

Implementasi Teknik Data Mining terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan

ANDI SADRI AGUNG, A. ARFAN FAUZI, ANDI AKRAM NUR RISAL, FHATIAH ADIBA

Universitas Negeri Makassar, Indonesia
Email : andisadriagung20@gmail.com

ABSTRAK

Cuaca cenderung tidak stabil dan susah diprediksi mempengaruhi beberapa sektor kehidupan manusia. Pemanfaatan teknologi dibidang data sains dalam memprediksi cuaca mendatang sangat dibutuhkan. Badan Meteorologi Klimatologi dan Geofisika (BMKG) merupakan organisasi pemerintahan, bertugas memprediksi cuaca di Indonesia. Namun dalam menentukan akurasi data, diperlukan metode yang memprediksi cuaca dengan akurat. Penelitian ini bertujuan menghasilkan informasi dari klasifikasi data prediksi curah hujan BMKG di Sulawesi Selatan. Klasifikasi dibagi menjadi 5 kategori dengan 1036 data dan 11 atribut. Pengumpulan data dimulai dari 1 Januari 2020 hingga 31 Oktober 2022. $K = 3$ menjadi nilai K terbaik dengan akurasi mencapai 82,21%. Sehingga disimpulkan bahwa data prediksi curah hujan BMKG memiliki tingkat akurasi data yang baik dan hasil penelitian ini dapat dijadikan acuan dalam menginformasikan curah hujan di Sulawesi Selatan.

Kata kunci: curah hujan, *data mining*, *K-Nearest Neighbor*, prediksi, Sulawesi Selatan

ABSTRACT

The weather tends to be unstable and difficult to predict, affecting several sectors of human life. The utilization of technology in the field of scientific data in predicting future weather is urgently needed. The Meteorology, Climatology, and Geophysics Agency (BMKG) is a government organization, tasked with predicting weather in Indonesia. But in determining the accuracy of data, we require a method that predicts the weather accurately. This study aims to produce information from the classification of BMKG rainfall prediction data in South Sulawesi. Classification is divided into 5 categories with 1036 data and 11 attributes. Data collection starts from 1 January 2020 to 31 October 2022. $K = 3$ is the best K value with an accuracy of 82.21%. So it was concluded that BMKG rainfall prediction data has a good data accuracy level, and this study's results can be used as a reference in informing rainfall in South Sulawesi.

Keywords: *rainfall, data mining, K-Nearest Neighbor, predictions, South Sulawesi*

1. PENDAHULUAN

Sulawesi Selatan merupakan provinsi yang terletak di semenanjung bagian selatan Sulawesi yang terletak di negara Indonesia. Terdapat 2 jenis musim di Indonesia yaitu musim kemarau dan musim hujan. Pergantian musim ini dipengaruhi oleh perubahan iklim global bumi yang mengalami kenaikan setiap tahun. Cuaca cenderung tidak stabil dan susah diprediksi sangat mempengaruhi beberapa sektor dalam kehidupan manusia **(Utami et al., 2021)**.

Cuaca merupakan kondisi atmosfer pada wilayah terbatas dan dalam jangka waktu tertentu **(Subagyo, 2015)**. Terdapat beberapa faktor yang mempengaruhi kondisi cuaca, diantaranya ialah kelembaban, suhu, angin, curah hujan, sinar matahari, dan tekanan atmosfer. Faktor tersebut dapat dijadikan atribut dalam memprediksi cuaca yang akan terjadi di waktu mendatang. Banyak orang yang akan terbantu dengan adanya prakiraan cuaca terkhusus pada pekerja di bidang pertanian dan kelautan. Prediksi cuaca yang akurat dan cepat dapat memudahkan *stakeholder* terkait dalam mengelola informasi perubahan cuaca **(Muhandhis et al., 2021)**.

Seiring dengan perkembangan teknologi saat ini dalam bidang data sains memprediksi cuaca di masa yang akan datang. Badan Meteorologi Klimatologi dan Geofisika (BMKG) merupakan *stakeholder* terkait dalam mengelola informasi prediksi cuaca di Indonesia, namun dalam menentukan tingkat akurasi data tersebut, diperlukan sebuah metode khusus.

