

Algoritma Naïve Bayes Dalam Penentuan Bantuan Renovasi Rumah Di Desa Sialang Buah

Marulak Lasron Siahaan¹, Zakarias Situmorang²

^{1,2} Universitas Katolik Santo Thomas Medan, Jl. Setia Budi No. 479 F Tanjung Sari, Medan, Indonesia

ARTICLE INFORMATION

Received: Maret, 2023

Revised: April, 2023

Available online: Oktober, 2023

KEYWORDS

Naïve Bayes, *Data Mining*, Bantuan RTLH, Sialang Buah

CORRESPONDENCE

Phone: +62 82166657074

E-mail: marulaksiahaan02@gmail.com¹

zakarias65@yahoo.com²

ABSTRACT

Rumah merupakan kebutuhan dasar dalam melaksanakan peran sosial bagi anggota keluarga, Fakir Miskin pun memerlukan rumah yang layak dan nyaman. Di Indonesia masih banyak sekali rumah yang masuk dalam kategori layak tak huni. Pemerintah memberikan bantuan Rumah Tidak Layak Huni (RTLH) berupa uang untuk pembelian bahan bangunan guna pemugaran Rumah Tidak Layak Huni dimana rumah tersebut tidak memenuhi persyaratan keselamatan bangunan dan kecukupan minimal bangunan. Fokus penelitian ini untuk mengetahui bagaimana penerapan metode klasifikasi dalam penentuan bantuan renovasi rumah dengan menggunakan algoritma naïve bayes, sehingga mengetahui berapa tingkat akurasi yang diperoleh. Algoritma naïve bayes merupakan salah satu metode data mining yang dikemukakan salah satu ilmuwan dari inggris bernama Thomas Bayes yang menggunakan konsep teorema bayes atau probabilitas untuk memprediksi kelas data. Naïve bayes dikenal sebagai algoritma dengan perhitungan yang sederhana. Data training data penerimaan RTLH pada tahun 2017 sampai 2019 sebanyak 218 Data Penerima Bantuan Renovasi Rumah di Desa Sialang Buah, dan telah dilakukan validasi yang menghasilkan tingkat akurasi sebesar 97,71%, recall 98,50%, dan precision 97,97%. Sedangkan pada testing merupakan data pengusul RTLH pada tahun 2022 sebanyak 58 Data, data testing yang diuji menggunakan algoritma naïve menghasilkan persentasi Menerima 81%, dan persentasi 19% Tidak Menerima.

PENDAHULUAN

Algoritma *Naive Bayes* merupakan sebuah pengklasifikasian probilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan Teorema *Bayes* dan mengamsumsikan sebuah atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Algoritma *Naive Bayes* banyak digunakan dalam menentukan sebuah Keputusan di antaranya pengklasifikasian kelayakan penerima PKH, Penentuan program studi yang tepat pada mahasiswa, yang menghasilkan akurasi yang sangat baik. Dalam hal ini Akan digunakan dalam menentukan Penentuan Bantuan Renovasi Rumah Di Desa Sialang Buah[1].

Rumah merupakan kebutuhan dasar dalam melaksanakan peran sosial bagi anggota keluarga. Fakir miskin pun memerlukan rumah yang layak dan nyaman. Rumah menjadi kebutuhan hidup untuk tempat tinggal dalam jangka waktu tertentu. Setiap keluarga seharusnya memiliki rumah yang layak dan sesuai dengan Undang-Undang Republik Indonesia tentang Perumahan dan kawasan permukiman bahwa negara bertanggung jawab melindungi segenap bangsa indonesia melalui penyelenggaraan perumahan dan kawasan permukiman agar masyarakat mampu bertempat tinggal serta menghuni rumah yang layak dan terjangkau di dalam perumahan yang sehat, aman, harmonis, dan berkelanjutan di seluruh wilayah indonesia.

Di Indonesia masih banyak sekali rumah yang masuk dalam kategori layak tak huni. Pemerintah memberikan bantuan Rumah Tidak Layak Huni (RTLH) berupa uang untuk pembelian bahan bangunan guna pemugaran Rumah Tidak Layak Huni dimana rumah tersebut tidak memenuhi persyaratan keselamatan 2 bangunan dan kecukupan minimal bangunan. Penyaluran bantuan Rumah Tidak Layak Huni ditangani oleh pemerintah daerah yang disalurkan kepada kelurahan/desa yang masuk dalam kategori kelurahan/desa miskin dengan prioritas tinggi[2].

Kelurahan / Desa Sialang Buah adalah satu desa yang terdapat di Kecamatan Teluk Mengkudu Kabupaten Serdang Bedagai Provinsi Sumatera Utara yang telah menjalankan Program bantuan renovasi Rumah Tidak Layak Huni (RTLH). Namun, dalam menentukan sasaran penangan bantuan Rumah Tidak Layak Huni, Pemerintah Desa Sialang Buah memiliki beberapa kendala. Kendala tersebut dikarenakan banyaknya data yang berubah sehingga memungkinkan proses penentuan penerima bantuan Rumah Tidak Layak Huni (RTLH) tidak tepat sasaran. Untuk itu mengatasi masalah tersebut, dibutuhkan Algoritma Klasifikasi *Naive Bayes* Dalam Penentuan Bantuan Renovasi Rumah Di Desa Sialang Buah, agar dapat mengetahui sudah layak menerima atau tidak layak menerima berdasarkan jumlah nilai akurasi yang diperoleh ketika data sudah diolah.

BAHAN DAN METODE PENELITIAN

2.1 Bahan

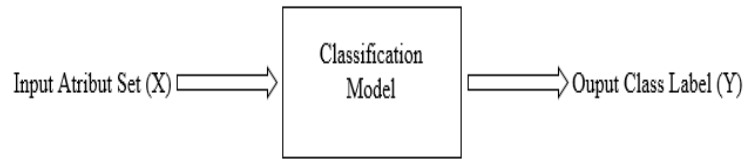
2.1.1 Data Mining

Data mining adalah rangkaian proses untuk menemukan nilai tambah berupa informasi yang selama ini belum diketahui secara manual dari database[3]. Data mining adalah ilmu yang tujuan utamanya adalah untuk mencari, menemukan atau mengekstrak pengetahuan dari data atau informasi yang kita miliki[4].

