

# Evaluasi Kinerja Machine Learning and Algoritma K-NN menggunakan bahasa R untuk Pengklasifikasian Persebaran Bunga Iris di Sumatera Utara

Shadikin Lubis<sup>1\*</sup>, Angga Putra Juledi<sup>2</sup>, Masrizal<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Universitas Labuhan Batu  
Corresponding Author Email: [sodikinlubis716@gmail.com](mailto:sodikinlubis716@gmail.com)

## Abstrak

Bunga Iris adalah salah satu object yang indah dan selalu ada dan hidup disekitar kita, bunga memberikan keindahan pada alam, rumah sekitar kita, sehingga enak dipandang dan bisa membuat hati nyaman dan senang. Pada penelitian ini, dilakukan ujicoba dalam mengevaluasi, mengklasifikasi berbagai jenis bunga, khususnya adalah bunga iris, khususnya yang ada di medan. bunga iris memiliki banyak tipe dan tersebar diseluruh Indonesia, atau wilayah tropis, pada penelitian ini disiapkan dataset sebanyak 150 dataset bunga iris yang dikelompokkan menjadi tiga bagian yaitu iris setosa, iris versicolor, dan iris virginica. Dan parameter yang digunakan ada empat yaitu Petal width, petal length, sepal width, dan sepal length. Evaluasi dan klasifikasi menggunakan bahasa R yang merupakan software Machine Learning yang powerfull yang digunakan pada riset ini ditambah dengan algoritma KNN yang juga powerfull dalam menyelesaikan suatu kasus prediksi tertentu. hasil riset ini menunjukkan system klasifikasi yang spesifik, dilihat dari indicator warna berdasarkan spesies berbeda sehingga pembaca akan lebih jelas dalam menentukan jenis bunga iris dan pengklasifikasiannya secara detail.

**Kata kunci:** bunga iris, machine learning, spesies, dataset, bahasa R

## 1. Pendahuluan

Bunga iris adalah bunga yang didapati disekitar kita, bunga ini memiliki banyak sekali variant, memiliki banyak warna dan bentuk, sehingga bunga iris sangat diminati sebagai salah satu wujud keindahan alam yang ada disekitar kita. Gambar 1 adalah contoh pencarian bunga iris di Google. Warna yang bisa didapatkan adalah biru, kuning, ungu, orange, dan warna lainnya. Teori ini dikembangkan untuk menganalisis jenis bunga lainnya, seperti mawar, melati, atau jenis bunga lain yang populer dan hidup di Indonesia. metode yang digunakan untuk mengklasifikasikan ini adalah KNN Algorithm yang didukung dengan R-Programming..



Figure 1. Bunga Iris

Banyak metode yang dapat digunakan untuk menyelesaikan problem atau klasifikasi suatu object, pada ilmu telekomunikasi, data-data yang didapatkan misalnya adalah Kekuatan sinyal, atau pelemahan sinyal [13-18]. Pada system klasifikasi ini kita fokus menggunakan KNN Algorithm, walaupun sebenarnya ada banyak metode lain yang bisa digunakan untuk klusterisasi.

## 2. Tinjauan Pustaka

### Algoritma K-NN

Algoritma k-nearest neighbor, juga dikenal sebagai KNN atau k-NN, adalah algoritma pengklasifikasi pembelajaran terawasi non-parametrik, yang menggunakan kedekatan untuk membuat klasifikasi atau prediksi tentang pengelompokan titik data individual. Meskipun dapat digunakan untuk masalah regresi atau klasifikasi, biasanya digunakan sebagai algoritma klasifikasi, bekerja dengan asumsi bahwa titik yang serupa dapat ditemukan berdekatan satu sama lain.

Perlu juga dicatat bahwa algoritma KNN juga merupakan bagian dari keluarga model "pembelajaran malas", yang berarti bahwa ia hanya menyimpan kumpulan data pelatihan versus menjalani tahap pelatihan. Ini juga berarti bahwa semua perhitungan terjadi ketika klasifikasi atau prediksi sedang dibuat. Karena sangat bergantung pada memori untuk menyimpan semua data pelatihannya, metode ini juga disebut sebagai metode pembelajaran berbasis instance atau berbasis memori.