Data mining merupakan salah satu metode yang cocok, dimana proses dengan mengimplementasikan matematika dan teknik statistik dalam kecerdasan buatan hingga *machine learning* terhadap berbagai *database*, kemudian diolah dan diekstraksi, lalu dilakukan identifikasi pengetahuan serta memanfaatkan data yang berpotensi menghasilkan sebuah informasi **(Mohamed & Ezzati, 2019)(Mulyati et al., 2020)**. Adapun *data mining* dapat dikategorikan berdasarkan tugas serta fungsi yang dapat dilakukan, antara lain prediksi, klasifikasi, pengelompokan (*clustering*), deskripsi, estimasi dan asosiasi.

Adapun klasifikasi merupakan ilmu yang terdapat pada *machine learning* dan juga sebuah metode yang dapat mengolah dan menangani *big data* **(Mulyati et al., 2020)**. Sebagaimana data iklim yang telah kami kumpulkan melalui *Data Online* BMKG (**DATA ONLINE - PUSAT DATABASE - BMKG, n.d.**). Diperoleh 1036 buah data cuaca harian *time series* di Provinsi Sulawesi Selatan dengan 11 jenis atribut dari bulan Januari 2020 hingga Oktober 2022.

Berdasarkan penelitian sebelumnya terkait prediksi cuaca menggunakan metode *data mining* berjudul "Algoritma *K-Nearest Neighbor* (KNN) dalam memprediksi dan menghitung tingkat akurasi Data Cuaca di Indonesia", diperoleh tingkat akurasi data sebesar 89% **(Rangkuti et al., 2021)**. Peneliti berikutnya, yang juga mengimplementasi Algoritma KNN menggunakan metode regresi dalam menentukan potensi hujan harian, diperoleh nilai RMSE 9.899 +/- 0.000 **(Rofiq et al., 2020)**.

Selain itu, terdapat penelitian yang membangun sebuah sistem deteksi kondisi cuaca berdasarkan analisis histogram HCL berbasis pengolahan citra dengan menerapkan algoritma KNN. Hasil yang didapatkan adalah tingkat akurasi sebesar 84,21%. Kondisi cuaca yang dapat diprediksi adalah cerah berawan, berawan, hujan, malam cerah dan malam hujan **(Hariani, 2020)**.

Selanjutnya, penelitian lain yang juga mengimplementasikan metode KNN dalam mengklasifikasi cuaca di wilayah Tanjungpinang, didapatkan persentase tertinggi yaitu dengan nilai $K = 31, 33, 35, 37, 39, 41, 45, 47, 50$ tingkat persentase kesamaan sebesar 70% (Syarifaturrahmah et al., 2022).

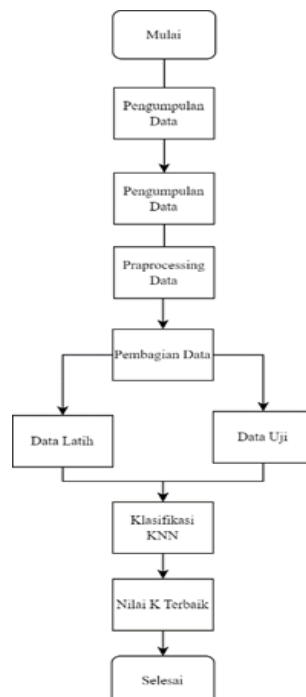
Terakhir, penelitian yang melakukan modifikasi terhadap Algoritma KNN, tingkat akurasi terbaik yang dihasilkan hanya mencapai 70,47%. Akurasi terbaik dalam memprediksi 12 bulan kedepan, dihasilkan oleh BPNN-lm mencapai 82,46%. Sedangkan dalam memprediksi 24 bulan kedepan, nilai MAPE terbaik dihasilkan BPNN-lm (Lareno & Muhammad, 2015).

Namun, berdasarkan hasil penelitian di atas, dapat disimpulkan bahwa dari beberapa metode yang digunakan, masih dapat dikembangkan, terutama dalam menghitung akurasi data cuaca harian. Selain itu, belum ada penelitian sebelumnya yang melakukan implementasi *data mining* terhadap prediksi data cuaca harian dan curah hujan, khususnya di provinsi Sulawesi Selatan.

