

Implementasi Algoritma *K-Nearest Neighbor* untuk Klasifikasi Cuaca

Dandy*¹, Daniel Udjulawa², Yohannes³

^{1,2,3}Universitas Multi Data Palembang; Jl. Rajawali No.14 Palembang, 0711-376400
Fakultas Ilmu Komputer dan Rekayasa, Informatika, Palembang
e-mail: *¹dandy@mhs.mdp.ac.id, ²daniel@mdp.ac.id, ³yohannesmasterous@mdp.ac.id

Abstrak

Cuaca merupakan kejadian alam singkat berkenaan dengan kondisi atmosfer yang berlangsung di bumi yang ditentukan oleh tekanan, kecepatan angin, suhu, dan fenomena udara. Penelitian ini melakukan klasifikasi terhadap 3 kelas cuaca yaitu cerah (*sunny*), berawan (*cloudy*), dan hujan (*rainy*) menggunakan algoritma *K-Nearest Neighbor* sebagai algoritma klasifikasi cuaca dengan parameter nilai *K* sebesar 3, 5, 7, dan 9. Dataset cuaca yang hendak diteliti berupa 96.453 data yang diambil dari website Kaggle. Dataset dilakukan pembagian atas data latih dan data uji dengan rasio 80:20. Implementasi algoritma *K-Nearest Neighbor* menghasilkan confusion matrix dan classification report dimana pada confusion matrix, jumlah data terprediksi benar terbanyak ada pada nilai *K* = 9 yaitu sebanyak 13.132 data terprediksi benar dengan jumlah data terprediksi benar terbesar ada pada kelas *cloudy* yaitu sebanyak 10.865 data. Adapun nilai akurasi tertinggi baik pada kelas cuaca *cloudy*, *rainy*, maupun *sunny* terdapat pada *K* = 9 yakni sebesar 68,073% dan nilai precision, recall, dan f1-score terbesar terdapat pada kelas *cloudy* pada *K* = 9 yang secara berturut-turut sebesar 72,095%, 89,288%, dan 79,775%.

Kata kunci—Cuaca, Klasifikasi, *K-Nearest Neighbor*

Abstract

Weather is a brief natural event concerning the atmospheric conditions that take place on Earth which are determined by pressure, wind speed, temperature, and air phenomena. This study classifies 3 weather classes, namely sunny, cloudy, and rainy using the *K-Nearest Neighbor* algorithm as a weather classification algorithm with *K* value parameters of 3, 5, 7, and 9. Weather dataset 96.453 data to be examined is data taken from the Kaggle website. The dataset is divided into training data and test data with a ratio of 80:20. The implementation of the *K-Nearest Neighbor* algorithm produces a confusion matrix and classification report where in the confusion matrix, the largest number of correctly predicted data is at the value *K* = 9, namely 13.132 correctly predicted data with the largest number of correctly predicted data in the cloudy class, namely 10.865 data. As for the classification report, the highest accuracy value for both the cloudy, rainy, and sunny weather classes is at *K* = 9, which is 68.073%, and the highest precision, recall, and f1-score values are found in the cloudy class at *K* = 9, respectively contributed 72.095%, 89.288%, and 79.775%.

Keywords—Classification, *K-Nearest Neighbor*, Weather



This is an open-access article under the [CC-BY-CA](https://creativecommons.org/licenses/by-sa/4.0/) license

1. PENDAHULUAN

Cuaca adalah kejadian pada alam berkenaan dengan kondisi atmosfer yang berlangsung di bumi dalam jangka waktu yang pendek. Cuaca terjadi karena reaksi suhu dan kelembaban yang berbeda antara satu tempat dengan tempat lainnya [1]. Cuaca ditentukan oleh beberapa faktor seperti tekanan, kecepatan angin, curah hujan, suhu, dan fenomena pada udara. Dengan berbagai macam faktor yang hendak diteliti, akan sangat sulit untuk membuat prakiraan cuaca yang akurat yang disebabkan oleh kondisi cuaca. Prakiraan cuaca yang tidak akurat mengakibatkan masyarakat kurang dapat mengantisipasi dampak cuaca dalam berbagai sektor, beberapa diantaranya yakni sektor pertanian, perkebunan, dan penerbangan [2].

Klasifikasi objek pada dataset dapat dilakukan dalam proses yang sederhana dan mudah dengan algoritma *K-Nearest Neighbor* karena dapat diimplementasikan dengan baik karena terdapat sedikit variabel pada dataset yang digunakan dalam proses klasifikasi [1]. Selain itu, algoritma *K-Nearest Neighbor* juga merupakan metode klasifikasi objek pemalas yang memiliki performa yang sangat baik. Performa yang sangat baik disebabkan oleh hasil klasifikasi dengan tingkat akurasi tinggi pada algoritma *K-Nearest Neighbor* karena mampu mengelompokkan objek baru berdasarkan data latih sebagai pembelajar dan data uji yang akan digunakan untuk proses klasifikasi [3].

