	>60	2	RENDAH	1	TIDAK	YA	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA	PKH
15	PEREMPUAN	41-50	2	RENDAH	3	TIDAK	TIDAK	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	YA	YA	PKH

Gambar 3. Data Testing

4. Modeling

Pemodelan merupakan tahapan yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan pemilihan teknik pemodelan, pembuatan model, serta pengujian model. Teknik pemodelan yang akan digunakan untuk melakukan klasifikasi penentuan jenis bantuan sosial adalah *Naive Bayes*. Sedangkan *tools* yang digunakan adalah *RapidMiner*.

Tahap selanjutnya yang perlu dilakukan adalah melakukan klasifikasi data dengan metode *Naive Bayes* pada *tools RapidMiner*.

a. Perhitungan Naïve Bayes

1. Perhitungan Probabilitas Prior (P(Ci))

Dari 1.025 data latih yang digunakan, diketahui kelas C0 (PKH) sebanyak 90 data, kelas C1 (BPNT) sebanyak 598 data, kelas C2 (BST) sebanyak 182 data, dan kelas C3 (BLT) sebanyak 155 data. Perhitungan probabilitas prior untuk kemungkinan setiap kelas dilakukan sebagai berikut:

Perhitungan probabilitas prior untuk kemungkinan PKH sebagai berikut:

$$P(C0) = \frac{90}{1025} = 0{,}088 \quad \textbf{(5)}$$

Perhitungan probabilitas prior untuk kemungkinan BPNT sebagai berikut:

$$P(C1) = \frac{598}{1025} = 0{,}583 \quad (6)$$

Perhitungan probabilitas prior untuk kemungkinan BST sebagai berikut:

$$P(C2) = \frac{182}{1025} = 0{,}178 \quad (7)$$

Sedangkan perhitungan probabilitas prior untuk kemungkinan BLT sebagai berikut:

$$P(C3) = \frac{155}{1025} = 0{,}151 \quad (8)$$

Hasil probabilitas prior dapat dilihat pada tabel berikut.

Tabel 3. Probabilitas Prior (P(Ci))

Kelas	Probabilitas Prior (P(Ci))
PKH (C0)	0,088
BPNT (C1)	0,583
BST (C2)	0,178
BLT (C3)	0,151

2. Perhitungan Probabilitasi Posterior (P(X|Ci))

Perhitungan probabilitas posterior dilakukan pada data latih sebanyak 1.025 data dengan menggunakan X sebagai vector pemilihan atribut penentuan jenis bantuan sosial yaitu Jenis Kelamin, Usia, Pekerjaan, Penghasilan, Jumlah Anggota Rumah Tangga (ART), Terdampak

e-ISSN: 2460-870X | p-ISSN: 2085-1367

Covid-19, ART Rentan Sakit Kronis, Bantuan Sosial Lain, ART Tunggal Lanjut Usia (>60 Tahun), ART Ibu Hamil, ART Anak Usia Dini (0-6 Tahun), ART Anak Sekolah (SD, SMP, SMA), ART Disabilitas, Memiliki KKS, dan Status DTKS, sehingga *P(X|Ci)* dapat dijabarkan menjadi P(Jenis Kelamin, Usia, Pekerjaan, Penghasilan, Jumlah Anggota Rumah Tangga (ART), Terdampak Covid-19, ART Rentan Sakit Kronis, Bantuan Sosial Lain, ART Tunggal Lanjut Usia (>60 Tahun), ART Ibu Hamil, ART Anak Usia Dini (0-6 Tahun), ART Anak Sekolah (SD, SMP, SMA), ART Disabilitas, Memiliki KKS, dan Status DTKS|Ci) untuk dihitung kemungkinannya pada setiap Ci.

3. Pendefinisian Probabilitas Data Baru
   Pendefinisian Probabilitas Data Baru dengan berdasarkan hasil dari persamaan (5), (6), (7), (8), yaitu :
   1. Kelas PKH (C0) dengan probabilitas prior 0,088
   2. Kelas BPNT (C1) dengan probabilitas prior 0,583
   3. Kelas BST (C2) dengan probabilitas prior 0,178
   4. Kelas BLT (C3) dengan probabilitas prior 0,151

