

## ANALISIS METODE *K-NEAREST NEIGHBOUR* (KNN) DALAM KLASIFIKASI DATA IRIS BUNGA

Permana Putra<sup>1)</sup>, Akim M H Pardede<sup>2)</sup>, Siswan Syahputra<sup>3)</sup>

<sup>1,2,3</sup> STMIK KAPUTAMA

Jl.Veteran No.4A-9A, Binjai, Sumatra Utara, Telp:(061)8828840, Fax: (061)8828845

Email: [permanaputra058@gmail.com](mailto:permanaputra058@gmail.com)

### ABSTRACT

*This study describes the use of the k-nearest neighbor (K-NN) method for the classification of iris data from UCI machine learning. In this research, several stages and methods were carried out. At this stage of research, the k-nearest neighbor (K-NN) method is used for testing with a certain K value. In measuring the results of the classification of iris data, it focuses on the level of accuracy of the iris data. In this study, the data used were 135 iris data from UCI machine learning. The measurement of the distance on the K-NN in this study uses the Euclidean distance with neighboring values of 3,4,5,6,7,8,9. From the implementation obtained, namely the k-nearest neighbor classification of the iris data. In this study, the process of determining the type of iris will be carried out, in determining the iris, 4 factors are used which are the characteristics of the iris itself. The four characteristics of the iris flower used in this research are sepal length, sepal width, petal length, and petal width. As for the results of testing the K-Nearest Neighbor method in data classification. The author's test uses variations in the K value of K-Nearest Neighbor 3,4,5,6,7,8,9. Has a very good percentage of accuracy compared to only K-NN. The test results show the K-Nearest Neighbor method in data classification has a good percentage accuracy when using random data. The percentage of variation in the value of K K-Nearest Neighbor 3,4,5,6,7,8,9 has a percentage of 100%.*

**Keywords:** K-Nearest Neighbor, Data Iris, flower iris

### ABSTRAK

Penelitian ini menjelaskan penggunaan metode k-nearest neighbor(K-NN) terhadap klasifikasi data iris bunga dari UCI machine learning. Di dalam penelitian ini dilakukan beberapa tahapan dan metode. Pada tahapan penelitian ini menggunakan metode k-nearest neighbor (K-NN) untuk pengujian dengan nilai K tertentu. Pada pengukuran hasil klasifikasi data iris bunga berfokus pada tingkat akurasi data iris bunga. Di dalam penelitian ini data yang digunakan yaitu sebanyak 135 data iris bunga yang berasal dari UCI machine learning. Pengukuran jarak pada K-NN dalam penelitian ini menggunakan euclidean distance dengan nilai ketetanggan 3,4,5,6,7,8,9. Dari implementasi yang diperoleh yaitu klasifikasi k-nearest neighbor terhadap data iris bunga. Dalam penelitian ini akan dilakukan proses penentuan jenis bunga iris, dalam penentuan bungairis tersebut digunakan 4 faktor yang merupakan karakteristik sendiri dari bunga iris tersebut. Keempatfaktor karakteristik bunga iris yang digunakan untuk penelitian yaitu *sepal length*, *sepal width*, *petal length*, dan *petal width*. Adapun dari hasil pengujian metode K-Nearest Neighbor dalam klasifikasi data. Adapun pengujian penulis menggunakan variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9. Memiliki persentase akurasi yang sangat baik di dibandingkan hanya K-NN. Hasil pengujian menunjukan metode K-Nearest Neighbor dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data random. Persentase variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9 memiliki persentase 100 % .

**Kata Kunci:** *K-Nearest Neighbor, Iris Data, bunga iris.*

## 1. PENDAHULUAN

Pada tahap ini, analisis dilakukan dengan menganalisis terhadap hasil studi literatur untuk mengetahui dan mendapatkan pemahaman mengenai metode yang penulis analisis yaitu pembagian data training dan data uji dengan Algoritma K-nearest Neighbor.

