

ANALISA DATA MINING DALAM MEMPREDIKSI MASYARAKAT KURANG MAMPU MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Nurdin^{1*}, Mela Rizki², Maryana³

¹Program Studi Magister Teknologi Informasi Universitas Malikussaleh; Jl. Batam, Kampus Bukit Indah - Lhokseumawe, Aceh.

^{2,3}Program Studi Teknik Informatika Universitas Malikussaleh; Jl. Batam, Kampus Bukit Indah - Lhokseumawe, Aceh

Riwayat artikel:

Received: 11 Februari 2024

Accepted: 30 Maret 2024

Published: 2 April 2024

Keywords:

Data Mining;

K-Nearest Neighbor;

Masyarakat Kurang Mampu;

Confusion Matrix.

Correspondent Email:

nurdin@unimal.ac.id

Abstrak. Kemiskinan merupakan salah satu persoalan mendasar yang menjadi pusat perhatian pemerintah di suatu negara. Salah satu aspek penting untuk mendukung Strategi Penanggulangan Kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran. Salah satu problem utama yang sering menghambat suksesnya program pemerintah tersebut adalah ketersediaan data sasaran masyarakat miskin yang sesuai. Penelitian ini bertujuan untuk membuat aplikasi yang dapat memprediksi masyarakat kurang mampu menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan lima kriteria atau indikator utama yaitu jenis pekerjaan, jumlah tanggungan, umur, penghasilan dan kondisi Rumah. Prediksi ini menyediakan data keluarga miskin yang sesuai untuk menerima berbagai bantuan dari pemerintah. Data yang digunakan untuk prediksi yaitu data sampel dari Kecamatan Pegasing. Pada penelitian ini algoritma *K-NN* dianalisa yang dikembangkan berbasis web. Prinsip kerja *K-Nearest Neighbor* adalah mencari jarak terdekat antara data yang dievaluasi dengan data latihnya. Hasil evaluasi menggunakan *confusion matrix* didapatkan akurasi yang dihasilkan untuk 216 data *training* dengan 93 data *testing* dengan rasio perbandingan 70:30 dan lima atribut yang digunakan menghasilkan akurasi sebesar 86,02%, Recall 61,90%, Presisi 72,22%, dan F1-Score 66,04%.

Abstract. Poverty is one of the fundamental issues that is center of attention of the government in a country. One important aspect to support the poverty reduction strategi is the availability of accurate and targeted poverty data. One of the main problems that often hinders the success of these government programs is the availability of appropriate data on the targeting of the poor. This study aims to design an application than can predict the poor using the *K-Nearest Neighbor Algorithm* with the five main indicators being the type of work, number of dependents, age income and condition of the household head of the family. This prediction provides data on poor families that are suitable for receiving various assistance from the government. The data used for predictions are sample data from Pegasing District. In this study, the *K-NN Algorithm* was analyzed which was developed based on the web. The working principle of *K-Nearest Neighbor* is to find the shortest distance between the evaluated data and training data. The results of the evaluation using the *confusion matrix* obtained the resulting accuracy for 216 training data with 93 testing data with a ratio of 70:30 and five attributes used produced an accuracy of 86,02%, Recall 61,90%, Precision 72,22%, and F1-Score 66,04%..

1. PENDAHULUAN

Masalah kemiskinan merupakan hal yang sering dilihat karena menyangkut berbagai macam aspek, seperti hak untuk terpenuhinya pangan, kesehatan, pendidikan dan pekerjaan. Penduduk yang masih berada di bawah garis kemiskinan mencakup mereka yang berpendapatan rendah, tidak berpendapatan tetap atau tidak berpendapatan sama sekali. Jumlah penduduk miskin di suatu wilayah dapat mencerminkan tingkat kemiskinan wilayah tersebut. Besar kecilnya jumlah penduduk miskin sangat di pengaruhi oleh garis kemiskinan. Karena penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran perbulan di bawah garis kemiskinan. Semakin tinggi garis kemiskinan, semakin banyak penduduk yang tergolong sebagai penduduk miskin [1].

Salah satu permasalahan yang sering menyebabkan program pemerintah terkait keluarga miskin kurang efektif dijalankan adalah terkait ketersediaan data yang sesuai, data keluarga miskin yang tepat sasaran. Ketersediaan data ini sering tidak dapat di verifikasi dan validasi karena luasnya wilayah dengan jumlah keluarga miskin yang sangat besar dan sering berubah. Status keluarga miskin yang di peroleh sering tidak akurat datanya baik menyangkut jumlah maupun lokasi penyebarannya. Menurut [2] Sering kali di suatu wilayah terjadi kekurangan jumlah keluarga penerima bantuan sementara di wilayah lain terjadi over penerima bantuan. Adanya data yang sesuai dapat membantu perencanaan program yang tepat sasaran.