### Equation Algoritma K-NN

Beberapa equations algoritma K-NN antara lain adalah Jarak Euclidean ( $P = 2$ ), Manhattan distance ( $p=1$ ), Minkowski distance, dan Hamming distance.

[1] Jarak Euclidean ( $p=2$ ): Ini adalah ukuran jarak yang paling umum digunakan, dan terbatas pada vektor bernilai nyata. Menggunakan rumus di bawah ini, ini mengukur garis lurus antara titik kueri dan titik lain yang diukur. Rumus yang digunakan pada Jarak Euclidean ( $P=2$ ) adalah seperti pada rumus 1.

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

[2] Manhattan distance ( $p=1$ ), Ini juga merupakan metrik jarak populer lainnya, yang mengukur nilai absolut antara dua titik. Ini juga disebut sebagai jarak taksi atau jarak blok kota karena biasanya divisualisasikan dengan kisi, yang menggambarkan bagaimana seseorang dapat menavigasi dari satu alamat ke alamat lain melalui jalan-jalan kota.

$$\text{Manhattan Distance} = d(x, y) = \left( \sum_{i=1}^m |x_i - y_i| \right) \quad [2]$$

[3] Minkowski distance: Ukuran jarak ini adalah bentuk umum dari metrik jarak Euclidean dan Manhattan. Parameter,  $p$ , dalam rumus di bawah ini, memungkinkan pembuatan metrik jarak lainnya. Jarak Euclidean diwakili oleh rumus ini ketika  $p$  sama dengan dua, dan jarak Manhattan dilambangkan dengan  $p$  sama dengan satu.

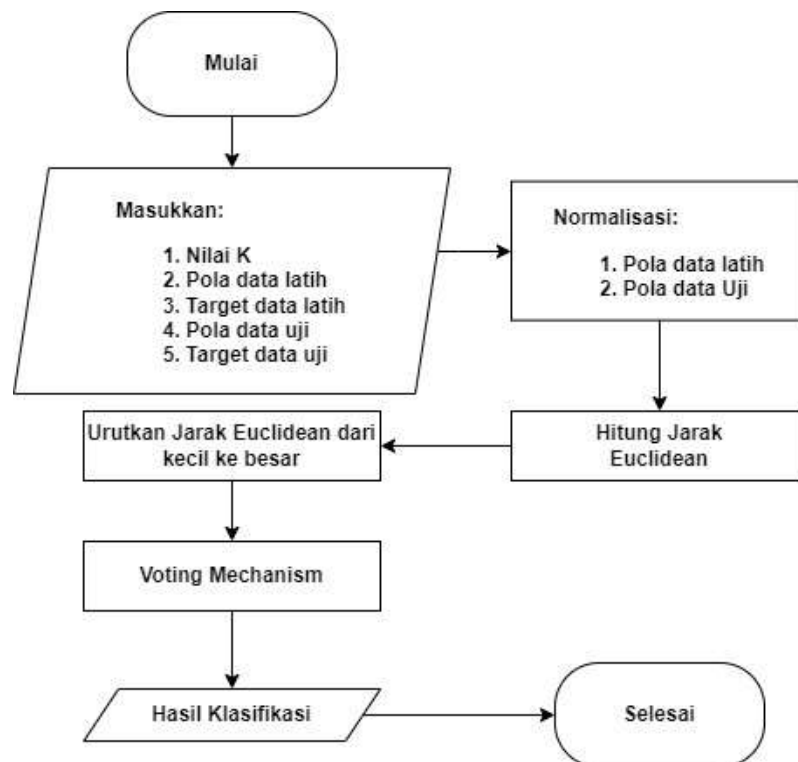
$$\text{Minkowski Distance} = \left( \sum_{i=1}^n |X_i - Y_i| \right)^{1/p} \quad [3]$$

[4] Hamming Distance: Teknik ini biasanya digunakan dengan vektor Boolean atau string, mengidentifikasi titik-titik di mana vektor tidak cocok. Akibatnya, ini juga disebut sebagai metrik tumpang tindih. Ini dapat diwakili dengan rumus berikut:

$$\text{Hamming Distance} = D_H = \left( \sum_{i=1}^k |X_i - Y_i| \right) \quad [4]$$

### 3. Metodologi

Metode Algoritma K-NN dapat dilihat pada Gambar 2 yaitu step-by-step untuk mendapatkan hasil klasifikasi KNN, yang pertama adalah penentuan nilai k, pola data latih, target data latih, pola data uji, target data uji, dan proses normalisasi. Proses normalisasi terdiri dari dua yaitu pola data latih dan pola data uji yang akan menghasilkan jarak Euclidean. Setelah jarak Euclidean diurutkan dari kecil ke besar maka proses dilanjutkan ke Voting untuk menentukan Hasil Klasifikasinya.