Penelitian terdahulu mengenai cuaca telah dilakukan dengan algoritma selain *K-Nearest Neighbor*. Penelitian mengenai cuaca dilakukan dengan metode *ensemble learning* dengan hasil pengujian berupa akurasi sebesar 81,21% dan MSE (*Means Square Error*) sebesar 18,79% [2]. Penelitian mengenai cuaca juga dilakukan dengan metode LSTM dengan hasil pengujian berupa nilai RMSE dan MAPE terbaik sebesar 1,7444% dan 1,9499% [4]. Terdapat pula penelitian mengenai cuaca dengan algoritma regresi linier berganda dengan hasil pengujian berupa nilai koefisien determinasi R² sebesar 25,5 persen [5].

Adapun terdapat juga penelitian terkait dengan objek yang berbeda yang menggunakan algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* digunakan untuk melakukan klasifikasi data iris bunga dengan hasil pengujian berupa tingkat akurasi sebesar 100% untuk nilai K sebesar 3, 4, 5, 6, 7, 8, dan 9 [3]. Algoritma *K-Nearest Neighbor* juga digunakan untuk melakukan klasifikasi citra belimbing berdasarkan fitur warna dengan hasil pengujian berupa tingkat akurasi terbaik sebesar 93,33% pada nilai K = 7 [6].

Selain itu, algoritma *K-Nearest Neighbor* juga digunakan untuk melakukan klasifikasi jurnal dengan implementasi perbandingan seleksi fitur yang menghasilkan hasil pengujian berupa *precision*, *recall*, *f1-score*, dan *accuracy* berturut-turut sebesar 81,2%, 80,3%, 81,6%, dan 86,6% [7]. Algoritma *K-Nearest Neighbor* juga diimplementasikan untuk melakukan prediksi pengelompokan tingkat risiko penyebaran *covid-19* Jawa Barat dengan hasil pengujian berupa *accuracy*, *precision*, dan *recall* berturut-turut sebesar 85%, 60%, dan 86,6% [8]. Algoritma *K-Nearest Neighbor* juga diimplementasikan untuk melakukan prediksi pasien yang terkena penyakit diabetes pada Puskesmas Manyampa Kabupaten Bulukumba dengan hasil pengujian berupa akurasi sebesar 68,30% [9].

Algoritma *K-Nearest Neighbor* dengan fitur bentuk *Simple Morphological Shape Descriptors* dan fitur warna *Grayscale Histogram* juga digunakan untuk melakukan klasifikasi citra makanan dengan hasil pengujian berupa tingkat akurasi untuk metode *Grayscale Histogram*, metode SMSD, serta gabungan antara metode *Grayscale Histogram* dan SMSD berturut-turut sebesar 60%, 54,8%, dan 77,8% [10]. Algoritma *K-Nearest Neighbor* juga digunakan untuk melakukan klasifikasi tingkat kelulusan pada siswa dengan hasil pengujian berupa rata-rata akurasi sebesar 96,49% [11]. Algoritma *K-Nearest Neighbor* dengan ekstraksi fitur warna RGB juga digunakan untuk melakukan klasifikasi bunga dengan hasil pengujian berupa akurasi sebesar 50%-60% pada percobaan pertama dan 90-100% pada percobaan kedua pada K = 5 [12].

Algoritma *K-Nearest Neighbor* juga digunakan untuk melakukan prediksi penjualan

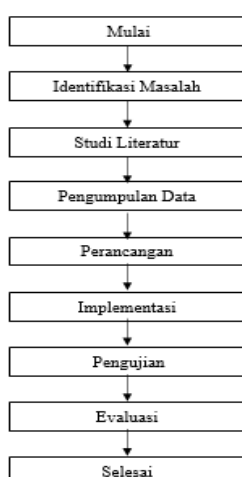
sepatu yang menghasilkan hasil pengujian berupa tingkat keberhasilan pada $K = 7$ dan $K = 9$ berturut-turut sebesar 74% dan 49% [13]. Algoritma *K-Nearest Neighbor* juga digunakan untuk melakukan klasifikasi kelulusan siswa dengan hasil pengujian berupa tingkat akurasi sebesar 85,28% pada $K = 7$ [14]. Algoritma *K-Nearest Neighbor* juga digunakan untuk melakukan sentimen analisis pembelajaran daring dengan hasil pengujian berupa *accuracy*, *precision*, *recall*, dan *f-measure* tertinggi pada $K = 10$ berturut-turut sebesar 84,65%, 87%, 86%, dan 87% [15].

Uraian ini menunjukkan bahwa penelitian terdahulu mengenai prediksi cuaca telah dilakukan dengan berbagai macam metode yang berbeda-beda. Juga, penelitian terdahulu dengan algoritma *K-Nearest Neighbor* telah dilakukan pada beberapa objek penelitian. Maka dari itu, penelitian ini menggunakan algoritma *K-Nearest Neighbor* sebagai suatu pendekatan baru dalam melakukan klasifikasi cuaca dengan hasil klasifikasi cuaca berupa tiga kategori yakni cerah, berawan, dan hujan yang didasarkan pada faktor temperatur, kelembaban udara, tekanan udara, dan kecepatan angin. Melalui pembagian data latih dan data uji menggunakan algoritma *K-Nearest Neighbor*, penelitian ini hendak memperoleh hasil klasifikasi cuaca berdasarkan keempat faktor tersebut. Berdasarkan permasalahan yang telah dipaparkan, terciptalah rumusan masalah dalam penelitian ini yakni bagaimana cara melakukan klasifikasi cuaca dengan menggunakan metode *K-Nearest Neighbor*.