Kemiskinan yang terjadi pada masyarakat Indonesia merupakan masalah utama yang menjadi perhatian di berbagai pemerintahan provinsi, kabupaten, kota, kecamatan dan desa. Berbagai cara penanggulangan kemiskinan telah dilakukan oleh pemerintah salah satunya adalah pemberian bantuan untuk masyarakat kurang mampu. Diharapkan melalui program pemberian bantuan untuk masyarakat kurang mampu dapat mengatasi permasalahan yang dialami setiap warga kurang mampu khususnya permasalahan ekonomi [3].

Data mining salah satu teknologi yang menggabungkan metode analisis tradisional dengan algoritma yang canggih untuk

memproses data dengan volume besar [4]. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi. Banyak metode atau algoritma data mining yang digunakan untuk menyelesaikan berbagai macam permasalahan, diantaranya penerapan metode naïve bayes untuk klasifikasi Karya Ilmiah Mahasiswa [5], penerapan metode Naive Bayes pada Sistem Pakar Diagnosis Dini COVID-19 [6], penerapan Algoritma Apriori untuk Menganalisis Kategori Kompetisi Mahasiswa [7]. Pada penelitian ini menggunakan algoritma K-Nearest Neighbor untuk memprediksi masyarakat kurang mampu.

Algoritma K-Nearest Neighbor merupakan salah satu algoritma atau yang digunakan dalam pengklasifikasian. Prinsip kerja K-Nearest Neighbor (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K-Nearest Neighbor terdekatnya dalam data pelatihan [8]. Algoritma ini memiliki beberapa kelebihan yaitu bahwa dia tangguh terhadap training data yang noise, lebih efektif dalam melakukan training data yang besar dan dapat menghasilkan data yang lebih akurat. Metode ini dapat melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [9].

Penelitian ini penting dilakukan untuk melakukan pendataan di masyarakat agar dapat membantu masyarakat kurang mampu mendapatkan bantuan dari pemerintah pada setiap Kecamatan. Maka dari itu peneliti tertarik melakukan penelitian membuat sebuah sistem yang bisa dapat membantu atau mempermudah mengetahui masyarakat kurang mampu di Kecamatan Pegasing dengan menggunakan metode *K-Nearest Neighbor* untuk mengetahui keakuratan hasil dari perhitungan dalam memprediksi masyarakat kurang mampu di Kecamatan Pegasing. Penelitian ini bertujuan untuk membuat aplikasi yang dapat memprediksi masyarakat kurang mampu menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan lima kriteria atau indikator utama yaitu jenis pekerjaan, jumlah tanggungan, umur, penghasilan dan kondisi Rumah. Prediksi ini menyediakan data

keluarga miskin yang sesuai untuk menerima berbagai bantuan dari pemerintah.

Disamping penelitian ini ada beberapa penelitian yang telah dilakukan oleh peneliti yang terkait dengan data mining diantaranya, penelitian penerapan data mining untuk menentukan penerima bantuan pangan non tunai menggunakan metode k-nearest neighbor [10], penelitian sistem pendukung keputusan penentuan penerimaan bantuan pkh menggunakan metode naïve bayes[11], dan penelitian penerapan metode Profile Matching Analisis pada sistem pendukung keputusan rekomendasi Program Studi [12].

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Penelitian terkait atau penelitian terdahulu menjadi salah satu acuan penulis dalam melakukan penelitian ini untuk mendapatkan gambaran atau perbandingan yang telah dilakukan oleh peneliti sebelumnya, sehingga akan terlihat perbedaan dengan penelitian ini. Ada beberapa penelitian yang menjadi studi literatur, diantaranya penelitian yang dilakukan oleh [2] penelitian tersebut dapat disimpulkan bahwa hasil pengujian UAT yang dilakukan menggunakan questioner dengan skala likert mendapatkan nilai 75% yang berada dalam kategori, sangat bagus. Pengujian K-NN untuk mengklasifikasikan keluarga miskin pada 100 contoh data diperoleh nilai akurasi tertinggi dengan perbandingan 90:10 pada nilai $K=5$, $K=7$ dan $K=9$ sebesar 90%.

Berdasarkan hasil pengujian dapat diketahui Algoritma Naïve Bayes terhadap data set yang telah diambil pada objek penelitian di peroleh tingkat akurasi sebesar 73% atau termasuk dalam kategori Good. Sementara nilai precision sebesar 92% dan Recall sebesar 86%. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi yang dibangun dapat digunakan sebagai bahan masukan bagi pengambil keputusan [13]. Sedangkan penelitian yang dilakukan oleh [14] hasil penelitiannya menunjukkan algoritma K-Nearest Neighbor cukup efektif dan efisien dalam mengklasifikasikan penerima beasiswa Aceh Carong dengan hasil 82,00%.