Data mining adalah metode yang dapat membantu para penggunanya untuk mengakses data yang besar dalam waktu yang relatif singkat. Dengan kata lain, data mining adalah alat dan aplikasi yang menggunakan analisis statistik pada data melalui suatu proses ekstraksi atau penggalian data dan informasi yang sebelumnya tidak diketahui. Data mining adalah proses mengekstraksi data untuk menemukan informasi terbaru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar, sehingga aktivitas penambangan adalah memeriksa database yang berukuran besar untuk menemukan pola atau bentuk yang baru sehingga berguna dalam proses pengambilan keputusan[5].

2.1.2 Klasifikasi

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya kedalam satu kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi terdapat dua pekerjaan utama yang dilakukan, yaitu pembangunan model prototipe untuk disimpan sebagai memori, dan penggunaan model tersebut untuk melakukan pengenalan/klasifikasi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpan[6]. Klasifikasi merupakan proses Pembelajaran suatu fungsi tujuan (target) f yang dimetakan tiap himpunan atribut x ke suatu label kelas y yang didefinisikan sebelumnya. Pembelajaran tersebut Akan menghasilkan suatu model yang kemudian disimpan sebagai memori[7].



Gambar 1. Klasifikasi Sebagai Pemetaan atribut input x ke label y

2.1.3 RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*). *RapidMiner* adalah perangkat lunak yang digunakan untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. Pada *RapidMiner* terdapat berbagai teknik deskriptif dan prediksi untuk memberikan wawasan kepada pengguna sehingga dapat membantu dalam pembuatan keputusan yang paling baik. *RapidMiner* memiliki kurang lebih 500 operator, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. *RapidMiner* ditulis dengan menggunakan bahasa *java* sehingga dapat bekerja di semua sistem operasi[8].

2.1.4 Flowchart

Flowchart atau bagan alur adalah bagan (diagram) alir yang merupakan sekumpulan simbol-simbol atau skema yang menunjukkan kegiatan-kegiatan program dari awal sampai akhir. Tujuan utama penggunaan *flowchart* adalah untuk menyederhanakan rangkaian proses atau prosedur untuk memudahkan pemahaman pengguna terhadap informasi tersebut. *Flowchart* menggunakan berbagai simbol yang terstandarisasi secara internasional. Hal ini ditujukan untuk memudahkan setiap orang memahami berbagai variasi *flowchart*. Simbol-simbol *flowchart* dapat dikategorikan ke dalam empat bagian, yaitu simbol *input/output*, proses, penyimpanan dan lainnya[9].

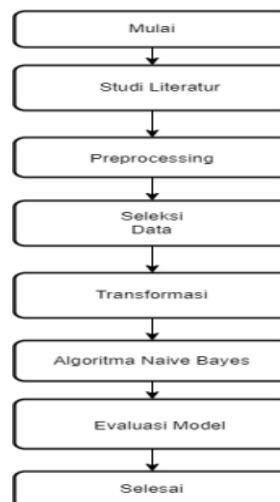
2.1.5 Bantuan Rumah Tidak Layak Huni

Menurut Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia Tentang Bantuan Stimulan Perumahan Swadaya Nomor 07/PRT / M/2018, Rumah adalah bangunan gedung yang berfungsi sebagai tempat tinggal yang layak huni, sarana pembinaan keluarga, cerminan harkat, dan martabat penghuninya, serta aset bagi pemiliknya. Rumah Tidak Layak Huni yang adalah rumah yang tidak memenuhi persyaratan keselamatan bangunan, kecukupan minimum luas bangunan dan kesehatan penghuni[10].

Bantuan Stimulan Perumahan Swadaya merupakan bantuan pemerintah bagi masyarakat berpenghasilan rendah untuk mendorong dan meningkatkan keswadayaan dalam peningkatan kualitas rumah dan pembangunan baru rumah beserta prasarana, sarana, dan utilitas umum. Masyarakat berpenghasilan rendah yang dimaksud merupakan masyarakat yang mempunyai keterbatasan daya beli sehingga perlu mendapat dukungan pemerintah untuk memperoleh rumah. 22 Pemerintah terus berupaya mengentaskan warga dari kemiskinan, termasuk pemberian bantuan-bantuan keuangan. Salah satunya adalah melalui Rumah Tidak Layak Huni. Bantuan sosial ini sumbernya tidak hanya dari APBN, tapi juga melalui APBD Provinsi, APBD Kabupaten, APBDDes, dan CSR,”[11]. Bentuk bantuan yang diberikan kepada penerima merupakan berupa uang sebagaimana dimaksud pada pasal 4 ayat (1), bentuk uang yang diberikan digunakan untuk memberi bahan bangunan dan membayar upah kerja

2.2 Metode

Adapun tahapan-tahapan metodologi penelitian yang diterapkan pada penelitian ini adalah seperti berikut:



Gambar 2 Kerangka Berpikir

2.2.1 Pengumpulan Data

Pengumpulan data Pada tahap ini sumber Data yang Akan digunakan dalam penelitian ini adalah Data Primer yang diperoleh dari pihak institusi yang bersangkutan dengan cara melakukan wawancara langsung dengan Kepala Desa, Sekertaris Desa Sialang Buah Kecamatan Teluk Mengkudu.

2.2.2 Studi Literatur

Studi literatur Pada Tahap ini penulis melakukan pengumpulan bahan referensi yaitu berupa buku, artikel, paper, jurnal, dan makalah, maupun situs-situs dari Internet. Studi literatur yang dilakukan berkaitan dengan Analisis Klasifikasi dan metode Naïve Bayes.

2.2.3 Analisis Dan Perancangan

Analisis dan perancangan Pada tahap ini dilakukan analisis dengan membagi menjadi dua menjadi data *training* dan data *testing*. Metode yang digunakan untuk pengolahan data menggunakan metode Klasifikasi *Naïve Bayes*. *Tools* yang akan digunakan untuk implementasi menggunakan software *RapidMiner* Pengujian yang Akan dilakukan menggunakan model *Confusion Matrix* untuk mengetahui akurasi dari yang diperoleh.