**Gambar 2.** Metodologi K-NN

### 4. Hasil dan Pembahasan

Ada hasil dan diskusi secara spesifik dilakukan pemanggilan dari parameter dataset iris untuk memastikan bahwa dataset dapat berjalan dengan baik. Perintah colnames (iris) akan memunculkan parameter spesifik pada bunga iris, seperti ditunjukkan pada gambar 7.

```

> colnames(iris)

 [1] "Sepal.Length" "Sepal.Width"  "Petal.Length"
 "Petal.Width"  "Species"
  
```

```

> iris[,"Petal.Length"]
[1] 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 1.5 1.6 1.4 1.1 1.2 1.5 1.3 1.4 1.7 1.5 1.7 1.5 1.0 1.7 1.9 1.6 1.6 1.5 1.4 1.6
[41] 1.3 1.3 1.3 1.6 1.9 1.4 1.6 1.4 1.5 1.4 4.7 4.5 4.9 4.0 4.6 4.5 4.7 3.3 4.6 3.9 3.5 4.2 4.0 4.7 3.6 4.4 4.5 4.1 4.5 3.9
[81] 3.8 3.7 3.9 5.1 4.5 4.5 4.7 4.4 4.1 4.0 4.4 4.6 4.0 3.3 4.2 4.2 4.2 4.3 3.0 4.1 6.0 5.1 5.9 5.6 5.8 6.6 4.5 6.3 5.8 6.1
[121] 5.7 4.9 6.7 4.9 5.7 6.0 4.8 4.9 5.6 5.8 6.1 6.4 5.6 5.1 5.6 6.1 5.6 5.5 4.8 5.4 5.6 5.1 5.1 5.9 5.7 5.2 5.0 5.2 5.4 5.1
> iris[,"Petal.Width"]
[1] 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 0.2 0.2 0.1 0.1 0.2 0.4 0.4 0.3 0.3 0.3 0.2 0.4 0.2 0.5 0.2 0.2 0.4 0.2 0.2 0.2
[41] 0.3 0.3 0.2 0.6 0.4 0.3 0.2 0.2 0.2 0.2 1.4 1.5 1.5 1.3 1.5 1.3 1.6 1.0 1.3 1.4 1.0 1.5 1.0 1.4 1.3 1.4 1.5 1.0 1.5 1.3
[81] 1.1 1.0 1.2 1.6 1.5 1.6 1.5 1.3 1.3 1.3 1.2 1.4 1.2 1.0 1.3 1.2 1.3 1.3 1.1 1.3 2.5 1.9 2.1 1.8 2.2 2.1 1.7 1.8 1.8 2.5
[121] 2.3 2.0 2.0 1.8 2.1 1.9 1.8 1.9 2.1 1.6 1.9 2.0 2.2 1.5 1.4 2.3 2.4 1.9 1.8 2.1 2.4 2.3 1.9 2.3 2.5 2.3 1.9 2.0 2.3 1.8
> iris[,"Sepal.Length"]
[1] 5.1 4.9 4.7 4.6 5.0 5.4 4.6 5.0 4.4 4.9 5.4 4.8 4.8 4.3 5.8 5.7 5.4 5.1 5.7 5.1 5.4 5.1 4.6 5.1 4.8 5.0 5.0 5.2 5.2 4.7
[41] 5.0 4.5 4.4 5.0 5.1 4.8 5.1 4.6 5.3 5.0 7.0 6.4 6.9 5.5 6.5 5.7 6.3 4.9 6.6 5.2 5.0 5.5 6.0 6.1 5.6 4.7 5.6 5.8 6.2 5.0
[81] 5.5 5.5 5.8 6.0 5.4 6.0 4.7 6.3 5.6 5.5 5.8 6.1 5.8 5.0 5.8 5.7 5.7 6.2 5.1 5.7 6.3 5.8 7.1 6.3 6.5 7.6 4.9 7.3 6.7 7.2
[121] 6.9 5.6 7.7 6.3 6.7 7.2 6.2 6.1 6.4 7.2 7.4 7.9 6.4 6.3 6.1 7.7 6.3 6.4 6.0 6.9 6.7 6.5 5.8 6.8 6.7 6.7 6.3 6.5 6.2 5.9
> iris[,"Sepal.Width"]
[1] 3.5 3.0 3.2 3.1 3.6 3.9 3.4 3.4 2.8 3.1 3.7 3.4 3.0 3.0 4.0 4.4 3.9 3.5 3.8 3.8 3.4 3.7 3.6 3.3 3.4 3.0 3.4 3.5 3.4 3.2
[41] 3.5 2.3 3.2 3.5 3.8 3.0 3.8 3.2 3.7 3.3 3.2 3.2 3.1 2.3 2.8 2.8 3.3 2.4 2.9 2.7 2.0 3.0 2.2 2.5 2.9 3.1 3.0 2.7 2.2 2.9
[81] 2.4 2.4 2.7 2.7 3.0 3.4 3.1 2.3 3.0 2.5 2.6 3.0 2.6 2.3 2.7 3.0 2.9 2.9 2.5 2.8 3.3 2.7 3.0 2.9 3.0 3.0 2.5 2.9 2.5 3.6
[121] 3.2 2.8 2.9 2.7 3.3 3.2 2.8 3.0 2.8 3.0 2.8 3.8 2.8 2.8 2.6 3.0 3.4 3.1 3.0 3.1 3.1 3.1 2.7 3.2 3.3 3.0 2.5 3.0 3.4 3.4

