

Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner Untuk Memprediksi Hujan Kota Surakarta

Alvian Ahmada Akhbar^{1*}, Dwi Hartanti²

^{1,2} Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}202030329@mhs.udb.ac.id, ²dwhartanti@udb.ac.id

Abstrak— Penelitian ini bertujuan untuk mengimplementasikan algoritma K-Nearest Neighbor (KNN) menggunakan aplikasi Rapid Miner untuk memprediksi hujan di Kota Surakarta. Curah hujan memiliki pola yang tidak menentu sehingga sulit dilakukan prediksi dengan cara manual. Curah hujan tidak dapat ditentukan secara pasti namun hal ini dapat diperkirakan. Dengan demikian, adanya Data Mining dapat memungkinkan mesin mengenali dan mempelajari pola data yang rumit. Maka dari itu pembelajaran program dapat mempelajari pola data curah hujan untuk melakukan suatu prediksi. Penelitian ini menggunakan tiga variable sebagai kriteria yaitu temperatur, kecepatan angin, dan kelembapan. Hasil pengujian menggunakan algoritma K-Nearest Neighbour (KNN) dan aplikasi Rapid Miner dengan nilai $K=3$, ditemukan akurasi sebesar 83,87%. Dari 31 data yang diambil pada bulan Juli 2023. Hasil analisa membuktikan bahwa metode KNN menggunakan aplikasi Rapid Miner dapat digunakan dalam memprediksi hujan di Kota Surakarta. **Kata kunci**— K-Nearest Neighbor, Hujan, Rapid Miner.

Abstract— This study aims to implement the K-Nearest Neighbor (KNN) algorithm using the Rapid Miner application to predict rain in Surakarta City. Rainfall has an erratic pattern making it difficult to predict manually. Rainfall cannot be determined with certainty but this can be estimated. Thus, the existence of Data Mining can enable machines to recognize and study complex data patterns. Therefore program learning can learn patterns of rainfall data to make some predictions. This study uses three variables as criteria, namely temperature, wind speed, and humidity. The test results using the K-Nearest Neighbor (KNN) algorithm and the Rapid Miner application with a value of $K = 3$, found an accuracy of 83.87%. From 31 data taken in July 2023. The results of the analysis prove that the KNN method using the Rapid Miner application can be used to predict rain in Surakarta City.

Keywords— K-Nearest Neighbor, Rain, Rapid Miner.

I. PENDAHULUAN

Iklim adalah salah satu hal yang cukup krusial di dalam dunia ini. Garis lintang, ketinggian, lereng, jarak dari perairan, dan kondisi arus air laut merupakan pengaruh dari pembentukan iklim di suatu tempat. Namun setiap daerah memiliki iklim yang berbeda-beda. Setiap daerah memiliki jenis iklim yang dipengaruhi oleh garis lintang, karakteristik pola iklim global dipelajari melalui Klimatologi. Karakteristik pola iklim global juga mempertimbangkan kondisi seperti hujan, suhu, angin atau penguapan. Iklim di permukaan bumi dapat dibedakan berdasarkan garis lintang menjadi iklim kutub, sedang, tropis, subtropis dan khatulistiwa. Indonesia adalah negara yang beriklim tropis. Iklim berpengaruh sangat besar bagi kelangsungan hidup manusia [1]. Daerah tropis seperti Indonesia, anomali iklim ini menimbulkan pergeseran pola curah hujan, perubahan besar-besaran curah hujan dan perubahan temperatur udara yang mengakibatkan timbulnya musim kemarau, dan kekeringan [2].

Curah hujan yang tidak menentu, membuat peramalan manual menjadi tidak mudah. Namun, dengan adanya data memungkinkan mesin untuk mengidentifikasi dan menjelajahi pola data yang kompleks [3]. Sehingga mesin dapat mempelajari pola pada data cuaca untuk melakukan prediksi [4].