Adapun ruang lingkup penelitian ini adalah berupa hasil klasifikasi cuaca adalah kelas cuaca berawan (*cloudy*), hujan (*rainy*), dan cerah (*sunny*) pada variabel *summary* (rangkuman cuaca) sebagai data uji dan variabel suhu, kelembaban, kecepatan angin dan tekanan udara sebagai data latih. Dataset berjumlah 96.453 data dengan judul *Weather Dataset* [16]. Dataset dapat diakses pada dalam format *csv* (*comma separated value*) dan dibagi menjadi 80% data latih dan 20% data uji. Klasifikasi dataset dilakukan dengan algoritma *K-Nearest Neighbor* dengan nilai k yang digunakan adalah 3, 5, 7, dan 9 [17].

2. METODE PENELITIAN

Klasifikasi cuaca dilakukan melalui beberapa tahapan. Tahapan-tahapan dalam melakukan klasifikasi cuaca yakni identifikasi masalah, studi literatur, pengumpulan data, perancangan, implementasi, pengujian, dan evaluasi. Tahapan-tahapan klasifikasi cuaca terdapat pada Gambar 1.



Gambar 1 Metode Penelitian

2.1 Identifikasi Masalah

Pada tahapan ini, dilakukan identifikasi masalah sesuai dengan fenomena yang hendak diteliti. Tujuan identifikasi masalah adalah untuk memperjelas cakupan penelitian yang hendak dikaji sehingga sesuai dengan fenomena penelitian. Fenomena yang hendak diteliti yaitu klasifikasi cuaca menggunakan algoritma *K-Nearest Neighbor* (KNN).

2.2 Studi Literatur

Pada tahapan ini, dilakukan pembelajaran mengenai fenomena yang hendak diteliti yaitu klasifikasi cuaca menggunakan algoritma *K-Nearest Neighbor* (KNN) dengan cara mencari berbagai referensi penelitian. Referensi penelitian berasal dari jurnal atau buku yang relevan dengan fenomena penelitian. Tujuan dilakukannya studi literatur adalah untuk memperkuat landasan penelitian melalui informasi dengan sumber-sumber literatur yang valid.

2.3 Pengumpulan Data

Pada tahapan ini, dilakukan pengumpulan data berupa dataset mengenai cuaca-cuaca terdahulu melalui *website Kaggle*. Dataset cuaca yang digunakan meliputi 3 kelas cuaca yakni berawan (*cloudy*), hujan (*rainy*), dan cerah (*sunny*) dengan jumlah data pada dataset sebanyak 96.453 baris data. Persebaran dataset untuk tiap-tiap kelas cuaca tertera pada Tabel 1.

Tabel 1 Dataset Cuaca

No	Kelas	Jumlah data
1.	<i>Cloudy</i>	60989
2.	<i>Rainy</i>	17289
3.	<i>Sunny</i>	18175

2.4 Perancangan

Pada tahapan ini, dilakukan perancangan penelitian mengenai proses klasifikasi cuaca. Perancangan dilakukan berdasarkan 3 kelas cuaca dengan memilih 4 variabel pada kolom yakni suhu (*temperature*), kelembaban udara (*humidity*), kecepatan angin (*wind speed*), dan tekanan udara (*pressure*) sebagai data latih dan 1 variabel pada kolom yaitu rangkuman cuaca (*summary*) sebagai data uji. Perbandingan data latih dan data uji yang digunakan adalah 80:20.

2.5 Implementasi

Pada tahapan ini, dilakukan implementasi terhadap hasil perancangan sistem. Implementasi yang dimaksud yakni implementasi algoritma *K-Nearest Neighbor* sebagai algoritma klasifikasi guna mengetahui hasil klasifikasi cuaca. Implementasi dilakukan dengan menggunakan data latih dan data uji yang telah dilakukan pembagian data dengan perbandingan data latih dan data uji 80:20 dari keseluruhan dataset cuaca berjumlah 96.453 data.

2.6 Pengujian

Pada tahapan ini, dilakukan pengujian terhadap sistem yang telah dirancang. Tahapan pengujian bertujuan untuk mengetahui perubahan pada hasil penelitian pada klasifikasi cuaca apabila nilai uji k diubah. Parameter uji yang digunakan yaitu nilai k sebesar 3, 5, 7, dan 9.

