

## Penerapan Algoritma K-Nearest Neighbor Pada Data Kategorik Untuk Klasifikasi Harga Jual Laptop

Citra Puspa Tria<sup>1</sup>, Aang Nuryaman<sup>1\*</sup>, Ahmad Faisol<sup>1</sup>, dan Eri Setiawan<sup>1</sup>

<sup>1</sup>Jurusan Matematika, Fakultas MIPA, Universitas Lampung  
Jl. Prof. Soemantri Brojonegoro No. 1 Bandar Lampung

\*Email korespondensi: [aang.nuryaman@fmipa.unila.ac.id](mailto:aang.nuryaman@fmipa.unila.ac.id)

---

### Abstrak

Salah satu algoritma klasifikasi yang banyak digunakan adalah *k-nearest neighbor* (KNN) di mana algoritma ini mengklasifikasikan objek berdasarkan kelas mayoritas dari *k* objek terdekat di sekitarnya. Pada umumnya, ukuran kedekatan antar objek dapat dihitung dengan menggunakan jarak Euclidean. Akan tetapi, apabila data yang digunakan adalah data kategorik, maka jarak tersebut dirasa kurang tepat untuk digunakan. Salah satu alternatif ukuran kedekatan antar objek yang dapat digunakan untuk data kategorik adalah jarak *weighted simple matching coefficient* (WSMC). Penelitian ini dilakukan untuk membangun model klasifikasi untuk penentuan kelas rentang harga laptop berdasarkan 4 fitur yang ditawarkan, yaitu merek laptop, prosesor, kapasitas RAM, dan kapasitas penyimpanan. Terdapat dua proporsi pembagian data latih dan data uji untuk membangun model klasifikasi, yaitu 80:20 dan 90:10. Berdasarkan hasil analisis, model klasifikasi terbaik diperoleh ketika proporsi data yang digunakan adalah 90:10 dan jumlah *k* sebanyak 17, dengan nilai akurasi yang diperoleh sebesar 86.96%, *recall* sebesar 50%, *presisi* sebesar 66.67%, dan *f1-score* sebesar 57.14%.

**Kata kunci :** Klasifikasi, K-Nearest Neighbor (KNN), Weighted Simple Matching Coefficient (WSMC)

---

### Abstract

One of the widely used classification algorithms is *k-nearest neighbor* (KNN), where it classifies objects based on the majority class of the *k* closest objects around it. In general, the measure of proximity between objects can be calculated using the Euclidean distance. However, if the data being used is categorical data, then the distance is not appropriate to use. One alternative measure of proximity between objects that can be used for categorical data is the *weighted simple matching coefficient* (WSMC) distance. This research was conducted to build a classification model for determining the class of laptop price range based on 4 features offered, namely laptop brand, processor, RAM capacity, and storage capacity. There are two proportions of training data and test data to build a classification model, namely 80:20 and 90:10. Based on the analysis's results, the best classification model is obtained when the proportion of data used is 90:10 and the number of *k* is 17, with accuracy value obtained of 86.96%, recall of 50%, precision of 66.67%, and *f1-score* of 57.14%.

**Keywords:** Classification, K-Nearest Neighbor (KNN), Weighted Simple Matching Coefficient (WSMC)

---

### 1. Pendahuluan

Kecerdasan buatan atau *artificial intelligence* merupakan salah satu bentuk teknologi yang saat ini banyak digunakan dalam berbagai aspek kehidupan. Statistika sebagai salah satu cabang ilmu yang banyak digunakan di kehidupan sehari-hari turut menerapkan konsep kecerdasan buatan ini. Salah satu metode statistika yang menggunakan kecerdasan buatan dalam prosesnya adalah *data mining*. *Data mining* merupakan salah satu tahapan dalam *knowledge discovery in database* (KDD) [1]. Proses *data mining* melibatkan metode statistika, matematika, kecerdasan buatan, dan *machine learning* untuk memperoleh informasi dari dataset yang besar [2].

*Machine learning* akan melatih komputer untuk mempelajari suatu dataset sehingga dapat menghasilkan suatu keputusan atau prediksi tanpa harus menggunakan kode program secara berulang [3]. Salah satu bentuk pembelajaran dalam *machine learning* adalah pembelajaran terawasi (*supervised learning*) di mana pembelajaran ini bekerja pada data masukan yang memiliki label sehingga keluaran yang diharapkan adalah label dari data yang belum memiliki label [4].