Penelitian lain yang dilakukan oleh [15] klasifikasi penerima dana bantuan desa

menggunakan metode K-Nearest Neighbor, dengan hasil penelitiannya dimana hasil prediksi keluarga penerima manfaat yang layak sebagai penerima BPNT pada tahun 2021 sebanyak 511 penerima dengan tingkat Akurasi 75,79%, Precision 76,17%, Recall 89,24%, F-measure 82,19%. Penelitian lainnya yang oleh Putri dkk (2021) tentang penerima manfaat bantuan non tunai Kartu Keluarga Sejahtera Menggunakan Metode Naïve Bayes dan KNN, dengan hasil penelitiannya klasifikasi penerima manfaat dari 6.491 penerima KKS dengan metode Algoritma K-NN menghasilkan nilai accuracy 66,46% dan nilai accurasi model algoritma Naïve Bayes sebesar 99,88% [16].

Penelitian yang dilakukan oleh Firasari dkk (2020) kombinasi K-NN dan Gradient Boosted Trees untuk klasifikasi penerima program bantuan sosial, dengan hasil penelitiannya yang diperoleh adalah akurasi Naïve Bayes lebih tinggi dibandingkan K-NN yaitu Naïve Bayes 89,04% dan K-NN 87,67%. Angka ini diklasifikasikan dalam kategori klasifikasi baik. Dari hasil penelitiannya dapat disimpulkan algoritma Naïve Bayes cocok diterapkan dalam perhitungan penerima bansos [17].

3. METODE PENELITIAN

3.1. Tahapan Penelitian

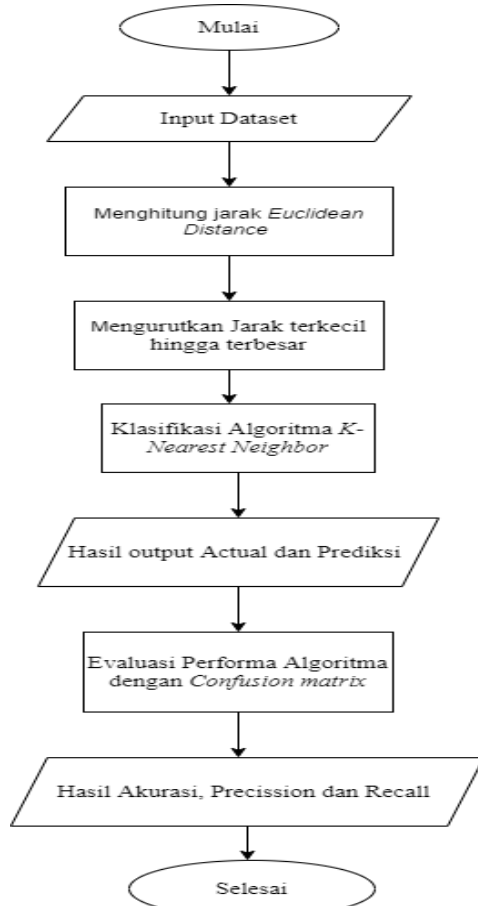
Metode penelitian analisa data mining untuk memprediksi masyarakat kurang mampu menggunakan metode KNN dengan tahapan penelitian sebagai berikut:

- a. Pengumpulan data
Pada penelitian ini data di ambil dari Kantor Dinas Sosial Kabupaten Aceh Tengah, dengan jenis data berupa jenis pekerjaan, jumlah tanggungan, umur, penghasilan dan kondisi rumah. Ada juga studi literatur yaitu merupakan kajian terhadap masalah melalui jurnal, internet untuk melengkapi data yang diperlukan dalam penelitian.
- b. Perancangan sistem
Perancangan sistem menggunakan UML. Perancangan ini untuk membantu proses pengembangan aplikasi yang akan dibangun dengan menggunakan bahasa pemrograman.

- c. Perhitungan manual metode K-Nearest Neighbor
Pada tahap ini melakukan perhitungan manual dengan menggunakan metode K-Nearest Neighbor berdasarkan data parameter/variabel yang sudah ditentukan.
- d. Implementasi sistem
Implementasi Sistem adalah tahap penerapan sistem yang akan dilakukan jika sistem disetujui termasuk program yang telah dibuat pada tahap perancangan sistem agar siap untuk operasikan.
- e. Pengujian sistem
Pengujian sistem merupakan tahapan yang membahas hasil uji coba sistem dari penelitian yang telah dilakukan yaitu berupa hasil pengujian aplikasi perangkat lunak.