Dalam penelitian ini akan dilakukan proses penentuan jenis bunga iris, dalam penentuan bungairis tersebut digunakan 4 faktor yang merupakan karakteristik sendiri dari bunga iris tersebut. Keempatfaktor karakteristik bunga iris yang digunakan untuk penelitan yaitu *sepal length, sepal width, petal length, dan petal width*. Sehingga dengan adanya banyak faktor yang diteliti dan record datayang ada dalam penentuan jenis bunga iris , maka diperlukan sebuah metode yang dapat digunakan untuk menghasilkan secara tepat dan akurat dalam menentukan jenis bunga pada umumnya, termasuk penentuan jenis bunga iris pada khususnya.

Dalam melakukan proses klasifikasi data mengacu pada metode kecerdasan buatan yang memfokuskan pada pembelajaran mesin (machine learning). Banyak metode lain untuk mengetahui mesin (machine learning) yang digunakan sebagai proses klasifikasi diantaranya K-Nearest Neighbor (k-NN) dan *Naïve Bayes Classifier*.

Klasifikasi adalah cara yang dilakukan sebagai teknik untuk membentuk model klasifikasi dari contoh data pelatihan. Klasifikasi akan menganalisis input datadan membentuk model dengan menggambarkan kelas data.

Dengan menggunakan metode *K-Nearst Neighbor* (KNN) merupakan salah satu metode klasifikasi terhadap sekumpulan data yang berdasarkan mayoritas dari kategori dan tujuannya untuk mengklasifikasikan obyek baru berdasarkan atribut dan *sample sample* dari *training* data. Sehingga target *output* yang diinginkan mendekati ketepatan dalam

melakukan pengujian pembelajaran.

### 1.1 KLASIFIKASI DATA

Klasifikasi data adalah proses pengurutan dan pengelompokan data ke dalam berbagai jenis, bentuk atau kelas berbeda lainnya. Klasifikasi data memungkinkan pemisahan dan klasifikasi data sesuai dengan persyaratan kumpulan data untuk berbagai tujuan bisnis atau pribadi. Ini terutama merupakan proses manajemen data.

Analisis kelompok sebagai suatu metode untuk melakukan klasifikasi data menjadi beberapa kelompok dengan menggunakan metode pengukuran ukuran asosiasi, sehingga data yang sama berada dalam satu kelompok dan data yang memiliki perbedaan yang besar diletakkan dalam kelompok data lainnya. Masukan untuk sistem analisis kelompok adalah sebuah data set dan kesamaan ukuran antara kedua data tersebut. Sedangkan hasil dari analisis kelompok adalah sejumlah kelompok yang membentuk sebuah partisi atau struktur partisi dari kumpulan data dan deskripsi umum dari setiap kelompok, dimana hal ini sangat penting untuk analisis yang lebih dalam pada karakteristik yang terdapat pada data tersebut.

Pengelompokan data harus menggunakan sebuah pendekatan untuk mencari kesamaan dalam data sehingga mampu menempatkan data ke dalam kelompok-kelompok yang tepat. Pengelompokan data akan membagi kumpulan data ke dalam beberapa kelompok dimana kesamaan dalam sebuah kelompok lebih besar jika dibandingkan dengan kelompok-kelompok lainnya.

Terdapat dua metode pembelajaran yang tersedia pada model klasifikasi yaitu :

- a. Eager learning adalah proses pembelajaran pada data latih secara intensif sehingga model dapat melakukan prediksi ke label kelas output dengan benar. Beberapa metode bersifat eager learning antara

lain, Neural Network, Bayesian, decision tree, Support Vector Machine.

- b. Lazy learning adalah proses pembelajaran tanpa melakukan pelatihan dan hanya menyimpan nilai data latih untuk digunakan pada proses prediksi. Beberapa metode bersifat lazy learning antara lain: K-Nearest Neighbor, Regresi Linear, Fuzzy K-Nearest Neighbor.

