E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

Algoritma *k-Nearest* Neighbor Classifier untuk Prediksi Curah Hujan di Kabupaten Bandung

Venia Restreva Danestiara
Fakultas Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia
Email: veniarestreva@unibi.ac.id

Abstrak

Implementasi algoritma *k-Nearest Neighbor Classifier* dengan teknik *Data Mining* pada prediksi curah hujan dapat membantu mayoritas masyarakat di Kabupaten Bandung yang memiliki mata pencaharian di bidang pertanian. Himpunan data berasal dari BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) Bandung dengan jangkauan waktu antara Januari 2005 hingga Desember 2016 yang terdiri dari enam atribut, yaitu: penyinaran matahari, kelembapan, angin, temperatur, uap dan curah hujan. Hasil pencarian tetangga terdekat adalah atribut curah hujan dari lima kategori kelas, yaitu: tidak hujan, hujan ringan, hujan sedang, hujan lebat dan hujan sangat lebat, yang dihitung menggunakan *euclidean distance*. Penelitian ini melakukan pengujian akurasi dengan tiga partisi himpunan data dan penggunaan sepuluh nilai *k* yang berbeda. Akurasi-akurasi yang didapatkan sangat baik dengan akurasi tertinggi sebesar 97,92% untuk nilai *k*={3,5} pada partisi himpunan data latih dan uji masing-masing 50%. Nilai *k*, jumlah himpunan data dan data latih berpengaruh dalam pencarian tetangga terdekat.

Kata Kunci: *k-Nearest Neighbor Classifier*, Prediksi Curah Hujan, *Data Mining*, Kabupaten Bandung.

Abstract

Implementation of the k-Nearest Neighbor Classifier algorithm with Data Mining techniques in predicting rainfall can help the majority of people in Bandung Regency who have currency in agriculture. The data set comes from BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) Bandung *with a range of time from January 2005 to December 2016 which consists of six attributes, namely: solar radiation, humidity, wind, temperature, vapor and rainfall. The closest neighbor search result is the rainfall attribute from five class categories, namely: no rain, light rain, moderate rain, heavy rain and very heavy rain, which are calculated using the Euclidean distance. This study conducted an accuracy test with three partition data sets and the use of ten different k values. The accuracies obtained were very good with the highest accuracy of* 97.92% *for the value of* k={3.5} *in the partition set training data and the test was* 50% *each. The value of k, the number of data sets and training data have an effect on the search for nearest neighbors.*

Keywords: *k-Nearest Neighbor Classifier, Rainfall Prediction, Data Mining,* Bandung *Regency.*

1 PENDAHULUAN

Sebagian besar masyarakat di Kabupaten Bandung memiliki mata pencaharian pada sektor agraris yang dipengaruhi oleh kondisi curah hujan. Curah hujan memacu kecepatan perkembangan individu hama dan penyakit sehingga memiliki andil yang sangat besar dalam pertumbuhan dan produksi tanaman. Prakiraan curah hujan dapat membantu perencanaan kegiatan pertanian agar sesuai

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

dengan harapan dan terhindar dari kegagalan panen. Dengan adanya prakiraan curah hujan diharapkan dapat membantu masyarakat di Kabupaten Bandung untuk mengantisipasi perubahan curah hujan yang mempengaruhi tingkat produktifitas pertanian.

Prakiraan curah hujan sangat kompleks karena banyak aspek berbeda yang mempengaruhi dan menimbulkan kondisi tertentu sehingga perlu dicari suatu metode yang sesuai, cepat dan tepat. Permasalahan prediksi dalam fenomena meteorologi ini dapat diatasi dengan menggunakan teknik *Data Mining*. Oleh sebab itu, penelitian ini menggunakan salah satu teknik klasifikasi menggunakan algoritma *k-Nearest Neighbor Classifier* yang bertujuan untuk memprediksi curah hujan di Kabupaten Bandung.

Beberapa penelitian mengenai peramalan cuaca di Kabupaten Bandung telah dilakukan menggunakan algoritma *Soft Computing* seperti *Envolving Neural Network* dengan akurasi sebesar 70% dan 84.34% (Nitha *et al*., 2015), *Grammatical Evolution and Adaptive Neuro-Fuzzy Inference System* dengan akurasi 91.67% (Nitha *et al*., 2015), *Local Regression Smoothing and Fuzzy-Grammatical Evolution* dengan akurasi 84.62% (Pratama *et al*., 2016), dan *Genetic Algorithm* dengan akurasi 90% (Nitha *et al*., 2013) dan 70% (Utama *et al*., 2016).