Klasifikasi merupakan salah satu bentuk *supervised learning* yang bertujuan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya berdasarkan data lampau dengan menggunakan suatu algoritma [5]. Salah satu algoritma klasifikasi yang banyak digunakan adalah *k-nearest neighbor* (KNN). Algoritma KNN mengklasifikasikan objek berdasarkan kelas mayoritas dari *k* objek terdekat di sekitarnya [6]. Oleh karena itu,

penentuan nilai  $k$  sangat penting karena dapat mempengaruhi hasil klasifikasi [7]. Pada umumnya, untuk melihat  $k$  objek terdekat dapat dilakukan dengan melihat jarak antar objek. Ukuran jarak yang dapat digunakan adalah jarak *Euclidean*, *Manhattan*, *Minkowski*, dan lain sebagainya. Akan tetapi, ukuran jarak tersebut dirasa kurang tepat apabila data yang digunakan adalah data kategorik [8]. Berdasarkan penelitian yang dilakukan oleh Chen L. & Guoi, G. (2005), perhitungan jarak pada data kategorik dapat dilakukan dengan menggunakan *weighted simple matching coefficient* (WSMC). Chen L. & Guoi, G. melakukan klasifikasi pada delapan dataset yang berbeda dengan jenis data kategorik dan menghasilkan akurasi tertinggi pada klasifikasi KNN dengan perhitungan jarak menggunakan WSMC untuk setiap dataset [9]. Penelitian dengan metode yang sama dilakukan oleh Wijaya, dkk tahun 2017 yang menerapkan algoritma *nearest neighbor* untuk enam dataset yang berbeda dan menghasilkan akurasi di atas 72% untuk setiap datasetnya [10].

Klasifikasi sebagai salah satu metode dalam *data mining*, dapat digunakan sebagai alat pendukung pengambilan keputusan, terutama dalam kehidupan sehari-hari. Contohnya adalah pengambilan keputusan dalam membeli barang, salah satunya laptop. Laptop merupakan salah satu bentuk perkembangan teknologi yang saat ini sangat dibutuhkan. Bentuknya yang ringan dan fleksibel sehingga pengguna dapat membawa dan menggunakannya di manapun. Perkembangan tipe dan jenis laptop yang diikuti oleh semakin beragamnya fitur yang ditawarkan menyebabkan orang kesulitan untuk memilih laptop yang sesuai dengan kebutuhan dan anggaran. Pada umumnya, harga laptop akan semakin mahal apabila fitur yang ditawarkan adalah fitur terbaru. Oleh karena itu, peneliti tertarik untuk mengetahui pola klasifikasi harga laptop berdasarkan fitur laptop, di mana fitur laptop berjenis data kategorik, dengan menggunakan algoritma KNN dan ukuran jarak WSMC.

## 2. Metodologi Penelitian

### 2.1 Landasan Teori

#### *Weighted Simple Matching Coefficient* (WSMC)

*Weighted simple matching coefficient* (WSMC) merupakan bentuk pengembangan dari *simple matching coefficient* (SMC) [9]. Perhitungan WSMC hampir sama dengan SMC yang melakukan perbandingan kategori antar fitur. Namun, SMC mengasumsikan seluruh fitur memiliki pengaruh yang sama dalam prediksi kelas hasil klasifikasi, sedangkan pada WSMC, setiap fitur akan diberikan bobot untuk mengurangi kesalahan prediksi kelas akibat fitur yang mungkin tidak terlalu berpengaruh pada hasil klasifikasi [10].

Pembobotan yang digunakan dalam WSMC terbagi menjadi 2, yaitu pembobotan global dan lokal. Pada metode global, bobot dihitung untuk tiap atribut, sedangkan bobot pada metode lokal dihitung untuk setiap kelas pada atribut yang digunakan. Kemudian, pembobotan WSMC ini dapat dihitung berdasarkan informasi entropi dan juga indeks gini, di mana keduanya merupakan ukuran penyimpangan distribusi kategori karena dalam data kategorik tidak memiliki varians.