Proses klasifikasi pada machine learning memiliki empat komponen (Saputra, 2018) yaitu :

1. Class  
Variabel dependen tetamu harus dalam bentuk yang mewakili label yang dimiliki oleh objek.
2. Energi  
Variabel independen yang diwakili oleh atribut karakteristik data. Misalnya, gaji, kehadiran, merokok, tekanan darah.
3. Dataset Pelatihan  
Satu set data yang memiliki kedua nilai komponen diatas yang digunakan untuk menentukan kelas yang sesuai berdasarkan energi.
4. Menguji dataset  
Satu set data baru yang akan diklasifikasi dengan model yang telah dibuat dan akan dievaluasi pada proses akurasi klasifikasi.

Pada proses klasifikasi, sebelum melakukan prediksi, perlu dilakukan proses pembelajaran terlebih dahulu. Proses pembelajaran tersebut memerlukan data. Data yang diperlukan pada saat proses klasifikasi terdiri atas dua jenis, yaitu :

- a. Data latih atau data training adalah data yang digunakan pada proses pembelajar dalam proses klasifikasi.
- b. Data uji atau data testing adalah data yang digunakan pada proses prediksi dalam proses klasifikasi.

## 1.2 K-NEAREST NEIGHBOUR

Klasifikasi KNN adalah metode non-

parametrik sederhana untuk klasifikasi. Terlepas dari kesederhanaan algoritme, kinerjanya sangat baik, dan merupakan metode tolok ukur yang penting. Klasifikasi KNN membutuhkan metrik dan integer positif  $K$  [42]. Aturan KNN memegang posisi sampel pelatihan dan kelas mereka. Saat memutuskan tentang data masuk baru. Tujuan algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan nilai atribut dan data latih (Okfalisa et al, 2017).

Cover dan Hart memperkenalkan K-Nearest Neighbor pada tahun 1968. K-Nearest Neighbor adalah metode klasifikasi pemalas karena algoritma ini menyimpan semua nilai data pelatihan dan menunda proses pembentukan model klasifikasi hingga data uji diberikan untuk prediksi. (Mulak & Talhar, 2015). Gagasan dalam metode k-Nearest Neighbor adalah untuk mengidentifikasi sampel  $k$  dalam set pelatihan yang variabel independen  $x$  mirip dengan  $u$ , dan menggunakan sampel  $k$  ini untuk mengklasifikasikan sampel baru ini ke dalam kelas,  $v$ .  $F$  adalah fungsi yang halus, sebuah ide yang masuk akal adalah mencari sampel dalam data pelatihan kami yang berada di dekatnya (dalam hal variabel independen) dan kemudian menghitung  $v$  dari nilai  $y$  untuk sampel ini. Jarak atau ukuran ketidaksamaan dapat dihitung antara sampel dengan mengukur jarak menggunakan jarak Euclidean. Jarak Euclidean antara titik-titik adalah (Hu et al, 2016).



Pada gambar dapat dilihat bahwa terlebih dahulu dilakukan perhitungan jarak data uji dengan data latih untuk

selanjutnya dilakukan diidentifikasi tetangga terdekat. Selanjutnya KNN akan memberikan titik ke kelas yang memiliki jarak terdekat dengan tetangganya. Pada gambar diatas data uji memiliki matrik jarak terdekat dengan kelas 1 sehingga data uji akan diklasifikasikan sebagai data kelas 1. Adapun langkah-langkah klasifikasi algoritma KNN adalah sebagai berikut:

1. Tentukan parameter nilai  $k$  = banyaknya jumlah tetangga terdekat
2. Hitung jarak antara data baru dengan semua data training
3. Urutkan jarak dan tetapkan tetangga terdekat berdasarkan jarak minimum ke- $k$
4. Periksa kelas dari tetangga terdekat
5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru

## 2. METODOLOGI PENELITIAN

Perlu dilakukan pengujian dalam penentuan pembagian data training dan data uji terhadap pengaruh jumlah  $K$  dalam K-Nearest Neighbor pada suatu dataset berukuran besar. Dataset yang ada menggunakan UCI Machine Learning Repository. UCI Machine Learning Repository adalah sebuah koleksi database, domain teori, dan data generator yang digunakan oleh komunitas yang mempelajari mesin pembelajaran (machine learning), untuk keperluan analisis empiris dari algoritma machine learning. Dataset yang tersedia pada UCI Machine Learning Repository digunakan oleh pelajar, pendidik, dan peneliti diseluruh dunia sebagai sumber utama dari data set pada machine learning. Jumlah data set yang tersedia pada UCI Machine Learning Repository pada saat ini sudah berjumlah 320 data set yang dapat digunakan sesuai dengan kebutuhan pada pembelajaran machine learning.

Machine learning memungkinkan manusia untuk memprogram komputer sehingga mesin dapat mengenali pola atau belajar dari apa yang dimasukkan ke dalamnya. Konsep pembelajaran adalah

proses mengubah informasi menjadi pengetahuan. Algoritma machine learning adalah data pelatihan, yang mewakili pengalaman, dan hasilnya adalah beberapa keahlian, yang biasanya berupa program komputer lain yang dapat melakukan beberapa tugas (Ion C. Freeman, Ashley J. et al, 2018).

Machine learning memungkinkan dalam klasifikasi data, aplikasi ini mengenali pola dalam data baik dengan pelatihan atau tanpa pelatihan. Di klasifikasi data disebut clustering dalam machine learning. Beberapa contoh algoritma pengelompokan termasuk K-Means, Farthest-First Maksimalisasi-Ekspektasi (EM), dan lainnya (Ion C. Freeman, Ashley J. et al, 2018).

Machine Learning merupakan cabang ilmu artificial intelligence yang memiliki konsep bahwa komputer sebagai mesin memiliki kemampuan untuk melakukan adaptasi terhadap lingkungan yang baru dan mampu mendeteksi pola dari fakta yang ada. Definisi pembelajaran yang dilakukan mesin adalah ketika mesin dari pengalaman  $E$  terhadap tugas  $T$  dan mengukur peningkatan kinerja  $P$ , jika kinerja tugas  $T$  diukur oleh kinerja  $P$ , meningkatkan pengalaman  $E$ .

Proses penelitian ini tedapat beberapa kegiatan, yaitu kegiatan-kegiatan yang terdapat pada penelitian. yaitu observasi lapangan pengumpulan data dan analisis data.

1. Observasi yang dilakukan pada penelitian ini adalah hal yang paling penting. Karena penulis dapat mengetahui tingkat visibilitas yang digunakan. Data-data yang telah dikumpulkan menjadi titik pantauan dalam observasi ini sehingga mendapat hasil yang digunakan.
2. Pengumpulan data literature, jurnal, paper, dan bacaan-bacaan lainnya yang berhubungan dengan algoritma klasifikasi data mining. Peneliti melakukan pengumpulan data dengan mengambil dataset UCI Machine Learning Repository. Peneliti menggunakan dataset iris.

3. Persiapan dan pemilihan data yang didapat seperti melihat struktur tabel yang ada pada database. Pemilihan data dilakukan karena tidak semua tabel serta data yang ada dalam database berhubungan dengan penelitian yang dilakukan, sehingga hanya data yang berkaitan dengan penelitian yang digunakan.
4. Pembersihan data dilakukan untuk memastikan bahwa tidak ada data yang terduplikasi, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Data yang telah bersih dari kesalahan dapat mempermudah penelitian dan mencegah adanya kesalahan pada penelitian.
5. Pada tahap ini menganalisis terhadap pembagian data training dan data uji dengan pengaruh jumlah K algoritma K-Nearest Neighbor.

Pada penelitian ini, untuk mengetahui kinerja dari metode yang digunakan maka akan digunakan sejumlah dataset. Adapun data yang digunakan terdiri dari 2 dataset yang berasal dari *UCI Machine Learning Repository* diantaranya adalah data *ionosphere*, iris dan wine.

