

Contents list available at www.jurnal.unimed.ac.id

CESS
(Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



Implementasi *Algorithm Naïve Bayes Classification* dalam Menentukan Bantuan Tepat Sasaran Rumah Tidak Layak Huni (RTLH) Di Kabupaten Bojonegoro

Implementation of Naive Bayes Classification Algorithm in Determining Appropriate Help Targets of Unlimited Houses (RTLH) in Bojonegoro District

Anissa Nurul Farida Tussholikhah¹, Nurissaidah Ulinnuha², Wika Dianita Utami³

^{1,2,3} Prodi Matematika UIN Sunan Ampel Surabaya

Gunung Anyar, Kec. Gunung Anyar, Surabaya, Indonesia

email: ¹faridaanisa4@gmail.com, ²nuris.ulinnuha@uinsby.ac.id, ³wikadianita@uinsby.ac.id

ABSTRAK

Rumah merupakan salah satu kebutuhan primer bagi setiap individu, dan termasuk kedalam aset terpenting yang harus dimiliki. Kelayakan rumah yang layak huni dan tidak layak huni harus dipertimbangkan. Rumah yang tidak memenuhi kecukupan minimum dari segi ruang dan luas ruangan dianggap sebagai Rumah tidak layak huni (RTLH). Untuk mengatasi terjadinya peningkatan RTLH maka pemerintah menanggulangnya dengan memberikan bantuan kepada masyarakat yang layak menerima dengan tepat sasaran. Penelitian ini bertujuan untuk menerapkan metode Naïve Bayes dalam menentukan bantuan tepat sasaran menggunakan dua kelas penelitian, yakni layak menerima bantuan RTLH dan tidak layak menerima bantuan RTLH. Dari analisis klasifikasi menggunakan confusion matrix didapatkan hasil akurasi sebesar 63%, recall 100% dan presisi 25% untuk 400 data training dan 100 data testing dari total 500 data dengan 10 atribut pengujian.

Kata Kunci: *RTLH, RLH, Naive Bayes, Klasifikasi, Confusion Matrix*

ABSTRACT

The house is one of the primary needs for every individual, and is included in the most important asset that must be owned. The livability of the livable and uninhabitable houses should be considered. A house that does not meet the minimum adequacy in terms of space and room area is considered an uninhabitable house (RTLH). To overcome the increase in RTLH, the government overcomes it by providing assistance to people who deserve to receive it on target. This study aims to apply the Naïve Bayes method in determining targeted assistance using two research classes, namely eligible to receive RTLH assistance and not eligible to receive RTLH assistance. From the classification analysis using the confusion matrix,

*Penulis Korespondensi:

email: faridaanisa4@gmail.com

the results obtained are 63% accuracy, 100% recall and 25% precision for 400 training data and 100 testing data from a total of 500 data with 10 test attributes.

Keywords: *RTLH, RLH, Naive Bayes, Classification, Confusion Matrix*

1. PENDAHULUAN

Rumah adalah satu diantara kebutuhan pokok bagi suatu individu, dan termasuk kedalam aset terpenting yang harus dimiliki. Kelayakan rumah yang ditempati pun harus dipertimbangkan. Rumah layak huni adalah rumah yang memenuhi keselamatan konstruksi bangunan, kecukupan luas minimum bangunan, dan standar kesehatan penghuni [1]. Disisi lain, rumah tidak layak huni merupakan rumah yang tidak memenuhi persyaratan keamanan bangunan dari kecukupan minimum permukaan bangunan. RTLH diatur dalam pasal 28H ayat (1) UUD 1945 dan UU No. 1 Tahun 2011 tentang implementasi kegiatan rehabilitasi rumah tidak layak huni [2].

Program rehabilitasi rumah tidak layak huni tidak hanya berfokus pada aspek fisik rumah saja, namun tentang bagaimana membangun kapasitas kalangan masyarakat miskin dalam memahami seberapa berartinya tempat tinggal yang layak huni dan aspek sosial di lingkungan keluarga, yang memiliki harapan dapat menjalin rasa kebersamaan dalam hal sosial serta semangat saling membantu di kalangan masyarakat yang telah memudar [3]. Harapan dari program ini adalah dapat memenuhi keperluan masyarakat fakir miskin guna mendapatkan pemukiman yang layak untuk ditempati supaya terhindar dari permasalahan sosial lainnya.