Tabel 4. ProbabilitasData Baru

Data X		**Probabilitas**			
Atribut	**Nilai**	**C0**	**C1**	**C2**	**C3**
Jenis Kelamin	Laki-laki	0	0,120	0,557	0,452
Usia	42	0,311	0,303	0,209	0,161
Pekerjaan	1	0	0,120	0,577	0,452
Penghasilan	Rendah	0,933	0,946	0,978	0,935
Jumlah ART	3	0,300	0,204	0,187	0.258
Terdampak Covid-19	Ya	0	0,455	0,544	0,206
ART Sakit Kronis	Tidak	0,156	0,018	0,016	0,071
ART Tunggal Lansia	Tidak	0,956	0,975	0,857	0,865
ART Ibu Hamil	Tidak	0,856	0,918	0,989	0,955
ART Anak Usia Dini	Tidak	0,622	0,957	0,962	0,942
ART Anak Sekolah	Ya	0,800	0,085	0,066	0,129
ART Disabilitas	Tidak	1	1	1	1
Memiliki KKS	Tidak	0	1	0,088	0,097
Bantuan Lain	Tidak	0,833	0,848	0,764	0,903
Status DTKS	Tidak	0	0,064	0,104	0,897

Setelah diketahui nilai probabilitas setiap atribut terhadap probabilitas tiap kelas atau P(X|Ci), maka langkah selanjutnya adalah menghitung total keseluruhan probabilitas tiap kelas dengan cara mengalikan nilai dari masing-masing atribut.

P(X|C0) = P(X|Jenis Bantuan=PKH)

= 0 * 0,311 * 0 * 0,933 * 0,300 * 0 * 0,156 * 0,956 * 0,856 * 0,622 * 0,800 * 1 * 0 * 0,833 * 0

= 0

P(X|C1) = P(X|Jenis Bantuan=BPNT)

= 0,120 * 0,303 * 0,120 * 0,946 * 0,204 * 0,455 * 0,018 * 0,975 * 0,918 * 0,957 * 0,085 * 1 * 1 * 0,848 * 0,064

= 0,00000029333675

P(X|C2) = P(X|Jenis Bantuan=BST)

= 0,557 * 0,209 * 0,577 * 0,978 * 0,187 * 0,544 * 0,016 * 0,857 * 0,989 * 0,962 * 0,066 * 1 * 0,088 * 0,764 * 0,104

= 0,00000004023297

P(X|C3) = P(X|Jenis Bantuan=BLT)

= 0,452 * 0,161 * 0,452 * 0,935 * 0.258 * 0,206 * 0,071 * 0,865 * 0,955 * 0,942 * 0,129 * 1 * 0,097 * 0,903 * 0,897

= 0,00000091531726

Berikutnya adalah melakukan perhitungan pemaksimal untuk masing masing kelas, yaitu dengan cara mengalikan P(X|Ci) dengan P(Ci). Maka didapat hasil sebagai berikut :

P(C0|X) = P(X|Jenis Bantuan=PKH) * P(PKH)

= 0 * 0,088

= 0

P(C1|X) = P(X|Jenis Bantuan=BPNT) * P(BPNT)

= 0,00000029333675 * 0,583

= 0,00000017101533

P(C2|X) = P(X|Jenis Bantuan=BST) * P(BST)

= 0,00000004023297 * 0,178

= 0,00000000716147

P(C3|X) = P(X|Jenis Bantuan=BLT) * P(BLT)

= 0,00000091531726 * 0,151

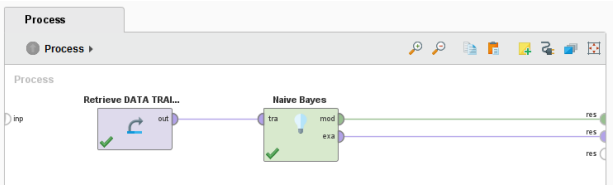
= 0,00000013821291

Berdasarkan nilai tersebut dapat diambil kesimpulan bahwa P(C1|X) memiliki nilai lebih besar daripada nilai P(C0|X), P(C2|X), dan P(C3|X), maka data baru tersebut diklasifikasikan ke dalam kelas BPNT untuk jenis bantuannya.