Iris merupakan salah satu dataset yang populer, dataset ini mungkin merupakan dataset yang paling terkenal dan dapat ditemukan dalam literatur pengenalan pola yang telah disusun oleh Fisher (1936). Dataset ini terdiri dari 150 *records* dengan 4 *attributes* dan 3 kelas, dimana setiap kelas data terdiri dari 50 *records* (Marshall. 1988).

Kumpulan data minat Iris atau kumpulan data Iris Fisher adalah kumpulan data multivarian yang diperkenalkan oleh ahli statistik dan biolog Inggris Ronald Fisher pada tahun 1936. Penggunaan beberapa pengukuran dalam masalah taksonomi misalnya analisis diskriminan linier. Disebut juga data Iris Anderson diatur karena Edgar Anderson mengumpulkan data untuk mengukur variasi morfologi bunga Iris dari tiga spesies terkait.

Dua dari tiga spesies dikumpulkan semua dari padang rumput yang sama, dan mengambil pada hari yang sama dan diukur pada saat yang sama oleh orang yang sama dengan peralatan yang sama, dan dipetik pada hari yang sama dan diukur pada waktu yang sama oleh orang yang sama dengan peralatan yang sama.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini menganalisis terhadap hasil studi literatur untuk mengetahui dan mendapatkan pemahaman mengenai metode yang penulis analisis yaitu pembagian data training dan data uji dengan pengaruh jumlah K dalam K-nearest Neighbor. Untuk lebih jelas dalam mendeskripsikan proses dalam penelitian ini maka akan dijelaskan tahapan demi tahapan dalam sub bab ini. Adapun tahapan-tahapannya secara garis besar dapat dilihat pada Gambar.



**Gambar.1 Flowchart Rancangan Penelitian**

#### Keterangan :

Adapun dari gambar dapat dilihat flowchart rancangan penelitian dalam penyelesaian algoritma K-NN dijelaskan sebagai berikut :

Langkah pertama dalam melakukan analisis penelitian dilakukan penentuan dataset,



adapun disini penulis menggunakan dataset Uci Machine Learning. Selanjutnya menentukan nilai K dan hitung jarak euclidean dari setiap data, setelah didapat hasil jarak setiap data maka diurutkan berdasarkan ascending. Setelah terurut maka dilakukan penentuan jarak terdekat sehingga di kelompokkan dan berpasangan. Maka selanjutnya cari kelas tetangga terbanyak sehingga mendapatkan hasil klasifikasi.

Keempat faktor karakteristik bunga iris yang digunakan untuk penelitian yaitu *sepal length*, *sepal width*, *petal length*, dan *petal width*. Sehingga dengan adanya banyak faktor yang diteliti dan record data yang ada dalam penentuan jenis bunga iris, maka diperlukan sebuah metode yang dapat digunakan untuk menghasilkan secara tepat dan akurat dalam menentukan jenis bunga pada umumnya, termasuk penentuan jenis bunga iris pada khususnya.

Adapun penelitian ini, menggunakan metode K-Nearest Neighbor dalam klasifikasi data. Pada sub bab ini penulis menggunakan data latih dari dataset iris dengan jumlah data sebanyak 10 record dan data uji sebanyak 1 record, yang mana jumlah atribut data yang diambil berjumlah 4 atribut.

Tabel 1. Data Latih

No	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
1	4.8	3	1.4	0.3	Setosa
2	5.1	3.8	1.6	0.2	Setosa
3	4.6	3.2	1.4	0.2	Setosa
4	5.3	3.7	1.5	0.2	Setosa
5	5.7	3	4.2	1.2	Versicolor
6	5.7	2.9	4.2	1.3	Versicolor
7	6.2	2.9	4.3	1.3	Versicolor
8	6.7	3	5.2	2.3	Virginica
9	6.3	2.5	5	1.9	Virginica
10	6.5	3	5.2	2	Virginica

Tabel 2. Data Uji

Bunga Iris	X1	X2	X3	X4	Species
X	5	3	4	1	???