Dilain pihak, RTLH merupakan bagian dari permasalahan pemukiman dan memiliki keterkaitan erat dengan kemiskinan, karena masih banyak masyarakat miskin yang tinggal di rumah tidak layak huni [4]. Menurut data yang terdapat Badan Pusat Statistik (BPS) persentase RTLH di Indonesia pada tahun 2020 mencapai 59,54%. Jika dibandingkan dengan tahun 2019 maka tahun 2020 memiliki persentase yang lebih tinggi, karena pada tahun 2019 persentase jumlah penduduk yang RTLH sebesar 56,51% [5].

Salah satu kabupaten yang mencanangkan program pengelolaan rumah tidak layak huni untuk mengurangi kemiskinan adalah Kabupaten Bojonegoro. Berdasarkan data di BPS Kabupaten Bojonegoro, tingkat kemiskinan hingga maret 2020 masih berjumlah 161,10 ribu jiwa. Karena jumlah penduduk Indonesia tak terkecuali di Kabupaten Bojonegoro tumbuh lebih cepat dari kebutuhan akan rumah layak huni, masyarakat ekonomi rendah yang membutuhkan RLH mengalami ketidakseimbangan dengan meningkatnya standar hidup sehingga mengharuskan mereka hidup dalam RTLH [6].

Beberapa masalah yang menghambat penyebaran bantuan untuk penduduk yang mempunyai rumah tidak layak huni masih ditemukan di Kabupaten Bojonegoro. Seperti, ketimpangan antara data yang dikumpulkan dengan data di lapangan menjadi salah satu penyebab masyarakat menerima bansos secara tidak merata. Selain itu, jumlah anggaran pemerintah daerah yang terbatas juga menjadi salah satu penghambat dalam penyebaran bantuan RTLH ini. Sehingga penyebaran bantuan RTLH pun kurang maksimal. Oleh karena itu perlu dilakukan klasifikasi dalam menentukan bantuan tepat sasaran rumah tidak layak huni.

Mengevaluasi suatu objek data untuk menempatkannya dalam kelas tertentu dari berbagai kelas yang tersedia disebut sebagai klasifikasi [7]. Dengan kata lain, proses menemukan model yang menggambarkan kelas data atau konsep dengan tujuan untuk memprediksi kelas objek yang labelnya tidak diketahui adalah apa yang disebut klasifikasi. *Decision/classification trees, Bayesian classifier/ Naive Bayes classifier, Neural Networks, Analisa statistik, Algoritma*

Genetika, Rough sets, K-Nearest Neighbor, Metode Rule Based, Memory Based Reasoning, dan Support Vector Machine (SVM) adalah jenis klasifikasi yang banyak digunakan [8].

Penelitian yang dilakukan oleh An Naas Shahifatun Na'iema, dkk yang berjudul "Klasifikasi penerima bantuan program rehabilitasi rumah tidak layak huni menggunakan algoritma K-Nearest Neighbor Classification of beneficiaries for the rehabilitation of uninhabitable houses using the K-Nearest Neighbor algorithm," menggunakan metode K-Nearest Neighbor, menghasilkan akurasi sebesar 97,93%, presisi 96,88%, sensitivitas 99,53%, dan AUC sebesar 0,964 pada k bernilai 5 [2]. Naisah Marito Putry, dkk juga Penelitian tentang "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," membandingkan metode KNN dengan metode Naïve Bayes. Didapatkan akurasi dari algoritma Naïve Bayes sebesar 80%, hasil ini lebih tinggi daripada akurasi dari algoritma KNN yaitu sebesar 75% [9]. Selain itu ada juga penelitian dalam pengklasifikasian rumah layak huni di Kabupaten Brebes dengan algoritma Learning Vector Quantization (LVQ) dan Naïve Bayes (NB) menghasilkan akurasi sebesar 71,43% dengan LVQ dan 95,24% dengan NB dari atribut yang digunakan berjumlah tujuh dengan 838 data [2].

Berdasarkan penelitian terdahulu, maka pada penelitian permasalahan bantuan tidak tepat sasaran untuk rumah tidak layak huni di Kabupaten Bojonegoro menerapkan algoritma Naïve Bayes dalam mengolah data RTLH. Untuk mengklasifikasi mana yang berhak menerima bantuan RTLH dan mana yang tidak prioritas untuk menerima RTLH. Penelitian ini diharapkan dapat menjadi panduan untuk pengambilan keputusan dalam penyebaran bantuan RTLH selanjutnya dan diharapkan penyebaran bantuan tersebut akan lebih maksimal daripada sebelumnya. Oleh karena itu, dilakukan penelitian dengan judul penelitian "Implementasi Algoritma Naïve Bayes Classification dalam Menentukan Bantuan Tepat Sasaran Rumah Tidak Layak Huni (RTLH) di Kabupaten Bojonegoro".