3.2. Skema Sistem

Skem sistem yang dirancang untuk analisa data mining untuk memprediksi masyarakat kurang mampu menggunakan metode KNN pada Gambar 1.



Gambar 1. Skema Sistem

3.3. Metode K-Nearest Neighbor

K-Nearest Neighbor merupakan salah satu algoritma yang digunakan dalam pengklasifikasian. Prinsip kerja *K-Nearest Neighbor* (KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan *K-Nearest Neighbor* terdekatnya dalam data pelatihan[18]. Adapun langkah-langkah penggunaan metode K-NN ini dijelaskan sebagai berikut:

1. Tentukan parameter K
2. Hitung jarak antara data yang akan dievaluasi dengan semua pelatihan
3. Urutkan jarak yang terbentuk (urutan dari nilai yang terkecil hingga nilai yang terbesar).
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

Dalam penelitian ini Metode Pendekatan yang digunakan adalah Metode Euclidean Distance dan Manhattan Distance. Berikut adalah rumus (1) untuk mencari nilai akar dari kuadrat dua vector.

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

Keterangan:

- d : jarak Euclidean
 x_{2i} : nilai pada data uji ke- i
 x_{1i} : nilai pada data latih ke- i
 p : banyaknya atribut.

4. HASIL DAN PEMBAHASAN

4.1. Variabel dan Dataset yang digunakan

Data yang digunakan dalam penelitian ini diperoleh dari Dinas Sosial Kabupaten Aceh Tengah. Variabel atau atribut penelitian yang digunakan sebagai dasar kebutuhan dari metode yang digunakan. Untuk menentukan nilai bobot, peneliti memasukkan nilai sampel perhitungan, bobot yang dimiliki masing-masing atribut atau kriteria akan dijumlahkan untuk menentukan kategori data tersebut "Mampu" atau "Tidak Mampu". Berikut bobot nilai setiap kriteria dapat pada Tabel 1.

Tabel 1. Bobot Nilai Setiap Kriteria

Kode	Nama Kriteria	Nilai	Bobot
x1	Pekerjaan	Tidak bekerja	4
		Buruh, petani	3
		Pedagang, wiraswasta	2
		pensiunan, pegawai.	
		PNS, Polri, TNI	1
x4	Penghasilan	<500.000	4
		500.000 - 1.000.000	3
		1.000.000 - 2.500.000	2
		>2.500.000	1
x5	Kondisi Rumah	Bambu / Triplek	4
		Kayu	3
		Batako	2
		Batu bata	1
		Beton	0

4.2. Penerapan Metode K-Nearest Neighbor

Dalam penelitian ini menggunakan data training sejumlah 20 data, data training ini diharapkan mampu menjadikan sistem yang akan dibangun memiliki hasil perhitungan tepat dan akurat. Tabel 2 berikut merupakan data training.

Tabel 2. Data Training

Nama kepala keluarga	x1	x2	x3	x4	x5	Label
xxxxxxxx	3	4	44	3	3	Kurang mampu
xxxxxxxx	3	3	49	3	3	Kurang mampu
xxxxxxxx	3	2	55	2	3	Kurang mampu
xxxxxxxx	3	3	40	3	2	Kurang mampu
xxxxxxxx	3	4	41	3	2	Kurang mampu
xxxxxxxx	3	4	39	3	3	Kurang mampu
xxxxxxxx	2	4	46	2	0	Mampu

xxxxxxxx	3	3	37	3	3	Kurang mampu
xxxxxxxx	3	5	43	3	3	Kurang mampu
xxxxxxxx	3	3	40	3	3	Kurang mampu
xxxxxxxx	3	1	33	3	3	Kurang mampu
xxxxxxxx	2	3	29	2	2	Mampu
.....
xxxxxxxx	3	4	50	3	3	Kurang mampu

Berikut ini data testing yang akan digunakan pada langkah perhitungan dengan metode K-NN. Sebagai sampel mengambil 1 data untuk melakukan perhitungan pada Tabel 3.

Tabel 3. Data Testing

Nama kepala keluarga	x1	x2	x3	x4	x5	Label
xxxxxxxx	3	3	59	3	3	Kurang mampu

Berdasarkan data pada Tabel 2 dan Tabel 3 dilakukan perhitungan dengan langkah-langkah berikut:

- Menentukan parameter K.

Dalam penyelesaian ini diambil beberapa nilai K untuk menentukan parameter K yang akan digunakan dalam penelitian ini adalah K=3.