### Penyelesaian

Berikut merupakan langkah-langkah proses perhitungan metode K-Nearest Neighbor dalam klasifikasi data.

### Keterangan :

X1 = Sepal length in cm

X2 = Sepal width in cm

X3 = Petal length in cm

X4 = Petal width in cm

1. Ditentukan nilai K = 4
2. Hitung jarak Euclidean data uji.

NO	X1	X2	X3	X4	Euclidean distance (5,3,4,1)
1	4.8	3	1.4	0.3	$\sqrt{(4.8-5)^2+(3-3)^2+(1.4-4)^2+(0.3-1)^2} = 2.7$
2	5.1	3.8	1.6	0.2	$\sqrt{(5.1-5)^2+(3.8-3)^2+(1.6-4)^2+(0.2-1)^2} = 24.02$
3	4.6	3.2	1.4	0.2	$\sqrt{(4.6-5)^2+(3.2-3)^2+(1.4-4)^2+(0.2-1)^2} = 2.75$
4	5.3	3.7	1.5	0.2	$\sqrt{(5.3-5)^2+(3.7-3)^2+(1.5-4)^2+(0.2-1)^2} = 2.73$
5	5.7	3	4.2	1.2	$\sqrt{(5.7-5)^2+(3-3)^2+(4.2-4)^2+(1.2-1)^2} = 0.75$
6	5.7	2.9	4.2	1.3	$\sqrt{(5.7-5)^2+(2.9-3)^2+(4.2-4)^2+(1.3-1)^2} = 0.79$
7	6.2	2.9	4.3	1.3	$\sqrt{(6.2-5)^2+(2.9-3)^2+(4.3-4)^2+(1.3-1)^2} = 1.27$
8	6.7	3	5.2	2.3	$\sqrt{(6.7-5)^2+(3-3)^2+(5.2-4)^2+(2.3-1)^2} = 2.45$
9	6.3	2.5	5	1.9	$\sqrt{(6.3-5)^2+(2.5-3)^2+(5-4)^2+(1.9-1)^2} = 1.93$
10	6.5	3	5.2	2	$\sqrt{(6.5-5)^2+(3-3)^2+(5.2-4)^2+(2-1)^2} = 2.16$

3. urutkan jarak secara ascending

NO	X1	X2	X3	X4	Euclidean distance (5,3,4,1)	Jarak ascending
1	4.8	3	1.4	0.3	$\sqrt{(4.8-5)^2+(3-3)^2+(1.4-4)^2+(0.3-1)^2}=2.7$	5
2	5.1	3.8	1.6	0.2	$\sqrt{(5.1-5)^2+(3.8-3)^2+(1.6-4)^2+(0.2-1)^2}=24.02$	10
3	4.6	3.2	1.4	0.2	$\sqrt{(4.6-5)^2+(3.2-3)^2+(1.4-4)^2+(0.2-1)^2}=2.75$	9
4	5.3	3.7	1.5	0.2	$\sqrt{(5.3-5)^2+(3.7-3)^2+(1.5-4)^2+(0.2-1)^2}=2.73$	8
5	5.7	3	4.2	1.2	$\sqrt{(5.7-5)^2+(3-3)^2+(4.2-4)^2+(1.2-1)^2}=0.75$	1
6	5.7	2.9	4.2	1.3	$\sqrt{(5.7-5)^2+(2.9-3)^2+(4.2-4)^2+(1.3-1)^2}=0.79$	2
7	6.2	2.9	4.3	1.3	$\sqrt{(6.2-5)^2+(2.9-3)^2+(4.3-4)^2+(1.3-1)^2}=1.27$	3
8	6.7	3	5.2	2.3	$\sqrt{(6.7-5)^2+(3-3)^2+(5.2-4)^2+(2.3-1)^2}=2.45$	7
9	6.3	2.5	5	1.9	$\sqrt{(6.3-5)^2+(2.5-3)^2+(5-4)^2+(1.9-1)^2}=1.93$	4
10	6.5	3	5.2	2	$\sqrt{(6.5-5)^2+(3-3)^2+(5.2-4)^2+(2-1)^2}=2.16$	6