Berikut merupakan langkah-langkah pembobotan fitur secara global dengan menggunakan informasi entropi:

- Menghitung  $p(m|s_d)$  atau peluang munculnya kategori di kelas tertentu dari keseluruhan munculnya kategori itu di data latih, yang didefinisikan sebagai berikut:

$$p(m|s_d) = \frac{\sum_{(x,y) \in c_m} I(x_d = s_d)}{\sum_{(x,y) \in tr} I(x_d = s_d)} \quad (1)$$

- Menghitung *global entropy* (GE) yang didefinisikan sebagai berikut:

$$GE(s_d) = - \sum_{m=1}^M p(m|s_d) \times \log_2 p(m|s_d) \quad (2)$$

- Menghitung nilai  $p(s_d)$  atau peluang munculnya suatu kategori dari keseluruhan data latih, yang dirumuskan sebagai berikut:

$$p(s_d) = \frac{\sum_{(x,y) \in tr} I(x_d = s_d)}{N} \quad (3)$$

- Menghitung bobot fitur ( $\omega_d^{GE}$ ), yang dirumuskan sebagai berikut:

$$\omega_d^{GE} = e^{-\frac{1}{\log_2 M} \sum_{s_d \in S_d} p(s_d) \times GE(s_d)} \quad (4)$$

Setelah bobot fitur sudah diketahui, langkah selanjutnya adalah mencari nilai SMC dengan membandingkan kategori di setiap fitur antara objek satu dan lainnya. SMC akan bernilai 1 apabila kategori antar objek berbeda

dan bernilai 0 apabila kategori antar objek sama. Selanjutnya, nilai SMC tersebut akan dikalikan dengan bobot fitur yang akan menghasilkan nilai WSMC, seperti dituliskan pada persamaan berikut:

$$WSMC_{global}(\mathbf{x}_i, \mathbf{x}_j, \bar{\omega}) = \sum_{d=1}^D \omega_d \times I(x_{id} \neq x_{jd}) \tag{5}$$

Dimana  $s_d, M, tr, c_m$  dan  $N$  secara berturut-turut menyatakan kategori ke- $d$  dalam sebuah fitur, jumlah kelas yang ada pada dataset, data latih, kelas ke- $m$  dan jumlah data latih.

**2.2 Metode Penelitian**

Data yang digunakan dalam penelitian ini adalah data sekunder mengenai harga jual laptop pada bulan November 2022 yang diperoleh dari laman Tokopedia dengan jumlah data sebanyak 1072. Variabel respon dalam penelitian ini adalah harga jual laptop, sedangkan variabel prediktor yang digunakan adalah merek laptop, prosesor, kapasitas RAM, dan kapasitas penyimpanan. Rincian variable yang digunakan lebih detail disajikan pada Tabel 1. Variabel harga yang berjenis data rasio akan diubah ke bentuk data nominal agar dapat dilakukan klasifikasi. Kategori data pada variabel harga akan ditentukan dengan melihat histogram dengan jumlah kelas sebanyak 2. Penelitian ini dilakukan dengan bantuan aplikasi Jupyter Notebook dengan langkah-langkah penelitian sebagai berikut:

- a) Melakukan pengumpulan data
- b) Melakukan analisis deskriptif melalui histogram untuk variabel harga jual laptop
- c) Melakukan praproses data, meliputi diskritisasi pada variabel harga dan pembersihan data hilang dan data duplikat
- d) Membagi data menjadi 2, yaitu data latih dan data uji
- e) Menentukan nilai  $k$  atau jumlah objek terdekat
- f) Menghitung jarak antar objek pada setiap data uji dengan data latih menggunakan ukuran WSMC yang terdiri atas bobot fitur dan SMC
- g) Menentukan kelas data uji berdasarkan kelas mayoritas  $k$  data latih terdekat, dalam hal ini yang memiliki jarak terdekat dengan data uji
- h) Mengevaluasi model berdasarkan nilai akurasi, presisi, *recall*, dan *f1-score*