```

Gambar 7. Pemanggilan parameter dataset iris

```

> summary(iris)
  Sepal.Length      Sepal.Width      Petal.Length      Petal.Width      Species
Min.   :4.300      Min.   :2.000      Min.   :1.000      Min.   :0.100      setosa   :50
1st Qu.:5.100      1st Qu.:2.800      1st Qu.:1.600      1st Qu.:0.300      versicolor:50
Median :5.800      Median :3.000      Median :4.350      Median :1.300      virginica :50
Mean   :5.843      Mean   :3.057      Mean   :3.758      Mean   :1.199
3rd Qu.:6.400      3rd Qu.:3.300      3rd Qu.:5.100      3rd Qu.:1.800
Max.   :7.900      Max.   :4.400      Max.   :6.900      Max.   :2.500

```

Gambar 8. Summary of dataset iris

Dari gambar 8, didapatkan kesimpulan species setosa =50, versicolor =50, and virginica = 50. Selanjutnya kita akan mem-plot iris data, dengan perintah berikut:

```

> plot(iris$Petal.Length, iris$Petal.Width, main="North
Sumatera's Iris Data")

```

```

> library(class)
> data(iris)
> set.seed(99)
> rannum<-sample(rep(1:150))
> rannum
 [1] 33 44 22 117 106 128 102 13 20 68 9 88 86 84 132 78 97 7
 [19] 94 87 58 54 147 46 149 112 130 65 14 60 31 129 63 38 103 79
 [37] 32 85 111 71 17 29 45 12 73 99 104 52 145 59 40 35 11 144
 [55] 50 141 15 109 75 96 55 121 3 124 53 56 28 114 136 110 72 39
 [73] 57 51 77 5 95 30 139 61 98 16 101 131 43 90 113 8 76 27
 [91] 81 19 133 142 91 6 82 134 10 150 116 89 49 138 135 18 83 4
 [109] 36 48 115 125 26 93 47 41 100 70 119 92 108 25 34 140 148 64
 [127] 21 146 66 69 127 23 80 107 126 120 37 122 62 123 74 137 67 42
 [145] 2 118 1 143 105 24

