

Classification of Heart Disease Using Modified K-Nearest Neighbor (MKNN) Method

Aulia Azzahra Ma'aruf Lubis^{1*}, Rozzi Kesuma Dinata², Hafizh Al-Kautsar Aidilof³

^{1,2,3} Universitas Malikussaleh, Indonesia

*Corresponding Author Email: *aulia.190170071@mhs.unimal.ac.id

ABSTRAK

Received: 11 March 2024
Revised: 20 March 2024
Accepted: 31 March 2024
Available online: 1 April 2024

Kata Kunci:

Penyakit Jantung, Modified K-Nearest Neighbor (MKNN), Klasifikasi, Nilai K

Penyakit jantung memiliki banyak jenis dan gejala yang dialami. Penyakit jantung adalah sebuah kondisi ketika organ jantung tidak dapat bekerja sebagaimana fungsinya dengan baik. Jantung adalah organ penting dalam tubuh manusia yang dimana fungsinya adalah memompa darah ke seluruh tubuh. Karena itu dibutuhkannya diagnosa awal untuk pencegahan penyakit jantung dengan memanfaatkan system yang dapat dibuat untuk diagnosa awal pada gejala yang dialami. Yang pada penelitian ini akan menggunakan metode *Modified K-Nearest Neighbor* (MKNN) dalam mengklasifikasikan penyakit jantung berdasarkan kriteria atau gejala yang ada. Penelitian ini menggunakan 6 kriteria penyakit dan 3 kelas diagnosa penyakit jantung. Dengan melewati beberapa langkah pengerjaan yaitu menghitung jarak *Euclidean*, menghitung nilai validitas dan terakhir menghitung *weight voting* dengan mengandalkan nilai K yang telah ditentukan sejak awal perhitungan. Pada penelitian ini telah ditentukan nilai K=5 dan didapat hasil pengujian akurasi sebesar 85%, dengan *recall* 90% dan *precision* 85%.

ABSTRACT

Keywords:

Heart Disease, Modified K-Nearest Neighbor (MKNN), Classification, K Value

Heart disease has many types and symptoms experienced. Heart disease is a condition when the heart organ cannot work as it functions properly. The heart is an important organ in the human body whose function is to pump blood throughout the body. Therefore, early diagnosis is needed for the prevention of heart disease by utilizing a system that can be made for early diagnosis of symptoms experienced. This study will use the Modified K-Nearest Neighbor (MKNN) method in classifying heart disease based on existing criteria or symptoms. This study used 6 disease criteria and 3 classes of heart disease diagnoses. By going through several steps, namely calculating the Euclidean distance, calculating the validity value and finally calculating the voting weight by relying on the K value that has been determined since the beginning of the calculation. In this study, the value of K = 5 was determined and obtained the results of accuracy testing of 85%, with 90% recall and 85% precision.

1. INTRODUCTION

Data *mining* merupakan serangkaian proses untuk mengekstrak pengetahuan atau pola dari sekumpulan data. Dengan menganalisis data yang tersedia dalam basis data, data *mining* bertujuan untuk memecahkan masalah dan memberikan kontribusi dalam pengambilan keputusan. *Output* dari proses data *mining* ini dapat digunakan untuk meningkatkan kualitas pengambilan Keputusan di berbagai bidang. Terdapat beberapa teknik yang dapat kita gunakan dalam data *mining* untuk menemukan sebuah informasi baru, salah satu tekniknya ialah klasifikasi.

Klasifikasi adalah Teknik mengelompokkan data secara terstruktur sesuai dengan aturan dan kaidah yang telah ditentukan sebelumnya. Dalam konteks diagnosis penyakit jantung, klasifikasi sendiri bertujuan untuk meningkatkan akurasi dan efektivitas dalam mengklasifikasikan kondisi pasien.

Dengan data dari *World Health Organization* (WHO) pada tahun 2021 yang memiliki perkiraan 17,8 juta kematian penyakit jantung dan juga berdasarkan data BPJS Kesehatan pada tahun 2021-2022 pembiayaan mencapai jumlah Rp.7,7 triliun dan masih menjadi penyebab utama kematian di Indonesia. Sehingga klasifikasi ini bertujuan untuk membuat sebuah sistem yang mengklasifikasikan data pasien dengan penyakit jantung. Dimana metode yang dipakai ialah metode *Modified K-Nearest Neighbor* (MKNN).

