



PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI RUMAH LAYAK ATAU TIDAK LAYAK HUNI (STUDI KASUS: DESA BULU KECAMATAN KRAKSAAN KABUPATEN PROBOLINGGO)

Deniyanto Muchlizin Wahidillah ¹⁾, Abu Tholib ²⁾, Muafi ³⁾

¹ Program Studi Teknik Informatika, Universitas Nurul Jadid

² Program Studi Teknik Informatika, Universitas Nurul Jadid

³ Program Studi Teknik Informatika, Universitas Nurul Jadid

email: ¹ deniyanmuchlizw@gmail.com, ² ebuenje@gmail.com, ³ muafiumar76@unuja.ac.id

ARTICLE INFO

Article History:

Received : 29 July 2022

Accepted : 17 January 2023

Published : 22 January 2023

Keywords:

Data Mining

Classification

K-Nearest Neighbor

K-Fold Cross Validation

Confusion Matrix

IEEE style in citing this article:

D. M. Wahidillah, A. Tholib and Muafi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Rumah Layak Atau Tidak Layak Huni (Studi Kasus: Desa Bulu Kecamatan Kraksaan Kabupaten Probolinggo)", *Jurnal.ilmiah.informatika*, vol. 7, no. 2, pp. 75-84, Dec. 2022.

ABSTRACT

The house is a need that must be met, in addition to food and clothing needs, as well as an indicator of community welfare, to create a safe, comfortable and healthy living environment. The government provides social rehabilitation assistance to the community for uninhabitable houses by helping to buy building materials so they can rehabilitate homes that are not proper. In Bulu Village there are still many places to live that are categorized as uninhabitable houses. This is because of the community's income factor and limited knowledge about the function of the house, as well as facilities and infrastructure that make it increasingly difficult to realize livable dwellings. This research uses the K-Nearest Neighbor algorithm for the classification of decent or uninhabitable homes which aims to provide convenience in determining prospective recipients of social rehabilitation assistance for uninhabitable houses with accurate results and minimizing mistargeting. The results of this study are an average accuracy of 96.25% with a standard deviation of 5.73% in the 7th model test using 10-fold cross-validation with odd k and validation and evaluation of results with the confusion matrix getting an accuracy value of 100%, the precision value is 100%, the recall value is 100%, and the f-measure value is 100% with k = 13.

Corresponding Author:

Deniyanto Muchlizin Wahidillah

Universitas Nurul Jadid

© 2022 Jurnal Ilmiah Informatika (Scientific Informatics Journal) with CC BY NC license

1. PENDAHULUAN

Menurut Peraturan yang dikeluarkan oleh Pemerintah Nomor 12 Tahun 2021 membahas Perubahan Atas Peraturan Pemerintah Nomor 14 Tahun 2016 Tentang Penyelenggaraan Perumahan dan Kawasan Permukiman, bahwa beberapa kriteria persyaratan rumah yang layak huni harus memenuhi beberapa aspek keselamatan bangunan, kebutuhan minimum ruang dan aspek kesehatan bangunan. "Aspek keselamatan bangunan merupakan kemampuan struktur bangunan rumah dihitung berdasarkan beban muatan, beban angin, dan beban gempa sesuai standar yang berlaku. Kebutuhan minimum ruang adalah jumlah kebutuhan minimum luas ruang dengan cakupan 9m² (sembilan meter persegi) per jiwa untuk rumah tapak dan dapat dipenuhi secara bertahap beserta ketinggian minimum langit-langit 2,7m (dua koma tujuh meter). Aspek kesehatan bangunan adalah merupakan ketentuan sistem penghawaan, sistem pencahayaan, sistem sanitasi, dan bahan bangunan yang sesuai dengan ketentuan standar."

Menurut Undang - Undang Nomor 1 Tahun 2011 membahas Perumahan dan juga Kawasan Permukiman yang disebutkan pada Pasal 1 Ayat 7, bahwa "rumah adalah bangunan gedung yang berfungsi sebagai tempat tinggal yang layak huni, sarana pembinaan keluarga, cerminan harkat dan martabat penghuninya, serta aset bagi pemiliknya."