Memprediksi curah hujan berdasarkan dataset yang di dapatkan sebelumnya. Penelitian ini menggunakan metode K-Nearest Neighbor (KNN). KNN termasuk kedalam bagian Data Mining yang mampu mengenali suatu pola data sequensial. Metoda klasifikasi ini menggunakan konsep perhitungan jarak terdekat dengan sebuah titik [5]. Penentuan jarak menggunakan rumus euclidian distance yang menghasilkan jarak antara data baru dengan seluruh data pada dataset yang sudah memiliki kelas. Penelitian sebelum ini dilakukan oleh Riza Indriani Rakhmalia di Universitas Islam Indonesia Yogyakarta dengan judul “*Perbandingan Hasil Metode Naïve Bayes Classifier dan Support Vector Machine dalam Klasifikasi Curah Hujan*”. Pada penelitian sebelumnya metode *Naïve Bayes* dapat memberikan klasifikasi dengan akurasi 78%.

Sedangkan dengan menggunakan metode SVM mendapatkan akurasi 80% [6].

Curah hujan membutuhkan sejumlah parameter yang terdiri dari suhu udara, kelembaban dan kecepatan angin yang dapat digunakan untuk melihat kecenderungan curah hujan di masa mendatang. dalam prediksi curah hujan. Namun saat ini, curah hujan sudah semakin sulit untuk di prediksi [7]. Oleh karena itu, diperlukan suatu model atau sistem yang dapat memprediksi curah hujan secara akurat berdasarkan data masa lalu dari sumber BMKG (Badan Meteorologi dan Geofisika) [8].

Maka dari itu penelitian ini telah memprediksi curah hujan dengan menggunakan metode K-Nearest Neighbor. Sebelumnya data dilakukan praproses untuk melengkapi data yang hilang menggunakan teknik interpolasi data, sehingga didapatkan data yang lengkap, selanjutnya dilakukan normalisasi terhadap data sehingga di dapatkan fitur yang dapat diproses menggunakan metode KNN untuk mengklasifikasikan data tersebut kedalam 2 kelas yaitu hujan atau tidak.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan menggunakan algoritma K-Nearest Neighbor untuk melakukan klasifikasi terhadap objek berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut [9]. Dimana kelas yang paling banyak muncul yang nantinya akan menjadi kelas hasil dari klasifikasi [10]. KNN dilakukan dengan mencari kelompok k objek yang paling dekat (mirip) dengan objek pada dataset sebelumnya [11]. Perhitungan menggunakan KNN ini adalah 31 data latih. Berikut adalah langkah-langkah dalam menghitung menggunakan metode KNN.

1. Algoritma K-Nearest Neighbor :

- Tentukan Parameter k jumlah tetangga paling dekat.
- Hitung Euclidean Distance masing masing objek terhadap sampel data yang ada dengan persamaan (1) :

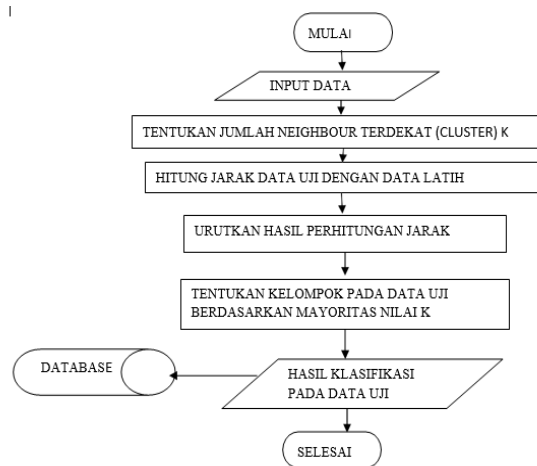
$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (1)$$

X_1 = data sample; X_2 = Data Uji; i = Variable Data; d = Jarak; p = Dimensi Data ;