```

Gambar 9. Rannum run

```

> iris<- iris[rannum,]
> iris
  Sepal.Length Sepal.Width Petal.Length Petal.Width  Species
33           5.2          4.1          1.5          0.1   setosa
44           5.0          3.5          1.6          0.6   setosa
22           5.1          3.7          1.5          0.4   setosa
117          6.5          3.0          5.5          1.8  virginica
106          7.6          3.0          6.6          2.1  virginica
128          6.1          3.0          4.9          1.8  virginica
102          5.8          2.7          5.1          1.9  virginica
13           4.8          3.0          1.4          0.1   setosa
20           5.1          3.8          1.5          0.3   setosa
68           5.8          2.7          4.1          1.0  versicolor
9           4.4          2.9          1.4          0.2   setosa
88           6.3          2.3          4.4          1.3  versicolor
86           6.0          3.4          4.5          1.6  versicolor
84           6.0          2.7          5.1          1.6  versicolor
132          7.9          3.8          6.4          2.0  virginica
78           6.7          3.0          5.0          1.7  versicolor
97           5.7          2.9          4.2          1.3  versicolor
7           4.6          3.4          1.4          0.3   setosa
94           5.0          2.3          3.3          1.0  versicolor

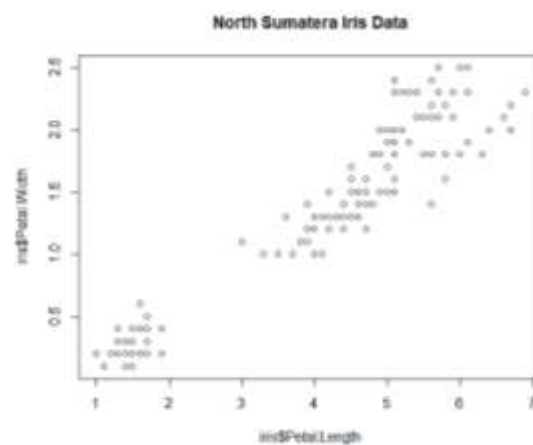
```

**Gambar 10.** Rannum iris run

```

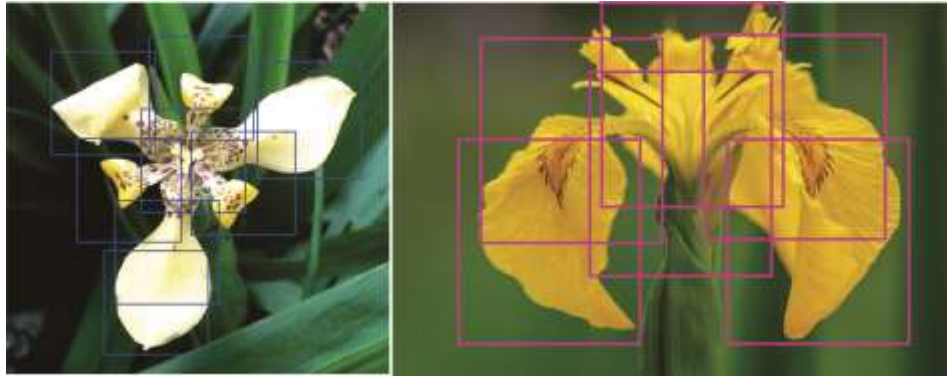
>plot(iris$Petal.Length, iris$Petal.Width, pch=21,
bg=c("red","green3",
"blue") [unclass(iris$Species)],main="North Sumatera's
Iris Data")

```

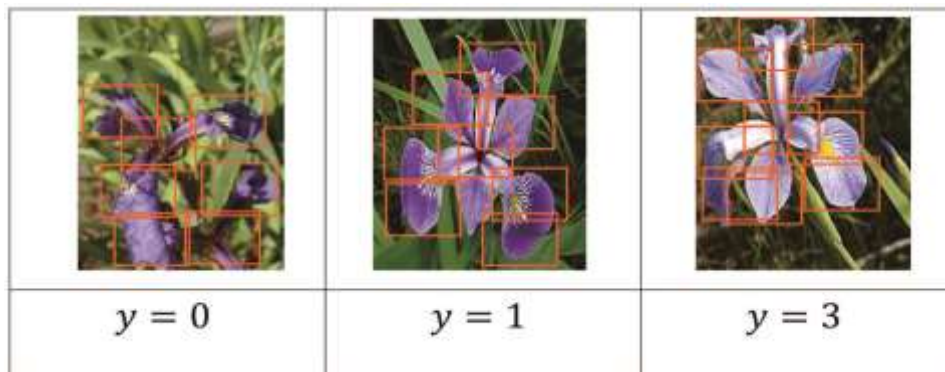


**Gambar 11.** Plot data North Sumatera Iris Data

Gambar 9 dan gambar 10 adalah perintah untuk rannum iris yang menampilkan values and parameter dari data dan library iris. Sedangkan gambar 11 adalah plot of North Sumatera Iris data. Secara spesifik dan lengkap terdapat pada gambar 14. Dengan tampilan yang berbeda ditunjukkan pada gambar 14 (dengan warna) dan gambar 15 dengan tampilan segitiga.