2. RESEARCH METHOD

Jantung adalah organ yang berdenyut secara ritmis dengan fungsi utama memompa dara ke sistem peredaran darah baik sistematis maupun pulmoner. Terletak di mediastinum, di antara kedua paru-paru, jantung merupakan pusat dari sistem peredaran darah dalam tubuh manusia. (Purnamasari & Setiyadi, 2019)

Data *mining* merupakan proses ekstraksi dan transformasi data menjadi informasi yang bermakna yang memungkinkan pengguna untuk mengakses data dalam jumlah besar secara cepat. Dengan menggunakan teknik yang sesuai, proses data *mining* dapat menghasilkan hasil yang optimal. Setiap entitas data dalam proses data *mining* terdiri dari kelas tertentu bersama dengan variabel dan faktor-faktor penentu yang menggambarkan kelas tersebut. (Dinata, Akbar, & Hasdyna, 2020)

Klasifikasi ialah suatu sistem untuk mengorganisasi pengetahuan (*for organizing knowledge*), yang wujud dari pengetahuan tersebut ialah berupa buku, dokumen (tercetak), ataupun sumber-sumber elektronik (tidak tercetak). (Mohamad, 2019)

Klasifikasi adalah salah satu aktivitas dalam data *mining* di mana data diekstraksi dan kemudian dilakukan prediksi, label atau kategori untuk setiap data berdasarkan fitur-fiturnya. (Pratama, Aidilof, Saputra, & Iswandi, 2020)

Metode *K-Nearest Neighbor* (KNN) dikenal juga dengan *lazy learner* (pembelajaran malas) karena hanya melakukan proses penyimpanan data, tanpa proses belajar (dari data) untuk menghasilkan model yang berupa aturan, formula matematis, maupun *hyperplane*. (Suyanto, Rachmawati, Sulistiyono, Wulandari, & Fachrie, 2022)

Algoritma *Modified K-Nearest Neighbor* adalah metode yang merupakan pengembangan dari algoritma KNN. Metode ini memiliki penambahan bagian langkah baru untuk pengklasifikasian, yang dimana ada pada perhitungan nilai validitas dan perhitungan nilai bobot. (Wafiyah, Hidayat, & Perdana, 2017)

Algoritma MKNN (*Modified K-Nearest Neighbor*) adalah metode yang merupakan pengembangan dari algoritma KNN. Metode ini memiliki penambahan bagian langkah baru untuk pengklasifikasian nilai validitas dan perhitungan nilai bobot. (Wafiyah, Hidayat, & Perdana, 2017).

Dibawah ini merupakan langkah-langkah proses metode *Modified K-Nearest Neighbor* : (Hartini, Lefanska, Ursia, Prasetyo, & Sugiharto, 2022)

1. Menghitung Jarak *Euclidean*

Perhitungan jarak *Euclidean* dalam metode *Modified K-Nearest Neighbor* (MKNN) dilakukan dengan dua kali pencarian jarak *Euclidean* yang digunakan. Pertama untuk mendefinisikan jarak antara dua titik antara data *training*, dan yang kedua jarak antara dua titik data *training* dengan data *testing*. Rumusnya dinyatakan dengan persamaan (1) :

$$d(x_i, y_j) = \sqrt{\sum_{i=0}^n (x_i - y_j)^2} \quad (1)$$

Keterangan :

x = data training

y = data testing

i = nilai atribut

n = dimensi atribut

2. Menghitung nilai validitas

Perhitungan nilai validitas digunakan untuk memvalidasi data training. Proses validasi dilakukan pada setiap jarak *Euclidean* antar data training dengan nilai yang paling terkecil dan dilihat sebanyak nilai K yang ditentukan. Dan hasil dari proses validasi yang telah didapat digunakan sebagai sumber informasi paada data tersebut. Nilai validitas ddapat dihitung dengan menggunakan rumus yang ditemukan dalam persamaan (2) :

$$Validity(x) = \frac{1}{H} \sum_{i=0}^n S(lbl(x), lbl(Ni(x))) \quad (2)$$

Keterangan :