Oleh karena itu, berdasarkan uraian permasalahan diatas, peneliti memutuskan untuk melakukan "Implementasi *Data Mining* Algoritma *K-Nearest Neighbor* terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan". Algoritma ini melakukan klasifikasi data berdasarkan 5 kategori klasifikasi curah hujan dengan kode 1 hingga 5 yaitu tidak hujan (1), hujan ringan (2), hujan sedang (3), hujan lebat (4), dan hujan sangat lebat (5) dengan 1036 data cuaca harian. Hasil akhir yang dicapai adalah nilai K (ketetanggaan terdekat) terbaik dengan tingkat akurasi tertinggi menggunakan Algoritma *K-Nearest Neighbor*.

2. METODE

Penelitian ini menggunakan metode yang terdiri dari beberapa tahap secara terurut sebagaimana pada Gambar 1, diawali dengan pengumpulan data hingga hasil akhir yaitu mencari akurasi hasil klasifikasi nilai K terbaik.



Gambar 1. Alur Metode Penelitian

2.1 Pengumpulan Data

Tahapan awal dalam penelitian ini adalah pengumpulan data berbentuk *dataset*. *Dataset* yang digunakan merupakan data cuaca harian yang tercatat di Stasiun Klimatologi Sulawesi Selatan dan diunduh dari *Database Online* BMKG. Terdiri atas 1036 *records* data dengan 11 total atribut dari 1 Januari 2020 hingga 31 Oktober 2022.

Data tersebut merupakan data cuaca harian dengan atribut seperti tanggal, *Tn* atau suhu minimum (°C), *Tx* atau suhu maksimum (°C), *Tavg* atau suhu rata-rata (°C), *RH_avg* atau kelembaban rata-rata(%), *RR* atau curah hujan (mm), *ss* atau lama penyinaran matahari (jam), *ff_x* atau Kecepatan angin maksimum (m/s), *ddd_x* atau Arah angin saat kecepatan maksimum (°), *ff_avg* atau Kecepatan angin rata-rata (m/s), dan *ddd_car* atau Arah angin terbanyak(°)

2.2 Preprocessing Data

Preprocessing data merupakan sebuah tahapan yang perlu dilalui sebelum memasukkan data yang akan diklasifikasi. Hal ini bertujuan agar data yang diklasifikasi bersih dari atribut kosong hingga jenis data yang berbeda, yang dapat mempengaruhi hasil klasifikasi dan akurasi yang akan diperoleh (Bertalya et al., 2021).

Preprocessing data yang dilakukan terhadap data cuaca harian yaitu membentuk data baru dengan melakukan penghapusan (*drop*) atribut tanggal dan arah angin terbanyak, dikarenakan keduanya bertipe data *object/string*. Informasi data setelah *drop* atribut ditunjukkan pada Tabel 1.