Tahapan data mining merupakan sebuah rangkaian proses yang dapat dibagi menjadi beberapa tahap. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantara knowledge base. Tahapan yang dilakukan pada proses data mining diawali dengan tahap pre-processing, menyeleksi data lalu, transformasi, Algoritma data mining serta tahap interpretasi atau evaluasi. Berikut merupakan gambar dari tahapan data mining[12].

1. Preprocessing

Merupakan proses menghilangkan noise dan data yang tidak konsisten atau data yang tidak relevan. Sebelum proses data mining dilaksanakan, perlu dilakukan sebuah proses pembersihan pada data mencakup pembuangan duplikasi data, pemeriksaan data yang inkonsisten, dan memperbaiki kesalahan pada data. Pembersihan data juga Akan mempengaruhi performansi dari teknik data mining karena data yang ditangani Akan berkurang jumlah dan kompleksitasnya.

2. Seleksi Data

Sekumpulan data operasional yang perlu dilakukan pemilihan (seleksi) sebelum tahap penggalian informasi dalam KDD (*knowledge discovery in database*) dimulai. Data hasil seleksi digunakan untuk proses data mining lalu disimpan dalam suatu berkas dan terpisah dari basis data operasional.

3. Transformasi

Merupakan proses merubah atau menggabungkan data kedalam format yang sesuai untuk diproses kedalam sebuah data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan.

4. Algoritma *Naïve Bayes*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pola atau informasi menarik yang ada didalam data terpilih dengan menggunakan teknik atau metode tertentu. Metode dalam data mining sangat bervariasi. Pemilihan metode yang tepat sangat bergantung pada tujuan dan proses secara keseluruhan. Adapun metode yang digunakan menggunakan algoritma *naïve bayes*. Algoritma ini menggunakan konsep teorema *bayes* atau probabilitas untuk memprediksi kelas data. *Naïve bayes* dikenal sebagai algoritma dengan perhitungan yang sederhana, jika digunakan untuk klasifikasi maka algoritma ini lebih dikenal dengan *naïve bayes classifier*[13]. Untuk melakukan klasifikasi *naïve bayes*, diperlukan langkah-langkah sebagai berikut:

- a. Menghitung probabilitas tiap label, pada tahap ini label ini biasa disebut juga dengan kelas atau kategori. Dan untuk perhitungannya jumlah tiap label Akan dijumlahkan lalu dibagi dengan jumlah banyak data.

$$P(Y_i) = \frac{\text{Jumlah Kelas } (Y_i)}{\text{Jumlah Data Latih}} \quad (1)$$

- b. Menghitung probabilitas kasus per label.

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \quad (2)$$

Keterangan:

- Y : data dengan kelas yang belum diketahui
- X : hipotesis data Y merupakan suatu kelas spesifik
- P(C|X) : probabilitas hipotesis C berdasarkan kondisi X
- P(C) : probabilitas hipotesis C
- P(X|C) : probabilitas X berdasarkan kondisi pada hipotesis C
- P(X) : probabilitas dari X

- c. Menghitung semua hasil variabel data testing pada tiap label

$$P(C|X) = P(X_1|C) \times P(X_2|C) \times \dots \times P(X_n|C) \quad (3)$$

- d. Membandingkan hasil probabilitas tiap label. Pada tahap ini data tiap label Akan dibandingkan dengan nilai yang Maksimum Akan menjadi hasil label untuk data *testing*.

5. Evaluasi Model

Merupakan tahapan untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Pola informasi yang dihasilkan dalam tahap ini berupa pola yang khas maupun model prediksi untuk dievaluasi apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya. Evaluasi dilakukan untuk mengetahui kinerja dan akurasi dari metode klasifikasi yang telah diterapkan. Confusion matrix merupakan tabel yang terdiri dari banyaknya baris data uji yang diprediksi benar dan salah oleh model klasifikasi yang kemudian digunakan untuk menentukan kinerja dan akurasi dari model klasifikasi [14].

Class	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP	FP
Negatif	FN	TN

Gambar 3. Confusion Matrix

(Sumber: Mitra Sunengsih Hasibuan, 2022)

Keterangan:

- TP adalah *True Positive*, yaitu jumlah data yang nilai aktualnya adalah positif dan terklasifikasi sebagai positif.
- TN adalah *True Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif dan terklasifikasi sebagai negatif.
- FN adalah *False Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif namun terklasifikasi salah.
- FP adalah *False Positive*, yaitu jumlah data yang nilai aktualnya adalah positif namun terklasifikasi salah.

Setelah mengetahui jumlah nilai *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), dan *True Positive* (TP), dapat dilakukan perhitungan selanjutnya untuk mencari nilai *precision*, *recall* dan *accuracy*.

- Precision* adalah sebuah tingkat keberhasilan antara informasi yang diminta dengan jawaban yang diberikan. Perhitungan *Precision* dinyatakan dengan persamaan.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

- Recall* adalah sebuah tingkat keberhasilan dalam menemukan kembali sebuah informasi, perhitungan *Recall* dinyatakan dengan persamaan.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

- Sedangkan *Accuracy* adalah sebuah tingkat kedekatan antara nilai prediksi dengan nilai sebenarnya. Perhitungan *Accuracy* dinyatakan dengan persamaan.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

HASIL DAN IMPLEMENTASI MODEL

3.1 Hasil

Dalam melakukan proses klasifikasi pertama kali melakukan proses pembersihan data yang terduplikasi atau Sama dengan data lain, proses ini pada tahap *preprocessing*. Setelah melakukan proses *preprocessing*, lalu melakukan menyeleksi data atribut yang akan digunakan, selanjutnya tahapan transformasi yang digunakan untuk merubah data dalam bentuk angka menjadi bentuk kategorial. Setelah melakukan tahap transformasi, Akan melakukan tahapan analisis data mining menggunakan metode *Naïve Bayes*, Setelah melakukan proses analisis datamining maka selanjutnya melakukan proses evaluasi untuk mengetahui tingkat akurasi yang dicapai.

3.1.1 Pengumpulan Data

Data yang diperoleh berupa data bantuan renovasi rumah Desa Sialang Buah Tahun 2017 sampai 2019 dalam bentuk format excel sebanyak 219 Data yang terdiri dari Data Penerima Bantuan RTLH dan Tidak Penerima Bantuan RTLH dimana data tersebut akan dijadikan sebagai data latih, dan Tahun 2022 terdapat 58 data yang sedang diusulkan yang dijadikan sebagai data uji.

