

KLASIFIKASI CURAH HUJAN HARIAN MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Muhammad Alviriza Ramadhan, Fetty Tri Anggraeny, Chrystia Aji Putra

Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur

Jalan Raya Rungkut Madya No. 1 Surabaya, Indonesia

20081010143@student.upnjatim.ac.id

ABSTRAK

Keberagaman iklim dan cuaca yang diakibatkan oleh letak geografis Indonesia yang dilalui oleh garis khatulistiwa menyebabkan perubahan cuaca yang tidak menentu yang berpengaruh kepada kehidupan manusia. Oleh karena itu, diperlukan sebuah solusi berupa pengelompokan atau klasifikasi curah hujan harian. Penelitian ini menggunakan metode K-Nearest Neighbor untuk melakukan proses klasifikasi. Yang dimana K-Nearest Neighbor adalah algoritma untuk menemukan kelompok K dalam data training yang paling dekat dengan data uji dan melakukan pencarian label pada kelas tertentu dalam lingkup data. Dengan menggunakan metrik jarak euclidean, metode ini dapat menghasilkan akurasi paling akurat sebesar 86%.

Kata kunci : *Cuaca, Curah Hujan Harian, Klasifikasi, K-Nearest Neighbor*

1. PENDAHULUAN

Indonesia merupakan sebuah negara yang memiliki letak geografis yang strategis. Dikarenakan posisinya yang terletak di daerah tropis, diantara Benua Asia dan Australia, diantara Samudera Pasifik dan Samudera Hindia, serta dilalui garis khatulistiwa, terdiri dari pulau dan kepulauan yang membujur dari barat ke timur, dikelilingi oleh luasnya lautan, menyebabkan wilayah Indonesia memiliki keragaman cuaca dan iklim [1]. Hal tersebut juga menyebabkan Indonesia memiliki iklim hutan hujan tropis dengan suhu udara yang hampir sama dan curah hujan yang tinggi yang terbagi rata sepanjang tahun.

Hujan merupakan bagian dari perubahan cuaca yang dipengaruhi oleh perubahan-perubahan yang terjadi pada beberapa variabel cuaca di atmosfer. Intensitas hujan yang turun, atau biasa disebut curah hujan, adalah salah satu informasi penting tentang hujan.

Curah hujan itu sendiri berpengaruh terhadap aktivitas manusia, seperti kegiatan bercocok tanam ataupun kegiatan sehari-harinya. Tak hanya itu intensitas curah hujan yang tinggi bisa menimbulkan bencana seperti banjir dan longsor [2].

Oleh sebab itu, perlu dilakukan klasifikasi curah hujan untuk bisa memprediksi seberapa besar tingkat curah hujan pada suatu waktu yang akan datang terutama pada prediksi curah hujan setiap harinya (curah hujan harian) agar dapat mempersiapkan rencana untuk menanggulangi bencana tersebut.

Salah satu metode paling sederhana dan tertua untuk klasifikasi adalah klasifikasi k tetangga terdekat (kNN). Metode ini mengklasifikasikan pengamatan yang tidak diketahui ke kelas mayoritas di antara pengamatan tetangga terdekatnya, yang diukur dengan metrik jarak, dalam data pelatihan [3]. Untuk menentukan kedekatan jarak antara tetangga biasanya

dilakukan perhitungan menggunakan Euclidean Distance [4].

Metode K-Nearest Neighbor (kNN) merupakan salah satu algoritma Machine Learning (ML) yang dianggap sebagai suatu metode yang sederhana untuk diterapkan dalam analisis data dengan dimensi variabel yang banyak [5].

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka penelitian ini dilakukan untuk mengetahui tingkat akurasi pada penerapan metode K-Nearest Neighbor untuk melakukan klasifikasi pada data curah hujan harian yang berasal dari pengamatan Stasiun Meteorologi Kelas III Sultan Muhammad Salahuddin - Bima selama lima tahun terakhir.