- Urutkan objek-objek kedalam kelompok dengan jarak Eucliden yang kecil. K-Nearest Neighbor dari sebuah instance x adalah K instance dengan jarak terkecil (kedekatan terbesar, nearest) dengan x . Kedekatan dan jauhnya tetangga dihitung berdasarkan Euclidean Distance menggunakan persamaan (2):

$$D_{(a,b)} = \sqrt{\sum_{k=1}^d (a_k - b_k)^2} \quad (2)$$

Dimana matriks $D(a,b)$ merupakan jarak skalar dari kedua vektor a dan b dari matriks. Untuk menentukan nilai K dilihat dari jumlah klasifikasi, jika jumlahnya genap, sebaiknya menggunakan nilai K ganjil, sebaliknya jika jumlah klasifikasi jumlahnya ganjil, sebaiknya menggunakan nilai K genap. Proses dari metode K-Nearest Neighbor dapat dilihat pada Gambar 1 :



Gambar 1. Flowchart Proses Pengujian KNN

2. Klasifikasi

Proses klasifikasi memiliki 4 komponen yaitu:

- Kelas yaitu variabel dependen berupa kategorikal yang merepresentasikan 'label' pada objek.
- Predictor yaitu variabel independen yang direpresentasikan oleh karakteristik (atribut) data.
- Training dataset yaitu satu set data yang berisi nilai dari kelas dan predictor yang berguna untuk menentukan kelas yang cocok berdasarkan predictor.

4. *Testing dataset*, berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat untuk akurasi klasifikasi dievaluasi.

3. Evaluasi dan Validasi Algoritma Data Mining

Langkah ini menggunakan tabel *Confussion Matrix* seperti pada tabel 1., jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif.

Tabel 1. Confussion Matrix

Klasifikasi yang benar	Di klasifikasi sebagai	
	Positive(+)	Negative(-)
+ (Positive)	True Positive	False Negative
- (Negative)	False Positive	True Positive

1. *True positives* adalah jumlah kumpulan data positif yang diklasifikasikan sebagai positif.
 2. *False positives* adalah jumlah kumpulan data negative yang diklasifikasikan sebagai positif,
 3. *False negatives* adalah jumlah kumpulan data positif yang diklasifikasikan sebagai negatif.
 4. *True negatives* adalah jumlah kumpulan data negative yang diklasifikasikan sebagai negatif
- Confussion matrix* adalah : Akurasi = $(A+D)/(A+B+C+D)$

4. Rapid Miner

Tahapan dari *metode K-Nearest Neighbor* dengan *Rapid Miner* adalah:

- 1) Menentukan jumlah parameter k .
Parameter k adalah jumlah label dalam prediksi hasil cuaca.
- 2) Hitung jarak antara data yang akan di evaluasi (data uji) dengan semua data training.
- 3) Urutkan jarak yang terbentuk (urut naik).
- 4) Pilih K data terdekat.
- 5) Pilih hasil prediksi terbanyak dalam data yang didapatkan.
- 6) Tampilkan Output.
- 7) Setelah semua kriteria tersebut lengkap, kemudian dilakukan perhitungan jarak antara data training dengan data sample prediksi cuaca yang sudah ada dengan rumus persamaan.

III. HASIL DAN PEMBAHASAN

1 Pengujian K-Nearest Neighbor

Pada proses ini yang dijadikan evaluasi data uji adalah data prediksi cuaca pada bulan Juli 2023 dengan *data training* yang sudah ada. *Data Set* yang digunakan sebanyak 31 data yang dapat dilihat pada Tabel 2.