3.1.2 Preprocessing

Berdasarkan hasil Preprocessing terdapat 2 data yang Sama dari 219 data, sehingga data yang sama harus dihapus salah satu dan data yang diperoleh menjadi 218 Data.

3.1.3 Seleksi data

Setelah dilakukan proses Pembersihan pada data yang terduplikat, lalu tahap ini akan menentukan atribut yang digunakan untuk klasifikasi penerimaan bantuan renovasi rumah tidak layak huni. Berikut variabel dan label yang dipakai Desa Sialang Buah dalam menentukan penerimaan bantuan renovasi rumah tidak layak huni.

Tabel 1. Variabel atau Atribut

Atribut	Klasifikasi Atribut
Kepemilikan Lahan	Milik Sendiri
	Tidak Milik Sendiri
Atap	Rumbia
	Seng

Atribut	Klasifikasi Atribut
Dinding	Papan
	Tepas
Lantai	Semen
	Papan
	Keramik
	Tanah
Penghasilan	Penghasilan Rendah
	Penghasilan Sedang
	Penghasilan Tinggi

Tabel 2. Label atau Class

Label	Klasifikasi Label
Status Menerima	Menerima
	Tidak Menerima

3.1.4 Transformasi Data

Tahapan untuk melakukan perubahan variabel menjadi dalam bentuk kategorial pada variabel penghasilan menjadi Penghasilan Rendah, penghasilan Sedang, Penghasilan Tinggi.

3.1.5 Algoritma Naïve Bayes

Jika terdapat contoh data uji seperti tabel dibawah, Maka

Tabel 3. Sampel Data Uji

Kepemilikan Lahan	Atap	Dinding	Lantai	Penghasilan	Status Menerima
Tidak Milik Sendiri	Seng	Papan	Semen	Sedang	?

- a. Menghitung probabilitas tiap label

Tabel 4. Hasil Hitung Probabilitas Class/Label

Total	Jumlah Data	Class / label		P(C X)	
		Menerima	Tidak Menerima	Menerima	Tidak Menerima
	218	134	84	0,615	0,385

- b. Menghitung probabilitas kasus per label

Tabel 5. Hasil Hitung Probabilitas kasus per label

Atribut / Variabel		Jumlah Data	Class / label		P(C X)	
			Menerima	Tidak Menerima	Menerima	Tidak Menerima
Total		218	134	84	0,615	0,385
Kepemilikan Lahan	Milik Sendiri	136	133	3	0,992	0,036
	Tidak Milik sendiri	82	1	81	0,007	0,964
Atap	Rumbia	58	41	17	0,306	0,202
Dinding	Seng	160	93	67	0,694	0,798
	Papan	126	74	52	0,552	0,619
	Tepas	92	60	32	0,448	0,381
Lantai	Keramik	1	0	1	0,000	0,012
	Papan	4	1	3	0,007	0,036
	Semen	206	129	77	0,963	0,916
	Tanah	7	4	3	0,030	0,036
Penghasilan	Rendah	55	42	13	0,313	0,155
	Sedang	119	69	50	0,515	0,595
	Tinggi	44	23	21	0,172	0,250

- c. Menghitung semua hasil variabel data testing pada tiap label

Tabel 6. Hasil Hitung Variabel Data Testing pada setiap class/label

Atribut / Variabel		P(C X)	
		Menerima	Tidak Menerima
Total		0,615	0,385
Kepemilikan Lahan	Tidak Milik Sendiri	0,007	0,964
Atap	Seng	0,694	0,798
Dinding	Papan	0,552	0,619
Lantai	Semen	0,963	0,916
Penghasilan	Sedang	0,515	0,595
Total		?	?

Selanjutnya mengkalikan semua nilai. Hasilnya sesuai dengan data pada Menerima dan Tidak Menerima.

$$P(\text{Menerima}) \times P(\text{Kepemilikan Lahan} = \text{Tidak Milik Sendiri} | \text{Menerima}) \times P(\text{Atap} = \text{Seng} | \text{Menerima}) \times P(\text{Dinding} = \text{Papan} | \text{Menerima}) \times P(\text{Lantai} = \text{Semen} | \text{Menerima}) \times P(\text{Penghasilan} = \text{Sedang} | \text{Menerima})$$

$$= 0,615 \times 0,007 \times 0,694 \times 0,552 \times 0,963 \times 0,515$$

$$= 0,000817909$$

$$P(\text{Tidak Menerima}) \times P(\text{Kepemilikan Lahan} = \text{Tidak Milik Sendiri} \mid \text{Tidak Menerima}) \times P(\text{Atap} = \text{Seng} \mid \text{Tidak Menerima}) \times$$

$$P(\text{Dinding} = \text{Papan} \mid \text{Tidak Menerima}) \times P(\text{Lantai} = \text{Semen} \mid \text{Tidak Menerima}) \times P(\text{Penghasilan} = \text{Sedang} \mid \text{Tidak Menerima})$$

$$= 0,385 \times 0,964 \times 0,798 \times 0,619 \times 0,916 \times 0,595$$

$$= 0,0999188002$$

- d. Membandingkan hasil probabilitas tiap label.

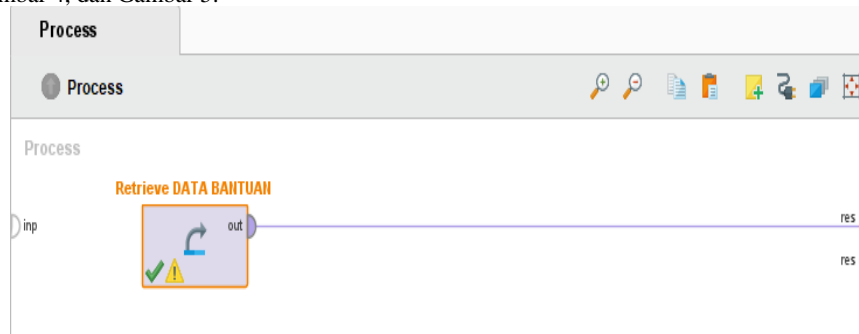
Hasil nilai tertinggi tersebut akan menjadi hasil dari label data uji. Dikarenakan Nilai P (Tidak Menerima) 0,0999188002 Lebih Besar dari P (Menerima) 0,000817909 yaitu maka hasil dari data uji **Tidak Menerima**.