2. TINJAUAN PUSTAKA

Berikut ini adalah pembahasan mengenai literatur yang relevan dengan topik penelitian. Hal ini bertujuan untuk mendapatkan gambaran menyeluruh tentang penelitian yang dilakukan.

2.1. Cuaca

Cuaca adalah keadaan atmosfer di suatu tempat pada waktu dan jangka waktu tertentu yang ditunjukkan oleh berbagai unsur atmosfer, seperti suhu udara, tekanan udara, kelembapan udara, angin, awan, curah hujan, dan sinar matahari. Cuaca adalah dinamis, dan dapat berubah dalam waktu singkat.

2.2. Curah Hujan

Curah hujan juga dapat didefinisikan sebagai jumlah air hujan yang terkumpul di daerah datar tanpa menguap, meresap, atau mengalir. Pada umumnya, curah hujan diukur menggunakan satuan per milimeter. Hal tersebut menunjukkan bahwa per satu milimeter curah hujan berarti dalam luasan satu meter persegi pada tempat yang datar tertampung air setinggi satu milimeter tempat yang datar tertampung air setinggi satu milimeter atau tertampung air setinggi 1 liter [6].

2.3. Prediksi

Prediksi merupakan pernyataan tentang nilai-nilai masa depan dari suatu variabel yang menarik minat. Prediksi ini dapat dibuat berdasarkan model statistik atau metode lain [7]. Berbagai teknik seperti jaringan saraf tiruan, model deret waktu, regresi linier, dan beberapa teknik lainnya dapat digunakan untuk membuat prediksi ini untuk berbagai tujuan, seperti perencanaan, pengambilan keputusan, dan pengendalian risiko.

2.4. Pra-Proses Data

Pra-proses data merupakan proses persiapan data untuk analisis yang melibatkan berbagai teknik, seperti pembersihan data, pendeteksian anomali, dan transformasi data. Pra-proses data penting untuk memastikan bahwa data yang digunakan dalam analisis akurat, konsisten, dan bebas dari kesalahan [8]. Pra-proses data merupakan langkah penting dalam proses *data mining*. Melakukan pemrosesan data pada data yang belum disaring secara hati-hati untuk suatu permasalahan dapat menyebabkan hasil yang menyimpang atau tidak sesuai tujuan [9].

2.5. Klasifikasi

Klasifikasi merupakan suatu proses yang bertujuan untuk mendapatkan suatu model yang dapat membedakan kelas data yang selanjutnya digunakan untuk kepentingan tertentu. Proses klasifikasi ini membantu menemukan label kelas yang tidak diketahui sebelumnya, yang kemudian label yang telah diperoleh dari proses klasifikasi ini dapat digunakan untuk melakukan prediksi [10].

2.6. K-Nearest Neighbor

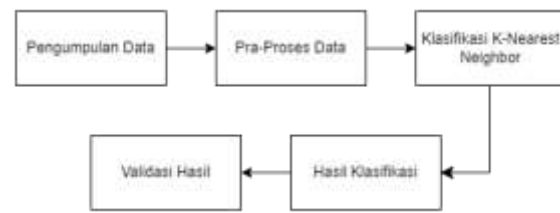
K-Nearest Neighbor adalah algoritma untuk menemukan kelompok K dalam data training yang paling dekat dengan data uji dan melakukan pencarian label pada kelas tertentu dalam lingkup data [11]. Metode K-Nearest Neighbor merupakan metode yang sering digunakan pada proses klasifikasi. Metode ini tergolong suatu metode supervised karena pada proses pengambilan keputusannya, KNN menemukan pola baru dalam data, dengan menghubungkan pola data yang sudah ada [12].

3. METODE PENELITIAN

Metodologi penelitian adalah proses penelitian yang terdiri dari langkah-langkah dan penerapan prinsip metode ilmiah. Setiap penelitian yang dilakukan memerlukan kerangka kerja agar hasil dan tujuan penelitian dapat dicapai secara optimal. Pada penelitian ini, yang menjadi tujuan capaian hasil penelitian adalah klasifikasi curah hujan harian menggunakan metode K-Nearest Neighbor.