4. Perhitungan Rapid Miner

Pada tahap ini dilakukan sebuah pengujian metode untuk menentukan nilai akurasi dengan menggunakan metode *Naïve Bayes* serta *software RapidMiner*

a. Proses Data Training

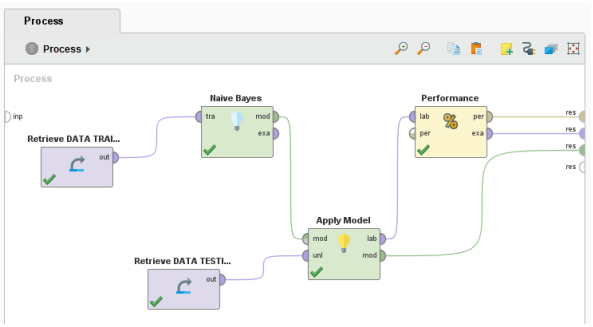


Gambar 4. Proses Data Training pada Rapid Miner menggunakan Naïve Bayes

b. Proses Prediksi Confusion Matrix

Proses klasifikasi oleh *RapidMiner* menggunakan metode *Naïve Bayes* dalam data penerima bantuan sosial ini untuk mencocokkan data *testing* dengan data *training* yang telah ditemukan sebelumnya. Pada tampilan *Process* masukkan masing-masing data *training* dan data *testing*, lalu masukkan operator *Naïve Bayes*, *Apply Model*, dan *Performance*, kemudian sambungkan seperti pada gambar berikut:

e-ISSN: 2460-870X | p-ISSN: 2085-1367



Gambar 5. Proses Prediksi Confusion Matrix

5. Evaluasi

Pada tahapan ini dilakukan pengukuran tingkat keakuratan hasil dan validasi yang dicapai oleh model CRISP-DM dengan menggunakan *Confusion Matrix*.

a. Hasil Performance Vektor

Proses klasifikasi dengan *RapidMiner* dengan metode *Naïve Bayes* yang digunakan untuk mengklasifikasikan data penerima bantuan sosial pada penelitian ini sehingga diperoleh nilai *accuracy*, *precision*, dan *recall*.

accuracy: 94.53%

	true PKH	true BPNT	true BST	true BLT	class precision
pred. PKH	19	0	0	1	95.00%
pred. BPNT	0	154	6	1	95.65%
pred. BST	1	1	36	0	94.74%
pred. BLT	0	0	4	33	89.19%
class recall	95.00%	99.35%	78.26%	94.29%	

Gambar 6. Hasil Performance Vektor

Dari hasil pengujian untuk nilai *accuracy* 94,53%, untuk nilai *precision* yaitu 95,00% pada class PKH, 95,65% pada class BPNT, 94,74% pada class BST, dan 89,19% pada class BLT, sedangkan nilai *recall* 95,00% pada class PKH, 99,35% pada class BPNT, 78,26% pada class BST, dan 94,29% pada class BLT.

Hasil *class accuracy*, *precision*, dan *recall* dari evaluasi *Confusion Matrix* untuk menentukan jenis bantuan sosial dapat dilihat pada tabel berikut.

Tabel 5. Hasil Class Accuracy, Precision, dan Recall

Class	***Accuracy***	***Precision***	***Recall***
PKH (C0)	94,53%	95,00%	95,00%
BPNT (C1)		95,65%	99,35%
BST (C2)		94,74%	78,26%
BLT (C3)		89,19%	94,29%

1) *Accuracy*

Dengan mengetahui jumlah data yang diklasifikasikan secara benar, maka dapat diketahui akurasi hasil prediksi yaitu 94,53% dari hasil data *testing*.

Tabel 6. Hasil Akurasi *(Evaluasi Confusion Matrix)*

	True PKH	True BPNT	True BST	True BLT
Pred. PKH	19	0	0	1
Pred. BPNT	0	154	6	1
Pred. BST	1	1	36	0
Pred. BLT	0	0	4	33
Accuracy	94,53%			

$$\begin{aligned} \textit{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \\ &= \frac{\text{jumlah kasus yang diidentifikasi benar}}{\text{jumlah kasus keseluruhan}} \text{ x } 100\% \\ &= \frac{19+154+36+33}{256} \text{ x } 100\% \\ &= 0{,}9453 \text{ x } 100\ \% \\ &= 94{,}53\% \end{aligned}$$

2) *Precision*

Precision adalah jumlah data positif yang dikenali secara benar sebagai positif (*True Positive*) dibagi dengan jumlah data yang dikenali sebagai positif (*False Positive*). Dari hasil pengujian nilai *precision* yaitu 95,00% pada *class* PKH, 95,65% pada *class* BPNT, 94,74% pada *class* BST, dan 89,19% pada *class* BLT.