3.2 Implementasi Model

Implementasi model merupakan tahapan yang dilakukan untuk menerapkan model yang telah dilakukan sebelumnya menggunakan perangkat lunak *RapidMiner Studio Educational 10.0.00*. Berikut langkah tampilan kinerja *rapidminer* disajikan pada gambar berikut.

3.2.1 Import Data

Pada tahap ini melakukan persiapan data yang akan diolah, yaitu data bantuan renovasi rumah di Desa Sialang Buah pada tahun 2017-2019 sebanyak 219 data. Data dibuat kedalam repository dengan format *Microsoft Excel*. Berikut tampilan import data dapat dilihat pada Gambar 4, dan Gambar 5.



Gambar 4. Gambar Import Data

<new process*> - RapidMiner Studio Educational 10.0.000 @ DESKTOP-9JCC8A2

File Edit Process View Connections Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Result History ExampleSet (Retrieve DATA BANTUAN) ExampleSet (/Temporary Repository/DATA BANTUAN)

Open in Turbo Prep Auto Model Filter (219 / 219 examples): all

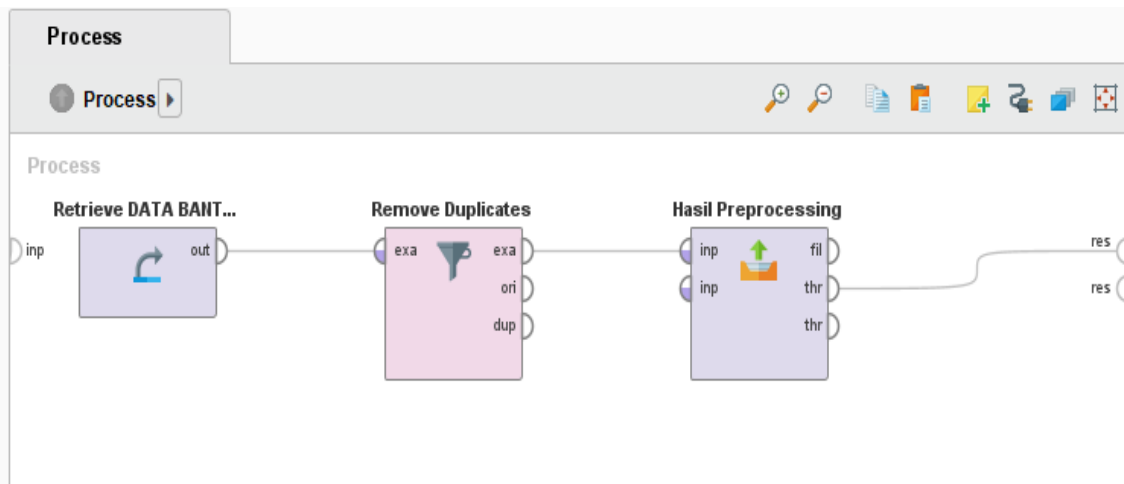
tatus Pene...	Nama Kepal...	Program	Tahun	Alamat	Kepemilikan...	Atap	Dinding	Lantai
lenerima	Horas Pardede	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
lenerima	Monang Sam...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
lenerima	Eliper Siahaan	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
lenerima	Lidia Br Siani...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
lenerima	Polman Pard...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
lenerima	Butti Manalu	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
lenerima	Tongam Nab...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
lenerima	Tomuanna Br...	BSPS	2018	Dusun 1	Milik Sendiri	Rumbia	Papan	Semen
lenerima	Marto Naingg...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
lenerima	Tionggar Sim...	BSPS	2018	Dusun 1	Milik Sendiri	Rumbia	Tepas	Semen
lenerima	Damson Par...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
lenerima	Marsinta Br P...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
lenerima	Efelina Br Par...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
lenerima	Evalina Br Si...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen

ExampleSet (219 examples, 1 special attribute, 9 regular attributes)



Gambar 5. Hasil Import Data

3.2.2 Preprocessing

Setelah data diimport sebanyak 219 data, maka tahap selanjutnya *preprocessing* untuk membersihkan data yang terduplikat, hasil *preprocessing* terdapat 2 data yang sama, selanjutnya akan dihapus menjadi 1 bagian, sehingga data menjadi sebanyak 218 data. Berikut tampilan tahap *preprocessing* dapat terlihat pada Gambar 6 dan Gambar 7.

Gambar 6. Gambar *Preprocessing*

ExampleSet (Remove Duplicates) ExampleSet (/Temporary Repository/DATA BANTUAN)

Open in  Turbo Prep  Auto Model Filter (218 / 218 examples): all

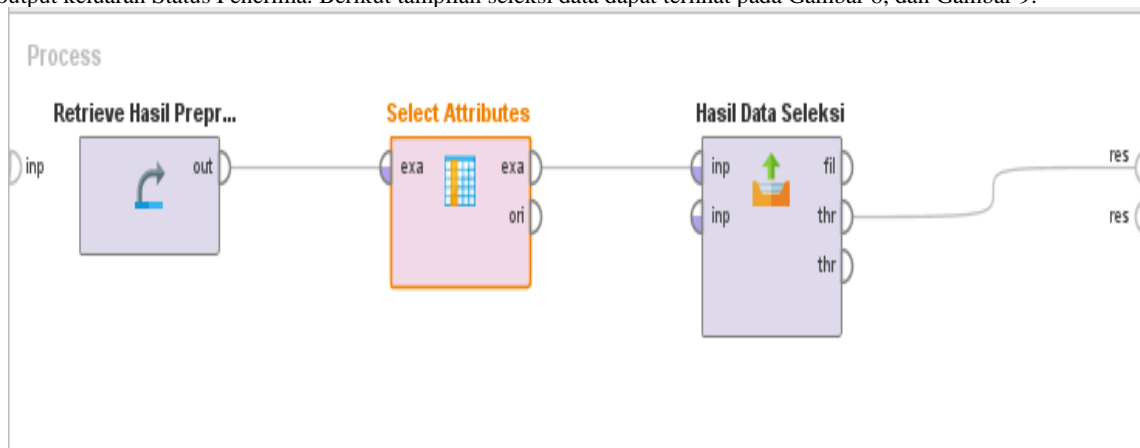
Pene...	Nama Kepala...	Program	Tahun	Alamat	Kepemilikan...	Atap	Dinding	Lantai
ma	Horas Pardede	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
ma	Monang Sam...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
ma	Eliper Siahaan	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
ma	Lidia Br Siani...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
ma	Polman Pard...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
ma	Butti Manalu	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
ma	Tongam Nab...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
ma	Tomuanna Br...	BSPS	2018	Dusun 1	Milik Sendiri	Rumbia	Papan	Semen
ma	Marito Naingg...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
ma	Tionggar Sim...	BSPS	2018	Dusun 1	Milik Sendiri	Rumbia	Tepas	Semen
ma	Damson Par...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
ma	Marsinta Br P...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen
ma	Efelina Br Par...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Tepas	Semen
ma	Evalina Br Si...	BSPS	2018	Dusun 1	Milik Sendiri	Seng	Papan	Semen