#### 4. tentukan jarak terdekat

NO	X1	X2	X3	X4	Euclidean distance (5,3,4,1)	Jarak ascending	species
1	4.8	3	1.4	0.3	$\sqrt{(4.8-5)^2+(3-3)^2+(1.4-4)^2+(0.3-1)^2}=2.7$	5	-
2	5.1	3.8	1.6	0.2	$\sqrt{(5.1-5)^2+(3.8-3)^2+(1.6-4)^2+(0.2-1)^2}=24.02$	10	-
3	4.6	3.2	1.4	0.2	$\sqrt{(4.6-5)^2+(3.2-3)^2+(1.4-4)^2+(0.2-1)^2}=2.75$	9	-
4	5.3	3.7	1.5	0.2	$\sqrt{(5.3-5)^2+(3.7-3)^2+(1.5-4)^2+(0.2-1)^2}=2.73$	8	-
5	5.7	3	4.2	1.2	$\sqrt{(5.7-5)^2+(3-3)^2+(4.2-4)^2+(1.2-1)^2}=0.75$	1	Versi color
6	5.7	2.9	4.2	1.3	$\sqrt{(5.7-5)^2+(2.9-3)^2+(4.2-4)^2+(1.3-1)^2}=0.79$	2	Versi color
7	6.2	2.9	4.3	1.3	$\sqrt{(6.2-5)^2+(2.9-3)^2+(4.3-4)^2+(1.3-1)^2}=1.27$	3	Versi color
8	6.7	3	5.2	2.3	$\sqrt{(6.7-5)^2+(3-3)^2+(5.2-4)^2+(2.3-1)^2}=2.45$	7	-
9	6.3	2.5	5	1.9	$\sqrt{(6.3-5)^2+(2.5-3)^2+(5-4)^2+(1.9-1)^2}=1.93$	4	virginia
10	6.5	3	5.2	2	$\sqrt{(6.5-5)^2+(3-3)^2+(5.2-4)^2+(2-1)^2}=2.16$	6	-

#### 5. cari kelas tetangga terbanyak

No	Sepal Length	Sepal Width	Petal Length	Petal Width	Species
1	4.8	3	1.4	0.3	setosa
2	5.1	3.8	1.6	0.2	setosa
3	4.6	3.2	1.4	0.2	setosa
4	5.3	3.7	1.5	0.2	setosa
5	5.7	3	4.2	1.2	versicolor
6	5.7	2.9	4.2	1.3	versicolor
7	6.2	2.9	4.3	1.3	versicolor
8	6.7	3	5.2	2.3	virginica
9	6.3	2.5	5	1.9	virginica
10	6.5	3	5.2	2	virginica
X	5	3	4	1	Versi color

#### 6. hasil klasifikasi

Jadi, dari hasil data uji yang di dapat maka hasil klasifikasi yang terbanyak mayoritasnya ialah versi color.

Pengujian metode K-NN dalam klasifikasi data menggunakan data set iris dengan variasi nilai 3, 5, 7, 8, 9.

The screenshot shows the K-Nearest Neighbor software interface. The 'Input data' field contains '2,3,2,1,3,8'. The 'Nilai K' is set to 3. The 'Classification' button is highlighted. The 'Training' table shows 10 data points with their species (setosa, versicolor, virginica) and codes. The 'Testing' table shows 10 data points with their predicted species and codes. The 'Classification Result' section shows the Euclidean distance for each data point and the predicted species for each data point.