H = jumlah titik terdekat

lbl (x) = kelas x

lbl (Ni(xi)) = label kelas terdekat x

diadaptasi dengan syarat dibawah ini. Pada fungsi S digunakan untuk menghitung kesamaan antara titi x, dan data ke-I dari tetangga terdekat yang dinyatakan pada rumus dibawah ini :

$$S(a, b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \quad (3)$$

Keterangan :

a = kelas a pada data training

b = kelas lainnya selain a pada data training

3. Menghitung nilai *weight voting*

Perhitungan *weight voting* berdampak pada data yang lebih dekat dan memiliki nilai validitas yang lebih tinggi. Mengalikan nilai validitas dengan *weight* dapat mengatasi kelemahan data yang lebih jauh dengan *weight*, rumus menghitung *weight voting* dinyatakan pada persamaan (4) :

$$W(i) = Validity(i) \times \frac{1}{de + \alpha} \quad (4)$$

Keterangan :

W(i) = perhitungan *weight voting*

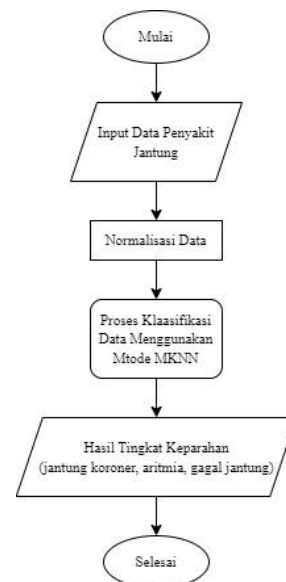
Validity (i) = nilai validitas

de = jarak *Euclidean*

α = nilai alpha (0,5)

3. RESULT AND DISCUSSION

3.1 Skema Sistem



Gambar 1. Skema Sistem

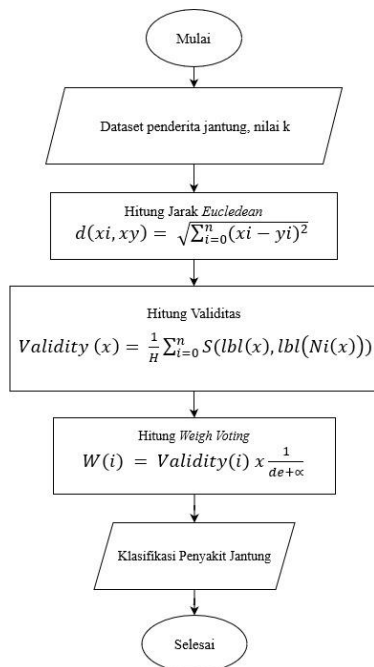
Data yang diambil pada proses penelitian ini dilakukan di Rumah Sakit Sari Mutiara Lubuk Pakam, Provinsi Sumatera Utara, pada bulan Mei 2023 sampai dengan bulan Juli 2023.

Dibawah ini merupakan penjelasan mengenai proses alur sistem melalui point-point sebagai berikut :

1. Proses sistem dimulai
2. User menginput berupa data-data pasien penyakit jantung dengan atributnya berupa jenis kelamin, umur, *Respiratory Rate* (RR), *Heart Rate* (HR), tekanan darah dan nyeri dada.
3. Selanjutnya user mengklik proses klasifikasi, lalu proses perhitungan algoritma *Modified K-Nearest Neighbor* dilakukan oleh sistem.
4. Setelah itu akan ditampilkannya hasil akhir penyakit jantung yang di golongkan dengan penyakit jantung koroner, aritmia dan gagal jantung.
5. Proses sistem selesai.

a. Perancangan Sistem Klasifikasi

Klasifikasi penyakit jantung dengan menggunakan metode MKNN yang memiliki beberapa tahap. Pertama, memasukkan dataset gejala penyakit jantung yang sudah dinormalisasikan. Kemudian, menghitung jarak *euclidean* antara setiap data dengan data uji yang baru. Selanjutnya, memilih sejumlah K tetangga terdekat berdasarkan hasil perhitungan jarak *Euclidean* tersebut. Setelah itu, menghitung nilai validitas data latih. Terakhir, menghitung *weighted voting* untuk menentukan klasifikasi akhir dari data uji. Agar lebih jelas berikut diagram alur proses klasifikasi *Modified K-Nearest Neighbor* (MKNN).