**Tabel 1. Variabel Penelitian**

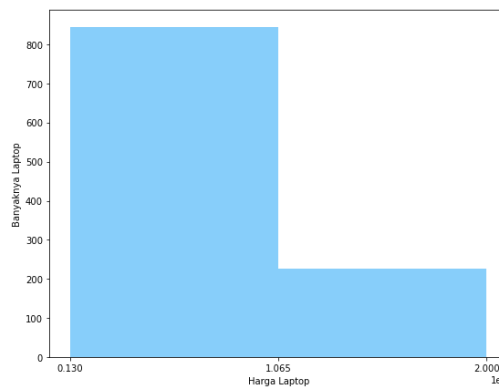
Nama Variabel	Skala Pengukuran	Keterangan
Y = Harga Jual Laptop	Rasio	Dalam juta rupiah
X <sub>1</sub> = Merek Laptop	Nominal	0 : Lenovo
		1 : HP
		2 : ASUS
		3 : Acer
		0 : Intel Pentium Silver
		1 : Intel Celeron
		2 : Intel Core I3 Gen 10
		3 : Intel Core I3 Gen 11
		4 : Intel Core I3 Gen 12
		5 : Intel Core I5 Gen 6
X <sub>2</sub> = Prozessor	Nominal	6 : Intel Core I5 Gen 7
		7 : Intel Core I5 Gen 10
		8 : Intel Core I5 Gen 11
		9 : Intel Core I5 Gen 12
		10 : Intel Core I7 Gen 6
		11 : Intel Core I7 Gen 7
		12 : Intel Core I7 Gen 8
		13 : Intel Core I7 Gen 10
		14 : Intel Core I7 Gen 11
		15 : Intel Core I7 Gen 12
		16 : AMD 3020E
		17 : AMD Athlon Silver
		18 : AMD Athlon Gold
		19 : AMD Ryzen 3 3000 Series
		20 : AMD Ryzen 3 5000 Series

Nama Variabel	Skala Pengukuran	Keterangan
X <sub>3</sub> = Kapasitas RAM	Nominal	21 : AMD Ryzen 3 7000 Series
		22 : AMD Ryzen 5 3000 Series
		23 : AMD Ryzen 5 4000 Series
		24 : AMD Ryzen 5 5000 Series
		25 : AMD Ryzen 5 6000 Series
		26 : AMD Ryzen 7 5000 Series
		27 : AMD Ryzen 7 6000 Series
		28 : AMD Ryzen 9 5000 Series
		0 : 4 GB
		1 : 8 GB
X <sub>4</sub> = Kapasitas Penyimpanan	Nominal	2 : 16 GB
		3 : 32 GB
		0 : 32 GB
		1 : 128 GB
		2 : 160 GB
		3 : 256 GB
		4 : 512 GB

### 3. Hasil dan Pembahasan

#### 3.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk melihat karakteristik data yang digunakan dalam penelitian. Analisis deskriptif dapat disajikan dalam bentuk tabel, diagram, maupun grafik. Dalam penelitian ini, analisis deskriptif akan dilakukan pada variabel harga jual laptop dengan menggunakan histogram. Hal ini bertujuan untuk melihat sebaran data harga jual laptop apabila dibagi ke dalam 2 kelas. Gambar 1 berikut merupakan histogram dari variabel harga jual laptop.



**Gambar 1. Histogram Variabel Harga**

Berdasarkan Gambar 1, terlihat bahwa kelas data terdiri atas kelas harga di rentang Rp1.300.000,- hingga Rp10.649.500,- dengan jumlah data sebanyak 846 data dan kelas harga di rentang Rp10.649.501,- hingga Rp19.999.000,- berjumlah 226 data. Jumlah data di masing-masing kelas tersebut tidak berbeda secara signifikan, sehingga dapat dikatakan bahwa tidak terjadi ketakseimbangan data yang ekstrem antara kedua kelas data tersebut..

### 3.2 Praproses Data

#### 3.2.1 Diskritisasi Data

Proses ini bertujuan untuk mengubah data pada variabel harga ke jenis data nominal, sehingga dapat dijadikan sebagai variabel respon dalam analisis klasifikasi. Data akan dibagi ke dalam 2 kelas, yaitu kelas dengan harga laptop di bawah 11 juta dan kelas dengan harga laptop di atas 11 juta. Tabel 2 berikut menunjukkan perubahan jenis data pada variabel harga ke jenis data nominal.