Pentingnya rumah yang layak huni untuk setiap masyarakat perlu diperhatikan. Oleh sebab itu, pemerintah bisa memberikan beberapa bantuan rehabilitasi sosial kepada masyarakat untuk rumah yang tidak layak huni dengan bantuan berupa uang untuk pembelian bahan bangunan, dengan tujuan rehabilitasi rumah yang tidak layak huni tersebut, dimana rumah dapat

dikatakan layak huni, jika memenuhi beberapa aspek kriteria persyaratan diantaranya, yaitu aspek keselamatan pada bangunan, kebutuhan minimum pada ruang dan aspek kesehatan di bangunan [1].

Pada Penelitian ini juga mengambil beberapa rujukan dari penelitian - penelitian sebelumnya tentang penerapan algoritma *K-Nearest Neighbor*. Adapun penelitian terkait sebagai berikut:

Penelitian dengan judul tentang "Analisis Performa Metode *K-Nearest Neighbor* Untuk Identifikasi Jenis Kaca". Hasil yang diperoleh yaitu nilai performa terbaik pada $k = 3$, dengan tingkat akurasi sebesar 64%, presisi sebesar 63%, recall sebesar 71% dan f-measure sebesar 67% [2].

Penelitian yang dilakukan oleh [3], tentang "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN (*K-Nearest Neighbor*)". Hasil yang diperoleh yaitu tingkat akurasi sebesar 100% dengan $k = 15$ dan $k = 30$ di data baru (D160) pada kategori "Tidak Layak". Kemudian tingkat akurasi sebesar 81,25% dengan $k = 45$, $k = 60$ dan $k = 75$ di data baru (D160) pada kategori "Layak".

Penelitian yang dilakukan oleh [4], tentang "Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode *K-Nearest Neighbor* (KNN)". Hasil yang diperoleh yaitu tingkat akurasi sebesar 94,12% dengan *error* sebesar 5,88%, dimana $k = 5$ sebagai nilai k optimal.

Penelitian juga dilakukan oleh [5], tentang "Penerapan Algoritma *K-Nearest Neighbor* Pada Klasifikasi Penentuan Gizi Balita (Studi Kasus di Posyandu Desa Bluto)". Hasil yang diperoleh yaitu tingkat akurasi sebesar 88% terhadap $k = 3$ dan tingkat akurasi sebesar 84% terhadap $k = 5$, maka hasil yang akan diterapkan merupakan nilai k optimal dengan tingkat akurasi tertinggi.

Penelitian yang dilakukan oleh [6], tentang "Algoritma *K-Nearest Neighbor* dengan *Euclidean Distance* dan *Manhattan Distance* Untuk Klasifikasi Transportasi Bus". Hasil yang diperoleh yaitu nilai rata-rata dari *accuracy* 81,96%, *precision* 44,94% dan *recall* 37,06% pada performansi K-NN pada *Euclidean Distance*. Sedangkan nilai rata - rata dari *accuracy* 84%, *precision* 45,49% dan *recall* 36,39% pada performansi K-NN dengan *Manhattan Distance* mempunyai selisih 2,04% lebih tinggi dibandingkan *Euclidean Distance*, dimana nilai k yang digunakan adalah $k = 3$.