2. DASAR/TINJAUAN TEORI

2.1. Rumah Tidak Layak Huni (RTLH)

Yang dimaksud dengan Rumah Layak Huni adalah rumah yang memenuhi kebutuhan kesejahteraan bangunan dan memenuhi wilayah struktur dasar dan kesehatan penghuninya. Sebaliknya, rumah yang tidak memenuhi persyaratan atau standar keselamatan bangunan, luas bangunan minimum, atau kesejahteraan penghuninya dianggap sebagai Rumah Tidak Layak Huni. Bantuan pembangunan rumah yang dijalankan pemerintah seringkali masih tidak tepat sasaran. Ada beberapa rumah yang tergolong layak huni bisa mendapatkan bantuan. Namun sebaliknya, rumah yang tergolong RTLH tak mendapatkan bantuan yang seharusnya didapatkan. Untuk meminimalisir hal tersebut, perlunya mengkaji kriteria-kriteria rumah yang tergolong RTLH.

Kriteria rumah masyarakat yang termasuk kedalam RTLH ialah:

1. Kondisi atap tergolong tidak baik atau dapat dikatakan tergolong kedalam kualitas rendah
2. Kualitas dari dinding rumah tergolong rendah (tidak layak), salah satunya ialah dinding dengan batu bata yang sudah lapuk
3. Lantai tergolong ke dalam kondisi tidak baik, rusak, dengan kata lain termasuk kualitas rendah. Misalnya lantai yang terbuat dari tanah/kayu/semen/tegel
4. Tidak terdapat ventilasi udara yang memadai di sekeliling rumah
5. Tidak terdapat jamban (toilet)/jamban tidak layak di dalam rumah [10]

Masyarakat dengan kriteria tersebut lah yang berhak mendapatkan bantuan sosial RTLH. Bantuan sosial RTLH ini dapat dimaknai sebagai prosedur pemulihan fungsi sosial masyarakat miskin dengan upaya memperbaiki Sebagian atau seluruh rumah yang tidak layak huni, dalam rangka menciptakan kondisi perumahan yang layak sebagai tempat tinggal [10].

RTLH sendiri dimaksudkan untuk memenuhi kebutuhan akan perumahan sebagai tempat tinggal, dengan meningkatkan kesejahteraan sosial masyarakat dan memperbaiki/merenovasi rumah yang tidak layak huni. Hal ini juga bertujuan untuk mengatasi berbagai masalah seperti kemiskinan, ketersediaan perumahan yang layak, kenyamanan hidup, peningkatan kapasitas keluarga untuk mengemban peran dan fungsi perlindungan keluarga, bimbingan dan Pendidikan [10].

2.2. Naive Bayes

Algoritma Naïve Bayes merupakan algoritma probabilistik sederhana berdasarkan teorema bayes. Algoritma Naïve Bayes didasarkan pada asumsi bahwa nilai atribut independent dan tidak terpengaruh oleh atribut lain. Naïve Bayes adalah varian langsung dari *Bayesian Network* dimana setiap atribut independent menerima kelas variabel [11]. Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk menemukan pengetahuan atau pola kesamaan karakteristik dalam kelas atau kelompok tertentu, dan dapat digunakan untuk memprediksi probabilitas suatu kelas.

Naïve Bayes merupakan metode yang tidak memiliki aturan. Dengan melihat frekuensi setiap klasifikasi pada data training, Naïve Bayes memanfaatkan probabilitas untuk memilih klasifikasi terbaik dari opsi yang tersedia. Dengan kata lain, metode yang digunakan untuk mengklasifikasi kelas dari setiap masalah yang dibagi menjadi masalah perhitungan numerik dengan menggunakan pendekatan grup [12].

Persamaan berikut menjelaskan persamaan umum dari Naïve Bayes [13]:

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

Dimana:

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

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

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

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

P(X) : Probabilitas *X*

Pengklasifikasian yang dilakukan dengan Naïve Bayes, perlu dilakukan dua proses antara lain, proses training dan proses testing. Proses training sendiri merupakan proses dimana dilakukannya training set dengan label yang sudah diketahui untuk membangun model pelatihan itu sendiri. Sedangkan proses testing digunakan untuk menguji keakuratan model yang dibuat selama proses training, umumnya data set merupakan data yang biasanya digunakan untuk memprediksi label [14].

Keuntungan lain dari metode Naïve Bayes adalah cepat, mudah, akurat, dan hanya membutuhkan sedikit data training untuk menghitung estimasi parameter yang diperlukan untuk proses klasifikasi [15].