Tabel 1. Informasi Data setelah Penghapusan (*drop*) Atribut

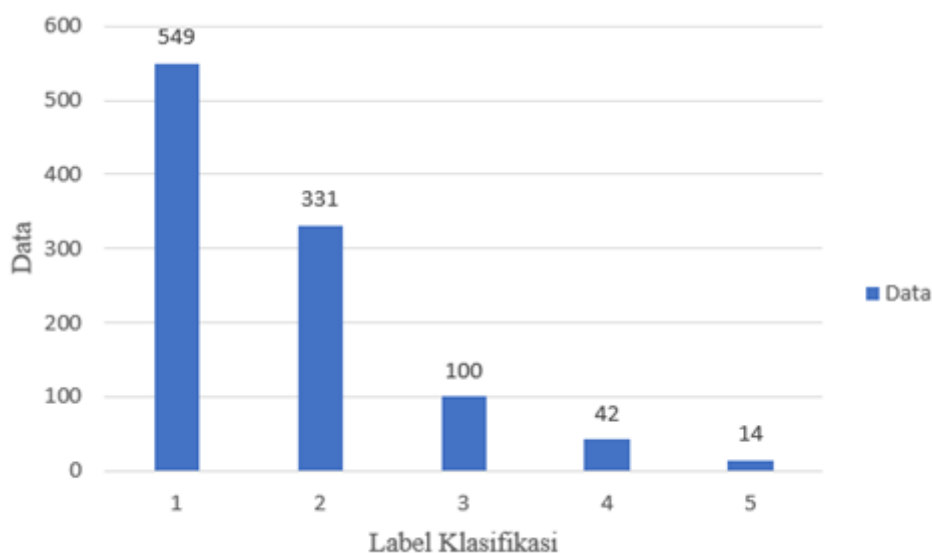
<i>Indeks</i>	<i>Column</i>	<i>Non-Null Count</i>	<i>Dtype</i>
0	<i>Tn</i>	1014 non-null	float64
1	<i>Tx</i>	1022 non-null	float64
2	<i>Tavg</i>	1022 non-null	float64
3	<i>RH_avg</i>	1022 non-null	float64
4	<i>RR</i>	934 non-null	float64
5	<i>Ss</i>	1021 non-null	float64
6	<i>ff_x</i>	1023 non-null	float64
7	<i>ddd_x</i>	1022 non-null	float64
8	<i>ff_avg</i>	1023 non-null	float64
9	<i>Rain</i>	1036 non-null	int64

Dari tabel diatas, diketahui bahwa data terdiri 1036 jumlah data dan terdapat beberapa data yang jumlah *non-null* atau kolom terisi tidak mencapai 1036 baris. Hal itu disebabkan karena adanya atribut kosong, yang akan mengganggu proses klasifikasi nantinya. Sehingga diperlukan proses pengisian atribut kosong dengan mengambil nilai mean atau rata-rata dari nilai keseluruhan atribut tersebut. Informasi data setelah pengisian atribut ditunjukkan pada Tabel 2.

Tabel 2. Informasi Data setelah Pengisian Atribut Kosong

<i>Indeks</i>	<i>Column</i>	<i>Non-Null Count</i>	<i>Dtype</i>
0	<i>Tn</i>	1036 non-null	float64
1	<i>Tx</i>	1036 non-null	float64
2	<i>Tavg</i>	1036 non-null	float64
3	<i>RH_avg</i>	1036 non-null	float64
4	<i>RR</i>	1036 non-null	float64
5	<i>ss</i>	1036 non-null	float64
6	<i>ff_x</i>	1036 non-null	float64
7	<i>ddd_x</i>	1036 non-null	float64
8	<i>ff_avg</i>	1036 non-null	float64
9	<i>rain</i>	1036 non-null	int64

Terakhir, atribut akan dibagi menjadi dua kategori variabel, yaitu *independen* (mempengaruhi) yang berupa masukan untuk hasil prediksi kategori curah hujan dan *dependen* (dipengaruhi) yaitu data target yang akan dicapai berupa hasil prediksi. Pada penelitian ini, yang menjadi variabel *dependen* adalah atribut *rain* (hujan) yang berisi data diskrit berupa angka 1 sampai 5 yang masing-masing mewakili label kelas klasifikasi kategori tidak hujan (1), hujan ringan (2), hujan sedang (3), hujan lebat (4), dan hujan sangat lebat (5) sebagaimana dengan ketentuan BMKG (*Probabilistik Curah Hujan 20 Mm (Tiap 24 Jam) | BMKG, n.d.*). Adapun rincian variabel *dependen* atau klasifikasi hujan ditunjukkan pada histogram Gambar 2.