**Tabel 2. Diskritisasi Data**

Harga	Range Harga
3949000	< 11JT
1895000	< 11JT
5043800	< 11JT
⋮	⋮
7999000	< 11JT
16799000	> 11JT
17499000	> 11JT

Selanjutnya, kedua kelas tersebut akan diberi label 0 untuk kelas harga di bawah 11 juta, dan label 1 untuk kelas harga di atas 11 juta.

**3.2.2 Pembersihan Data Hilang dan Data Duplikat**

Adanya data hilang dan data duplikat dapat mempengaruhi hasil akurasi klasifikasi, sehingga perlu dilakukan penanganan masalah ini, salah satunya dengan melakukan penghapusan baris data yang memiliki permasalahan tersebut. Berdasarkan hasil pemeriksaan, ditemukan bahwa terdapat 172 cell data hilang pada variabel prosesor dan 672 data yang sama atau terduplikat. Oleh karena itu, baris data yang memiliki cell data hilang dan yang merupakan data duplikat akan dihapus. Total data setelah melalui proses pembersihan data adalah 228 data, di mana data ini merupakan data yang akan digunakan dalam proses klasifikasi.

**3.3 Klasifikasi**

**3.3.1 Pembagian Data**

Klasifikasi sebagai salah satu metode *data mining* akan melalui proses pembelajaran untuk memperoleh suatu keputusan atau prediksi. Proses pembelajaran tersebut akan dilakukan oleh data latih agar model terbaik diperoleh dan dapat diimplementasikan pada data uji. Oleh karena itu, perlu dilakukan pembagian data menjadi data latih dan data uji. Dalam penelitian ini, proporsi pembagian data latih dan data uji yang digunakan adalah sebesar 90:10 dan 80:20 sebagaimana yang disajikan pada Tabel 3.

**Tabel 3. Pembagian Data**

Proporsi	Jumlah Data	
	Data Latih	Data Uji
90% Data Latih dan 10% Data Uji	205	23
80% Data Latih dan 20% Data Uji	182	46

**3.3.2 Penentuan Nilai k dan Perhitungan Jarak Menggunakan WSMC**

Akan dipilih nilai *k* sebesar 13 untuk simulasi. Selanjutnya, akan dihitung jarak antara sampel data latih dengan seluruh data uji yang dimulai dengan menghitung bobot setiap fitur. Dengan menggunakan persamaan (1) hingga (5), diperoleh bobot fitur untuk setiap proporsi data latih seperti sebagai berikut:

**Tabel 4. Bobot Fitur**

Proporsi Data Latih	Bobot Fitur			
	Merek Laptop	Prosesor	Kapasitas RAM	Kapasitas Penyimpanan
90%	0,4407	0,6844	0,4994	0,4881
80%	0,4378	0,6677	0,4987	0,4924

Berdasarkan Tabel 4, terlihat bahwa fitur dengan bobot terbesar untuk setiap proporsi data latih adalah fitur prosesor. Hal itu menandakan bahwa fitur prosesor memberikan pengaruh paling besar dalam menentukan range harga laptop. Fitur dengan bobot terbesar kedua hingga terakhir secara berturut-turut adalah kapasitas RAM, penyimpanan, dan merek laptop.

Langkah selanjutnya adalah mencari hasil fungsi indikator  $I(x_{id} \neq x_{jd})$  dengan cara membandingkan kategori untuk setiap variabel prediktor antara data latih dan data uji, dengan hasil 1 apabila  $I(true)$  dan 0 apabila  $I(false)$ . Berikut ditampilkan data latih dan sampel data uji untuk simulasi perhitungan WSMC.