Gambar 2. Alur Sistem MKNN

b. Manajemen Basis Data

3.3.1 Rancangan Struktur Tabel Admin

Tabel 1. Admin

Field	Type	Width
admin_id	Integer	11
admin_nama	Varchar	150

admin_username	Varchar	150
admin_password	Varchar	255
admin_level	Tiny integer	4

3.3.2 Rancangan Struktur Tabel Kriteria

Tabel 2. Kriteria

Field	Type	Width
kriteria_id	Integer	11
kriteria_nama	Varchar	150

3.3.3 Rancangan Struktur Tabel Testing

Tabel 3. Testing

Field	Type	Width
testing_id	Integer	11
Testing_jenis_kelamin	Varchar	35
Testing_umur	Integer	3
Testing_tekanan_darah	Integer	3
Testing_hr	Integer	3
Testing_rr	Integer	3
Testing_nyeri_dada	Integer	3

3.3.4 Rancangan Struktur Tabel Training

Tabel 4. Training

Field	Type	Width
training_id	Integer	11
training_jenis_kelamin	Varchar	35
training_umur	Integer	3
training_tekanan_darah	Integer	3
training_hr	Integer	3
training_rr	Integer	3
training_nyeri_dada	Integer	3
Training_klasifikasi	Varchar	100

3.3.5 Rancangan Struktur Tabel Hasil

Tabel 5. Hasil

Field	Type	Width
id	Integer	11
admin_id	Integer	11
testing_id	Integer	11
urutan	Small integer	7
nilai_k	Tiny integer	3
weight	Varchar	255
posisi	Varchar	255
klasifikasi	Varchar	255
klasifikasi_hasil	Varchar	255
tanggal	Datetime	

c. Perhitungan Manual

Perhitungan manual menggunakan nilai $K=1$ dengan $K=5$ diuji pada nilai $\alpha=0,5$, yang memiliki 80 data latih dan 20 data uji. Kemudian dihitung dengan perhitungan metode MKNN yang dibandingkan dengan nilai akurasi nantinya.

3.4.1 Jarak Euclidean antar data training

Mengurutkan jarak *Euclidean* antar data *training* ini berfungsi untuk mengambil data kecil dari data *training* agar bisa menghitung nilai validitas untuk langkah perhitungan selanjutnya. Berikut 10 hasil perhitungannya :

Tabel 6. Jarak Euclidean data training

No	Jarak Euclidean	Klasifikasi K=5	Kelas Tetangga
1	0	PJK	PJK
	4,58257	PJK	
	5,09902	PJK	
	5,2915	PJK	
	7	PJK	
2	0	PJK	PJK
	1,14142	PJK	
	3,16227	PJK	
	4,69042	PJK	
	4,89898	PJK	
3	0	PJK	PJK
	4,1231	PJK	
	4,58258	PJK	
	4,89898	PJK	
	6,08276	PJK	
4	0	PJK	PJK
	4,24264	PJK	
	5,2915	PJK	
	6,08276	PJK	
	7	PJK	
5	0	PJK	PJK
	6,7082	GJ	
	7	PJK	
	7,54983	PJK	
	8	GJ	
6	0	GJ	GJ
	3,4641	GJ	
	3,60555	GJ	
	4,58258	GJ	
	4,79583	GJ	
7	0	PJK	PJK
	1,41421	PJK	
	2,44949	PJK	
	4,69042	PJK	
	5,91608	PJK	
8	0	GJ	GJ

9	2,82842	GJ	GJ
	3,60555	GJ	
	3,60555	GJ	
	4,24264	GJ	
	0	GJ	
10	4,69042	GJ	GJ
	4,79583	GJ	
	5,38516	GJ	
	5,38516	GJ	
	0	GJ	
10	3,60555	GJ	GJ
	4,24264	GJ	
	4,58258	GJ	
	4,69042	GJ	
	0	GJ	

3.4.2 Nilai Validitas

Ada dua syarat menghitung validitas :

- 1) Jika kelas (klasifikasi) data sama maka nilai validitas yang diambil adalah 1
- 2) Jika kelas (klasifikasi) data tidak sama maka nilai validitas yang diambil adalah 0

Berikut 30 tabel hasil nilai validitas yang diambil dari jarak *Euclidean* antar data *training* yang paling terkecil dan dilihat dari kelas datanya maka dimasukkan nilai validitas dari syarat diatas, untuk nilai validitas 31-80 nilainya sama yaitu 0,3333. Maka tabelnya sebagai berikut :