2.7 Evaluasi Hasil Klasifikasi

Pada tahapan ini, dilakukan kajian analisis hasil pengujian pada klasifikasi cuaca yang didapat yaitu *Confusion Matrix* yang menyatakan hasil evaluasi pembelajaran supervised learning, yang kemudian menghasilkan nilai TP (*true positive*), TN (*true negative*), FP (*false positive*), dan FN (*false negative*) yang masing-masing menyatakan jumlah data yang terklasifikasi oleh sistem. Dari nilai TP, TN, FP, dan FN, didapatkanlah ukuran kinerja algoritma *K-Nearest Neighbor* berupa *classification report* yang berisikan nilai *accuracy*, *precision*, *recall*,

dan *f1-score* yang secara berturut-turut dirumuskan pada Persamaan (1), (2), (3), dan (4) dalam bentuk persentase. Setelah dilakukan penghitungan terhadap nilai *accuracy*, *precision*, *recall*, dan *f1-score*, ditentukanlah kelas cuaca baru berdasarkan nilai *accuracy*, *precision*, *recall*, dan *f1-score* terbesar.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \tag{3}$$

$$F1-Score = 2 \times \frac{precision \times recall}{precision + recall} \times 100\% \tag{4}$$

Keterangan rumus :

TP = *True Positive*, yakni jumlah data positif yang terklasifikasi benar oleh sistem.

TN = *True Negative*, yakni jumlah data negatif yang terklasifikasi benar oleh sistem.

FP = *False Positive*, yakni jumlah data positif yang terklasifikasi salah oleh sistem.

FN = *False Negative*, yakni jumlah data negatif yang terklasifikasi salah oleh sistem.

3. HASIL DAN PEMBAHASAN

Pengujian melakukan klasifikasi pada 3 kelas cuaca, yakni berawan (*cloudy*), hujan (*rainy*), dan cerah (*sunny*) pada variabel *summary* (rangkuman cuaca) dengan parameter suhu, kelembaban, kecepatan udara, dan tekanan udara dengan total dataset cuaca sebanyak 96.453 data yang kemudian dilakukan pembagian data latih dan data uji dengan perbandingan 80:20. Setelah itu, digunakanlah nilai ketetanggaan K sebesar 3, 5, 7, dan 9 untuk melakukan pembelajaran algoritma *K-Nearest Neighbor* (KNN) pada kelas *cloudy*, *rainy*, dan *sunny* melalui sampel dataset cuaca yang terdapat pada Tabel 2. Sampel data terdiri dari 10 data dengan kelas cuaca yang telah ditentukan dan 1 data baru yang hendak ditentukan kelasnya yang terdapat pada baris terakhir Tabel 2.

Tabel 2 Sampel Dataset Cuaca

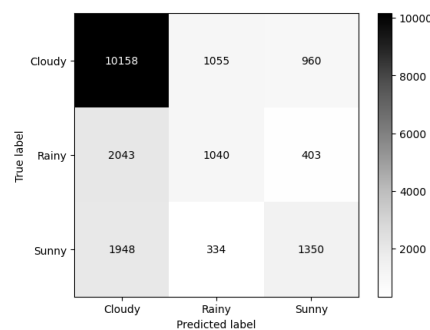
No	Temperature (C)	Humidity	Wind Speed	Pressure (millibars)	Summary
1.	17,18	0,93	12,54	1.013,05	<i>Rainy</i>
2.	12,88	0,93	3,61	1.013,92	<i>Sunny</i>
3.	17,19	0,9	6,41	1.021,54	<i>Cloudy</i>
4.	17,22	0,9	10,92	1.021,77	<i>Rainy</i>
5.	17,22	0,9	11,08	1.021,87	<i>Rainy</i>
6.	13,87	0,93	4,75	1.014,66	<i>Sunny</i>
7.	16,07	0,88	2,78	1.015,25	<i>Sunny</i>
8.	24,58	0,48	10,09	1.015,16	<i>Cloudy</i>
9.	22,04	0,56	8,98	1.015,66	<i>Cloudy</i>
10.	21,52	0,6	10,53	1.015,95	<i>Cloudy</i>
11.	20,44	0,61	5,88	1.016,16	<i>Cloudy</i>

Setelah itu dilakukan perhitungan pada urutan jarak antara data baru dan 10 sampel data lainnya pada data cuaca. Kemudian dilakukan pengurutan jarak Euclidean dari jarak terkecil hingga jarak terbesar. Pengurutan jarak *euclidean* menyebabkan perubahan pada urutan sampel data seperti tertera pada Tabel 3.

Tabel 3 Urutan Jarak *Euclidean*

Data ke-	Temperature (C)	Humidity	Wind Speed	Pressure (millibars)	Urutan Jarak dengan Data Baru	Summary
9	22,04	0,56	8,98	1.015,66	3,52	Cloudy
10	21,52	0,6	10,53	1.015,95	4,78	Cloudy
7	16,07	0,88	2,78	1.015,25	5,44	Sunny
8	24,58	0,48	10,09	1.015,16	5,99	Cloudy
3	17,19	0,9	6,41	1.021,54	6,31	Cloudy
6	13,87	0,93	4,75	1.014,66	6,84	Sunny
1	17,18	0,93	12,54	1.013,05	8,05	Rainy
2	12,88	0,93	3,61	1.013,92	8,21	Sunny
4	17,22	0,9	10,92	1.021,77	8,21	Rainy
5	17,22	0,9	11,08	1.021,87	8,37	Rainy