Gambar 2. Histogram Data Label Klasifikasi Hujan (*Dependen*)

Gambar 2 menjelaskan atribut *Rain*, data target yang akan diproses menggunakan algoritma KNN. Pada kategori 1 terdapat 549 data (Tidak Hujan), 2 sebanyak 331 data (Hujan Ringan), 3 memiliki data 100 data (Hujan Sedang), 4 memiliki 42 data (Hujan Lebat) dan terakhir kategori 5 terdapat 14 data (Hujan Sangat Lebat). Sedangkan untuk variabel *independen* menggunakan atribut *Tavg* atau suhu rata-rata (°C), *RH_avg* atau kelembaban rata-rata (%),

RR atau curah hujan (mm) dan *ff_avg* atau Kecepatan angin rata-rata (m/s) sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Sample Data Atribut Independen

<i>Indeks</i>	<i>Tavg</i>	<i>RH_avg</i>	<i>ff_avg</i>	<i>RR</i>
0	27	89	3	30
1	25,2	97	2	15,5
2	25,1	95	1	69,3
3	27,2	86	3	32,9
4	26,1	88	3	1,5

2.3 Pembagian Data

Tahapan berikutnya adalah pembagian data pelatihan dan data pengujian. Proses ini dilakukan menggunakan *sklearn.model_selection* dengan data latih sebesar 80% atau 828 data dari keseluruhan data, dan sisanya 20% atau 208 data menjadi data uji nantinya. Selain itu juga ditentukan nilai *random state* yang akan melakukan proses acak untuk pembagian data. Hal ini berguna agar tingkat akurasi nantinya tidak terpengaruhi oleh urutan data *time series* yang disediakan.

2.4 Klasifikasi menggunakan Algoritma K-Nearest Neighbor (KNN)

Metode yang digunakan pada penelitian ini adalah klasifikasi menggunakan *K-Nearest Neighbor* atau KNN. Merupakan salah satu *supervised learning* yang dikenal juga sebagai *lazy learning*. Hal tersebut dikarenakan KNN hanya melakukan proses pembelajaran (*learning*) dan klasifikasi tanpa membangun model (Brawijaya et al., 2020). Seluruh data latih hanya disimpan ke dalam memori (Fajri et al., 2020). Sehingga proses klasifikasi yang dilakukan juga lebih mudah dan cepat.

Proses klasifikasi pada algoritma ini terhadap objek, berdasarkan data pelatihan yang memiliki jarak terdekat dengan objek tersebut atau dikenal sebagai nilai K (ketetanggaan terdekat), sehingga akan dicari nilai K dengan akurasi terbaik (Widaningsih, 2019).

Proses penerapan algoritma KNN secara berurut dibagi kedalam 3 proses seperti dibawah ini: (1) Menentukan nilai K awal atau jumlah ketetanggaan yang digunakan sebagai penentuan kelas awal pada proses *training* atau pelatihan. (2) Menghitung jarak antar data baru pada proses pengujian terhadap data lainnya pada data pelatihan menggunakan metrik jarak, dan (3) Memilih sejumlah K data dengan jarak terdekat dan menentukan kelas dari data baru tersebut.

2.5 Pengukuran Akurasi Nilai K Terbaik

Tahap terakhir dalam penelitian ini adalah melakukan pengukuran tingkat akurasi nilai K terbaik dari proses klasifikasi menggunakan algoritma KNN. Proses pengukuran ini menggunakan metode *Confusion Matrix*, merupakan metode analitik prediktif yang dapat menampilkan serta dapat memberikan perbandingan nilai aktual dengan nilai hasil prediksi dari proses model *training* (Martia Nanda et al., n.d.). Nilai perbandingan tersebut digunakan untuk menghitung metrik evaluasi beberapa nilai diantaranya nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* atau *F-Measure*.

Accuracy atau akurasi adalah nilai yang diperoleh berdasarkan jumlah data yang bernilai benar atau positif dan data hasil prediksi benar ditambah data yang bernilai salah atau negatif dan

hasil prediksi negatif. Kemudian dibagi dengan total seluruh data yang berada pada *dataset*. Perhitungannya menggunakan Persamaan 1.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

Keterangan :

TP = *True Positive* yaitu jumlah data benar dan hasil prediksi benar.

TN = *True Negative* yaitu jumlah data benar namun hasil prediksi salah.

FP = *False Positive* yaitu jumlah data salah dan hasil prediksi salah.

FN = *False Negative* yaitu jumlah data salah namun hasil prediksi benar.