**Tabel 5. Data Latih dan Sampel Data Uji**

Data Uji					
No.	Merek	Prosesor	Kapasitas Laptop	Kapasitas Penyimpanan	Range Harga
1.	0	20	2	4	0
Data Latih					
1.	1	8	1	4	0
2.	2	3	1	4	0
⋮	⋮	⋮	⋮	⋮	⋮
181	1	19	1	3	0
182.	0	26	2	5	1

Selanjutnya, kategori pada setiap fitur antara seluruh data latih dan sampel data uji dibandingkan dan diperoleh hasil fungsi indikator seperti sebagai berikut:

**Tabel 6. Hasil SMC**

No.	Merek	Prosesor	Kapasitas Laptop	Kapasitas Penyimpanan
1.	1	1	1	0
2.	1	1	1	0
⋮	⋮	⋮	⋮	⋮
181	1	1	1	1
182.	0	1	0	1

Setelah bobot fitur dan SMC diperoleh, kedua nilai tersebut akan dikalikan, seperti yang ditunjukkan pada persamaan (5). Hasil perkalian tersebut akan menghasilkan nilai WSMC.

**Tabel 7. Hasil WSMC**

i	Jarak antara Data Uji dan Data Latih ke-i
1	1,6216
2	1,6216
⋮	⋮
182	1.1725

### 3.3.3 Penentuan Kelas

Setelah nilai WSMC antara data latih dan sampel data uji diperoleh, akan dipilih  $k$  data latih terdekat atau  $k$  data latih dengan nilai WSMC terkecil, dalam hal ini telah dipilih nilai  $k$  sebesar 13.

**Tabel 8. Tiga Belas Data dengan Nilai WSMC Terkecil**

i	Jarak (Nilai WSMC) antara Data Uji dan Data Latih ke-i	Label Kelas
80	0,4378	0
74	0,4994	0
4	0,6844	1
23	0,6844	1
36	0,6844	0
41	0,6844	1
51	0,6844	1
64	0,6844	0
87	0,6844	0
130	0,6844	1
177	0,6844	0
63	0,9372	0
140	0,9875	0

Kelas hasil klasifikasi akan ditentukan berdasarkan kelas mayoritas dari ketiga belas data latih terdekat, sehingga berdasarkan Tabel 8, terlihat bahwa kelas mayoritas adalah kelas 0. Oleh karena itu, dapat dikatakan bahwa laptop merek Lenovo dengan fitur prosesor AMD Ryzen 3 5000 Series, kapasitas RAM sebesar 16 GB, dan kapasitas penyimpanan sebesar 512 GB, memiliki harga di bawah 11 juta.

### 3.4 Evaluasi Model dan Pemilihan Model Terbaik

Setelah algoritma KNN diterapkan pada data uji, akan diperoleh prediksi kelas untuk setiap data uji. Namun, hasil prediksi tersebut mungkin saja tidak sesuai dengan kelas sebenarnya. Oleh karena itu, perlu dilakukan perbandingan dan evaluasi ketepatan hasil. Perbandingan antara kelas hasil prediksi dan kelas sebenarnya ini dapat dilihat dengan menggunakan *confusion matrix*.

Tabel 9. *Confusion Matrix* Hasil Klasifikasi Berdasarkan 80% Data Latih dan Jumlah  $k$  Sebanyak 13

Klasifikasi	Kelas Prediksi		
	1	0	
Kelas Sebenarnya	1	5	4
	0	4	33

Berdasarkan *confusion matrix* pada Tabel 9, dapat dilihat performa model klasifikasi yang telah dibangun melalui beberapa perhitungan metrik, seperti tingkat akurasi, presisi, *recall*, dan *f1-score* sebagaimana yang disajikan pada Tabel 10 dan Tabel 11. Berdasarkan Tabel 10, nilai akurasi tertinggi untuk proporsi data latih 80% terjadi ketika nilai  $k$  yang digunakan adalah 13, yaitu sebesar 82.61%. Sedangkan, untuk proporsi data latih 90%, akurasi tertinggi diperoleh ketika nilai  $k$  yang digunakan adalah 17, dengan akurasi sebesar 86,96%. Apabila akurasi antara kedua proporsi data latih dibandingkan, akurasi tertinggi terjadi ketika data latih yang digunakan adalah sebesar 90%. Hal tersebut menandakan bahwa proporsi data latih 90% menghasilkan klasifikasi dengan tingkat ketelitian terbaik dibandingkan dengan klasifikasi berdasarkan data latih 80%. Tabel 11 menunjukkan perbandingan performa klasifikasi dengan data latih 80% dan data latih 90%. Selanjutnya, apabila dilihat dari nilai presisi dan *f1-score*, proporsi data latih sebesar 90% juga menghasilkan klasifikasi dengan nilai performa tertinggi. Oleh karena itu, dapat disimpulkan bahwa klasifikasi dengan data latih sebesar 90% merupakan klasifikasi terbaik dibandingkan dengan klasifikasi dengan data latih sebesar 80%.