Kemudian dilakukan klasifikasi berdasarkan k data terdekat yaitu 3, 5, 7, dan 9 dengan hasil data baru untuk nilai k sebesar 3, 5, 7, dan 9 masuk ke dalam kelas cuaca *cloudy* dikarenakan data baru tersebut memiliki jumlah kelas mayoritas *cloudy* untuk tiap-tiap nilai k tetangga terdekat. Setelah hasil klasifikasi didapatkan, dilakukan evaluasi untuk mengetahui jumlah data terprediksi benar dan salah sebagai suatu kelas tertentu pada *confusion matrix* dan nilai *accuracy*, *precision*, *recall*, dan *f1-score* pada *classification report* untuk mengetahui ukuran kinerja algoritma *K-Nearest Neighbor* pada masing-masing kelas. Hasil pengujian algoritma *K-Nearest Neighbor* berupa *confusion matrix* dapat dilihat pada Gambar 2, Gambar 3, Gambar 4, dan Gambar 5, dimulai pada *confusion matrix* untuk nilai $K = 3$ yang dapat dilihat pada Gambar 2.

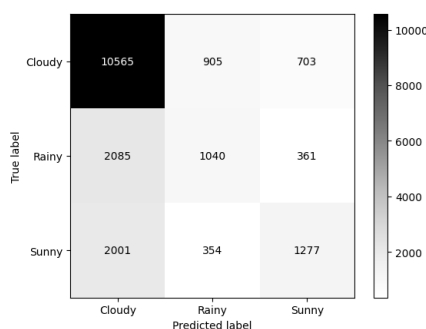
Gambar 2 *Confusion Matrix* untuk Nilai $K = 3$

Dari *confusion matrix* pada Gambar 2 tampak *true label* yang menyatakan nilai aktual untuk masing-masing kelas cuaca dinyatakan sebagai baris dan *predicted label* yang menyatakan nilai prediksi untuk masing-masing kelas cuaca dinyatakan sebagai kolom. Prediksi yang benar dinyatakan pada diagonal (10.158, 1.040, 1.350). Adapun prediksi untuk tiap-tiap kelas cuaca yakni sebagai berikut.

Untuk kelas *cloudy*, baris pertama *confusion matrix* menunjukkan kelas *cloudy* aktual. Dari 12.173 data kelas *cloudy* aktual, 10.158 data kelas *cloudy* terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 1.055 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 960 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom pertama *confusion matrix* menunjukkan kelas *cloudy* prediksi. Dari total 14.149 data kelas *cloudy* prediksi, 10.158 data terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 2.043 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 1.948 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy*.

Kemudian untuk kelas *rainy*, baris kedua *confusion matrix* menunjukkan kelas *rainy* aktual. Dari 3.486 data kelas *rainy* aktual, 1.040 data kelas *rainy* terprediksi secara tepat sebagai kelas *rainy*, sedangkan 2.043 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 403 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom kedua *confusion matrix* menunjukkan kelas *rainy* prediksi. Dari total 2.429 data kelas *rainy* prediksi, 1.040 data terprediksi secara tepat sebagai kelas *rainy*, sedangkan 1.055 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 334 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*.

Dan yang terakhir untuk kelas *sunny*, baris ketiga *confusion matrix* menunjukkan kelas *sunny* aktual. Dari 3.632 data kelas *sunny* aktual, 1.350 data kelas *sunny* terprediksi secara tepat sebagai kelas *sunny*, sedangkan 1.948 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 334 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*. Adapun kolom ketiga *confusion matrix* menunjukkan kelas *sunny* prediksi. Dari total 2.713 data kelas *sunny* prediksi, 1.350 data terprediksi secara tepat sebagai kelas *sunny*, sedangkan 960 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny* dan 403 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun *confusion matrix* untuk nilai $K = 5$ dapat dilihat pada Gambar 3.



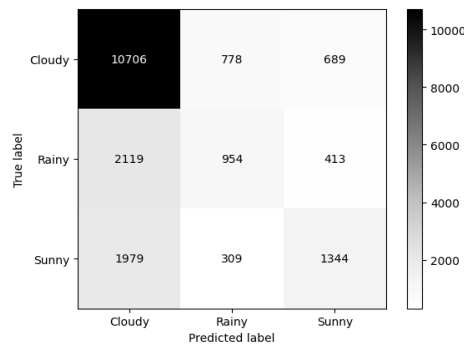
Gambar 3 *Confusion Matrix* untuk Nilai $K = 5$

Dari *confusion matrix* pada Gambar 3 tampak *true label* yang menyatakan nilai aktual untuk masing-masing kelas cuaca dinyatakan sebagai baris dan *predicted label* yang menyatakan nilai prediksi untuk masing-masing kelas cuaca dinyatakan sebagai kolom. Prediksi yang benar dinyatakan pada diagonal (10.565, 1.040, 1.277). Adapun prediksi untuk tiap-tiap kelas cuaca yakni sebagai berikut.

Untuk kelas *cloudy*, baris pertama *confusion matrix* menunjukkan kelas *cloudy* aktual. Dari 12.173 data kelas *cloudy* aktual, 10.565 data kelas *cloudy* terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 905 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 703 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom pertama *confusion matrix* menunjukkan kelas *cloudy* prediksi. Dari total 14.651 data kelas *cloudy* prediksi, 10.565 data terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 2.085 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 2.001 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy*.