Tabel 7. Nilai Validitas

No	Kelas	Kelas Tetangga	Validitas
1	PJK	PJK	0,33333333
2	PJK	PJK	0,33333333
3	PJK	PJK	0,33333333
4	PJK	PJK	0,33333333
5	PJK	PJK	0,33333333
6	GJ	GJ	0,33333333
7	PJK	PJK	0,33333333
8	GJ	GJ	0,33333333
9	GJ	GJ	0,33333333
10	GJ	GJ	0,33333333
11	PJK	PJK	0,33333333
12	PJK	PJK	0,33333333
13	PJK	PJK	0,33333333
14	GJ	GJ	0,33333333
15	GJ	GJ	0,33333333
16	GJ	GJ	0,33333333
17	GJ	GJ	0,33333333
18	GJ	PJK	0
19	AR	AR	0,33333333
20	AR	AR	0,33333333
21	AR	AR	0,33333333
22	AR	AR	0,33333333

23	PJK	PJK	0,33333333
24	PJK	PJK	0,33333333
25	PJK	PJK	0,33333333
26	GJ	GJ	0,33333333
27	AR	PJK	0
28	AR	AR	0,33333333
29	AR	AR	0,33333333
30	GJ	GJ	0,33333333

3.4.3 Jarak Euclidean data training dengan data testing

Jarak *Euclidean* antar data *training* dengan total 80 data dihitung masing-masing objek dengan setiap 20 data *testing* yang ada. Berikut tabel 10 data hasil perhitungan jarak *Euclidean* untuk setiap data *training* dan *testing* :

Tabel 8. Jarak Euclidean data Training dan Testing

No	Euclidean
1	30,38091506
2	32,21800739
3	30,757113
4	25,74878638
5	29,18903904
6	31,70173497
7	32,71085447
8	30
9	30,16620626
10	27,40437921

3.4.4 Perhitungan weight voting

Untuk memperoleh nilainya, perhitungan *weight voting* dilakukan dengan rumus mengkalikan nilai validitas yang telah didapat dengan pembagian dari satu per jumlah jarak *Euclidean* antar data *training* dan data *testing* dengan nilai *alpha* yaitu 0,5. Berikut 10 contoh perhitungan manual untuk *weight voting* :

Tabel 9. Nilai Validitas

No	d	Euclidean	Kelas
1	(1, 1)	0,010794153	PJK
2	(1, 2)	0,01018807	PJK
3	(1, 3)	0,010653575	PJK
4	(1, 4)	0,012686301	PJK
5	(1, 5)	0,011216261	PJK
6	(1, 6)	0,010341058	GJ
7	(1, 7)	0,010026842	PJK
8	(1, 8)	0,010918033	GJ
9	(1, 9)	0,010858859	GJ
10	(1, 10)	0,011933611	GJ

3.4.1 Menentukan kelas dari data testing

Langkah akhir menentukan kelas dari data *testing* sebanyak nilai *K* yang telah ditentukan yaitu *K*=5 dengan syarat kelas dari data *testing* adalah total *weight voting* terbesar. Berikut 10 tabel urutannya :

Tabel 10. Jarak Euclidean data training

No	Jarak Euclidean	Klasifikasi K=5	Kelas Tetangga
1	(46, 1)	0,084	AR
	(70, 1)	0,07615	AR
	(71, 1)	0,06552	AR
	(67, 1)	0,03977	AR
	(37, 1)	0,06055	AR
2	(37, 2)	0,1332	AR
	(70, 2)	0,1129	AR
	(46, 2)	0,04824	AR
	(71, 1)	0,04318	AR
3	(21, 1)	0,03568	AR
	(4, 3)	0,07203	PJK
	(12, 3)	0,07021	PJK
	(11, 3)	0,05846	PJK
	(5, 1)	0,0526	PJK
4	(13, 1)	0,04358	PJK
	(6, 4)	0,05658	GJ
	(9, 4)	0,0477	GJ
	(59, 4)	0,04243	GJ
	(53, 4)	0,04038	GJ
5	(58, 4)	0,03731	GJ
	(4, 5)	0,05947	PJK
	(50, 5)	0,0575	PJK
	(38, 5)	0,05571	PJK
	(1, 5)	0,04527	PJK
6	(45, 5)	0,04398	PJK
	(4, 6)	0,06853	PJK
	(38, 6)	0,0519	PJK
	(55, 6)	0,0519	PJK
	(12, 6)	0,04997	PJK
7	(50, 6)	0,04937	PJK
	(9, 7)	0,09514	GJ
	(79, 7)	0,05947	GJ
	(58, 7)	0,05846	GJ
	(59, 7)	0,05846	GJ
8	(6, 7)	0,05571	GJ
	(14, 8)	0,084	GJ
	(77, 8)	0,0428	GJ
	(72, 8)	0,03202	GJ
9	(76, 8)	0,03187	GJ
	(32, 8)	0,02829	GJ
	(55, 9)	0,1332	PJK