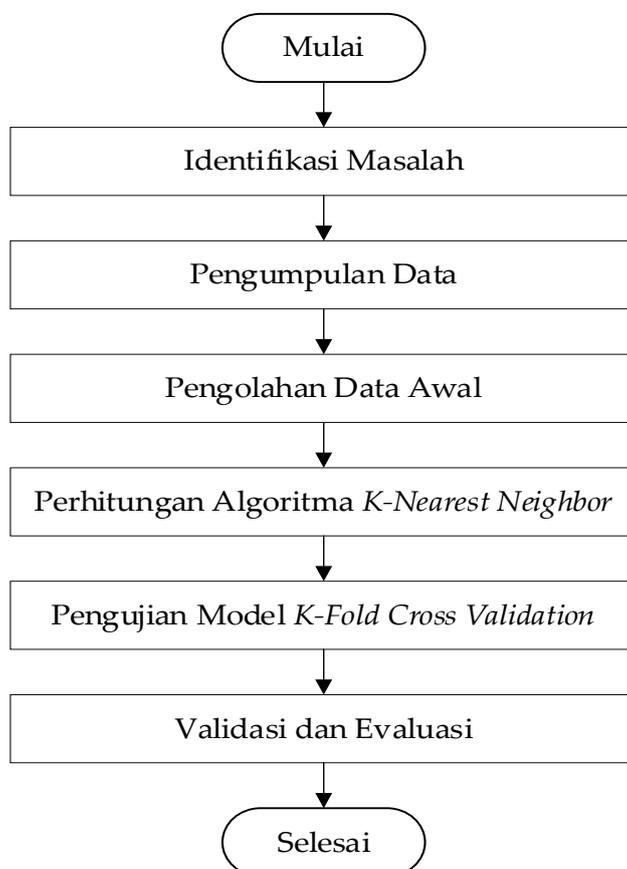
Desa Bulu merupakan salah satu Desa dari 18 Desa dan Kelurahan di Kecamatan Kraksaan yang ada di Kabupaten Probolinggo dan termasuk Provinsi Jawa Timur yang terdiri dari 7 Dusun, 30 RT dan 7 RW. Pada Desa tersebut masih banyak rumah warga yang terkategori

rumah tidak layak huni. Hal ini, disebabkan dari faktor pendapatan masyarakat dan keterbatasan yang dimiliki oleh warga terkait fungsi rumah yang layak, serta sarana dan prasarana yang menyebabkan sehingga sulit terwujudnya rumah yang layak huni.

Sesuai permasalahan di atas, maka penelitian ini menggunakan algoritma *K-Nearest Neighbor* karena merupakan salah satu algoritma terbaik dalam menentukan tingkat akurasi pada permasalahan *Data Mining*, terutamanya untuk penyelesaian persoalan klasifikasi. Penelitian ini bertujuan untuk menerapkan algoritma *K-Nearest Neighbor* untuk mengklasifikasi rumah layak atau tidak layak huni

2. METODE PENELITIAN

Tahapan pada penelitian ini yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Identifikasi Masalah

Pada Tahapan ini menentukan rumusan masalah yang ada, sehingga melakukan beberapa peninjauan dan melakukan eksplorasi untuk menemukan permasalahan yang terjadi.

2.2 Pengumpulan Data

Pada tahap ini merupakan langkah yang sangat penting pada penelitian ini dengan metode pengumpulan data akan didapatkan melalui tahap survei dan studi literatur.

2.3 Pengolahan Data Awal

Pengolahan data awal dibutuhkan supaya data yang akan diolah mempunyai kualitas yang bagus.[7] Hal ini dilakukan, sebab tidak seluruh data dan atribut bisa dipakai. Berikut beberapa teknik pengelolaan data pada penelitian ini, yaitu seleksi atribut, transformasi data, pembersihan data, serta normalisasi data.[8]

2.4 Perhitungan Algoritma K-Nearest Neighbor

Pada tahapan ini semua data dianalisis, dikelompokkan variabel mana yang mempunyai keterhubungan dengan satu sama lainnya. Setelah data dianalisis lalu Pembagian data ke dalam data latihan (*training data*) dan data uji (*testing data*) juga diperlukan untuk pembuatan model.

Pengklasifikasian algoritma *K-Nearest Neighbor* dilakukan menggunakan cara menetapkan nilai k , kemudian menghitung jarak terdekat antara data latihan dan data uji. Agar dapat menghindari hasil klasifikasi yang memenuhi 2 kategori atau lebih, maka nilai k harus dibuat ganjil. Adapun langkah-langkah algoritma *K-Nearest Neighbor*:

1. Menentukan parameter k untuk menentukan jumlah tetangga terdekat.

2. Menghitung *Euclidean Distance* antara data latihan dan data uji. Rumus *Euclidean Distance* dapat dilihat pada Persamaan 1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

dimana $d(x,y)$ jarak *euclidean* antara data latihan titik x dan data uji y , x data latihan, y data uji, i variabel data, serta n dimensi atribut.