2.3. Confusion Matrix

Confusion Matrix adalah tabel yang menunjukkan berapa banyak baris data uji yang diprediksi model klasifikasi akan benar atau salah. Tabel ini diperlukan untuk mengetahui seberapa baik model klasifikasi bekerja [16]. Model klasifikasi dari *Confusion Matrix* dapat dilihat pada tabel berikut [17].

Tabel 1. Model Klasifikasi *Confusion Matrix*

		Actual Class	
		True = 1	False = 0
Predict Class	True = 1	True Positif (TP)	False Negatif (FN)
	False = 0	False Negatif (FN)	True Negatif (TN)

Dimana:

1. *True Positif* (TP) adalah banyaknya data positif yang diklasifikasikan dengan benar oleh sistem.
2. *True Negatif* (TN) adalah banyaknya data negatif yang diklasifikasikan dengan benar oleh sistem.
3. *False Negatif* (FN) adalah banyaknya data negatif yang diklasifikasikan dengan salah oleh sistem.
4. *False Positif* (FP) adalah banyaknya data positif yang diklasifikasikan dengan salah oleh sistem.

Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *precision*, dan *recall*. Presentase catatan data yang diklasifikasikan dengan benar oleh algoritma dikenal sebagai nilai akurasi, dihitung dengan membandingkan data yang diklasifikasikan benar dengan seluruh kumpulan data [18]. Nilai akurasi dapat diperoleh dengan persamaan sebagai berikut [17]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Proporsi jumlah kasus yang di prediksi positif yang juga benar-benar terjadi sebagai positif sejati dalam data adalah nilai presisi [18]. Dengan kata lain, nilai presisi dapat didefinisikan sebagai rasio jumlah total kategori data positif yang diklasifikasikan dengan benar ke angka tersebut. Nilai presisi dapat diperoleh dengan persamaan sebagai berikut [17]:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

Sedangkan nilai *recall* juga dikenal sebagai nilai *sensitivity* adalah proporsi kasus positif yang diprediksi secara tepat menjadi positif [18], dengan kata lain *recall* menunjukkan bahwa sistem dengan benar mengklasifikasikan beberapa persen dari data kategori positif. Persamaan nilai *recall* dapat dituliskan sebagai berikut [17]:

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

3. METODE

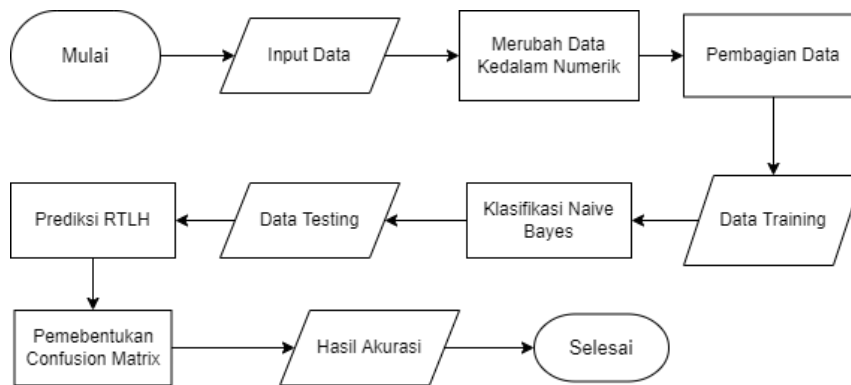
Penelitian ini adalah jenis penelitian kuantitatif. Data yang digunakan ialah data indikator rumah tidak layak huni yang diubah kedalam bentuk data numerik sebelum akhirnya diolah dengan menggunakan metode Naïve Bayes. Data indikator rumah tidak layak huni adalah sebagai berikut:

- a) Jumlah Anggota Rumah tangga (Jumlah_ART) merupakan variabel jumlah anggota rumah tangga dalam satu rumah.
- b) Jenis lantai (JL) merupakan variabel jenis lantai dengan 10 karakteristik yang berbeda, termasuk marmer/granit, keramik, parket/vinil/permadani, ubin/tegel/teraso, kayu/papan kualitas tinggi, semen/bata merah, bambu, kayu/papan kualitas rendah, tanah, dan lainnya.
- c) Jenis dinding (JD) merupakan variabel jenis dinding dengan 7 spesifikasi yaitu, tembok, plesteran anyaman bambu/kawat, kayu, anyaman bambu, batang kayu, bambu, dan lainnya.
- d) Keadaan dinding (KD) merupakan variabel kualitas dinding yang dibedakan dalam kriteria jelek = 0 dan bagus = 1.
- e) Jenis atap (JA) merupakan variabel jenis atap dengan 10 spesifikasi berbeda yaitu, beton/genteng beton, genteng keramik, genteng metal, genteng tanah liat, asbes, seng, sirap, bambu, Jerami/ijuk/daun – daunan/rumbia, dan lainnya.
- f) Keadaan atap (KA) merupakan variabel kualitas atap yang dibedakan dalam kriteria jelek = 0 dan bagus = 1.
- g) Sumber air minum (SAM) merupakan variabel sumber air minum berdasarkan kriteria RTLH dan BPS sumber air minum berasal dari sumur atau mata air yang tak terlindungi atau sungai atau air hujan. Berikut adalah 12 spesifikasi dari sumber air minum yaitu, air kemasan bermerk, air isi ulang, leding meteran, leding eceran, sumur bor/pompa, sumur terlindungi, sumur tak terlindungi, mata air terlindungi, mata air tak terlindungi, air sungai/danau/waduk, air hujan, dan lainnya.
- h) Sumber penerangan (SP) merupakan variabel sumber penerangan yang dibagi berdasarkan kriteria bukan listrik = 0, listrik non-PLN = 1, listrik PLN = 2.
- i) Fasilitas BAB (FB) merupakan variabel fasilitas BAB dengan 4 spesifikasi yaitu, sendiri, bersama, umum, dan tidak ada.
- j) Luas lantai per orang (LL) merupakan variabel luas lantai yang didapatkan dari luas bangunan/lantai dibagi jumlah penghuni dalam satuan meter.
- k) Kelas merupakan variabel kelas yang dibedakan menjadi Rumah Tidak Layak Huni (RTLH) = 1 dan Rumah Layak Huni (RLH) = 0.

500 data indikator rumah tidak layak huni yang digunakan dalam penelitian ini diperoleh dari Data Terpadu Kesejahteraan Sosial (DTKS) di Kabupaten Bojonegoro tahun 2020. Berdasarkan 500 data tersebut akan dilakukan 3 kali percobaan *training* dan *testing*. Pertama, data dibagi menjadi data *training* dan data *testing* dengan perbandingan 90:10 persen. Kedua, dengan perbandingan 80:20 persen, dan ketiga, dengan perbandingan 70:30 persen. Tujuan dilakukan uji coba ini ialah agar diketahui mana hasil akurasi terbaik dalam menentukan bantuan RTLH tersebut. Selain itu, diharapkan diperoleh hasil analisis klasifikasi yang sesuai sehingga dapat diketahui kelompok mana yang termasuk dalam RLH dan mana yang termasuk dalam kategori RTLH.

Berikut adalah alur penelitian dari metode Naïve Bayes dengan visualisasi diagram alur seperti pada Gambar 1:

1. Melakukan penginputan data RTLH Kabupaten Bojonegoro tahun 2020
2. Mengubah total 500 data di setiap masing-masing atribut yang berbentuk indikator ke dalam numerik
3. Membagi 500 data menjadi data *testing* dan *training* untuk setiap variabel x dan y
4. Memasukkan data *training* pada fungsi klasifikasi Naïve Bayes
5. Melakukan prediksi dari data *testing* RTLH
6. Melakukan pembentukan *confusion matrix* yang berisi ketepatan prediksi
7. Menganalisis hasil klasifikasi agar diketahui akurasi dari penelitian yang telah dilakukan.



Gambar 1. Diagram Alur Penelitian

4. HASIL DAN PEMBAHASAN

Data pada penelitian ini adalah data indikator rumah tidak layak huni Kabupaten Bojonegoro tahun 2020 sebanyak 500 data seperti pada Tabel 2 di bawah ini.

Tabel 3. Sampel Data Rumah Tidak Layak Huni Kabupaten Bojonegoro

No	Jumlah	Luas_ _ART	Luas_ antai	JL	JD	KD	JA	KA	SAM	SP	FB	LL	Kelas
1	2	60.0	Tanah	Kayu	Jelek/Kualitas Rendah	Genteng Tanah liat	Jelek/Kualitas Rendah	Sumur Bor/Pompa	Listrik PLN	Sendiri	30.0	RLH	
2	3	60.0	Tanah	Kayu	Jelek/Kualitas Rendah	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur Bor/Pompa	Listrik PLN	Sendiri	20.0	RLH	
3	3	90.0	Tanah	Kayu	Jelek/Kualitas Rendah	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur bor/Pompa	Listrik PLN	Sendiri	30.0	RLH	
4	2	60.0	Tanah	Aanyaman	Bagus/Kualitas Tinggi	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur Bor/Pompa	Listrik PLN	Bersama	30.0	RTLH	
5	6	120.0	Tanah	Kayu	Jelek/Kualitas Rendah	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur Bor/Pompa	Listrik PLN	Sendiri	20.0	RLTH	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