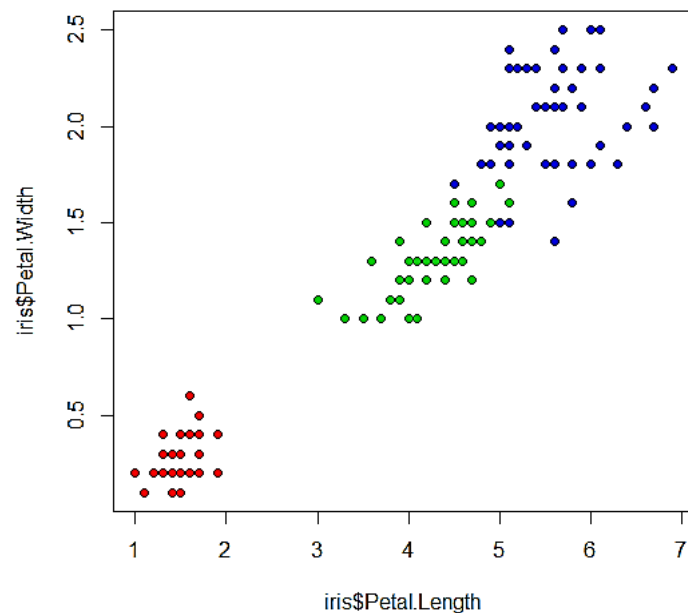
**Gambar 12.** Deteksi bunga iris



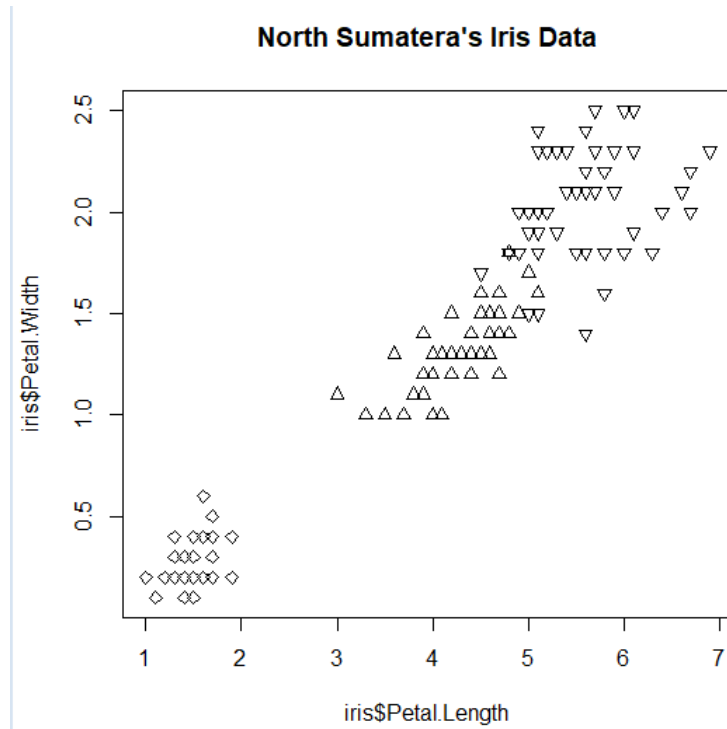
**Gambar 13.** Deteksi dan pengklasifikasian bunga iris

Selanjutnya persebaran iris di Sumatera utara berdasarkan perbandingan iris petal width and Length nya adalah seperti ditunjukkan pada Gambar 14 dan 15.