3. Mengurutkan jarak hasil perhitungan langkah ke 2, yaitu dari nilai terendah ke nilai tertinggi.
4. Mengumpulkan kategori y sesuai nilai k .
5. Mencari jumlah kelas dari tetangga yang mempunyai jarak terdekat, serta memutuskan kelas tersebut menjadi kelas data yang akan diprediksi sesuai hasil kelas yang paling dominan [9].

2.5 Pengujian Model K-Fold Cross Validation

Tahapan ini, model yang diusulkan diuji buat memperkirakan perkiraan akurasi. Pada penelitian ini, pengujian model memakai *k-fold cross validation* yang akan di uji sebanyak nilai k . Nilai k yang dipakai ialah 10, atau bisa disebut juga *10-fold cross validation* dan hasil dari pengujian model tersebut, kemudian ditetapkan nilai k terbaik lalu diaplikasikan pada metode yang diusulkan.

2.6 Validasi dan Evaluasi

Tahapan ini dilakukan validasi dan evaluasi terhadap model yang ditetapkan untuk mengetahui tingkat keakurasian model.[10] Selanjutnya dilakukan validasi dan evaluasi hasil memakai *confusion matrix* untuk menghasilkan performa, yaitu nilai *accuracy*, *precision*, *recall* dan *f-measure*. Rumus *confusion matrix* dapat dilihat pada persamaan berikut:

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

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (4)$$

$$F - Measure = 2 \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (5)$$

3. HASIL DAN PEMBAHASAN

Hasil penelitian ini diperoleh sejumlah 100 data yang merupakan hasil survei terhadap masyarakat Desa Bulu. Data terdiri dari 80 data yang kategori hasilnya telah diketahui, serta 20 data yang kategori hasilnya belum diketahui.

3.1 Pengolahan Data Awal

Pengolahan data diawali dengan seleksi atribut untuk diambil atribut tertentu, kemudian transformasi data untuk mengkonversikan data yang berupa kategorikal atau nominal menjadi data numerik. Data selanjutnya dikelompokkan berdasarkan hasil seleksi atribut dan transformasi data ke dalam tabel kriteria variabel yang dapat dilihat pada tabel 1.

Tabel 1. Kriteria Variabel

No	Atribut	Keterangan	Bobot
1	Jumlah Penghasilan	>Rp. 3.000.000	1
		Rp. 2.000.000 - Rp. 3.000.000	2
		Rp. 1.000.000 - Rp. 2.000.000	3
		Rp. 500.000 - Rp. 1.000.000	4
		< Rp. 500.000	5
2	Jumlah Anggota Keluarga	1	1
		2	2
		3	3
		4	4
		5	5
		6	6
3	Pondasi	Baik	1
		Kurang Baik	2
4	Balok	Baik	1
		Kurang Baik	2
5	Luas Ruang	> 16 m ²	1
		12 m ² - 15,9 m ²	2
		9 m ² - 11,9 m ²	3
		7 m ² - 8,9 m ²	4
		< 7 m ²	5
6	Ketinggian Langit-langit	> 3 m	1
		2,7 m - 3 m	2
		< 2,7 m	3
7	Penghawaan	Baik	1
		Kurang Baik	2
8	Pencahayaan	Baik	1

		Kurang Baik	2
9	Sanitasi	Layak	1
		Kurang Layak	2
10	Kondisi Atap	Bagus	1
		Usang	2
11	Kondisi Dinding	Bagus	1
		Usang	2
12	Kondisi Lantai	Bagus	1
		Rusak	2
13	Hasil	RLH	0
		RTLH	1

Selanjutnya data dilakukan pembersihan data untuk mengidentifikasi *missing value* pada variabel independen. Kemudian data dinormalisasi menggunakan normalisasi *min-max*. Berikut beberapa perhitungan normalisasi *min-max* dengan rumus dapat dilihat pada Persamaan 6.

$$V' = \frac{V - \min A}{\max A - \min A} (\text{newmax}A - \text{newmin}A) + \text{newmin}A \quad (6)$$

dimana V' nilai yang didapatkan dari data baru hasil dari normalisasi, V nilai dari data sebelum dinormalisasi, $\max A$ nilai maksimum pada kolom, $\min A$ nilai minimum pada kolom, $\text{new max}A$ batas nilai maksimum terbaru, serta $\text{new min}A$ batas nilai minimum terbaru.