- Menghitung kuadrat jarak *Euclidean*.

Kuadrat jarak euclidean masing-masing objek terhadap data sampel yang diberikan. Berikut ini proses perhitungan 20 data training pada data testing 1 adalah sebagai berikut.

$$\begin{aligned}
 d1 &= \sqrt{(3-3)^2 + (3-4)^2 + (59-44)^2 + (3-3)^2 + (3-3)^2} \\
 &= \sqrt{0^2 + 1^2 + 15^2 + 0^2 + 0^2} \\
 &= 15.03329638 \\
 d2 &= \sqrt{(3-3)^2 + (3-3)^2 + (59-49)^2 + (3-3)^2 + (3-3)^2} \\
 &= \sqrt{0^2 + 0^2 + 10^2 + 0^2 + 0^2} \\
 &= 10 \\
 d3 &= \sqrt{(3-3)^2 + (3-2)^2 + (59-55)^2 + (3-2)^2 + (3-3)^2} \\
 &= \sqrt{0^2 + 1^2 + 4^2 + 1^2 + 0^2} \\
 &= 4.242640687 \\
 d4 &= \sqrt{(3-3)^2 + (3-3)^2 + (59-40)^2 + (3-3)^2 + (3-2)^2} \\
 &= \sqrt{0^2 + 0^2 + 19^2 + 0^2 + 1^2} \\
 &= 19.02629759
 \end{aligned}$$

d5

$$= \sqrt{(3-3)^2 + (3-4)^2 + (59-41)^2 + (3-3)^2 + (3-2)^2}$$

$$= \sqrt{0^2 + 1^2 + 18^2 + 0^2 + 1^2}$$

$$= 18.05547009$$

d20 =

$$= \sqrt{(3-3)^2 + (3-4)^2 + (59-50)^2 + (3-3)^2 + (3-3)^2}$$

$$= \sqrt{0^2 + 1^2 + 9^2 + 0^2 + 0^2}$$

$$= 9.848857802$$

Hasil perhitungan jarak d1 sampai d20 pada Tabel 4.

Tabel 4. Hasil Perhitungan Jarak *Euclidean Distance*

No	Jarak Euclidean	Label
1	15.03329638	Kurang Mampu
2	10	Kurang Mampu
3	4.242640687	Kurang Mampu
4	19.02629759	Kurang Mampu
5	18.05547009	Kurang Mampu
6	20.02498439	Kurang Mampu
7	13.45362405	Mampu
8	22	Kurang Mampu
9	16.40121947	Kurang Mampu
10	40	Kurang Mampu
11	26.07680962	Kurang Mampu
12	30.0499584	Mampu
...
20	9.848857802	Kurang Mampu

- c. Mengurutkan objek-objek dari jarak terkecil untuk mengurutkan nya kita hanya perlu membuat urutan dari data yang mempunyai jarak terkecil ke yang terbesar.

Tabel 5. Hasil Pengurutan Jarak *Euclidean*

No	Jarak Euclidean	Label
3	4.242640687	Kurang Mampu
14	6.08276253	Kurang Mampu
18	6.08276253	Kurang Mampu
19	9.219544457	Kurang Mampu
20	9.848857802	Kurang Mampu
2	10	Kurang Mampu
7	13.45362405	Mampu
1	15.03329638	Kurang Mampu
9	16.40121947	Kurang Mampu
17	17	Kurang Mampu
5	18.05547009	Kurang Mampu
4	19.02629759	Kurang Mampu

- d. Mengumpulkan kategori Y (Klasifikasi K-NN), pada tahap ini kita hanya mengambil data sesuai dengan jumlah k yang tentukan dilangkah 1, pada langkah 1, k yang ditentukan adalah K=3, maka kita memilih 3 angka data terbaik saja seperti pada Tabel 6.

Tabel 6. Jarak Terdekat dan Nilai K

Perhitungan jarak			
Dataset	Jarak	Klasifikasi	K
3	4.242640687	Kurang Mampu	1
14	6.08276253	Kurang Mampu	2
18	6.08276253	Kurang Mampu	3

Pada tahap ini mengelompokkan hasil *class* data uji dengan K=3 berdasarkan hasil *class* yang didapat. Dengan menggunakan 216 data training dan 93 data testing pada tahun 2021 sampai 2022 sehingga diperoleh hasil pada Tabel 7.