Tabel 2. Dataset Prediksi Hujan pada Bulan Juli

Tanggal	Temp (C°)	Kec Angin (M/S)	Kelembapan (%)	Keterangan
1 juli	28	17	77	tidak
2 juli	28	15	76	hujan
3 juli	28	14	75	tidak
4 juli	27	12	78	hujan
5 juli	28	12	79	hujan
6 juli	27	10	81	hujan
7 juli	27	11	80	hujan
8 juli	27	9	80	hujan
9 juli	28	9	79	hujan
10 juli	27	6	81	hujan
11 juli	27	7	82	hujan
12 juli	27	8	82	hujan
13 juli	27	8	81	tidak
14 juli	27	9	81	hujan
15 juli	28	17	77	tidak
16 juli	28	15	76	hujan
17 juli	28	14	75	tidak
18 juli	27	12	78	hujan
19 juli	28	12	79	hujan
20 juli	27	10	81	hujan
21 juli	27	11	80	hujan
22 juli	27	9	80	hujan
23 juli	28	9	79	hujan
24 juli	27	6	81	hujan
25 juli	27	7	82	hujan
26 juli	27	8	82	hujan
27 juli	27	8	81	tidak
28 juli	27	9	81	hujan
29 juli	28	14	75	tidak
30 juli	27	12	78	hujan
31 juli	28	17	77	tidak

Data uji yang didapatkan untuk memprediksi hujan di bulan Juli yang memuat Temperatur, Kecepatan Angin dan Kelembapan. Kemudian ya atau tidaknya adalah Label untuk menentukan prediksi hujan yang memungkinkan terjadi dan sebagai informasi terhadap para pengguna.

Dari tabel 2 diketahui informasi data yang disajikan untuk para pengguna memuat beberapa data yang dapat menjadi pandangan mengenai cuaca hari tersebut. Hal pertama yang dilakukan adalah dengan menentukan nilai K yang terbaik. Dimana K

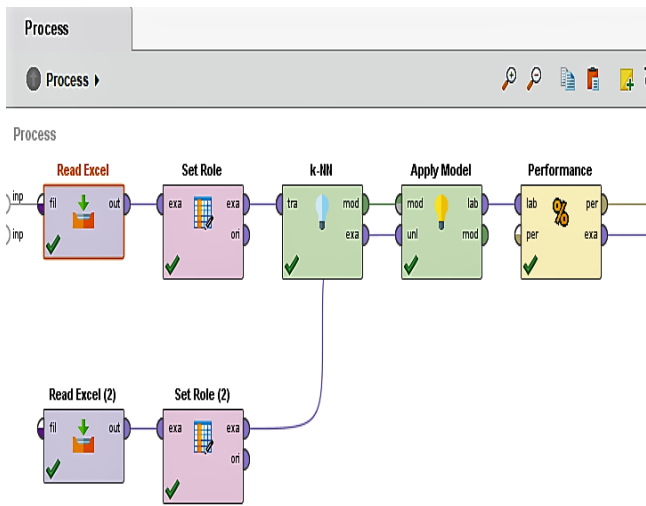
di sini merupakan jumlah klasifikasi Prediksi Cuaca yang beracuan terhadap data Temperatur, Kecepatan Angin dan Kelembapan.

2. Klasifikasi

Nilai K dari tetangga terdekat adalah 2 (K=3), berarti akan dikelompokkan 2 hasil prediksi cuaca (Label) yang akan diambil untuk di informasikan kepada pengguna. Setelah ditentukan nilai K maka langkah selanjutnya adalah dengan menghitung kuadrat jarak eucliden objek terhadap data training

3. Rapid Miner

Desain untuk pengujian dengan menggunakan Rapid miner seperti Gambar 2 sedangkan hasil dari Pengujian K-NN dapat dilihat pada Gambar 3 dan Gambar 4.