ExampleSet (218 examples, 1 special attribute, 9 regular attributes)

Gambar 7. Hasil *Preprocessing*

3.2.3 Seleksi Data

Selanjutnya data hasil *preprocessing* akan dimport kembali untuk dilakukan tahap pemilihan *class* atau *label*, dan atribut yang akan digunakan. Data atribut yang akan digunakan berupa Kepemilikan lahan, Atap, Dinding, Lantai, Penghasilan, dan *class* atau output keluaran Status Penerima. Berikut tampilan seleksi data dapat terlihat pada Gambar 8, dan Gambar 9.



Gambar 8. Seleksi Data

ExampleSet (Select Attributes)

Open in Turbo Prep Auto Model Filter (218 / 218 examples): all

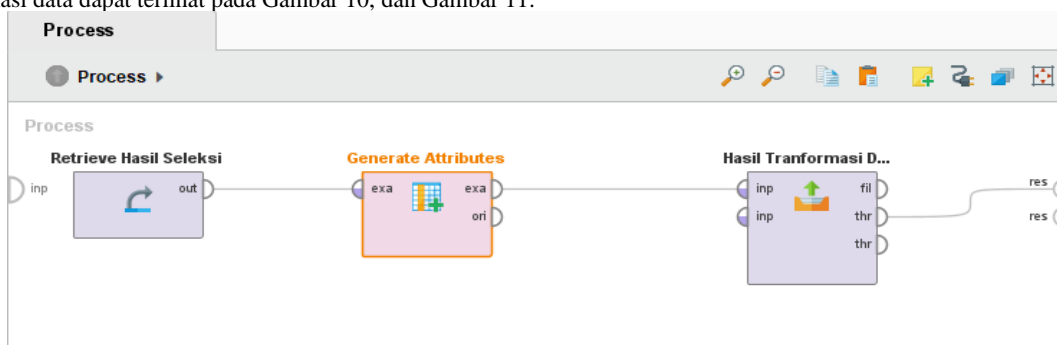
Row No.	Status Pene...	Kepemilikan...	Atap	Dinding	Lantai	Penghasilan
1	Menerima	Milik Sendiri	Seng	Papan	Semen	1300000
2	Menerima	Milik Sendiri	Seng	Tepas	Semen	2000000
3	Menerima	Milik Sendiri	Seng	Papan	Semen	1500000
4	Menerima	Milik Sendiri	Seng	Papan	Semen	1200000
5	Menerima	Milik Sendiri	Seng	Tepas	Semen	2000000
6	Menerima	Milik Sendiri	Seng	Tepas	Semen	3500000
7	Menerima	Milik Sendiri	Seng	Papan	Semen	3000000
8	Menerima	Milik Sendiri	Rumbia	Papan	Semen	1500000
9	Menerima	Milik Sendiri	Seng	Tepas	Semen	1800000
10	Menerima	Milik Sendiri	Rumbia	Tepas	Semen	1800000
11	Menerima	Milik Sendiri	Seng	Papan	Semen	1500000
12	Menerima	Milik Sendiri	Seng	Papan	Semen	1000000
13	Menerima	Milik Sendiri	Seng	Tepas	Semen	1000000
14	Menerima	Milik Sendiri	Seng	Papan	Semen	1500000
15	Tidak Meneri...	Tidak Milik Se...	Seng	Papan	Semen	2500000

ExampleSet (218 examples, 1 special attribute, 5 regular attributes)

Gambar 9. Hasil Seleksi Data

3.2.4 Transformasi Data

Tahap selanjutnya tahapan transformasi, data yang akan ditransformasi diimport kembali menggunakan hasil data seleksi dengan atribut data penghasilan akan dirubah menjadi kategorial jenis Penghasilan Rendah, Sedang, dan Tinggi. Berikut tampilan transformasi data dapat terlihat pada Gambar 10, dan Gambar 11.



Gambar 10. Transformasi Data

ExampleSet (Generate Attributes)

Open in Turbo Prep Auto Model Filter (218 / 218 examples): all

Row No.	Status Pene...	Kepemilikan...	Atap	Dinding	Lantai	Penghasilan
1	Menerima	Milik Sendiri	Seng	Papan	Semen	rendah
2	Menerima	Milik Sendiri	Seng	Tepas	Semen	sedang
3	Menerima	Milik Sendiri	Seng	Papan	Semen	sedang
4	Menerima	Milik Sendiri	Seng	Papan	Semen	rendah
5	Menerima	Milik Sendiri	Seng	Tepas	Semen	sedang
6	Menerima	Milik Sendiri	Seng	Tepas	Semen	tinggi
7	Menerima	Milik Sendiri	Seng	Papan	Semen	tinggi
8	Menerima	Milik Sendiri	Rumbia	Papan	Semen	sedang
9	Menerima	Milik Sendiri	Seng	Tepas	Semen	sedang
10	Menerima	Milik Sendiri	Rumbia	Tepas	Semen	sedang
11	Menerima	Milik Sendiri	Seng	Papan	Semen	sedang
12	Menerima	Milik Sendiri	Seng	Papan	Semen	rendah
13	Menerima	Milik Sendiri	Seng	Tepas	Semen	rendah
14	Menerima	Milik Sendiri	Seng	Papan	Semen	sedang
15	Tidak Meneri...	Tidak Milik Se...	Seng	Papan	Semen	tinggi