3.1. Tahapan Penelitian

Berikut ini adalah flowchart mengenai tahapan penelitian yang dilakukan pada penelitian ini.



Gambar 1. Diagram tahapan penelitian

Berdasarkan diagram diatas, penjelasan dari masing-masing tahapan penelitian adalah sebagai berikut.

a. Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah data yang diambil pada website pusat database data online BMKG (dataonline.bmkg.go.id) yang meliputi data cuaca harian yang merupakan hasil pengamatan Stasiun Meteorologi Kelas III Sultan Muhammad Salahuddin - Bima selama lima tahun terakhir dengan jumlah data sebesar 1826 records, serta jumlah atribut sebesar 10 variabel. Berikut ini adalah informasi dari dataset yang digunakan.

```

RangeIndex: 1826 entries, 0 to 1825
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Tanggal                1826 non-null   int64
1   temp_min               1783 non-null   float64
2   temp_max               1806 non-null   float64
3   temp_avg               1781 non-null   float64
4   kelembapan_avg        1781 non-null   float64
5   curah_hujan            1710 non-null   float64
6   penyinaran_matahari    1803 non-null   float64
7   kec_angin              1809 non-null   float64
8   arah_angin             1809 non-null   float64
9   kec_angin_avg          1809 non-null   float64
dtypes: float64(9), int64(1)
  
```

Gambar 2. Informasi Dataset

Berdasarkan keterangan seperti yang tertera pada gambar, dapat disimpulkan bahwasanya terdapat beberapa *missing values* pada dataset. Oleh karena itu, perlu dilakukannya tahapan pra-proses data untuk memperbaikinya.

b. Pra-Proses Data

Tahapan pra-proses data ini terdiri atas seleksi data, pembersihan data, serta pelabelan data menurut kelasnya. Pada tahap seleksi data, data yang digunakan pada penelitian ini selanjutnya hanya sebanyak 9 variabel saja yaitu temp_min, temp_max, temp_avg, kelembapan_avg, curah_hujan, penyinaran_matahari, kec_angin, arah_angin, dan kec_angin_avg.

Lalu untuk proses pembersihan data, data yang dibersihkan disini merupakan data yang mengandung *missing values*. Jadi, apabila dalam sebuah data tertentu terdapat satu variabel didalamnya yang mengandung missing values, maka satu baris data tersebut akan dihilangkan. Setelah melalui proses ini, data yang semula terdiri atas 1826 record telah dipangkas menjadi 1510 record.

Pada proses pelabelan data, data yang telah melalui proses seleksi dan pembersihan lalu dikelompokkan sesuai dengan intensitas curah hujannya. Berikut ini adalah kategori tingkat curah hujan beserta labelnya menurut BMKG.

Tabel 1. Kategori curah hujan

Tingkat Curah Hujan	Keterangan
0 mm/hari	Berawan
0,5 – 20 mm/hari	Hujan Ringan
20 – 50 mm/hari	Hujan Sedang
50 – 100 mm/hari	Hujan Lebat
100 – 150 mm/hari	Hujan Sangat Lebat
>150 mm/hari	Hujan Ekstrem

Setelah dikelompokkan menurut kategori curah hujan masing-masing, berikut ini adalah persebaran curah hujan menurut kategorinya.

Tabel 2. Persebaran curah hujan

Kelas	Jumlah Data
Berawan	993
Hujan Ringan	436
Hujan Sedang	71
Hujan Lebat	10

Sebagaimana yang terlihat pada tabel diatas, karena pada dataset tidak mengandung data dengan kategori “Hujan Sangat Lebat”, dan “Hujan Ekstrem”, maka untuk pelabelan kelas pada dataset ini hanya menggunakan empat kelas saja untuk melakukan penelitian ini.

c. Klasifikasi K-Nearest Neighbor

Tahapan ini merupakan tahapan inti dari penelitian ini. Berikut ini merupakan algoritma dari metode K-Nearest Neighbor.