Precision adalah nilai peluang data yang hasil prediksinya benar atau positif dan pada kenyataannya termasuk data kategori benar atau positif. Perhitungannya menggunakan Persamaan 2.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan nilai peluang data dengan kategori benar atau positif dan dengan tepat hasil prediksinya benar atau positif. Perhitungannya menggunakan Persamaan 3.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score atau *F-Measure* merupakan nilai yang diperoleh dari hasil nilai *Precision* dan *Recall* antara kategori hasil prediksi dengan kategori sebenarnya. Perhitungannya menggunakan Persamaan 4.

$$F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

Sedangkan untuk mencari nilai K dengan akurasi terbaik, dilakukan proses perulangan kemudian dicari nilai akurasi dari tiap nilai K yang diinisiasi.

3. HASIL DAN PEMBAHASAN

Hasil proses prediksi dan perhitungan akurasi dari data curah hujan di provinsi Sulawesi Selatan dilakukan melalui beberapa tahapan, yaitu mulai dari tahap pengumpulan data, *preprocessing* data, klasifikasi data, hingga dengan pengukuran dan pencarian tingkat akurasi yang terbaik.

Data awal yang dikumpulkan terdiri atas 1036 buah data dengan 11 total atribut awal dari 1 Januari 2020 hingga 31 Oktober 2022. Data belum siap melalui proses klasifikasi, sehingga perlu dilakukan tahap berikutnya yaitu *preprocessing* data, dimana dilakukan pembersihan data dan pengisian atribut kosong serta penentuan atribut *independen* dan atribut *dependen*. Selanjutnya, dilakukan proses pembagian data latih dan data uji menggunakan model *sklearn* pada *python* dengan data latih sebanyak 828 atau 80% dari 1036 total data keseluruhan dan sisanya 208 atau 20% data cuaca sebagai data latihnya.

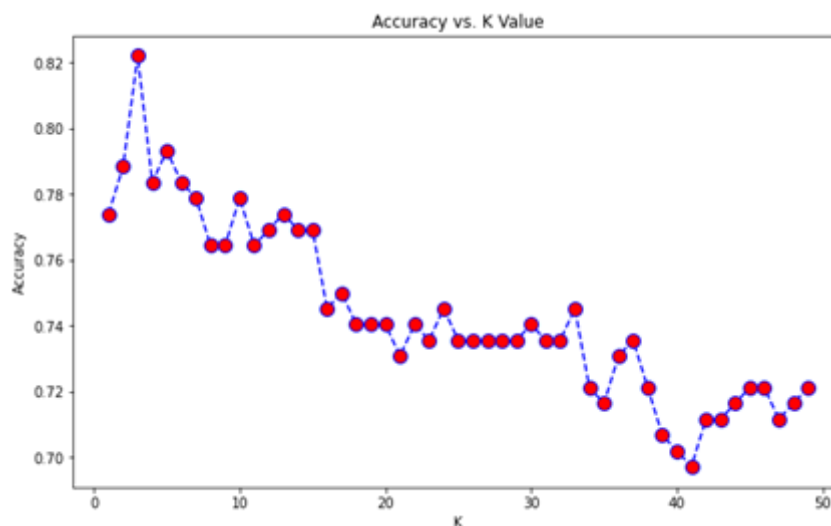
Tahap berikutnya adalah proses klasifikasi menggunakan algoritma KNN, sebagaimana langkah awal yaitu menentukan nilai K awal. Dalam penelitian ini, nilai K awal yang digunakan adalah 5. Dari proses pelatihan, juga diukur tingkat akurasi terhadap data latih dengan nilai K = 5 yaitu sebesar 84,54%. Selanjutnya, dilakukan pengukuran akurasi terhadap data uji

dengan nilai $K = 5$. Diperoleh hasil pengujian dengan akurasi sebesar 79,32% yang ditampilkan dalam *confusion matrix classification report* sebagaimana pada Gambar 3.

	precision	recall	f1-score	support
1	0.81	0.87	0.84	122
2	0.68	0.63	0.66	62
3	1.00	0.69	0.82	13
4	1.00	1.00	1.00	9
5	1.00	1.00	1.00	2
accuracy			0.79	208
macro avg	0.90	0.84	0.86	208
weighted avg	0.79	0.79	0.79	208