Telah banyak dilakukan penelitian menggunakan algoritma *k-Nearest Neighbor* dengan menggunakan nilai *k* yang beragam. Nilai *k* digunakan sebagai parameter untuk menentukan jumlah tetangga terdekat dari data uji. Kelas dari data uji didapatkan berdasarkan jumlah kelas tetangga terdekat terbanyak. Beberapa penelitian menggunakan *k-Nearest Neighbor* di antaranya adalah penelitian prediksi curah hujan di Chennai dengan menggunakan 3-NN menghasilkan nilai MAPE sebesar 9.98 (Mallika *et al*., 2016), peneletian prediksi kecepatan angin jangka pendek menggunakan parameter 4-NN memperoleh nilai MAPE sebesar 9.01 (Budak *et al*., 2013), memprediksi land cover dengan citra sentinel-2 menggunakan 1-NN menghasilkan akurasi sebesar 94.32% (Noi *et al*., 2017), pada penelitian mengenai klasifikasi citra motor berdasarkan sinyal EEG dengan 15-NN menghasilkan akurasi sebesar 70.08% (Isa *et al*., 2017).

2 KAJIAN PUSTAKA

Algorima *k-Nearest Neighbor* merupakan suatu metode yang menggunakan algoritma *supervised* dengan hasil dari *query instance* yang baru diklasifikasikan menggunakan parameter *k* untuk menyatakan jumlah tetangga terdekat yang dilibatkan dalam menentukan kelas pada data uji. Dari *k* tetangga terdekat yang terpilih kemudian dilakukan *voting* untuk mendapatkan kelas dengan jumlah suara tetangga terbanyak sebagai hasil prediksi pada data uji.

2.1 Pencarian Nilai k Optimum

Nilai *k* terbaik pada algoritma *k-Nearest Neighbor* tergantung pada data yang digunakan sehingga banyak penelitian menggunakan parameter *k* yang beragam untuk mengetahui hasil prediksi terbaik. Nilai *k* biasanya kecil dan integer dengan nilai positif. Jika *k* terlalu kecil, maka hasil prediksi yang akan didapat bisa sensitif terhadap keberadaan *noise*. Sedangkan apabila *k* terlalu besar, maka tetangga terdekat yang terpilih mungkin terlalu banyak dari kelas lain yang tidak relevan karena jaraknya terlalu jauh, sehingga nilai *k* perlu diobservasi. Jika *k*=1 maka disebut juga *Nearest Neighbor*.

2.2 *Euclidean Distance*

Perhitungan ukuran ketidakmiripan yang paling sering digunakan untuk menghitung jarak di antara beberapa dimensi yang diinginkan. Cara kerja ukuran ketidakmiripan yaitu jika antara satu data dengan data yang lain semakin dekat maka semakin kecil ketidakmiripannya dan jika antara satu data dengan yang lainnya semakin jauh maka semakin besar ketidakmiripannya. Berikut adalah ukuran ketidakmiripan untuk *euclidean distance* pada data *A* yang dinyatakan oleh fungsi:

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

$$a: A \times A \rightarrow B \quad (1)$$

B merupakan set bilangan riil yang memenuhi syarat persamaan:

$$\exists a_0 \in B : -\infty < a_0 \leq a\,(c,d) < +\infty, \quad \forall c,d \in A \quad (2)$$

Nilai a adalah metrik ukuran ketidakmiripan. Nilai a_0 adalah nilai *level* terendah ketidakmiripan yang bisa didapat dua vektor dalam *A* ketika nilai keduanya identik. Nilai *c* adalah nilai sumbu horizontal dan nilai *d* adalah nilai sumbu vertikal dalam began .

Euclidean distance sebagai salah satu ukuran ketidakmiripan diformulasikan dalam persamaan berikut:

$$a_2(c,d) = \|c-d\|_2 = \sqrt{\sum_{i=1}^{e}(c_i - d_i)^2} \quad (3)$$

$c, d \in A$. Nilai c_i dan d_i adalah nilai fitur ke-i dari *c* dan *d*. Nilai *e* adalah jumlah fitur dalam vektor. Ukuran ketidakmiripan pada *euclidean distance* akan memberikan nilai $d_0 = 0$, maka jarak minimal yang mungkin di antara data-data bernilai 0. Jarak dari *c* ke *d* akan sama dengan *d* ke *c*, $a(c,d) = a(d,c)$.

3 METODE PENELITIAN

Secara umum, penelitian ini memprediksi curah hujan pada bulan di tahun mendatang menggunakan informasi yang ada pada bulan-bulan di tahun-tahun sebelumnya. Data yang digunakan memiliki jangkauan waktu antara Januari 2005 hingga Desember 2016. Pada kasus yang sederhana, jika akan memprediksi curah hujan pada bulan Maret 2011 maka bulan tersebut menjadi data uji dan nilai pada atribut-atribut adalah ciri uji, dengan menggunakan data latih dari Januari 2005 hingga Februari 2011 diharapkan dapat memprediksi curah hujan pada Maret 2011 dengan mencari tetangga terdekat dari titik-titik ciri uji pada titik-titik ciri latih yang terdapat pada data latih. Rancangan sistem dari penelitian ini tertera pada Tabel 1.