Kemudian untuk kelas *rainy*, baris kedua *confusion matrix* menunjukkan kelas *rainy* aktual. Dari 3.486 data kelas *rainy* aktual, 1.040 data kelas *rainy* terprediksi secara tepat sebagai kelas *rainy*, sedangkan 2.085 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 361 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom kedua *confusion matrix* menunjukkan kelas *rainy* prediksi. Dari total 2.299 data kelas *rainy* prediksi, 1.040 data terprediksi secara tepat sebagai kelas *rainy*, sedangkan 905 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 354 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*.

Dan yang terakhir untuk kelas *sunny*, baris ketiga *confusion matrix* menunjukkan kelas *sunny* aktual. Dari 3.632 data kelas *sunny* aktual, 1.277 data kelas *sunny* terprediksi secara tepat sebagai kelas *sunny*, sedangkan 2.001 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 354 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*. Adapun kolom ketiga *confusion matrix* menunjukkan kelas *sunny* prediksi. Dari total 2.341 data kelas *sunny* prediksi, 1.277 data terprediksi secara tepat sebagai kelas *sunny*, sedangkan 703 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny* dan 361 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun *confusion matrix* untuk nilai $K = 7$ dapat dilihat pada Gambar 4.



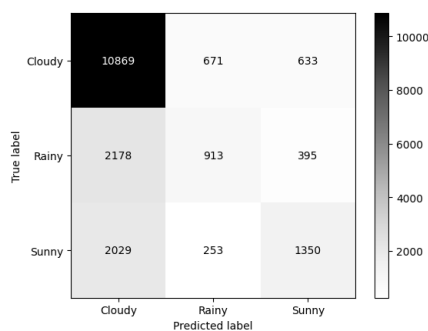
Gambar 4 *Confusion Matrix* untuk Nilai $K = 7$

Dari *confusion matrix* pada Gambar 4 tampak *true label* yang menyatakan nilai aktual untuk masing-masing kelas cuaca dinyatakan sebagai baris dan *predicted label* yang menyatakan nilai prediksi untuk masing-masing kelas cuaca dinyatakan sebagai kolom. Prediksi yang benar dinyatakan pada diagonal (10.706, 954, 1.344). Adapun prediksi untuk tiap-tiap kelas cuaca yakni sebagai berikut.

Untuk kelas *cloudy*, baris pertama *confusion matrix* menunjukkan kelas *cloudy* aktual. Dari 12.173 data kelas *cloudy* aktual, 10.706 data kelas *cloudy* terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 778 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 689 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom pertama *confusion matrix* menunjukkan kelas *cloudy* prediksi. Dari total 14.804 data kelas *cloudy* prediksi, 10.706 data terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 2.119 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 1.979 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy*.

Kemudian untuk kelas *rainy*, baris kedua *confusion matrix* menunjukkan kelas *rainy* aktual. Dari 3.486 data kelas *rainy* aktual, 954 data kelas *rainy* terprediksi secara tepat sebagai kelas *rainy*, sedangkan 2.119 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 413 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom kedua *confusion matrix* menunjukkan kelas *rainy* prediksi. Dari total 2.041 data kelas *rainy* prediksi, 954 data terprediksi secara tepat sebagai kelas *rainy*, sedangkan 778 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 309 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*.

Dan yang terakhir untuk kelas *sunny*, baris ketiga *confusion matrix* menunjukkan kelas *sunny* aktual. Dari 3.632 data kelas *sunny* aktual, 1.344 data kelas *sunny* terprediksi secara tepat sebagai kelas *sunny*, sedangkan 1.979 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 309 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*. Adapun kolom ketiga *confusion matrix* menunjukkan kelas *sunny* prediksi. Dari total 2.341 data kelas *sunny* prediksi, 1.344 data terprediksi secara tepat sebagai kelas *sunny*, sedangkan 689 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny* dan 413 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun *confusion matrix* untuk nilai $K = 9$ dapat dilihat pada Gambar 5.



Gambar 5 *Confusion Matrix* untuk Nilai $K = 9$

Dari *confusion matrix* pada Gambar 5 tampak *true label* yang menyatakan nilai aktual untuk masing-masing kelas cuaca dinyatakan sebagai baris dan *predicted label* yang menyatakan nilai prediksi untuk masing-masing kelas cuaca dinyatakan sebagai kolom. Prediksi yang benar dinyatakan pada diagonal (10.869, 913, 1.350). Adapun prediksi untuk tiap-tiap kelas cuaca yakni sebagai berikut.

Untuk kelas *cloudy*, baris pertama *confusion matrix* menunjukkan kelas *cloudy* aktual. Dari 12.173 data kelas *cloudy* aktual, 10.869 data kelas *cloudy* terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 671 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 633 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom pertama *confusion matrix* menunjukkan kelas *cloudy* prediksi. Dari total 15.076 data kelas *cloudy* prediksi, 10.869 data terprediksi secara tepat sebagai kelas *cloudy*, sedangkan 2.178 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 2.029 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy*.