Tabel 3. Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor
<p>MULAI Impor dataset Pilih Nilai k Hitung jarak antar setiap titik data menggunakan rumus euclidean distance Urutkan titik data berdasarkan jarak yang telah dihitung Pilih jarak minimum Tentukan titik data pada kelas yang paling sering END</p>

Untuk menghitung jarak antar titik pada metode K-Nearest Neighbor seperti yang tertera pada tabel diatas menggunakan rumus jarak euclidean. Rumusnya adalah sebagai berikut.

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

d. Hasil Klasifikasi

Hasil klasifikasi yang dimaksud disini merupakan output yang dihasilkan setelah dataset melalui proses klasifikasi dengan metode K-Nearest Neighbor.

e. Validasi Hasil

Pada proses ini dilakukannya pencocokan sekaligus perhitungan data asli pada dataset terhadap data prediksi yang dihasilkan melalui metode K-Nearest Neighbor.

3.2. Skenario Uji Coba

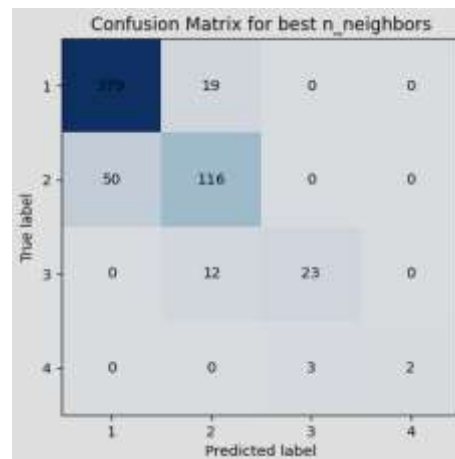
Untuk skenario uji coba pada penelitian ini dilakukan dengan menambahkan variasi dari nilai k tetangga terdekat serta perbandingan data training dan testing. Untuk nilai k yang akan diujicobakan pada penelitian ini adalah 3, 5, dan 7. Sedangkan untuk perbandingan data training dan data testing yang digunakan adalah 60:40, 70:30, serta 80:20.

4. HASIL DAN PEMBAHASAN

Pada bagian ini terdapat penjelasan terkait dengan hasil dari keseluruhan skenario percobaan yang dilakukan dalam penelitian ini. Berikut ini adalah Penjelasannya secara satu persatu.

4.1. Uji Coba Nilai K 3 dan Perbandingan 60:40

Setelah dilakukannya uji coba dengan variasi nilai k senilai 3 serta perbandingan data training dan data testing sebesar 60:40, maka didapatkanlah hasil berupa akurasi sebesar 86,09%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

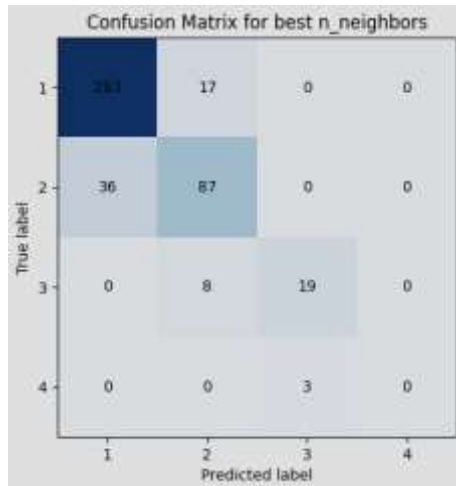


Gambar 3. Confusion Matrix Uji Coba Pertama

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 379 prediksi benar dan hanya 19 kesalahan pada kelas 2. Kelas 2 memiliki 116 prediksi benar dengan 50 kesalahan pada kelas 1. Kelas 3 memiliki 23 prediksi benar dengan 12 kesalahan pada kelas 2, sedangkan kelas 4 memiliki 2 prediksi benar dengan 3 kesalahan pada kelas 3.