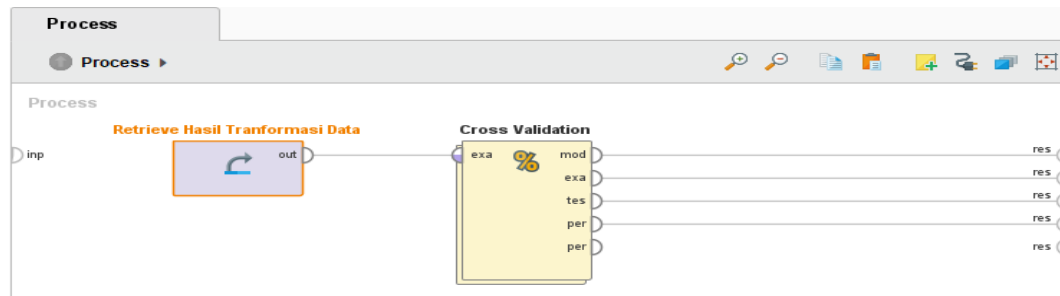
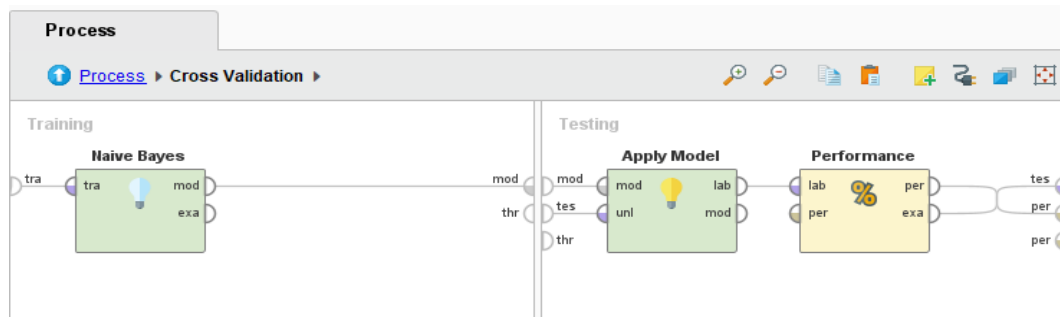
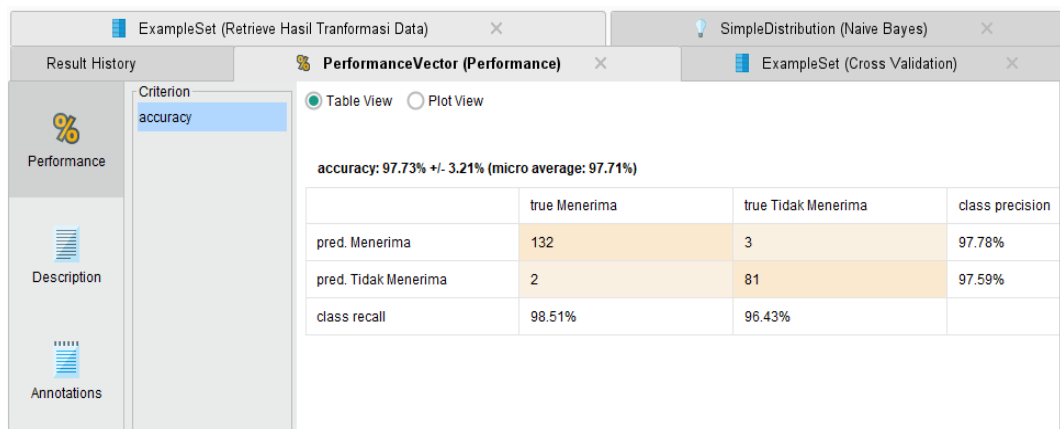
ExampleSet (218 examples, 1 special attribute, 5 regular attributes)

Gambar 11. Hasil Transformasi Data

3.2.5 Algoritma Naïve Bayes

1. Implementasi Algoritma Naïve Bayes pada Data Training

Tahapan selanjutnya melakukan validasi data *training* menggunakan *cross validation*, sebelum melakukan validasi terlebih dahulu melakukan import data terhadap data yang sudah di transformasi pada tahap sebelumnya. Selanjutnya menghubungkan data transformasi ke operator *cross validation* untuk mengetahui tingkat akurasi yang diperoleh pada data *training*. Berikut tampilan dapat terlihat pada Gambar 12, dan Gambar 13.

Gambar 12. Menghubungkan Hasil Transformasi ke operator *Cross Validation*Gambar 13. Model *Cross Validation*

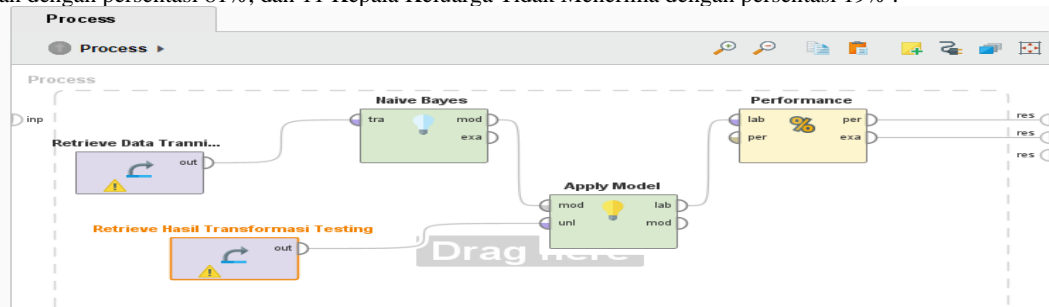
Gambar 14. Hasil Performance Tranning

Berdasarkan pada Gambar 14 diperoleh akurasi sebesar 97,71 % dengan luaran terdapat sebanyak 81 data yang dinyatakan Tidak Menerima diprediksi benar, sebanyak 132 data dinyatakan Menerima diprediksi dengan benar, 2 data dinyatakan Menerima diprediksi salah, dan 3 data dinyatakan Tidak menerima diprediksi benar, dan luaran *simpleDistribution label* pada Status Penerima sebesar 0,615 pada *class* Menerima, dan 0,385 pada *class* Tidak Menerima.