Gambar 3. Confusion Matrix Pengujian $K = 5$

Berdasarkan Gambar 3, dapat diamati, bahwa untuk klasifikasi terhadap 5 kelas, untuk kategori 1 (tidak hujan) sebesar 84%, kategori 2 (hujan ringan) sebesar 66%, kategori 3 (hujan sedang) sebesar 82% dan masing-masing kategori 4 (hujan lebat) dan kategori 5 (hujan sangat lebat) sebesar 100%. Selanjutnya akan mencari nilai K terbaik untuk mendapatkan akurasi yang jauh lebih baik. Sehingga proses terakhir adalah melakukan perbandingan terhadap semua nilai K dengan *range* 1 hingga 50. Hasil perbandingan nilai K terbaik ditunjukkan Gambar 4.



Gambar 4. Perbandingan Akurasi Nilai K terbaik

Hasil pengujian untuk akurasi terendah berada pada nilai $K = 41$ sebesar 69,71%, sedangkan untuk akurasi tertinggi dicapai pada nilai $K = 3$ sebesar 82,21%. Sehingga dapat disimpulkan bahwa nilai $K = 3$ merupakan nilai K terbaik dalam melakukan pengujian terhadap data prediksi curah hujan di provinsi Sulawesi Selatan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan terhadap data prediksi curah hujan di provinsi Sulawesi Selatan yang berasal dari Stasiun Klimatologi Sulawesi Selatan dan masalah serta tujuan penelitian yang telah dibahas sebelumnya. Maka kami dapat menarik beberapa kesimpulan sebagaimana berikut ini :

1. Implementasi *data mining* dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) terhadap data prediksi curah hujan BMKG di provinsi Sulawesi Selatan dapat diterapkan dengan baik.
2. Akurasi pelatihan awal nilai K = 5 sebesar 84,54% dan pengujian awal sebesar 79,32%, maka model klasifikasi yang dirancang dikategorikan sangat baik dengan akurasi yang tergolong tinggi.
3. Selanjutnya dalam melakukan perbandingan nilai K terbaik, diperoleh nilai K = 3 sebagai nilai K terbaik dengan akurasi tertinggi sebesar 82,21%.
4. Hasil penelitian diatas, dapat disimpulkan juga bahwa data prediksi curah hujan BMKG memiliki tingkat akurasi data yang cukup tinggi, sehingga hasil prediksi curah hujan dapat menjadi informasi penting BMKG dalam menginformasikan tingkat curah hujan di Sulawesi Selatan.

Adapun saran peneliti terhadap penelitian selanjutnya yaitu diharapkan agar menggunakan data yang lebih banyak, semakin banyak data yang digunakan, maka akurasi dari model prediksi yang diteliti akan semakin tinggi. Selain itu juga disarankan menggunakan atribut lain dalam menentukan prediksi, serta perbandingan dengan metode lain menggunakan data yang sama maupun dengan lokasi yang berbeda.

DAFTAR RUJUKAN

- Bertalya, Prihandoko, Setyowati, L., Irawan, F. I., & Irlianti, S. R. (2021). Formulation of city health development index using data mining. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(1), 362–369. <https://doi.org/10.11591/ijeecs.v23.i1.pp362-369>
- Brawijaya, J., Djohan, F., & Suryaningrum, K. M. (2020). Aplikasi Pendeteksi Dan Analisa Cuaca Menggunakan Metode K-Nearest Neighbor Berbasis Android. *JSTIE (Jurnal Sarjana Teknik Informatika)*, 8(2), 1–10. <https://doi.org/10.12928/jstie.v8i1.14623>
- DATA ONLINE - PUSAT DATABASE - BMKG. (n.d.). Retrieved January 17, 2022, from <https://dataonline.bmkg.go.id/home>
- Fajri, M. S., Septian, N., & Sanjaya, E. (2020). Evaluasi Implementasi Algoritma Machine Learning K-Nearest Neighbors (kNN) pada Data Spektroskopi Gamma Resolusi Rendah. *Al-Fiziya: Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics*, 3(1), 9–14. <https://doi.org/10.15408/fiziya.v3i1.16180>