Kemudian untuk kelas *rainy*, baris kedua *confusion matrix* menunjukkan kelas *rainy* aktual. Dari 3.486 data kelas *rainy* aktual, 913 data kelas *rainy* terprediksi secara tepat sebagai kelas *rainy*, sedangkan 2.178 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 395 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*. Adapun kolom kedua *confusion matrix* menunjukkan kelas *rainy* prediksi. Dari total 1.837 data kelas *rainy* prediksi, 913 data terprediksi secara tepat sebagai kelas *rainy*, sedangkan 671 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *rainy* dan 253 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*.

Dan yang terakhir untuk kelas *sunny*, baris ketiga *confusion matrix* menunjukkan kelas *sunny* aktual. Dari 3.632 data kelas *sunny* aktual, 1.350 data kelas *sunny* terprediksi secara tepat sebagai kelas *sunny*, sedangkan 2.029 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *cloudy* dan 253 data kelas *sunny* terprediksi secara tidak tepat sebagai kelas *rainy*. Adapun kolom ketiga *confusion matrix* menunjukkan kelas *sunny* prediksi. Dari total 2.378 data kelas *sunny* prediksi, 1.350 data terprediksi secara tepat sebagai kelas *sunny*, sedangkan 633 data kelas *cloudy* terprediksi secara tidak tepat sebagai kelas *sunny* dan 395 data kelas *rainy* terprediksi secara tidak tepat sebagai kelas *sunny*.

Adapun hasil pengujian algoritma *K-Nearest Neighbor* berupa *classification report* menghasilkan nilai akurasi yang meningkat seiring dengan bertambahnya nilai K , yakni nilai akurasi sebesar 65,046% untuk nilai $K = 3$, 66,777% untuk nilai $K = 5$, 67,410% untuk nilai $K =$

7, dan 68,073% untuk nilai $K = 9$. *Classification report* secara keseluruhan dapat dilihat pada Tabel 4.

Tabel 4 *Classification Report*

No.	Nilai K	Label Kelas	Precision	Recall	F1-Score	Accuracy
1.	3	Cloudy	71,793%	83,447%	77,183%	65,046%
		Rainy	42,816%	29,834%	35,165%	
		Sunny	49,760%	37,170%	42,553%	
2.	5	Cloudy	72,111%	86,790%	78,773%	66,777%
		Rainy	45,237%	29,834%	35,955%	
		Sunny	54,549%	35,160%	42,759%	
3.	7	Cloudy	72,318%	87,949%	79,371%	67,410%
		Rainy	46,742%	27,367%	34,521%	
		Sunny	54,947%	37,004%	44,225%	
4.	9	Cloudy	72,095%	89,288%	79,775%	68,073%
		Rainy	49,701%	26,190%	34,304%	
		Sunny	56,770%	37,170%	44,925%	

Berdasarkan hasil pengujian algoritma *K-Nearest Neighbor* berupa *confusion matrix* pada Gambar 2, Gambar 3, Gambar 4, dan Gambar 5 untuk masing-masing nilai $K = 3$, $K = 5$, $K = 7$, dan $K = 9$ dan *classification report* pada Tabel 4 yang memuat nilai *precision*, *recall*, *f1-score*, dan *accuracy*, semakin besar parameter uji nilai K , semakin banyak pula jumlah data yang terprediksi benar pada *confusion matrix*. Selain itu, pada *confusion matrix*, jumlah data terprediksi benar terbanyak ada pada kelas *cloudy*, yakni 10.158 data terprediksi benar untuk nilai $K = 3$, 10.565 data terprediksi benar untuk nilai $K = 5$, 10.706 data terprediksi benar untuk nilai $K = 7$, dan 10.869 data terprediksi benar untuk nilai $K = 9$. Adapun *classification report* menunjukkan kesamaan tingkat akurasi untuk masing-masing kelas *cloudy*, *rainy*, maupun *sunny* baik pada $K = 3$, $K = 5$, $K = 7$, dan $K = 9$. Selain itu, kelas *rainy* memiliki nilai *precision*, *recall*, dan *f1-score* terendah, sedangkan kelas *cloudy* memiliki nilai *precision*, *recall*, dan *f1-score* tertinggi dan kelas *sunny* memiliki nilai *precision*, *recall*, dan *f1-score* yang mendekati nilai *precision*, *recall*, dan *f1-score* pada kelas *rainy* baik pada $K = 3$, $K = 5$, $K = 7$, dan $K = 9$. Hal ini disebabkan karena persebaran data uji pada kelas *cloudy* jauh lebih tinggi dibandingkan persebaran data uji baik pada kelas *rainy* maupun kelas *sunny*.