498	7	60.0	Tanah	Anyaman	Bagus/Kualitas Tinggi	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur Terlindung	Listrik PLN	Sendiri	8.58	RTLH
499	4	105.0	Tanah	Plesteran	Jelek/Kualitas Rendah	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur Bor/Pompa	Listrik PLN	Sendiri	26.25	RLH
500	4	60.0	Tanah	Anyaman	Bagus/Kualitas tinggi	Genteng Tanah Liat	Jelek/Kualitas Rendah	Sumur Bor/Pompa	Listrik PLN	Sendiri	15.0	RTLH

Setelah dilakukan penginputan, data indikator tersebut diubah kedalam bentuk data numerik. Berikut hasil data numerik yang disajikan dalam Tabel 2.

Tabel 3. Sampel Data Numerik Rumah Tidak Layak Huni Kabupaten Bojonegoro

No	Jumlah_ART	Luas_lantai	JL	JD	KD	JA	KA	SAM	SP	FB	LL	Kelas
1	2	60.0	6	3	2	4	2	7	1	1	30.0	0
2	3	60.0	6	3	2	4	2	7	1	1	20.0	0
3	3	90.0	6	3	2	4	2	7	1	1	30.0	0
4	2	60.0	6	0	0	4	2	7	1	0	30.0	1
5	6	120.0	6	3	2	4	2	7	1	1	20.0	1
:	:	:	:	:	:	:	:	:	:	:	:	:
498	7	60.0	6	0	0	4	2	9	1	1	8.58	1
499	4	105.0	6	5	2	4	2	7	1	1	26.25	0
500	4	60.0	6	0	0	4	2	7	1	1	15.0	1

Selanjutnya, dilakukan pembagian data dari keseluruhan sampel. Data dikategorikan sebagai data *training* dan *testing* yang akan digunakan dalam metode Naïve Bayes. Berdasarkan 12 indikator, ada 9 indikator yang digunakan sebagai variabel prediktor yaitu variabel x terdiri dari indikator JL, JD, KD, KA, SAM, SP, FB, LL. Lalu satu variabel yaitu variabel "Kelas" yang digunakan sebagai variabel penentu yaitu variabel y.

Setelah dilakukan pemisahan variabel x dan y selanjutnya dilakukan analisis uji coba data *training* dan data *testing*. Akan ada tiga uji coba dalam pengujian *training testing*, dengan rasio masing-masing 90:10, 80:20, dan 70:30 persen. Dimana akan diambil hasil akurasi terbaik dari masing-masing uji coba *training* dan *testing*.

Pada uji coba pertama yang memiliki perbandingan 90:10 persen, terdapat 450 data termasuk kedalam data *training* dan 50 data termasuk kedalam data *testing*. Pada uji coba kedua memiliki perbandingan 80:20 persen, terdapat 400 data termasuk kedalam data *training* dan 100 data termasuk kedalam data *testing*. Pada uji coba ketiga memiliki

perbandingan 70:30 persen, terdapat 350 data yang termasuk kedalam data *training* dan 150 data termasuk kedalam data *testing*.

Selanjutnya akan dilakukan proses klasifikasi pada data *training* dan proses prediksi pada data *testing*. Hasil analisis klasifikasi Naïve Bayes dapat diketahui melalui *confusion matrix* yang berasal dari pengolahan *y testing* dan *y prediksi*. Berikut adalah tabel *confusion matrix* dari ketiga uji coba.

Tabel 4. Hasil *Confusion Matrix* Pada Uji Coba Pertama

	RLH	RTLH
RLH	22	0
RTLH	23	5

Pada Tabel 4 dengan perbandingan 90:10 didapatkan ada sebanyak 22 sampel data yang berhasil di prediksi kedalam RLH, sehingga tidak berhak menerima bantuan yang akan diberikan oleh pemerintah. Lalu, ada sebanyak 5 sampel yang berhasil di prediksi kedalam RTLH, sehingga sebanyak 5 sampel ini berhak mendapatkan bantuan RTLH yang diberikan oleh pemerintah. Selain itu, ada 23 sampel data yang hasilnya salah dalam proses prediksi.

Tabel 5. Hasil *Confusion Matrix* Pada Uji Coba Kedua

	RLH	RTLH
RLH	51	0
RTLH	37	12

Pada Tabel 5 dengan perbandingan 80:20 didapatkan ada sebanyak 51 sampel data yang berhasil di prediksi kedalam RLH, sehingga tidak berhak menerima bantuan yang akan diberikan oleh pemerintah. Lalu, ada sebanyak 12 sampel yang berhasil di prediksi kedalam RTLH, sehingga sebanyak 12 sampel ini berhak mendapatkan bantuan RTLH yang diberikan oleh pemerintah. Selain itu, ada 37 sampel data yang hasilnya salah dalam proses prediksi.