Gambar 2. Desain Pengujian Menggunakan Rapid Miner

Row No.	Keterangan	prediction(K...	confidence(...	confidence(...	Tanggal	Temperatur(...	Kecepatan ...	Kelembapan...
1	Tidak	Tidak	0.667	0.333	Jul 1, 2023	28	17	77
2	Hujan	Tidak	0.500	0.500	Jul 2, 2023	28	15	76
3	Tidak	Tidak	0.500	0.500	Jul 3, 2023	28	14	75
4	Hujan	Hujan	0.250	0.750	Jul 4, 2023	27	12	78
5	Hujan	Hujan	0	1	Jul 5, 2023	28	12	79
6	Hujan	Hujan	0	1	Jul 6, 2023	27	10	81
7	Hujan	Hujan	0	1	Jul 7, 2023	27	11	80
8	Hujan	Hujan	0	1	Jul 8, 2023	27	9	80
9	Hujan	Hujan	0	1	Jul 9, 2023	28	9	79
10	Hujan	Hujan	0	1	Jul 10, 2023	27	6	81
11	Hujan	Hujan	0	1	Jul 11, 2023	27	7	82
12	Hujan	Hujan	0.250	0.750	Jul 12, 2023	27	8	82
13	Tidak	Tidak	0.500	0.500	Jul 13, 2023	27	8	81
14	Hujan	Tidak	0.500	0.500	Jul 14, 2023	27	9	81
15	Tidak	Tidak	0.500	0.500	Jul 15, 2023	28	17	77

ExampleSet (31 examples: 4 special attributes: 4 regular attributes)

Gambar 3. Hasil Pengujian K-NN

Row No.	Keterangan	prediction(K...	confidence(...	confidence(...	Tanggal	Temperatur(...	Kecepatan ...	Kelembapan...
17	Tidak	Tidak	0.500	0.500	Jul 17, 2023	28	14	75
18	Hujan	Hujan	0.250	0.750	Jul 18, 2023	27	12	78
19	Hujan	Hujan	0	1	Jul 19, 2023	28	12	79
20	Hujan	Hujan	0	1	Jul 20, 2023	27	10	81
21	Hujan	Hujan	0	1	Jul 21, 2023	27	11	80
22	Hujan	Hujan	0	1	Jul 22, 2023	27	9	80
23	Hujan	Hujan	0	1	Jul 23, 2023	28	9	79
24	Hujan	Hujan	0	1	Jul 24, 2023	27	6	81
25	Hujan	Hujan	0	1	Jul 25, 2023	27	7	82
26	Hujan	Hujan	0.250	0.750	Jul 26, 2023	27	8	82
27	Tidak	Tidak	0.500	0.500	Jul 27, 2023	27	8	81
28	Hujan	Tidak	0.500	0.500	Jul 28, 2023	27	9	81
29	Tidak	Tidak	0.500	0.500	Jul 29, 2023	28	14	75
30	Hujan	Tidak	0.500	0.500	Jul 30, 2023	27	12	78
31	Tidak	Tidak	0.667	0.333	Jul 31, 2023	28	17	77

Gambar 4. Hasil Pengujian K-NN

Hasil Pengujian pada Gambar 3 dengan nilai K=3, kolom Keterangan dan Prediction merupakan prediksi cuaca hujan atau tidaknya Kota Surakarta pada hari itu. Data Temperatur, Kecepatan Angin dan Kelembapan juga telah diinputkan ke dalam proses pengujian. Hasil Performance vektor dapat dilihat pada gambar 5.

	true Tidak	true Hujan	class precision
pred. Tidak	8	5	61.54%
pred. Hujan	0	18	100.00%
class recall	100.00%	78.26%	

Gambar 5. Hasil Performace Vektor.

IV. KESIMPULAN

Berdasarkan hasil dari penelitian ini adalah memprediksi curah hujan di Kota Surakarta menggunakan metode K- Nearest Neighbor dan menggunakan Rapid Miner dengan parameter temperatur rata-rata, kelembapan rata-rata, dan kecepatan angin rata-rata. Menghasilkan suatu prediksi hujan harian dengan melihat hasil prediksi curah hujan menggunakan metode K-Nearest Neighbor dan aplikasi Rapid Miner. Parameter yang sangat berpengaruh dalam keakuratan hasil prediksi yaitu kecepatan angin rata-rata karena menurut BMKG angin berfungsi dalam menggeser awan dari satu tempat ke tempat yang lain. Wilayah yang mempunyai angin yang tidak terlalu kuat mempunyai kemungkinan memiliki intensitas curah hujan yang kecil. Berdasarkan nilai K yang telah dilakukan pengujian menggunakan Rapid Miner maka didapatkan hasil terbaik dari 31 data dengan nilai akurasi sebesar 83,87 %.