Tabel 7. Hasil Pengujian Nilai K=3

No	Nama Kepala Keluarga	Actual	Prediksi
1	xxxxxxxx	Kurang Mampu	Kurang Mampu
2	xxxxxxxx	Kurang Mampu	Kurang Mampu
3	xxxxxxxx	Kurang Mampu	Kurang Mampu
4	xxxxxxxx	Mampu	Mampu
5	xxxxxxxx	Kurang Mampu	Kurang Mampu

6	xxxxxxxx	Kurang Mampu	Kurang Mampu
7	xxxxxxxx	Mampu	Mampu
8	xxxxxxxx	Kurang Mampu	Kurang Mampu
9	xxxxxxxx	Mampu	Mampu
10	xxxxxxxx	Kurang Mampu	Kurang Mampu
11	xxxxxxxx	Kurang Mampu	Kurang Mampu
12	xxxxxxxx	Mampu	Mampu
13	xxxxxxxx	Kurang Mampu	Mampu
14	xxxxxxxx	Kurang Mampu	Kurang Mampu
15	xxxxxxxx	Kurang Mampu	Mampu
16	xxxxxxxx	Kurang Mampu	Kurang Mampu
17	xxxxxxxx	Kurang Mampu	Kurang Mampu
18	xxxxxxxx	Mampu	Kurang Mampu
19	xxxxxxxx	Kurang Mampu	Kurang Mampu
20	xxxxxxxx	Kurang Mampu	Kurang Mampu
21	xxxxxxxx	Kurang Mampu	Kurang Mampu
22	xxxxxxxx	Kurang Mampu	Mampu
23	xxxxxxxx	Mampu	Mampu
24	xxxxxxxx	Kurang Mampu	Mampu
25	xxxxxxxx	Kurang Mampu	Kurang Mampu
..
93	xxxxxxxx	Kurang Mampu	Kurang Mampu

4.3. Hasil Perhitungan Confusion Matrix

Pada tahap *confussion matrix* ini kita melakukan pengujian untuk memperkirakan obyek yang benar dan salah.

Tabel 8. Confussion matrix

Aktual/ Sebenarnya	Prediksi	
	Positif	Negatif
Positif	13	8
Negatif	5	67

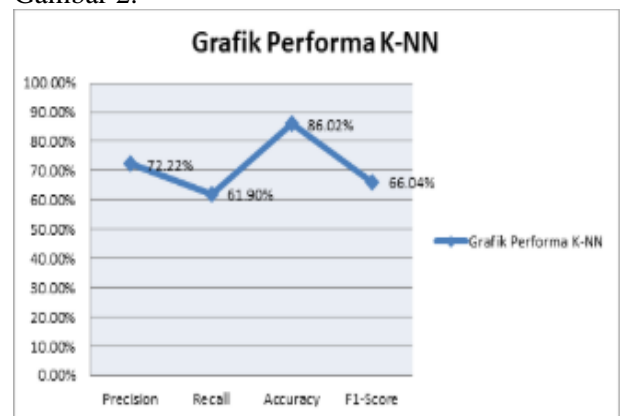
$$\begin{aligned} \text{Precision} &= (\text{TP}) / (\text{TP} + \text{FP}) \\ &= (13) / (13 + 5) \times 100 \\ &= 72,22 \% \end{aligned}$$

$$\begin{aligned} \text{Recall} &= (\text{TP}) / (\text{TP} + \text{FN}) \\ &= (13) / (13 + 8) \times 100 \\ &= 61,90 \% \end{aligned}$$

$$\begin{aligned} \text{Accuracy} &= (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \\ &= (13 + 67) / (13 + 5 + 8 + 67) \times 100 \\ &= 86,02\% \end{aligned}$$

$$\begin{aligned} F-1 \text{ Score} &= (2 \times 0,61 \times 0,72) / (0,61 + 0,72) \\ &= (0,8784) / (1,33) \\ &= 0,6604 \times 100\% \\ &= 66,04\% \end{aligned}$$

Berikut ini grafik data performa K-NN pada Gambar 2.



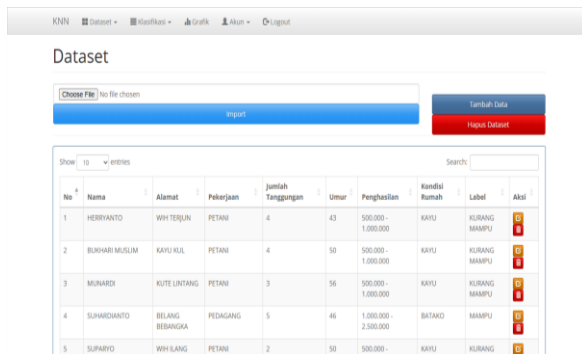
Gambar 2. Grafik Performa K-NN

4.4. Hasil Implementasi Sistem

Berdasarkan perancangan sistem analisa data mining untuk memprediksi masyarakat kurang mampu menggunakan metode KNN sebagai berikut:

1. Halaman dataset

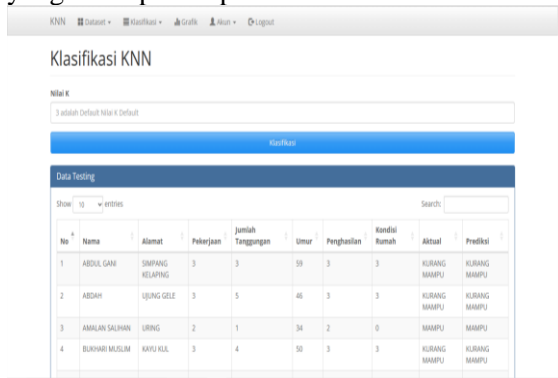
Pada halaman dataset ini menampilkan tabel hasil dari dataset yang diimport ke dalam sistem. Berikut merupakan tampilan halaman dataset yang ditampilkan pada Gambar 3.



Gambar 3. Halaman Menu Dataset

2. Halaman menu klasifikasi

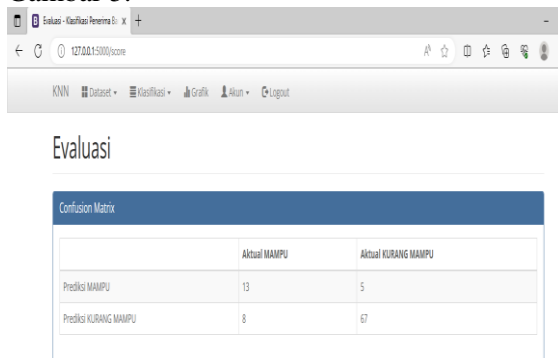
Pada menu klasifikasi ini menampilkan hasil klasifikasi algoritma K-Neares Neighbor yang sudah diprediksi dengan algoritma K-Nearest Neighbor dan juga menampilkan akurasi algoritma K-Nearest Neighbor. Berikut merupakan tampilan halaman menu klasifikasi yang ditampilkan pada Gambar 4.



Gambar 4. Halaman Menu Klasifikasi

3. Halaman menu evaluasi

Pada menu evaluasi ini menampilkan hasil performa algoritma K-Nearest Neighbor dari analisa prediksi masyarakat kurang mampu yang sudah dilakukan klasifikasi pada tabel *confusion matrix*. Berikut merupakan tampilan halaman menu evaluasi yang ditampilkan pada Gambar 5.



Gambar 5. Halaman Menu Evaluasi

4. Halaman menu sebaran data

Pada menu sebaran data ini terdapat kumpulan data berdasarkan data yang ada pada sistem ini. Berikut merupakan tampilan halaman menu sebaran data yang ditampilkan pada Gambar 6.



Gambar 6. Halaman Menu Sebaran Data

5. KESIMPULAN

- Variabel yang digunakan pada penelitian ini adalah data Kepala Keluarga yaitu: Jumlah pekerjaan, Jumlah tanggungan, Umur, Penghasilan dan Kondisi rumah.
- Penelitian ini menggunakan data training dan data testing dengan rasio perbandingan 70:30 dari dataset yang berjumlah 309 data yaitu data training sebanyak 216 data, dan data testing sebanyak 93 data.
- Hasil penelitian ini menghasilkan sebuah output prediksi metode *K-Nearest Neighbor* dengan menggunakan 2 class, yaitu: Mampu dan Tidak Mampu
- Besarnya nilai akurasi yang diperoleh menunjukkan Algoritma *K-Nearest Neighbor* dengan menggunakan pendekatan *Euclidean Distance* dan $K=3$ didapatkan tingkat keakurasian algoritma sebesar 86,02%.
- Berdasarkan penelitian prediksi Masyarakat Kurang Mampu di Kecamatan Pegasing, menggunakan algoritma K-Nearest Neighbor yang menghasilkan nilai performa algoritma yang cukup baik yaitu: *accuracy* sebesar 86,02%, *precision* sebesar 72,22%, *recall* sebesar 61,90% dan *F1-Score* sebesar 66,04%