4.2. Uji Coba Nilai K 3 dan Perbandingan 70:30

Setelah dilakukannya uji coba dengan variasi nilai k senilai 3 serta perbandingan data training dan data testing sebesar 70:30, maka didapatkanlah hasil berupa akurasi sebesar 85,87%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

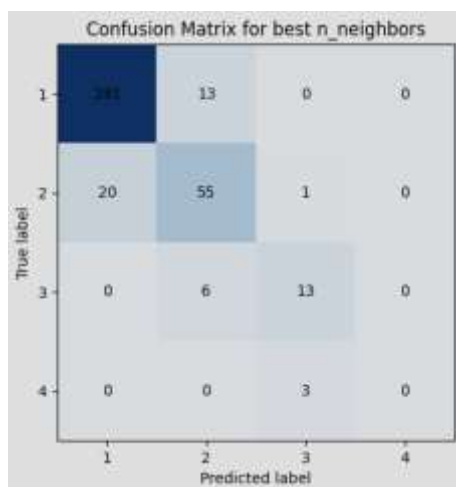


Gambar 4. Confusion Matrix Uji Coba Kedua

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 283 prediksi benar dan hanya 17 kesalahan pada kelas 2. Kelas 2 memiliki 87 prediksi benar dengan 36 kesalahan pada kelas 1. Kelas 3 memiliki 19 prediksi benar dengan 8 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 3 kesalahan pada kelas 3.

4.3. Uji Coba Nilai K 3 dan Perbandingan 80:20

Setelah dilakukannya uji coba dengan variasi nilai k senilai 3 serta perbandingan data training dan data testing sebesar 80:20, maka didapatkanlah hasil berupa akurasi sebesar 85,76%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

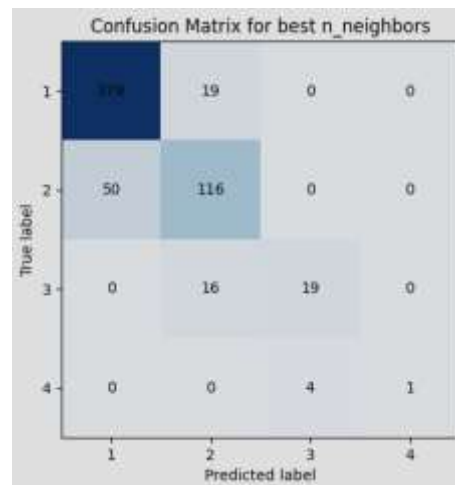


Gambar 5. Confusion Matrix Uji Coba Ketiga

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 191 prediksi benar dan hanya 13 kesalahan pada kelas 2. Kelas 2 memiliki 55 prediksi benar dengan 20 kesalahan pada kelas 1. Kelas 3 memiliki 13 prediksi benar dengan 6 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 3 kesalahan pada kelas 3.

4.4. Uji Coba Nilai K 5 dan Perbandingan 60:40

Setelah dilakukannya uji coba dengan variasi nilai k senilai 5 serta perbandingan data training dan data testing sebesar 60:40, maka didapatkanlah hasil berupa akurasi sebesar 85,26%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

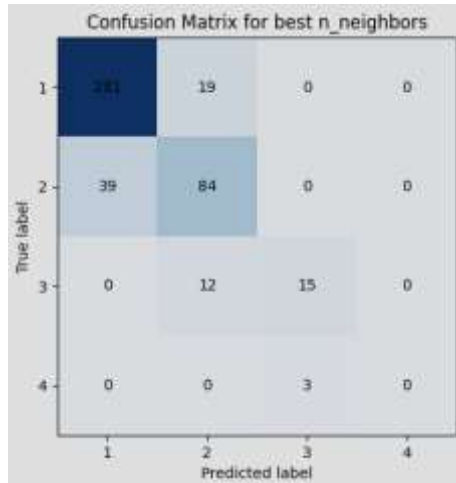


Gambar 6. Confusion Matrix Uji Coba Keempat

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 379 prediksi benar dan hanya 19 kesalahan pada kelas 2. Kelas 2 memiliki 116 prediksi benar dengan 50 kesalahan pada kelas 1. Kelas 3 memiliki 19 prediksi benar dengan 16 kesalahan pada kelas 2, sedangkan kelas 4 memiliki 1 prediksi benar dengan 4 kesalahan pada kelas 3.