Tabel 7. Hasil Precision (Evaluasi Confusion Matrix)

	True PKH	True BPNT	True BST	True BLT	*Precision*
Pred. PKH	9	0	0	1	95,00%
Pred. BPNT	0	154	6	1	95,65%
Pred. BST	1	1	36	0	94,74%
Pred. BLT	0	0	4	33	89,19%

3) *Recall*

Recall adalah jumlah data yang *True Positive* dibagi dengan jumlah data yang sebenarnya positif (*True Positive* + *True Negative*). Untuk nilai *recall* yaitu 95,00% pada *class* PKH, 99,35% pada *class* BPNT, 78,26% pada *class* BST, dan 94,29% pada *class* BLT.

Tabel 8. Hasil Recall (Evaluasi Confusion Matrix)

	True PKH	True BPNT	True BST	True BLT
Pred. PKH	19	0	0	1
Pred. BPNT	0	154	6	1
Pred. BST	1	1	36	0
Pred. BLT	0	0	4	33
Recall	95,00%	99,35%	78,26%	94,29%

b. Hasil Klasifikasi *Class* dengan Metode *Naïve Bayes*

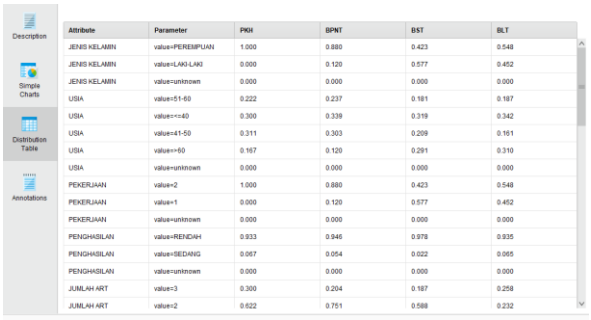
untuk menganalisis tabel data penerima bantuan sosial dengan menggunakan metode *Naïve Bayes* dapat menghasilkan 4 (empat) kelas utama pembagian, didapatkan hasil jumlah dari model distribusi kelas PKH sebanyak 0,088, kelas BPNT sebanyak 0,583, kelas BST sebanyak 0,178, dan kelas BLT sebanyak 0,151



Gambar 8. *Simple Distribution Model*

e-ISSN: 2460-870X | p-ISSN: 2085-1367

Sedangkan distribusi hasil analisa dengan metode *Naïve Bayes* terhadap tabel data penerima bantuan sosial dapat dilihat pada gambar dibawah ini .



Attribute	Parameter	PKH	BPNT	BST	BLT
JENIS KELAMIN	value=PEREMPUAN	1.000	0.880	0.423	0.548
JENIS KELAMIN	value=LAKI-LAKI	0.000	0.120	0.577	0.452
JENIS KELAMIN	value=unknown	0.000	0.000	0.000	0.000
USIA	value=51-60	0.222	0.237	0.181	0.187
USIA	value=<=40	0.300	0.339	0.319	0.342
USIA	value=41-50	0.311	0.303	0.209	0.161
USIA	value=>60	0.167	0.120	0.291	0.310
USIA	value=unknown	0.000	0.000	0.000	0.000
PEKERJAAN	value=2	1.000	0.880	0.423	0.548
PEKERJAAN	value=1	0.000	0.120	0.577	0.452
PEKERJAAN	value=unknown	0.000	0.000	0.000	0.000
PENGHASILAN	value=RENDAH	0.933	0.946	0.978	0.935
PENGHASILAN	value=SEDANG	0.067	0.054	0.022	0.065
PENGHASILAN	value=unknown	0.000	0.000	0.000	0.000
JUMLAH ART	value=3	0.300	0.204	0.187	0.258
JUMLAH ART	value=2	0.622	0.751	0.588	0.232

Gambar 9. Tabel Distribusi

4. KESIMPULAN

1. Metode *Naïve Bayes Classifier* dapat digunakan untuk pembuatan model secara cepat dan mudah dipahami dengan akurasi yang tinggi sehingga dapat digunakan sebagai acuan dalam menentukan jenis bantuan sosial pada data baru.
2. Metode *Naïve Bayes Classifier* sangat cocok diterapkan dalam memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga memudahkan pengurus desa dalam menentukan jenis bantuan sosial yang sesuai untuk masyarakat.
3. Dengan pengujian yang dilakukan menggunakan *tools RapidMiner* dengan menggunakan 80% data *training* dan 20% data *testing*, serta evaluasi menggunakan *Confusion Matrix* menghasilkan tingkat akurasi tinggi sebesar 94,53%. Jumlah banyaknya data *testing* yang digunakan sangat mempengaruhi hasil dari akurasi.