DAFTAR PUSTAKA

- [1] T. Asril, "Prediction of Students Study Period using K-Nearest Neighbor Algorithm," *International Journal of Emerging Trends in Engineering Research*, vol. 8, No.6, pp. 2585–2593, 2020. <https://doi.org/10.30534/ijeter/2020/60862020>
- [2] F. Kurnia, J. Kurniawan, I. S. Fahmi, I. and S. Monalisa, "Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor Berbasis Euclidean Distance" Vol. 12, 2019.
- [3] S. Yani, F. S. Jumeilah, and M. Kadafi, "Algoritma K-Nearest Neighbor Untuk Menentukan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus : Kelurahan Karya Jaya)," *J. Inf. Technol. Ampera*, vol. 1, no. 2, pp. 75–87, 2020, doi: 10.51519/journalita.volume1.issue2.year2020.page75-87.
- [4] Nurdin, and D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe," *Jurnal Techsi.*, vol. 6, no. 1, pp. 133–155, 2015.
- [5] N. Nurdin, M. Suhendri, Y. Afrilia, and R. Rizal, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC)," *Sistemasi*, vol. 10, no. 2, p. 268, 2021, doi: 10.32520/stmsi.v10i2.1193.
- [6] Nurdin, E. Susanti, H. Al-Kautsar Aidilof, and D. Priyanto, "Comparison of Naive Bayes and Dempster Shafer Methods in Expert System for Early Diagnosis of COVID-19," *Teknik Informatika dan Rekayasa Komputer*, vol. 22, no. 1, pp. 217–230, 2022.
- [7] N. Nurdin, C. C. Pradita, and F. Fadlisyah, "Implementasi Data Mining Untuk Menganalisis Kategori Kompetisi Mahasiswa Menggunakan Algoritma Apriori," *Sisfo: Jurnal Sistem Informasi*, vol. 7, no. 2, pp. 28–45, 2023
- [8] A. Khairi, A. F. Ghozali, and A. D. N. Hidayah, "Implementasi K-Nearest Neighbor (KNN) untuk Mengklasifikasi Masyarakat Pra-Sejahtera Desa Sapikerep Kecamatan Sukapura," *TRILOGI J. Ilmu Teknol. Kesehatan, dan Hum.*, vol. 2, no. 3, pp. 319–323, 2021, doi: 10.33650/trilogi.v2i3.2878.
- [9] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *ILKOM Jurnal Ilmiah*, vol. 12, no. 2, pp. 104–111, 2020, <https://doi.org/10.33096/ilkom.v12i2.539.104-111>.
- [10] N. Nurdin, Z. Aulia, Risawandi, and L. Rosnita, "Application of the K-Nearest Neighbor Method to Determine Recipients of Non-Cash Food Assistance," *Jurnal Ilmu Komputer*, vol. 16, no.2, pp. 115–126, 2023.
- [11] M. Qamal, I. Sahputra, N. Nurdin, M. Maryana, and M. Mukarramah, "Sistem Pendukung Keputusan Penentuan Penerimaan Bantuan PKH Menggunakan Metode Naïve Bayes," *TECHSI - J. Tek. Inform.*, vol. 14, no. 1, pp. 21, 2023, doi: 10.29103/techsi.v14i1.6960.
- [12] R. D. Rasyada, N. Nurdin, and F. Fajriana, "Application of the Profile Matching Analysis Method in Decision Support Systems for Study Program Recommendations," *Sistemasi*, vol. 13, no. 1, pp. 83-95, 2024, doi:10.32520/stmsi.v13i1.3161
- [13] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System model for prediction analytics using / (K-nearest neighbors algorithm. *Journal of Computational and Theoretical Nanosceince*, vol. 16, no.10, pp. 4425–4430, 2019. <https://doi.org/10.1166/jctn.2019.8536>
- [14] A. Razi, "Klasifikasi Penerima Beasiswa Aceh Carong (Aceh Pintar) Di Universitas Malikussaleh Menggunakan Algoritma Knn (K-Nearest Neighbors)," *Jurnal Tika*, vo. 7, no.1, pp. 79–84, 2022. <https://doi.org/10.51179/tika.v7i1.1116>
- [15] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, and W. Gata, "Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K- Nearest Neighbor)" *Jurnal Techno Nusa Mandiri*, vol. 16, no. 1, pp. 1–6, 2019, <https://doi.org/10.33480/techno.v16i1.25>
- [16] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, "Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3, pp. 331–337, 2021, doi: 10.47065/bits.v3i3.1093.
- [17] E. Firasari, U. Khultsum, M. N. Winnarto, and R. Risnandar, "Kombinasi K-NN dan Gradient Boosted Trees untuk Klasifikasi Penerima Program Bantuan Sosial," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1231, 2020, doi: 10.25126/jtiik.0813087.
- [18] A. Khairi, A. F. Ghozali, and A. D. N. Hidayah, "Implementasi K-Nearest Neighbor (KNN) untuk Mengklasifikasi Masyarakat Pra-Sejahtera Desa Sapikerep Kecamatan Sukapura," *TRILOGI J. Ilmu Teknol. Kesehatan, dan Hum.*, vol. 2, no. 3