**Tabel Hasil Variasi Nilai K Metode K-NN dan Cross Validation**

Dataset Iris	Jumlah Nilai (K) K-NN	Hasil Analisis Metode K-
--------------	-----------------------	--------------------------

	NN
3, 5, 7, 8, 9	85%
3, 5, 7, 8, 9	86%
3, 5, 7, 8, 9	77.3%
3, 5, 7, 8, 9	77%
3, 5, 7, 8, 9	81%
3, 5, 7, 8, 9	73.6%
3, 5, 7, 8, 9	68%
3, 5, 7, 8, 9	73%
3, 5, 7, 8, 9	83%
3, 5, 7, 8, 9	92%

Adapun analisis dari Tabel menyajikan informasi tingkat akurasi specificity algoritma K-Nearest Neighbor. Analisis dilakukan dengan melakukan perhitungan  $\text{Jumlah benar} / \text{jumlah data} * 100\%$ .

#### 4. KESIMPULAN

kesimpulan yang di dapat sebagai berikut:

1. Sistem yang dibangun dapat memudahkan dalam proses K-Nearest Neighbor untuk mengetahui kinerja dan peningkatan akurasi.
2. Nilai K 1 hingga 10 yang di uji persentase dari hasil analisis metode K-NN lebih tinggi dari hasil analisis metode K-NN. Dan dari nilai K yang telah di ujikan nilai K 2 dan nilai K 9 memiliki persentase terbesar sehingga keakuratannya juga semakin tepat.
3. Adapun dari hasil pengujian metode K-Nearest Neighbor dalam klasifikasi data. Adapun pengujian penulis menggunakan variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9. Memiliki persentase akurasi yang sangat baik di dibandingkan hanya K-NN. Hasil pengujian menunjukan metode K-Nearest Neighbor dalam klasifikasi data memiliki akurasi persentase yang baik ketika menggunakan data random. Persentase variasi nilai K K-Nearest Neighbor 3,4,5,6,7,8,9 memiliki persentase 100 % .

#### 5. SARAN

adapun penulis memberikan saran

untuk dapat dikembangkan untuk masa yang akan datang sebagai berikut:

1. Diharapkan kedepannya perlu dilakukan perbandingan terhadap algoritma lain, untuk menguji sejauh mana algoritma K-Nearest Neighbor masih dapat diandalkan untuk memproses dan menemukan pola hubungan asosiasi antar item pada database skala besar.
2. Diharapkan kedepannya adanya pengembangan baik berupa penambahan *fitur-fitur* aplikasi yang sesuai dibutuhkan dan *anti virus* yang berguna mengamankan sistem untuk melindungi data-data yang berada di dalam *database*.
3. Diharapkan sebaiknya ditambahkan fasilitas untuk *backup* data. Jadi, jika terjadi kerusakan pada *server* data tidak akan terhapus.

#### DAFTAR PUSTAKA

- [1]. Boiculese, V.L., Dimitru, G., & Moscalu, M. 2013. Improving Recall of K-Nearest Neighbor Algorithm for Classes of Uneven Size. *The 4th IEEE International Conference on E-Health and Bioengineering-EHB* : 1-4.
- [2]. Havaluddin, "Memahami Penggunaan UML (Unified Modelling Language)", Jurnal Informatika Mulawarman, Vol. 6, No. 1, Februari 2011.
- [3]. Jaafar, H., Mukahar, N., & Ramli, D.A. 2016. Methodology of Nearest Neighbor: Design and Comparison of Biometric Image Database. *IEEE Student Conference on Research and Development (SCORED)* : 1-6.
- [4]. Pan, D., Zhao, Z., Zhang, L., & Tang, C. 2017. Recursive Clustering



K-Nearest  
Neighbors Algorithm and the  
Application in the Classification of  
Power  
Quality Disturbance. IEEE  
Conference on Energy Internet and  
Energy  
System Integration (EI2) : 1-5.

- [5]. Priyanto Hidayatullah, 2012, “*Visual Basic .NET Membuat Aplikasi Database dan Program Kreatif*”. Bandung : Informatika.
  
- [6]. Wardani, Sistria H. (2016). Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode *K-Nearest Neighbor* (KNN) Berbasis Web. Pontianak : Jurusan Sistem Komputer Fakultas MIPA Universitas Tanjungpura. ISSN : 2338-493X