1) Normalisasi data pada baris pertama kolom pertama.

$$V' = \frac{5-1}{5-1} (1 - 0) + 0$$

$$V' = 1$$

2) Normalisasi data pada baris pertama kolom pertama.

$$V' = \frac{3-1}{7-1} (1 - 0) + 0$$

$$V' = 0.3333333333333333$$

3) Normalisasi data pada baris pertama kolom pertama.

$$V' = \frac{2-1}{3-1} (1 - 0) + 0$$

$$V' = 0.5$$

Perhitungan normalisasi data di atas juga berlaku terhadap baris dan kolom seterusnya hingga data ke 100, sehingga hasil dari perhitungan normalisasi data dapat dilihat pada Tabel 2.

Tabel 2. Hasil Normalisasi Data

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Hasil
1	1	0.3333333333	0.5	0.5	0.75	1	1	1	0	0.5	0.5	0	1
2	1	0.5	1	1	1	1	0	0.5	0.5	0	0.5	1	1
3	1	0.3333333333	0.5	0.5	0.75	0.5	0	0	0.5	0.5	0.5	1	1
4	0.5	0.5	0	0	0.75	0.5	0	0	0	0	0	0	0
5	0.5	0.6666666667	0	0	0.5	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	0	0.5	0	0	0.5	0	0	0	0	0.5	0	0	?
97	0.25	0.3333333333	0	0	0.5	0	0	0	0	0	0	0	?
98	0.25	0.6666666667	0	0	0	0	0	0	0	0	0	0	?
99	1	0	0.5	0.5	0.75	0.5	0.5	1	0	0.5	0.5	1	?
100	0.25	0.8333333333	0.5	1	0.75	1	0	0	0	1	0.5	1	?

Keterangan:

- X1 : Jumlah Penghasilan
- X2 : Jumlah Anggota Keluarga
- X3 : Pondasi
- X4 : Balok
- X5 : Luas Ruang
- X6 : Ketinggian Langit-langit
- X7 : Penghawaan
- X8 : Pencahayaan
- X9 : Sanitasi
- X10 : Kondisi Atap
- X11 : Kondisi Dinding
- X12 : Kondisi Lantai

$$d(1,81) = \sqrt{(1 - 0.75)^2 + (0.333333333 - 0.666666667)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0.75 - 0.75)^2 + (1 - 1)^2 + (1 - 0.5)^2 + (1 - 0.5)^2 + (0 - 0)^2 + (0.5 - 0.5)^2 + (0.5 - 0.5)^2 + (0 - 1)^2}$$

$$d(1,81) = 1.293681225$$

2) Perhitungan data uji pertama pada data latih kedua.

$$d(2,81) = \sqrt{(1 - 0.75)^2 + (0.5 - 0.666666667)^2 + (1 - 0.5)^2 + (1 - 0.5)^2 + (1 - 0.75)^2 + (1 - 1)^2 + (0 - 0.5)^2 + (0.5 - 0.5)^2 + (0.5 - 0)^2 + (0 - 0.5)^2 + (0.5 - 0.5)^2 + (1 - 1)^2}$$

$$d(2,81) = 1.1843892$$

Perhitungan *Euclidean Distance* di atas juga berlaku terhadap data latih selanjutnya hingga data ke 80 dan data uji hingga data ke 100, sehingga hasil dari perhitungan *Euclidean Distance* yang dapat dilihat pada Tabel 3.