4.5. Uji Coba Nilai K 5 dan Perbandingan 70:30

Setelah dilakukannya uji coba dengan variasi nilai k senilai 5 serta perbandingan data training dan data testing sebesar 70:30, maka didapatkanlah hasil berupa akurasi sebesar 83,89%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

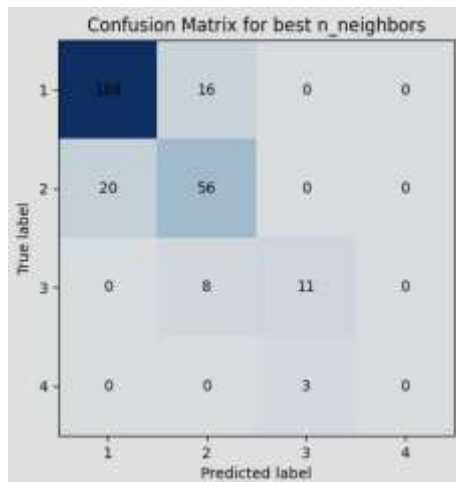


Gambar 7. Confusion Matrix Uji Coba Kelima

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 281 prediksi benar dan hanya 19 kesalahan pada kelas 2. Kelas 2 memiliki 84 prediksi benar dengan 39 kesalahan pada kelas 1. Kelas 3 memiliki 15 prediksi benar dengan 12 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 3 kesalahan pada kelas 3.

4.6. Uji Coba Nilai K 5 dan Perbandingan 80:20

Setelah dilakukannya uji coba dengan variasi nilai k senilai 5 serta perbandingan data training dan data testing sebesar 80:20, maka didapatkanlah hasil berupa akurasi sebesar 84,44%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

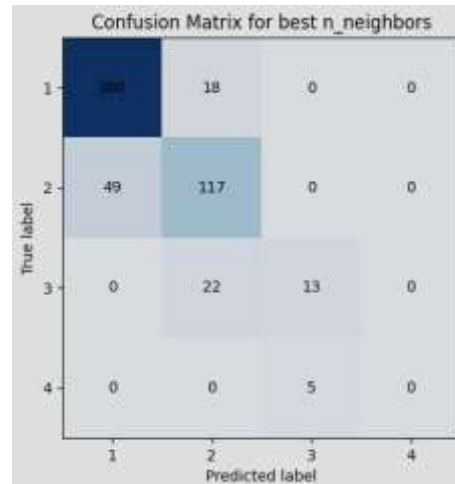


Gambar 8 Confusion Matrix Uji Coba Keenam

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 188 prediksi benar dan hanya 16 kesalahan pada kelas 2. Kelas 2 memiliki 56 prediksi benar dengan 20 kesalahan pada kelas 1. Kelas 3 memiliki 11 prediksi benar dengan 8 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 3 kesalahan pada kelas 3.

4.7. Uji Coba Nilai K 7 dan Perbandingan 60:40

Setelah dilakukannya uji coba dengan variasi nilai k senilai 7 serta perbandingan data training dan data testing sebesar 60:40, maka didapatkanlah hasil berupa akurasi sebesar 84,44%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