Saran

1. Untuk penelitian selanjutnya menggunakan metode selain *Naïve Bayes*, atau dapat membandingkan antara dua metode dalam satu studi kasus. Contohnya bisa menggunakan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* dan dapat diimplementasikan ke dalam sebuah aplikasi program yang dapat bermanfaat untuk membantu petugas desa dalam menganalisa jenis bantuan sosial bantuan yang diberikan sesuai berdasarkan data ajuan penerima bantuan sosial.

UCAPAN TERIMA KASIH

Penulis ucapkan terima kasih kepada semua pihak yang telah membantu dalam proses penelitian ini, dan semoga apa yang menjadi hasil dari penelitian ini bermanfaat untuk tempat penelitian dan dunia penelitian pada umumnya.

REFERENSI

[1] M. Parida, “Sistem Pendukung Pengambilan Keputusan Seleksi Penerimaan Beras (Raskin) Menggunakan Metode AHP,” *J. Inf. Dan Komput.*, vol. 7, pp. 1–8, 2019.

[2] U. Lestari and M. Targiono, “Sistem Pendukung Keputusan Klasifikasi Keluarga Miskin Menggunakan Metode Simple Additive Weighting (Saw) Sebagai Acuan Penerima Bantuan Dana Pemerintah (Studi Kasus: Pemerintah Desa Tamanmartani, Sleman),” J. TAM (Technology Accept. Model., vol. 8, no. 1, pp. 70–78, 2017, [Online]. Available: http://www.ojs.stmikpringsewu.ac.id/index.php/JurnalTam/article/view/97

[3] B. Humas, “Kemensos Lanjutkan Perlindungan Sosial untuk Masyarakat Terdampak Pandemi Melalui Program Bansos,” Kementerian Sosial Republik Indonesia, 2021. https://kemensos.go.id/kemensos-lanjutkan-perlindungan-sosial-untuk-masyarakat-terdampak-

pandemi-melalui-program-bansos (accessed Apr. 08, 2022).

[4] "BLT Desa, Jaring Pengaman Sosial Masyarakat Desa di Masa Pandemi," Kementerian Keuangan Republik Indonesia, 2020. https://www.kemenkeu.go.id/publikasi/berita/blt-desa-jaring-pengaman-sosial-masyarakat-desa-di-masa-pandemi (accessed Apr. 08, 2022).

[5] N. Humas Kemendes PDTT, "40 Persen Dana Desa Untuk BLT, Gus Halim: Wujud Keberpihakan Kepada Warga Miskin," Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi Republik Indonesia, 2021. https://www.kemendesa.go.id/berita/view/detil/4115/40-persen-dana-desa-untuk-blt-gus-halim-wujud-keberpihakan-kepada-warga-miskin (accessed Apr. 08, 2022).

[6] A. Ifon Purnama, A. Aziz, A. Sartika Wiguna, and K. Kunci, "Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan PKH Desa Wae Jare Menggunakan Metode Naïve Bayes," Kurawal J. Teknol. Inf. dan Ind., vol. 3, pp. 1–8, 2020, [Online]. Available: https://jurnal.machung.ac.id/index.php/kurawal

[7] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," J. Ris. Komput., vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.

[8] Bustami, "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi," TECHSI J. Penelit. Tek. Inform., vol. 3, no. 2, pp. 128–146, 2014, doi: 10.26555/jifo.v8i1.a2086.

[9] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," J. Teknol. Inf., vol. XVI, no. 1, pp. 32–40, 2021.

[10] F. Ariani, Amir, N. Alam, and K. Rizal, "Klasifikasi Penetapan Status Karyawan Dengan Menggunakan Metode Naïve Bayes," Paradig. - J. Komput. dan Inform., vol. 20, no. 2, pp. 33–38, 2018, doi: 10.31294/p.v.

[11] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," Energy - J. Ilm. Ilmu-Ilmu Tek., vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: https://ejournal.upm.ac.id/index.php/energy/article/view/111

[12] L. Qadrini, A. Seppewali, and A. Aina, "Decision Tree dan Adaboost Pada Klasifikasi Penerima Program Bantuan Sosial," J. Inov. Penelit., vol. 2, no. 7, pp. 1959–1966, 2021

[13] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," Sist. J. Sist. Inf., vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.