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan kehadirat Allah SWT, karena atas rahmat dan ridha-Nya penulis dapat menyelesaikan penyusunan ini, artikel jurnal judul “(Analisis Metode K-Nearest Neighbor Menggunakan Rapid Miner Untuk Memprediksi Hujan Kota Surakarta)”.

Penyusunan artikel jurnal ini bertujuan untuk menganalisis metode K-NN menggunakan Rapid Miner untuk memprediksi hujan di Kota Surakarta yang dimana sangat sulit untuk memprediksi hujan secara manual.

Segala upaya telah dilakukan dalam penyusunan skripsi ini. Namun, penulis menyadari akan kekurangan karena keterbatasan kemampuan dan pengetahuan penulis.

Oleh karena itu, penulis mengharapkan kritik dan saran yang bersifat membangun guna memperbaiki kekurangan tersebut di masa yang akan datang. Penulis berharap artikel jurnal ini dapat bermanfaat bagi pembaca pada umumnya dan khususnya bagi penulis.

REFERENSI

- [1] B. Tjasyono H. K., *Meteorologi Indonesia Volume I -Karakteristik dan Sirkulasi Atmosfer*, I. 2012
- [2] F B. Irawan, “Fenomena Anomali Iklim El Nino dan La Nina: Kecenderungan Jangka Panjang dan Pengaruhnya terhadap Produksi Pangan,” *Forum Penelit. Agro Ekon.*, vol. 24, no. 1, p. 28, 2016.
- [3] F. R. Lumbanraja, R. S. Sani, D. Kurniawan, and A. R. Irawati, “Implementasi Metode Support Vector Machine Dalam Prediksi Persebaran Demam Berdarah Di Kota Bandar Lampung,” *J. Komputasi*, vol. 7, no. 2, 2019.
- [4] S. B. Navathe, W. Wu, S. Shekhar, X. Du, X. Sean Wang, and H. Xiong, “Database Systems for Advanced Applications: 21st International Conference, DASFAA 2016 Dallas, TX, USA, April 16–19, 2016 Proceedings, Part I,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9642, pp. 214–228, 2016.
- [5] A. Bode, “K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika,” *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 188–195, 2017.
- [6] P. S. Statistika, F. Matematika, D. A. N. Ilmu, P. Alam, and U. I. Indonesia, “Perbandingan Hasil Metode Naïve Bayes Classifier dan Support Vector Machine Dalam,” 2018.
- [7] D. Howard and B. Mark, “Neural Network Toolbox Documentation,” *Neural Netw. Tool*, p. 846, 2004.
- [8] N. Ritha, M. Bettiza, and A. Dufan, “Prediksi Curah Hujan dengan Menggunakan Algoritma Levenberg-Marquardt dan Backpropagation,” *J. Sustain.*, vol. 5, no. 2, pp. 11–16, 2016.
- [9] A. J. T, D. Yanosma, and K. Anggriani, “Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka,” *Pseudocode*, vol. 3, no. 2, pp. 98–112, 2017
- [10] M. Fansyuri, “Analisa algoritma klasifikasi k-nearest neighbor dalam menentukan nilai akurasi terhadap kepuasan pelanggan (study kasus pt. Trigatra komunikatama),” *Humanika J. Ilmu Sos. Pendidikan, dan Hum.*, vol. 3, no. 1, pp. 29–33, 2020.
- [11] N. Reflan, A. Aflahah, Kusri, and Juwari, “Implementasi Metode K-Nearest Neighbor (Knn) Untuk Memprediksi Varietas Padi Yang Cocok Untuk Lahan Pertanian,” *J. Inf. Politek. Indonusa Surakarta*, vol. 4, pp. 2–8, 2018.