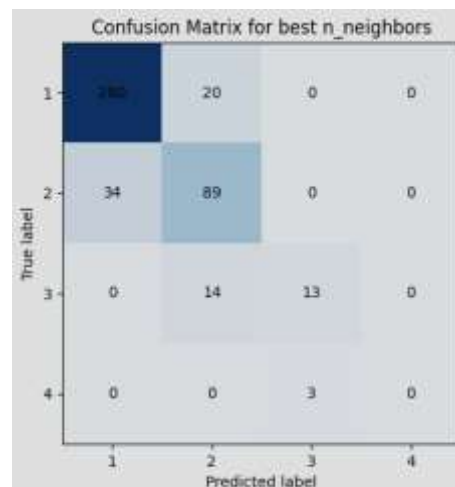


Gambar 9. Confusion Matrix Uji Coba Ketujuh

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 380 prediksi benar dan hanya 18 kesalahan pada kelas 2. Kelas 2 memiliki 117 prediksi benar dengan 49 kesalahan pada kelas 1. Kelas 3 memiliki 13 prediksi benar dengan 22 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 5 kesalahan pada kelas 3.

4.8. Uji Coba Nilai K 7 dan Perbandingan 70:30

Setelah dilakukannya uji coba dengan variasi nilai k senilai 7 serta perbandingan data training dan data testing sebesar 70:30, maka didapatkanlah hasil berupa akurasi sebesar 84,33%. Berikut ini adalah hasil pengujian berupa confusion matrix dari percobaan ini.

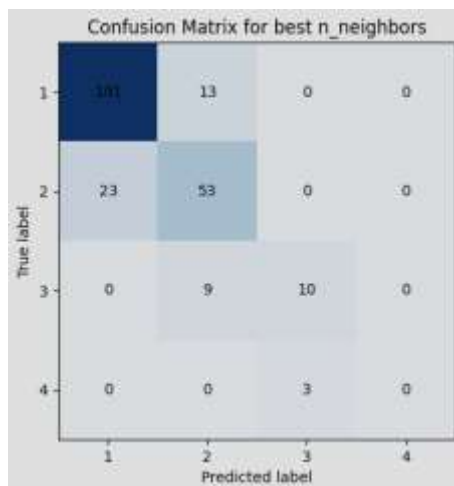


Gambar 10. Confusion Matrix Uji Coba Kedelapan

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 280 prediksi benar dan hanya 20 kesalahan pada kelas 2. Kelas 2 memiliki 89 prediksi benar dengan 34 kesalahan pada kelas 1. Kelas 3 memiliki 13 prediksi benar dengan 14 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 3 kesalahan pada kelas 3.

4.9. Uji Coba Nilai K 7 dan Perbandingan 80:20

Setelah dilakukannya uji coba dengan variasi nilai k senilai 7 serta perbandingan data training dan data testing sebesar 80:20, maka didapatkanlah hasil berupa akurasi sebesar 84,11%. Berikut ini adalah confusion matrix dari percobaan ini.



Gambar 11. Confusion Matrix Uji Coba Kesembilan

Dari confusion matrix diatas dapat disimpulkan bahwasanya Kelas 1 memiliki akurasi tertinggi dengan 191 prediksi benar dan hanya 13 kesalahan pada kelas 2. Kelas 2 memiliki 53 prediksi benar dengan 23 kesalahan pada kelas 1. Kelas 3 memiliki 10 prediksi benar dengan 9 kesalahan pada kelas 2, sedangkan kelas 4 tidak memiliki nilai prediksi benar dengan 3 kesalahan pada kelas 3.

Pada keseluruhan hasil percobaan yang telah dilakukan cenderung menunjukkan bahwasanya terdapat ketimpangan hasil klasifikasi yang didapatkan, terutama pada kelas 1. Hal tersebut terjadi dikarenakan terdapat ketidakseimbangan data pada dataset yang digunakan yang dimana kelas 1 memiliki jauh lebih banyak data daripada kelas lainnya. Oleh karena itu, model yang dihasilkan cenderung lebih akurat dalam memprediksi kelas 1 karena memiliki lebih banyak contoh untuk belajar.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan bahwasanya klasifikasi curah hujan harian apabila dilakukan menggunakan metode K-Nearest Neighbor mendapatkan hasil akurasi yang cukup akurat. Hal ini dapat dilihat dari serangkaian percobaan yang telah dilakukan, akurasi yang dihasilkan rata-rata diatas 80%. Dengan nilai akurasi