Tabel 6. Hasil *Confusion Matrix* Pada Uji Coba Ketiga

	RLH	RTLH
RLH	73	1
RTLH	58	18

Pada Tabel 6 dengan perbandingan 70:30 didapatkan ada sebanyak 73 sampel data yang berhasil di prediksi kedalam RLH, sehingga tidak berhak menerima bantuan yang akan diberikan oleh pemerintah. Lalu, ada sebanyak 18 sampel yang berhasil di prediksi kedalam RTLH, sehingga sebanyak 18 sampel ini berhak mendapatkan bantuan RTLH yang diberikan oleh pemerintah. Selain itu, ada 59 sampel data yang hasilnya salah dalam proses prediksi.

Berdasarkan ketiga uji coba tersebut didapatkan akurasi masing-masing sebesar 54% pada uji coba pertama, 63% pada uji coba kedua, dan 61% pada uji coba ketiga. Sehingga dapat diketahui akurasi terbaik terletak pada uji coba kedua, disaat perbandingan data *training* dan *testing* sebesar 80:20 persen. Tabel 6 di bawah ini menunjukkan nilai *recall* dan presisi dari ketiga uji coba.

Tabel 7. Ringkasan Tiga Kali Percobaan Prediksi

	90:10	80:20	70:30
Akurasi	54%	63%	61%
Recall	100%	100%	94%
Presisi	18%	25%	24%

Evaluasi menggunakan *confusion matrix* didapatkan hasil akurasi terbaik dari tiga kali percobaan pada Tabel 7. Akurasi terbaik adalah pada uji *training* dan *testing* dengan perbandingan 80:20 persen, dengan kata lain disaat jumlah data *training* sebesar 400 dan data *testing* 100, didapatkan akurasi sebesar 63%, *recall* 100% dan presisi 25%.

Pada model terbaik dengan pembagian data 80:20, diketahui bahwa jumlah sampel yang termasuk kedalam prediksi RLH dan aktual RLH sebesar 51 data. Hal ini berarti terdapat 51 data yang tidak berhak mendapat bantuan RTLH yang diberikan pemerintah. Lalu jumlah sampel yang termasuk ke dalam prediksi RTLH dan aktual RTLH sebesar 12 data, artinya terdapat 12 data yang benar-benar berhak mendapat bantuan RTLH yang diberikan pemerintah. Selain itu, terdapat 37 sampel yang hasilnya salah dalam proses prediksi.

5. KESIMPULAN

Penerapan data mining dalam menentukan klasifikasi penerimaan bantuan rumah tidak layak huni di Kabupaten Bojonegoro adalah untuk mengetahui layak atau tidaknya suatu keluarga mendapatkan bantuan RTLH tersebut dengan menggunakan algoritma Naïve Bayes. Terdapat dua kelas yang digunakan dalam prediksi klasifikasi RTLH yaitu, kelompok yang layak/berhak menerima bantuan RTLH dan kelompok yang tidak layak/tidak berhak menerima bantuan RTLH.

Berdasarkan hasil evaluasi menggunakan *confusion matrix* didapatkan akurasi terbaik dari tiga kali percobaan yaitu pada uji *training* dan *testing* dengan perbandingan 80:20 persen, dengan kata lain disaat jumlah data *training* sebesar 400 dan data *testing* 100. Dengan ini didapatkan akurasi sebesar 63%, *recall* 100% dan presisi 25%.

Semakin besar data *training* semakin tinggi akurasi, artinya semakin banyak data yang masuk sebagai data *training* mengakibatkan model semakin akurat dalam mengenali pola klasifikasi. Jumlah data *testing*, jumlah data *training*, serta atribut yang digunakan semuanya dapat berdampak pada keakuratan analisis klasifikasi Naïve Bayes. Untuk penelitian selanjutnya, dapat menggunakan variasi perbandingan data *training* dan data *testing* serta memilih atribut yang sesuai untuk mendapatkan analisis dengan akurasi yang lebih akurat.