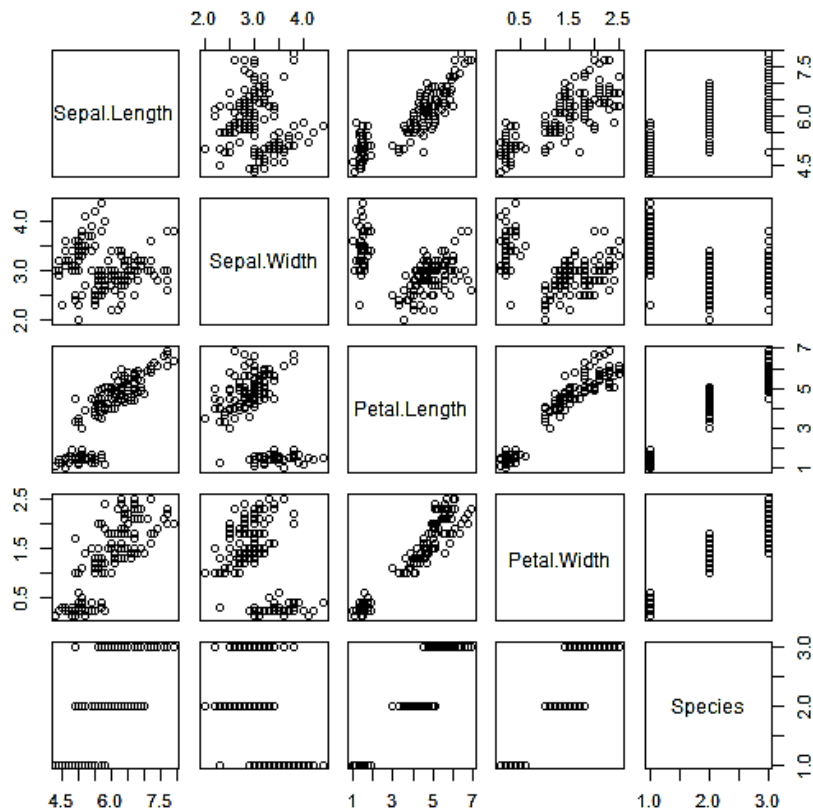
**North Sumatera's Iris Data**



**Gambar 14.** Plot data North Sumatera Iris Data



**Gambar 15.** Plot data North Sumatera Iris Data



**Gambar 14.** Plot data North Sumatera Iris

## 5. Simpulan

Dari dataset bunga iris ini mampu menentukan berapa data yang digunakan sebagai data training dan persentase data uji, perbandingan data training dan data uji adalah 80:20. Dari hasil training bunga iris ini didapatkan hasil perbandingan spesies dan parameter lainnya secara spesifik. dengan Algoritma K-Nearest Neighbor (KNN), proses klasifikasi terhadap objek berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek yaitu spesies bunga iris berhasil diklasifikasikan. Jika ingin running KNN dengan data yang besar, maka jumlah k juga besar, misalnya data >100, nilai k=19. Penelitian ini akan terus dikembangkan untuk spesies bunga lainnya.