- Hariani, S. (2020). Sistem Deteksi Kondisi Cuaca Berdasarkan Analisis Histogram HCL Menggunakan Algoritma k-Nearest Neighbor (KNN). *Jurnal EECCIS*, 14(1), 27–30. <https://doi.org/10.21776/jeeccis.v14i1.626>
- Lareno, B., & Muhammad. (2015). Modified Nearest Neighbor Untuk Prediksi Curah Hujan. *Proceedings Konferensi Nasional Sistem Dan Informatika (KNS&I)*, 272–277.
- Martia Nanda, D., Hendro Pudjiantoro, T., Nurul Sabrina, P., & Yani, A. (n.d.). *Metode K-Nearest Neighbor (KNN) dalam Memprediksi Curah Hujan di Kota Bandung*. 387. <https://doi.org/10.31284/p.snestik.2022.2750>
- Mohamed, S., & Ezzati, A. (2019). A data mining process using classification techniques for employability prediction. *Indonesian Journal of Electrical Engineering and Computer Science*, 14(2), 1025–1029. <https://doi.org/10.11591/ijeecs.v14.i2.pp1025-1029>
- Muhandhis, I., Ritonga, A. S., & Murdani, D. M. H. (2021). Implementasi Metode Inferensi Fuzzy Tsukamoto Untuk Memprediksi Curah Hujan Dasarian Di Sumenep. *Jurnal Ilmiah Edutic*, 8(1), 1–10. <https://doi.org/10.21107/edutic.v8i1.8907>
- Mulyati, S., Husein, S. M., & Ramdhan. (2020). Rancang Bangun Aplikasi Data Mining Prediksi Kelulusan Ujian Nasional Menggunakan Algoritma (KNN) K-Nearest Neighbor dengan Metode Euclidean Distance pada SMPN 2 Pagedangan. *Jurnal Teknik Informatika (JIKA) Universitas Muhammadiyah Tangerang*, 4(1), 65–73. <https://doi.org/10.31000/jika.v4i1.2288>
- Probabilistik Curah Hujan 20 mm (tiap 24 jam) | BMKG*. (n.d.). Retrieved January 21, 2023, from <https://www.bmkg.go.id/cuaca/probabilistik-curah-hujan.bmkg>
- Rangkuti, M. Y. R., Alfansyuri, M. V., & Gunawan, W. (2021). Penerapan Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Dan Menghitung Tingkat Akurasi Data Cuaca Di Indonesia. *Hexagon Jurnal Teknik Dan Sains*, 2(2), 11–16. <https://doi.org/10.36761/hexagon.v2i2.1082>
- Rofiq, H., Pelangi, K. C., & Lasena, Y. (2020). Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma K Nearest Neighbor (KNN). *MISI (Jurnal Manajemen Informatika Dan Sistem Informasi)*, 3(1), 8–15. <https://doi.org/10.36595/misi.v3i1.125>
- Subagyo, A. (2015). Cuaca Panas Berpengaruh terhadap Terjadinya Kebakaran di Perumahan Padat Penduduk. *Orbith: Majalah Ilmiah Pengembangan Rekayasa Dan Sosial*, 11(3), 153–160. <https://doi.org/10.32497/orbith.v11i3.88>

- Syarifaturrahmah, Uperiati, A., & Suswaini, E. (2022). Penerapan Metode K-Nearest Neighbor (KNN) Untuk Klasifikasi Cuaca Di Wilayah Tanjungpinang. *Student Online Journal (SOJ) UMRAH - Teknik*, 2(1), 250–256.
- Utami, A. S., Rini, D. P., & Lestari, E. (2021). Prediksi Cuaca di Kota Palembang Berbasis Supervised Learning Menggunakan Algoritma K-Nearest Neighbour. *JUPITER (Jurnal Penelitian Ilmu Dan Teknologi Komputer)*, 13(1), 9–18. <https://jurnal.polsri.ac.id/index.php/jupiter/article/view/3243><https://jurnal.polsri.ac.id/index.php/jupiter/article/download/3243/1390>
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, KNN Dan SVM. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>