REFERENSI

- [1] Darussalam and G. Arief, "Jurnal Resti," *Resti*, vol. 1, no. 1, pp. 19–25, 2017.
- [2] A. Naas, S. Na'iema, H. Mulyo, and A. Widiastuti, "Klasifikasi penerima bantuan program rehabilitasi rumah tidak layak huni menggunakan algoritme K-Nearest Neighbor Classification of beneficiaries for the rehabilitation of uninhabitable houses using the K-Nearest Neighbor algorithm," *J. Teknol. dan Sist. Komput.*, vol. 10, no. 1, pp. 32–37, 2022, doi: 10.14710/jtsiskom.2022.14110.
- [3] F. Kuhua, A. R. Dilapanga, and J. Mantiric, "Jurnal Administro," *J. Adm.*, vol. 1, no. 1, pp. 05–09, 2019.
- [4] K. Oleh, D. Perumahan, R. Dan, P. Di, K. Magelang, and R. A. Kafa, "Manajemen Program Bantuan Sosial Penanganan Rumah Tidak Layak Huni Apbd Kabupaten (BANSOS-RTLH APBD)," vol. 6, no. 1, pp. 1–18, 2022.
- [5] B. P. S. (BPS), "Data RTLH Nasional," 2020. <https://www.bps.go.id/indicator/29/1241/1/persentase-rumah-tangga-yang-memiliki-akses-terhadap-hunian-yang-layak-dan-terjangkau-menurut-provinsi.html>
- [6] H. Nalatisifa and Y. Ramdhani, "Sistem Penunjang Keputusan Menggunakan Metode Topsis Untuk Menentukan Kelayakan Bantuan Rumah Tidak Layak Huni (RTLH)," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 2, pp. 246–256, 2020, doi: 10.30812/matrik.v19i2.638.
- [7] F. J. Simatupang, T. Wuryandari, M. Jurusan Statistika FSM Universitas Diponegoro, and S. Pengajar Jurusan Statistika, "Klasifikasi Rumah Layak Huni Di Kabupaten Brebes Dengan Menggunakan Metode Learning Vector Quantization Dan Naive Bayes," *J. Gaussian*, vol. 5, no. 1, pp. 99–111, 2016, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [8] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [9] N. M. Putry, "Komparasi Algoritma Knn Dan Naive Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [10] A. Sofian, M. Ulum, M. A. Rozaq, M. Nasikin, and ..., "Analisis Penyaluran Program RTLH dalam Pandangan Ekonomi Islam di Desa Banyuwangi Kecamatan Bandongan Kabupaten Magelang," ... *Ekon. Islam*, vol. 1, no. 1, pp. 60–66, 2022, [Online]. Available: <https://journal.unisnu.ac.id/jrei/article/view/88%0Ahttps://journal.unisnu.ac.id/jrei/article/download/88/55>
- [11] B. Pratama, A. S. Akbar, P. Studi, and T. Informatika, "Klasifikasi Penentuan Warga Penerima Bantuan Sosial Di Masa Pandemi Menggunakan Metode Naive Bayes (Studi Kasus : Rt 002 Rw 01 Kel.Jagakarsa Kec.Jagakarsa Jakarta Selatan)," *J. Nas. Inform.*, vol. 2, no. 2, pp. 143–157, 2021, [Online]. Available: <http://ejournal-ibik57.ac.id/index.php/junif/article/view/338%0Ahttps://ejournal-ibik57.ac.id/index.php/junif/article/download/338/240>
- [12] E. Fitriani, "Perbandingan Algoritma C4.5 Dan Naive Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan," *Sistemasi*, vol. 9, no. 1, p. 103, 2020, doi: 10.32520/stmsi.v9i1.596.
- [13] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," *Respati*, vol. 16, no. 1, p. 32, 2021, doi: 10.35842/jtir.v16i1.386.

- [14] A. P. Wijaya and H. A. Santoso, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government Naïve Bayes Classification on Document Classification to Identify E-Government Content," *J. Appl. Intell. Syst.*, vol. 1, no. 1, pp. 48–55, 2016.
- [15] T. Imandasari, E. Irawan, A. P. Windarto, and A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 750, 2019, doi: 10.30645/senaris.v1i0.81.
- [16] Y. Rahman and H. Wijayanto, "Klasifikasi Batik Menggunakan Metode K-Nearest Neighbour Berdasarkan Gray Level Co-Occurrence Matrices (GLCM)," *Jur. Tek. Inform. FIK UDINUS*, vol. 244, no. Ecpe, pp. 1–7, 2015.
- [17] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.
- [18] A. Andriani, "Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik ' BSI Yogyakarta ,' " *Semin. Nas. Teknol. Inf. dan Komun. 2013 (SENTIKA 2013)*, vol. 2013, no. SENTIKA, pp. 163–168, 2013, [Online]. Available: https://repository.bsi.ac.id/index.php/unduh/item/48930/Sentika_2013Anik-Andriani.pdf