## Daftar Referensi

- [1] J. A. Riquelme, R. J. Barrientos, R. Hernández-García and C. A. Navarro, "An exhaustive algorithm based on GPU to process a kNN query," 2020 39th International Conference of the Chilean Computer Science Society (SCCC), 2020, pp. 1-8, doi: 10.1109/SCCC51225.2020.9281231.
- [2] S. Du and J. Li, "Parallel Processing of Improved KNN Text Classification Algorithm Based on Hadoop," 2019 7th International Conference on Information, Communication and Networks (ICICN), 2019, pp. 167-170, doi: 10.1109/ICICN.2019.8834973.
- [3] H. Zhai, "Improving KNN Algorithm Efficiency Based on PCA and KD-tree," 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), 2022, pp. 83-87, doi: 10.1109/MLKE55170.2022.00021.
- [4] A. P. Pawlovsky and H. Matsuhashi, "The use of a novel genetic algorithm in component selection for a kNN method for breast cancer prognosis," 2017 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE), 2017, pp. 1-5, doi: 10.1109/GMEPE-PAHCE.2017.7972084.
- [5] W. Lv, H. Huang, W. Tang and T. Chen, "Research and Application of Intersection Similarity Algorithm Based on KNN Classification Model," 2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA), 2021, pp. 141-144, doi: 10.1109/CAIBDA53561.2021.00037.
- [6] L. Yao and T. Cao, "Research on Talent Classification Based on Improved KNN Algorithm," 2022 International Symposium on Control Engineering and Robotics (ISCER), 2022, pp. 252-256, doi: 10.1109/ISCER55570.2022.00050.
- [7] A. P. Salim, K. A. Laksitowening and I. Asror, "Time Series Prediction on College Graduation Using KNN Algorithm," 2020 8th International Conference on Information and Communication Technology (ICoICT), 2020, pp. 1-4, doi: 10.1109/ICoICT49345.2020.9166238.
- [8] D. Wang, H. Wei and B. Bai, "Teaching Design and Implementation Based on R Language Under the Background of Big Data," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 49-52, doi: 10.1109/ICBAIE52039.2021.9390011.
- [9] D. Wang, Y. Ji, B. Bai and Y. Shi, "Practice of ideological and political education in the course of "R language foundation"," 2020 International Conference on Information Science and Education (ICISE-IE), 2020, pp. 75-78, doi: 10.1109/ICISE51755.2020.00023.
- [10] A. M. Olney and S. D. Fleming, "A Cognitive Load Perspective on the Design of Blocks Languages for Data Science," 2019 IEEE Blocks and Beyond Workshop (B&B), 2019, pp. 95-97, doi: 10.1109/BB48857.2019.8941224.
- [11] S. T. Al-Amin, S. Uday Sampreeth Chebolu and C. Ordonez, "Extending the R Language with a Scalable Matrix Summarization Operator," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 399-405, doi: 10.1109/BigData50022.2020.9378399.
- [12] N. O. Strelkov and V. V. Krutskikh, "Usage of R Language for Universal Calculations and Simulations," 2021 3rd International Youth Conference on Radio Electronics,



- 
- Electrical and Power Engineering (REEPE), 2021, pp. 1-5, doi: 10.1109/REEPE51337.2021.9387989.
- [13] Y. A. Liani et al., "The Broiler Chicken Coop Temperature Monitoring Use Fuzzy Logic and LoRAWAN," 2021 3rd International Conference on Electronics Representation and Algorithm (ICERA), 2021, pp. 161-166, doi: 10.1109/ICERA53111.2021.9538771.
- [14] P. D. P. Adi, A. Kitagawa, D. A. Prasetya and A. B. Setiawan, "A Performance of ES920LR LoRa for the Internet of Things: A Technology Review," 2021 3rd East Indonesia Conference on Computer and Information Technology (EIconCIT), 2021, pp. 1-7, doi: 10.1109/EIconCIT50028.2021.9431912.
- [15] P. D. P. Adi et al., "ZigBee and LoRa performances on RF Propagation on the Snow Hills area," 2021 International Conference on Converging Technology in Electrical and Information Engineering (ICCTEIE), 2021, pp. 36-41, doi: 10.1109/ICCTEIE54047.2021.9650623.
- [16] P. D. P. Adi et al., "Application of IoT-LoRa Technology and Design in irrigation canals to improve the quality of agricultural products in Batu Indonesia," 2021 2nd International Conference On Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS), 2021, pp. 88-94, doi: 10.1109/ICON-SONICS53103.2021.9617175.
- [17] P. D. Prasetyo Adi et al., "ECG-LPWAN based for Real-time monitoring Patient's Heart Beat Status," 2022 International Seminar on Application for Technology of Information and Communication (iSemantic), 2022, pp. 7-14, doi: 10.1109/iSemantic55962.2022.9920379.
- [18] P. D. Prasetyo Adi, Y. Wahyu and A. Kitagawa, "Analyzes of Chirps Spread Spectrum of ES920LR LoRa 920 MHz," 2022 11th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS), 2022, pp. 139-144, doi: 10.1109/EECCIS54468.2022.9902922.
- [19] H. Xin, S. Liu and T. Li, "A multilogarithm linear regression traffic congestion prediction model based on R language New York City's five administrative districts (mainly in urban main roads and highways)," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2020, pp. 284-287, doi: 10.1109/ICBASE51474.2020.00066.
- [20] E. Ramalakshmi and N. Kompala, "Multi-threading image processing in single-core and multi-core CPU using R language," 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2017, pp. 1-5, doi: 10.1109/ICECCT.2017.8117873.