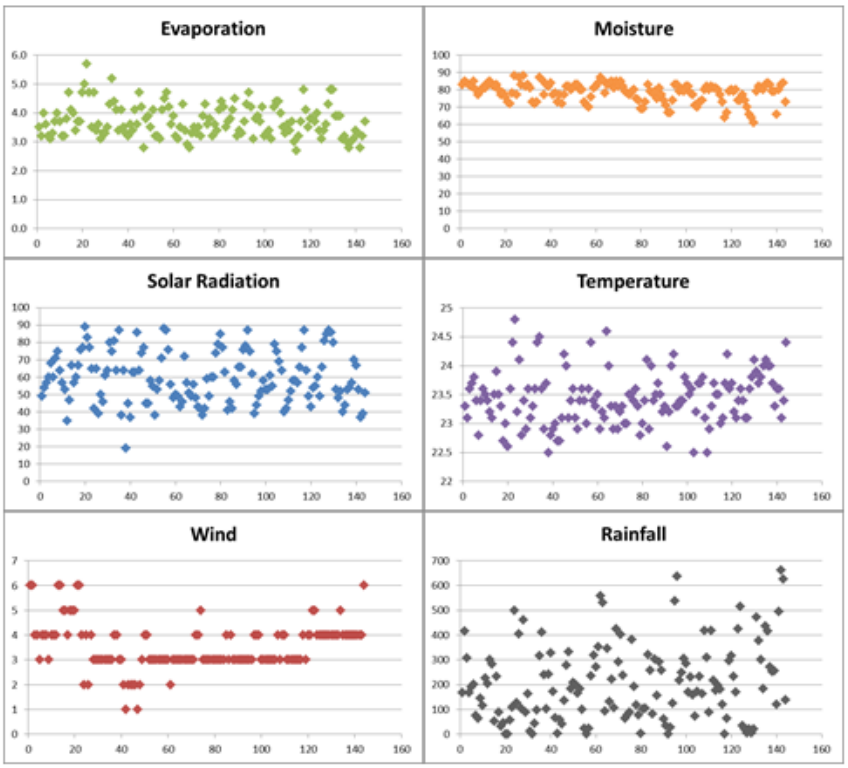
3.1 Himpunan Data Cuaca

Data yang digunakan untuk penelitian berupa data cuaca per bulan selama 12 tahun dimulai dari Januari 2005 – Desember 2017 dengan 864 *record* untuk Kabupaten Bandung. Data cuaca didapatkan dari BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) Bandung. Pada data tersebut terdapat enam buah variabel, yaitu: Temperatur (C) Curah Hujan (mm), Lama Penyinaran Matahari (%), Lembab Nisbi (%), Kecepatan Angin (Knot) dan Penguapan (mm). Sebaran data digambarkan pada Gambar 1.

Tabel 1. Rancangan Sistem Algoritma *k-Nearest Neighbor Classifier*

Rancangan Sistem
***Input* :**
Himpunan data cuaca BMKG Kabupaten Bandung tahun 2005-2016
***Output* :**
Prdiksi Curah Hujan
Akurasi Prediksi Curah Hujan
***Preprocessing* :**
Diskritisasi Data
Data Partisi
Learning Algoritma *k-Nearest Neighbor Classifier* :
Penentuan Parameter *k*
Perhitungan Jarak Menggunakan *euclidean distance*
Pengkuran Perfomansi Pada Klasifikasi :
Confussion Matrix, Precission, Recall, Fscore

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023



Gambar 1. Sebaran Data Cuaca

3.2 Himpunan Data Cuaca

Preprocessing data merupakan tahapan awal yang dilakukan untuk mengolah data yang belum sesuai dengan bentuk data yang diharapkan, seperti: *incomplete*, *noisy* dan *inconsistent* dengan tujuan dilakukannya penyeragaman untuk mempermudah kegiatan pemrosesan data. Pada data cuaca di Kabupaten Bandung dilakukan diskritisasi pada atribut Curah Hujan untuk mengubah nilai kontinu menjadi kategorikal dengan acuan pembagian data berdasarkan infomasi yang didapat dari BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) pada Tabel 2. Hal ini dilakukan untuk memperoleh label kelas prediksi.

Kemudian, dilakukan pembagian data menjadi data latih dan data uji. Hal ini dilakukan dengan tujuan untuk mendapatkan analisis akurasi berdasarkan jumlah data yang diklasifikasi. Berikut adalah data partisi pada Tabel 3.