Tabel 10. Hasil Akurasi Untuk Setiap Nilai  $k$

$k$	Akurasi	
	Data Latih	Data Latih
	80%	90%
1	0.6522	0.7391
2	0.7391	0.8261
3	0.7174	0.7391
4	0.8043	0.8696
5	0.7609	0.8261
6	0.8043	0.8261
7	0.7826	0.7826
8	0.7826	0.8261
9	0.8043	0.8261
10	0.8261	0.7826
11	0.7826	0.7826
12	0.7391	0.8261
13	0.8261*	0.7826
14	0.7826	0.7826
15	0.8261	0.8261
16	0.7609	0.7826
17	0.7391	0.8696*
18	0.7609	0.8261
19	0.8043	0.7826
20	0.7609	0.7826

Tabel 11. Performa Klasifikasi

Proporsi Data Latih	Akurasi	Recall	Presisi	F1-Score
80%	0,8261	0,5556	0,5556	0,5556
90%	0,8696	0,5000	0,6667	0,5714

### 3. Kesimpulan

Berdasarkan hasil dan pembahasan mengenai klasifikasi harga laptop berdasarkan 4 fitur yang ditawarkan dengan menggunakan algoritma KNN, dapat disimpulkan bahwa model klasifikasi terbaik dalam pengklasifikasian harga laptop adalah model dengan proporsi data latih dan data uji sebesar 90:10 dengan jumlah  $k$  objek terdekat sebanyak 17. Nilai akurasi yang dihasilkan adalah sebesar 86,96%, *recall* sebesar 50%, presisi sebesar 66,67%, dan *f1-score* sebesar 57,14%.

### Daftar Pustaka:

- [1] Fitriani, E.E., & Yustanti, W. 2022. Perbandingan Kinerja Metode Problem Transformation-KNN dan Algorithm Adaptation-KNN pada Klasifikasi Multi-Label Pertanyaan Kotakode. *Journal of Emerging Information Systems and Business Intelligence*. **3**(3): 122–129.
- [2] Turban, E., Aronson, J.E., & Liang, T.P. 2005. *Decision Support System and Intellegent System (Sistem Pendukung Keputusan dan Sistem Cerdas)*. Andi, Yogyakarta.
- [3] Lombu, A.S., Hidayat, S., & Hidayatullah, A.F. 2022. Pemodelan Klasifikasi Gaji Menggunakan Support Vector Machine. *Journal of Computer System and Informatics (JoSYC)*. **3**(4): 363–370.
- [4] Retnoningsih, E., & Pramudita, R. 2020. Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*. **7**(2): 156.
- [5] Indrayanti, Sugianti, D., & Al Karomi, M.A. 2017. Optimasi Parameter K Pada Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Mellitus. *Prosiding SNATIF Ke-4 Tahun 2017*. 153–160.
- [6] Afandie, M.N., Cholissodin, I., & Supianto, A.A. 2014. Implementasi Metode K-Nearest Neighbor Untuk Pendukung Keputusan Pemilihan Menu Makanan Sehat dan Bergizi. *DORO: Repository Jurnal Mahasiswa FILKOM Universitas Brawijaya*. **3**(1).
- [7] Al Karomi, M.A. 2015. Optimasi Parameter K pada Algoritma KNN untuk Klasifikasi Heregistrasi Mahasiswa. *Jurnal IC-Tech*.
- [8] Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. Jon Wiley & Sons, Inc., New Jersey.
- [9] Chen, L., & Guoi, G. 2015. Nearest Neighbor Classification of Categorical Data by Attributes Weighting. *Expert System with Applications*. **42**(6): 3142–3149.
- [10] Wijaya, R., Suciati, N., & Khotimah, W.N. 2017. Implementasi Nearest Neighbor pada Data Kategorik dengan Pembobotan Atribut Menggunakan Weighted Simple Matching Coefficient. *Jurnal Teknik ITS*. **6**(2).