Gambar 15. Hasil Probabilitas *Class/Label*

2. Implementasi Algoritma *Naïve Bayes* pada data *testing*

Pada tahapan selanjutnya menerapkan model pada data *testing* seperti Gambar 16, sehingga menghasilkan class atau *label* baru. Berdasarkan jumlah data *testing* yang akan diuji sebanyak 58 data, diperoleh sebanyak 47 Kepala Keluarga Menerima bantuan dengan persentasi 81%, dan 11 Kepala Keluarga Tidak Menerima dengan persentasi 19% .



Gambar 16. Tampilan Data *Testing*

Row No.	Status Mene...	prediction(S...	confidence(...	confidence(...	Kepemilikan...	Atap	Dinding	Lan
1	?	Menerima	0.976	0.024	Milik Sendiri	Seng	Tepas	Sen
2	?	Menerima	0.976	0.024	Milik Sendiri	Seng	Tepas	Sen
3	?	Menerima	0.970	0.030	Milik Sendiri	Seng	Tepas	Sen
4	?	Menerima	0.976	0.024	Milik Sendiri	Seng	Tepas	Sen
5	?	Menerima	0.982	0.018	Milik Sendiri	Rumbia	Papan	Sen
6	?	Tidak Meneri...	0.007	0.993	Tidak Milik Se...	Seng	Papan	Sen
7	?	Menerima	0.992	0.008	Milik Sendiri	Rumbia	Papan	Sen
8	?	Menerima	0.987	0.013	Milik Sendiri	Seng	Papan	Sen
9	?	Menerima	0.994	0.006	Milik Sendiri	Rumbia	Tepas	Sen
10	?	Menerima	0.994	0.006	Milik Sendiri	Rumbia	Tepas	Sen
11	?	Menerima	0.986	0.014	Milik Sendiri	Rumbia	Tepas	Sen
12	?	Menerima	0.987	0.013	Milik Sendiri	Seng	Papan	Sen
13	?	Menerima	0.994	0.006	Milik Sendiri	Rumbia	Tepas	Sen
14	?	Menerima	0.976	0.024	Milik Sendiri	Seng	Tepas	Sen

Gambar 17. Hasil Data *Testing*

KESIMPULAN

Pada implementasi data tranning, data yang digunakan sebanyak 218 Data Penerima Bantuan Renovasi Rumah di Desa Sialang Buah, dan telah dilakukan validasi yang menghasilkan tingkat akurasi sebesar 97, 71 %, dengan Probabilitas kelas Menerima sebesar 0,615 dan Probabilitas kelas Tidak Menerima sebesar 0,385. Di mana sebanyak 81 data yang dinyatakan Tidak Menerima diprediksi benar, sebanyak 132 data dinyatakan Menerima diprediksi dengan benar, 2 data dinyatakan Menerima diprediksi salah, dan 3 data dinyatakan Tidak menerima diprediksi benar.

Implementasi pada data testing dengan menggunakan algoritma naïve bayes mampu menampilkan prediksi pada data testing yaitu data pengusul penerimaan rumah di Desa Sialang Buah. Sebanyak 58 data testing dengan hasil persentasi Menerima 81 %, dan persentasi 19 % Tidak Menerima.

DAFTAR PUSTAKA

- [1] R. A. Tyas, M. Anggraini, I. A. Sulasyah, and Q. Aini, "Implementasi Algoritma Naïve Bayes Dalam Penentuan Rating Buku," *Sistemasi*, vol. 9, no. 3, p. 557, 2020, doi: 10.32520/stmsi.v9i3.915.
- [2] A. Muhidin and I. Baragigiratri, "PEMETAAN PENDUDUK CALON PENERIMA BANTUAN RENOVASIRUMAH DESA PESANGKALAN MENGGUNAKAN ALGORITMACLUSTERINGK-MEANS," *J. Teknol. Pelita Bangsa SIGMA*, vol. 8, no. 2407–3903, pp. 147–152, 2018.
- [3] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [4] S. Susanto and D. Suryani, "Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data," *Buku*, pp. 1–3, 2010.
- [5] Y. Ardilla et al., *DATA MINING DAN APLIKASINYA*, 1st ed. Bandung: Penerbit Widina, 2021.
- [6] E. Prasetyo, *DATA MINING - Konsep dan Aplikasi Menggunakan MATLAB*, 1st ed. Yogyakarta: CV. ANDI OFFSET, 2012.
- [7] F. Astuti Hernawati, *DATA MINING*, 1st ed. Yogyakarta: CV. ANDI OFFSET, 2013.
- [8] Aprilla Dennis, "Belajar Data Mining dengan RapidMiner," *Innov. Knowl. Manag. Bus. Glob. Theory Pract. Vols 1 2*, vol. 5, no. 4, pp. 1–5, 2013.
- [9] A. P. Pratama, "Aplikasi Reminder Sms Service Order Kendaraan Di Bengkel Ummagelang Authorized," *J. .Pijar MIPA*,

- vol. 11, p. 12, 2019.
- [10] *Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat Republik Indonesia Tentang Bantuan Stimulan Perumahan Swadaya Nomor 07/PRT / M/2018 Tentang Bantuan Stimulan Perumahan Swadaya*. 2018.
 - [11] Portal Berita Pemerintah Provinsi Jawa Tengah, “Bantuan Program RTLH Bebas Pajak,” 2020. <https://jatengprov.go.id/beritadaerah/bantuan-program-rtlh-bebas-pajak/> (accessed Nov. 21, 2022).
 - [12] M. Z. PRATAMA, “Penentuan Kriteria Dan Penerima Bantuan Pangan Non Tunai Di Desa Parungkuda Dengan Menggunakan Algoritma Naïve Bayes,” 2020.
 - [13] A. P. Ayudhitama and U. Pujianto, “Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer,” *J. Inform. Polinema*, vol. 6, no. 2, pp. 1–9, 2020, doi: 10.33795/jip.v6i2.274.
 - [14] Mitra Sunengsih Hasibuan, “PERBANDINGAN METODE NAÏVE BAYES DENGAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN KEPUASAN PENGGUNA TIKTOK PROGRAM STUDI TEKNIK INFORMATIKA,” 2022.