tertinggi sebesar 86,09% yang dimana hal tersebut didapatkan dengan variasi nilai k sebesar 3 dan perbandingan data training dan testing sebesar 60:40. Namun, torehan angka akurasi tersebut bukanlah merupakan angka akurasi tertinggi. Masih terdapat banyak metode lain yang dapat dikembangkan dengan penelitian ini untuk mendapatkan hasil yang lebih akurat. Selain itu, jumlah dimensi dataset juga dapat diperluas sehingga memungkinkan variasi dalam nilai prediksi yang dihasilkan. Selanjutnya, penerapan metode ini ke dataset lain juga dapat digunakan untuk pekerjaan masa depan.

DAFTAR PUSTAKA

- [1] M. Ridwan, (06 Maret 2019). Prakiraan Musim Kemarau Tahun 2019 di Indonesia. Link : https://www.bmkg.go.id/iklim/prakiraan-musim_bmkg. Diakses pada 12 Mei 2024.
- [2] I. G. J. Adi Putra and I. M. A. Oka Gunawan, "Klasifikasi Curah Hujan Harian Menggunakan Learning Vector Quantization (Studi Kasus : Stasiun Pengamatan Ngurah Rai)," *Majalah Ilmiah Untab*, vol. 19, no. 2, pp. 183–187, Sep. 2022.
- [3] T. Cover and P. Hart, "Nearest neighbor Pattern Classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967. doi:10.1109/tit.1967.1053964
- [4] F. M. Fathoni, C. A. Putra, and A. L. Nurlaili, "Klasifikasi Penyakit Daun Anggur menggunakan metode k-nearest neighbor Berdasarkan Gray level co-occurrence matrix," *Biner : Jurnal Ilmiah Informatika dan Komputer*, vol. 3, no. 1, pp. 8–15, Jan. 2024. doi:10.32699/biner.v3i1.6332
- [5] K. Alkhatib, H. Najadad, and I. Hmeidi, "Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm," *International Journal of Business, Humanities and Technology*, vol. 3, no. 3, pp. 32–44, Mar. 2013.
- [6] B. Triatmodjo, *Hidrologi Terapan*. Yogyakarta, Indonesia: Beta Offset, 2008.
- [7] G. E. P. Box, (1971). "Statistical control: A review and preview," *Journal of the Royal Statistical Society. Series A (General)*, vol.134, no. 3, pp. 251-297, Jan 1971.
- [8] J. Johnson, P. Denning, K. A. Delic, and D. Sousa-Rodrigues, "Big data: big data or big brother? that is the question now.," *Ubiquity*, vol. 2018, no. August, Aug. 2018, doi: 10.1145/3158352.
- [9] D. Pyle, *Data Preparation for Data Mining*. Los Altos: Morgan Kaufmann Publishers, 1999
- [10] T. A. Pratiwi, M. Irsyad, R. Kurniawan, S. Agustian, and B. S. Negara, "Klasifikasi Kebakaran Hutan Dan Lahan Menggunakan Algoritma Naïve Bayes Di Kabupaten Pelalawan," *CESS (Journal of Computer Engineering, System and Science)*, vol. 6, no. 1,

Art. no. 1, Feb 2021, doi:
10.24114/cess.v6i1.22555.

- [11] X. W. Steinberg, "Top 10 algorithms in data mining," *Knowl Inf Syst* (2008) 14:1–37 DOI 10.1007/s10115-007-0114-2, p. 14, 2008.
- [12] Y. V. Via, F. T. Anggraeny, and R. A. Jorgie, "Penerapan algoritma case based reasoning Dan K-nearest neighbor Untuk diagnosa Penyakit Ayam," *Prosiding Seminar Nasional Informatika Bela Negara*, vol. 2, pp. 192–195, Nov. 2021. doi:10.33005/santika.v2i0.140)