3.2 Perhitungan Algoritma *K-Nearest Neighbor*

Berikut langkah-langkah perhitungan algoritma *K-Nearest Neighbor*:

1. Menentukan parameter *k* untuk menentukan jumlah tetangga terdekat. Pada penelitian ini parameter *k* yang ditetapkan yaitu *k* = 13.
2. Menghitung *Euclidean Distance* antara data latih dan data uji.
 - 1) Perhitungan data uji pertama pada data latih pertama.

Tabel 3. Hasil *Euclidean Distance* Data Uji Pertama

No	Euclidean Distance	Ranking	Hasil
1	1.293681225	23	1
2	1.1843892	13	1
3	1.083333333	7	1
4	1.685312368	52	0
5	1.903943276	63	0
⋮	⋮	⋮	⋮
76	1.346291202	27	1
77	1.529796646	41	1
78	2.083333333	74	0
79	1.166666667	12	1
80	2.150581317	77	0

3. Mengurutkan jarak hasil perhitungan langkah ke 2, yaitu dari nilai terendah ke nilai tertinggi. Mengumpulkan kategori *y* sesuai nilai *k* dan mencari jumlah kelas dari tetangga yang

mempunyai jarak terdekat, serta memutuskan kelas tersebut menjadi kelas data yang akan diprediksi sesuai hasil kelas yang paling dominan yang dapat dilihat pada Tabel 4.

Tabel 4. Penentuan Kelas Mayoritas Data Uji Pertama

No	Euclidean Distance	Ranking	Hasil
37	0.527046277	1	1
17	0.881917104	2	1
6	0.901387819	3	1
30	0.935414347	4	1
13	0.961046883	5	1
29	1.060660172	6	1
3	1.083333333	7	1
53	1.118033989	8	1
71	1.145643924	9	1
70	1.148670729	10	1
66	1.157703666	11	1
79	1.166666667	12	1
2	1.1843892	13	1

Dari hasil penentuan kelas mayoritas pada data uji pertama, maka data *testing* pertama masuk ke dalam kategori 1 yaitu kategori "RTLH".

Perhitungan tersebut juga berlaku di data uji berikutnya. Sehingga didapatkan hasil perhitungan prediksi otomatis pada data uji yang dapat dilihat pada Tabel 5.

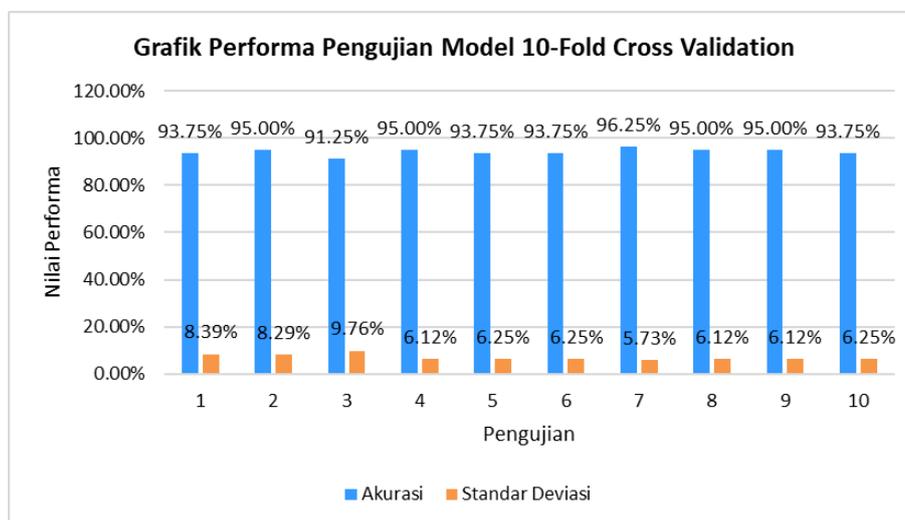
Tabel 5. Hasil Perhitungan Prediksi Otomatis Pada Data Uji