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, diperoleh hasil penelitian dari implementasi algoritma *K-Nearest Neighbor* pada klasifikasi cuaca dan diperoleh kesimpulan sebagai berikut :

1. Algoritma *K-Nearest Neighbor* untuk klasifikasi cuaca memiliki tingkat akurasi tertinggi pada nilai $K = 9$ sebesar 68,073%.
2. Nilai *precision* tertinggi untuk kelas *cloudy* sebesar 72,318% pada $K = 7$ serta nilai *precision* tertinggi untuk kelas *rainy* dan *sunny* berturut-turut sebesar 49,701% dan 56,770% pada $K = 9$, nilai *recall* tertinggi untuk kelas *cloudy* dan *sunny* berturut-turut sebesar 89,288% dan 37,170% pada $K = 9$ serta nilai *recall* tertinggi untuk kelas *rainy* sebesar 29,834% pada $K = 3$, dan nilai *f1-score* tertinggi untuk kelas *cloudy* dan *sunny* berturut-turut sebesar 79,775% dan 44,925% pada $K = 9$ dan nilai *f1-score* tertinggi untuk kelas *rainy* sebesar 35,955% pada $K = 5$.

5. SARAN

Penelitian ini memiliki saran pengembangan berupa penggunaan algoritma lain seperti *deep learning* pada klasifikasi cuaca pada penelitian selanjutnya. Penggunaan *deep learning* bertujuan untuk meningkatkan nilai akurasi pada hasil pengujian. Peningkatan nilai akurasi disebabkan karena nilai akurasi pada penelitian ini rendah.

DAFTAR PUSTAKA

- [1] J. Brawijaya, F. and K. M. Suryaningrum, "Aplikasi Pendeteksi dan Analisa Cuaca Menggunakan Metode K-Nearest Neighbor Berbasis Android," *Jurnal Sarjana Teknik Informatika*, pp. 1-10, 2020.
- [2] A. M. Siregar, T. S. Faisal, A. Fauzi and I. Kadori, "Klasifikasi untuk Prediksi Cuaca Menggunakan Ensemble Learning," *Jurnal Pengkajian dan Penerapan Teknik Informatika*, pp. 138-147, 2020.
- [3] P. Putra, A. M. H. Pardede and S. Syahputra, "Analisis Metode K-Nearest Neighbor (KNN) dalam Klasifikasi Data Iris Bunga," *Jurnal Teknik Informatika Kaputama*, pp. 297-305, 2022.
- [4] T. Lattifia, P. W. Buana and N. K. D. Rusjyanthi, "Model Prediksi Cuaca Menggunakan Metode LSTM," *Jurnal Ilmiah Teknologi dan Komputer*, pp. 994-1000, 2022.
- [5] A. Luthfiarta, A. Febriyanto, H. Lestiawan and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *Journal of Information System*, pp. 10-17, 2020.
- [6] D. I. Muhammad, E. and N. Falih, "Penggunaan K-Nearest Neighbor (KNN) untuk Mengklasifikasikan Citra Belimbing Berdasarkan Fitur Warna," *Jurnal Informatik*, pp. 9-16, 2021.
- [7] F. Istighfarizky, N. A. S. ER, I. M. Widiartha, L. G. Astuti, I. G. N. A. C. Putra and I. K. G. Suhartana, "Klasifikasi Jurnal menggunakan Metode KNN dengan Mengimplementasikan Perbandingan Seleksi Fitur," *Jurnal Elektronik Ilmu Komputer Udayana*, pp. 167-176, 2022.
- [8] D. M. U. Atmaja, A. R. Hakim, D. Haryadi and N. Suwaryo, "Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Pengelompokkan Tingkat Risiko Penyebaran Covid-19 Jawa Barat," *Prosiding Seminar Nasional Teknologi Energi dan Mineral*, pp. 1218-1226, 2021.
- [9] M. S. Mustafa and I. W. Simpen, "Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Pasien Terkena Penyakit Diabetes Pada Puskesmas Manyampa Kabupaten Bulukumba," *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, pp. 1-10, 2019.
- [10] M. R. Setiawan, Y. A. Sari and P. P. Adikara, "Klasifikasi Citra Makanan Menggunakan K-Nearest Neighbor dengan Fitur Bentuk Simple Morphological Shape Descriptors dan Fitur Warna Grayscale Histogram," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, pp. 2726-2731, 2019.
- [11] E. Purwaningsih and E. Nurelasari, "Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa," *Syntax: Jurnal Informatika*, pp. 46-56, 2021.
- [12] L. Farokhah, "Implementasi K-Nearest Neighbor Untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, pp. 1129-1136, 2020.

-
- [13] B. Hardiyanto and F. Rozi, "Prediksi Penjualan Sepatu Menggunakan Metode K-Nearest Neighbor," *Journal of Education and Information Communication Technology*, pp. 13-18, 2020.
- [14] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) Algorithm with Euclidean and Manhattan in Classification of Student Graduation," *Journal of Engineering and Applied Technology*, pp. 86-91, 2021.
- [15] A. R. Isnain, J. Supriyanto and M. P. Kharisma, "Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning," *Indonesian Journal of Computing and Cybernetics Systems*, pp. 121-130, 2021.
- [16] M. J, "Weather Dataset," 2017. [Online]. Available: <https://www.kaggle.com/datasets/muthuj7/weather-dataset>. [Accessed 23 May 2023].
- [17] A. Amalia, A. Zaidiah and I. N. Isnainiyah, "Prediksi Kualitas Udara Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, pp. 496-507, 2022.