	(49 , 9)	0,07615	PJK	
	(48 , 9)	0,07021	PJK	
	(13 , 9)	0,05571	PJK	
	(50 , 9)	0,04243	PJK	
	(37 , 10)	0,04937	AR	
	(70 , 10)	0,04137	AR	
10	(46 , 10)	0,03977	AR	AR
	(74 , 10)	0,03099	AR	
	(34 , 10)	0,03031	GJ	

3.4.2 Akurasi

Dalam menghitung nilai akurasi penulis menggunakan *confusion matrix* berikut tabel hasil perhitungan akurasi nya :

Tabel 11. Perbandingan Kelas

No	Status Benar	Hasil Prediksi
1	AR	AR
2	AR	AR
3	PJK	PJK
4	GJ	GJ
5	PJK	PJK
6	PJK	PJK
7	GJ	GJ
8	GJ	GJ
9	PJK	PJK
10	AR	AR
11	PJK	PJK
12	PJK	PJK
13	GJ	GJ
14	GJ	GJ
15	PJK	AR
16	PJK	AR
17	PJK	AR
18	PJK	PJK
19	PJK	PJK
20	AR	AR

Maka dari tabel diatas didapat nilai *precision*, *recall* dan *accuracy* yang dapat dilihat pada tabel dibawah ini:

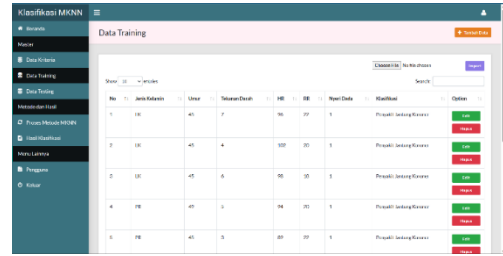
Tabel 12. Confusion Matrix

No	Kelas	Precision	Recall	Accuracy
1	PJK	100%	72%	
2	GJ	100%	100%	
3	AR	57%	100%	85%
	All	85%	90%	

d. Implementasi Sistem

3.5.1 Halaman data training

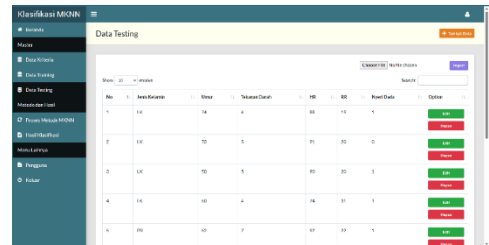
Berikut tampilan halaman dari 80 data *training* yang akan menampilkan seluruh data *training* yang telah di *input* berdasarkan kriteria yang ada :



Gambar 3. Data Training

3.5.2 Halaman data testing

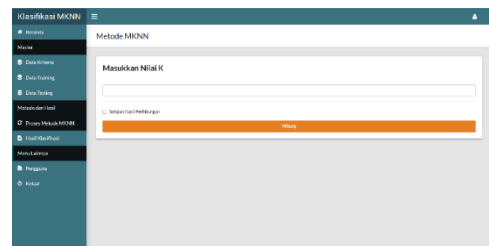
Berikut tampilan halaman dari 20 data *testing* pada sistem klasifikasi penyakit jantung yang telah dimasukkan datanya. Data bisa dihapus dan diedit oleh pengguna.



Gambar 4. Data Testing

3.5.3 Proses perhitungan MKNN

Halaman proses metode *Modified K-Nearest Neighbor* (MKNN) ini akan muncul ketika pengguna menekan menu "proses metode MKNN" dan akan diarahkan untuk memasukkan nilai k untuk melanjutkan proses klasifikasi, tampilan sistem nya seperti gambar berikut.



Gambar 5. Proses Metode MKNN

3.5.4 Halaman hasil klasifikasi

Setelah dimasukkan nilai K pada halaman proses metode MKNN maka akan ditampilkannya halaman hasil klasifikasi.