No	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	Hasil
81	0.75	0.666666667	0.5	0.5	0.75	1	0.5	0.5	0	0.5	0.5	1	1
82	0.5	0.5	0.5	0.5	0.75	0.5	0.5	0.5	0	0.5	0.5	0	1
83	0.25	0.5	0.5	0.5	0.75	0.5	0	0	0	0.5	0	0	0
84	0.75	0.5	0	0	0.5	0	0.5	1	0.5	0.5	0.5	1	1
85	0.5	0.333333333	0	0	0.5	0	0.5	0.5	0.5	0.5	0.5	1	1
86	0.25	0.833333333	0	0	0.5	0	0	0	0	0.5	0	0	0
87	0.25	0.833333333	0	0	0.5	0	0	0	0	0.5	0.5	0	0
88	0.75	0.666666667	0	0	0.75	0.5	0	0	0	0.5	0.5	0	0
89	0	0.166666667	0	0	0.5	0	0	0	0	0.5	0	0	0
90	0	0.5	0	0	0.75	0.5	0	0	0	0.5	0	0	0
91	1	0	0.5	0.5	0.75	0.5	0	0	0	0.5	0.5	0	0
92	0.25	0.5	0	0	0.5	0	0	0	0	0.5	0.5	0	0
93	0	0.5	0	0	0	0	0	0	0	0	0	0	0
94	0.25	0.5	0	0	0.75	0.5	0.5	0.5	0	0.5	0	0	0
95	0.5	0.5	0.5	0.5	0.75	0.5	1	1	0.5	1	0.5	0	1
96	0	0.5	0	0	0.5	0	0	0	0	0.5	0	0	0
97	0.25	0.333333333	0	0	0.5	0	0	0	0	0	0	0	0
98	0.25	0.666666667	0	0	0	0	0	0	0	0	0	0	0
99	1	0	0.5	0.5	0.75	0.5	0.5	1	0	0.5	0.5	1	1
100	0.25	0.833333333	0.5	1	0.75	1	0	0	0	1	0.5	1	1

3.3 Pengujian Model K-Fold Cross Validation

Tahapan ini, dataset akan terbagi secara otomatis menjadi 10 bagian, serta dilakukan pengujian sebanyak 10 kali

atau biasa dianggap *10-fold cross validation* pada nilai k ganjil. Hasil pengujian model *10-fold cross validation* berupa grafik performa dapat lihat pada Gambar 2.



Gambar 2. Grafik Performa Pengujian Model 10-Fold Cross Validation

Berdasarkan hasil pengujian model *10-fold cross validation*, maka dihasilkan nilai rata-rata tingkat akurasi tertinggi sebesar 96,25% dengan standar deviasi 5,73% dari setiap iterasi diperoleh saat pengujian ke 7 sebanyak $k = 13$.

3.4 Validasi dan Evaluasi

Validasi dan evaluasi hasil dilakukan memakai *confusion matrix* untuk menghasilkan performa, yaitu nilai akurasi, presisi, *recall* dan *f-measure*.

Tabel 6. Confusion Matrix

	True RLH	True RTLH
Pred. RLH	11	0
Pred. RTLH	0	9

Berdasarkan data latih yang diprediksi dengan benar melalui algoritma *K-Nearest Neighbor* sebanyak 80 terhadap 20 data uji, 11 orang yang benar "RLH" dan 9 orang yang benar "RTLH". Lalu 0 orang "RLH" tapi diprediksi menjadi "RTLH" dan 0 orang "RTLH" tapi diprediksi menjadi "RLH". Perhitungan nilai akurasi, presisi, *recall* dan *f-measure* yang dihasilkan adalah sebagai berikut:

$$Accuracy = \frac{11+9}{11+9+0+0} \times 100\% = 100\%$$

$$Precision = \frac{11}{11+0} \times 100\% = 100\%$$

$$Recall = \frac{11}{11+0} \times 100\% = 100\%$$

$$F - Measure = 2 \frac{100 \times 100}{100 + 100} \times 100\% = 100\%$$

Berdasarkan validasi dan evaluasi hasil pada algoritma *K-Nearest Neighbor* di atas didapatkan nilai prosentase tingkat *accuracy* sebesar 100%, *precision* sebesar 100%, *recall* sebesar 100% dan *f-measure* sebesar 100%.

4. UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih banyak kepada Bapak Abu Tholib, M.Kom selaku pembimbing I dan Bapak

Muafi, S.Ag, M.Kom selaku pembimbing II, karena atas bimbingannya dari beliau, sehingga penulis akhirnya dapat menyelesaikan tugas penelitiannya tepat waktu.

5. KESIMPULAN

Berdasarkan hasil uji coba yang telah dilakukan pada bab sebelumnya dan sekaligus menjawab dari rumusan masalah dan tujuan penelitian ini, maka kesimpulan yang didapat sebagai berikut: Penelitian ini menghasilkan sebuah analisis metode *data mining* dengan menerapkan algoritma *K-Nearest Neighbor* untuk kasus klasifikasi rumah layak atau tidak layak huni di Desa Bulu Kecamatan Kraksaan Kabupaten Probolinggo dengan nilai akurasi sebesar 100%, nilai *precision* sebesar 100%, nilai *recall* sebesar 100%, dan nilai *f-measure* sebesar 100% menggunakan *confusion matrix*. Kemudian pada pengujian model *10-fold cross validation* sebanyak 10 kali pengujian dengan *k* ganjil dan menghasilkan nilai rata-rata akurasi tertinggi dari setiap iterasi pada pengujian model ke 7 sebanyak *k* = 13 sebesar 96,25% dengan standar deviasi sebesar 5,73%.

6. REFERENSI

- [1] A. Muhidin and I. Baragigiratri, "Pemetaan Penduduk Calon Penerima Bantuan Renovasi Rumah Desa Pesangkalan Menggunakan Algoritma Clustering K-Means," *J. SIGMA*, vol. 9, no. 3, pp. 75–82, Mar. 2019.
- [2] M. M. Baharuddin, H. Azis, and T. Hasanuddin, "Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca," *Ilk. J. Ilm.*, vol. 11, no. 3, pp. 269-274, Dec. 2019.
- [3] R. Hasanah, M. Hasan, W. Pangesti, F. Wati, and W. Gata, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN (K-Nearest Neighbor)," *Techno*, vol. 16, no. 1, pp. 1-6, Mar. 2019.
- [4] S. Diansyah, "Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN)," *J. Sistim Inf. Dan Teknol.*, pp. 7–12, Feb. 2022.
- [5] R. Wahyudi, M. Orisa, and N. Vandyansyah, "Penerapan Algoritma K-Nearest Neighbors Pada Klasifikasi Penentuan Gizi Balita (Studi Kasus Di Posyandu Desa Bluto)," *JATI J. Mhs. Tek. Inform.*, vol. 5, no. 2, pp. 750-757, Oct. 2021.
- [6] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 104-111, Aug. 2020.
- [7] P. Soleh, A. Tholib and M. N. F Hidayat, "Penerapan Data Mining Untuk Analisa Pola Pembelian Produk Menggunakan Algoritma Frequent Pattern–Growth," *Rekayasa*, vol. 14, no. 3, pp. 456-460, 2021.
- [8] F. N. Fajri, A. Tholib and W Yuliana, "Application of Machine Learning Algorithm for Determining Elective Courses in Informatics Study Program," *J. Tek. Inform. Dan Sist. Inf.*, vol. 8, no. 3, pp. 485-496, Dec. 2022.
- [9] K. Malik, Y. Pratama, and K. Nisa', "Diagnosa Pharyngitis Menggunakan Metode K-Nearest Neighbor (K-NN) di Puskesmas Leces Probolinggo," *TRILOGI J. Ilmu Teknol. Kesehat. Dan Hum.*, vol. 2, no. 3, pp. 199-206, Dec. 2021.
- [10] S. Sudriyanto, R. Rizaldi, and M. A. R. Hariri, "Implementasi Algoritme Decision Tree (C4. 5) dengan Optimize Weights (PSO) untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu," *J. Informatika*, vol. 6, no. 2, pp. 252-257, Jun. 2021.