Tabel 2. Kategori Curah Hujan

No.	Kategori Curah Hujan
1	Curah Hujan <= 0.1mm, Tidak Hujan
2	0.1<= 20mm, Hujan Ringan
3	20< Curah Hujan <= 50mm, Hujan Sedang
4	50< Curah Hujan <= 100mm, Hujan Lebat
5	Curah Hujan > 100mm, Hujan Sangat Lebat

Tabel 3. Partisi Data

No.	Partisi Data	
	Data Latih	Data Uji
1	Jan 2005 – Des 2010	Jan 2011 – Des 2016
2	Jan 2009 – Des 2012	Jan 2013 – Des 2016
3	Jan 2013 – Des 2014	Jan 2015 – Des 2016

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

Pada partisi pertama data latih dan data uji masing-masing menggunakan jumlah data sebanyak 6 tahun. Pada partisi kedua menggunakan jumlah data masing-masing 3 tahun dan partisi ketiga menggunakan jumlah data masing-masing 1 tahun.

3.3 *Learning Algoritma k-Nearest Neighbor*

Metode *k-Nearest Neighbor Classifier* digunakan untuk klasifikasi yang merupakan metode *non-parametric*. Algoritma ini berbasis *Nearest Neighbor* yang termasuk ke dalam algoritma *lazy learning*. Alur *k-Nearest Neighbor* pada penelitian pada Gambar 2.

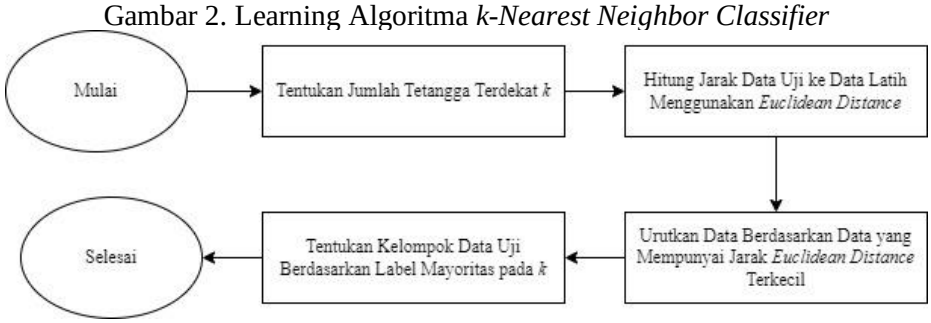
a. Penentuan Paramerer *k*
   Nilai *k* yang digunakan yaitu: *k* = {1,3,5,7,9,11,13,15,17,19}. Berdasarkan hasil pengujian yang dilakukan terlihat bahwa nilai *k* sangat berpengaruh terhadap akurasi yang dihasilkan seperti ditunjukkan pada Tabel 4 dan Tabel 5.

b. Perhitungan Jarak Menggunakan *Euclidean Distance*
   Mencari jarak pada tetangga-tetangga pada data uji menggunakan metode *euclidean distance* yang bekerja untuk mencari kedekatan berdasarkan ukuran ketidaksamaan. Jika semakin kecil nilainya maka jarak antara dua data semakin dekat. Apabila nilai *k* kecil maka hanya tetangga terbaiklah yang akan menjadi hasil dari klasifikasi.

3.4 Pengkuran Perfomansi Pada Klasifikasi

Ketepatkan klasifikasi dapat dievaluasi dengan menghitung jumlah hasil prediksi bernilai c terhadap data aktual yang bernilai c dan juga terhadap bernilai d, serta hasil prediksi bernilai d terhadap data aktual bernilai c dan juga bernilai d. Keempat perhitungan tersebut merupakan *confusion matrix* yang direpresentasikan pada Tabel 4. Dengan menggunakan *confusion matrix* dapat terlihat dengan jelas hasil dari klasifikasi. f_{cc} dan f_{dd} merupakan jumlah hasil prediksi yang sesuai dengan kelas aktual. f_{cd} dan f_{dc} merupakan jumlah hasil prediksi yang tidak sesuai dengan kelas akual, hasil prediksi bernilai c namun data aktual bernilai d begitu pula sebaliknya. Pengukuran performansi pada klasifikasi dilakukan dengan mengevaluasi pada fokus yang berbeda. Beberapa pengukuran performansi berdasarkan nilai pada *confusion matrix* terdapat pada Tabel 5.

Gambar 2. Learning Algoritma *k-Nearest Neighbor Classifier*



Tabel 4. *Confusion Matrix*

f_{ab}		Kelas Prediksi (b)	
		Kelas = c	Kelas = d
Kelas Aktual (a)	Kelas = c	f_{cc}	f_{cd}
	Kelas = d	f_{dc}	f_{dd}

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

Tabel 5. Pengukuran Transformasi

Pengukuran Performansi	Rumus	Fokus Evaluasi
Akurasi	$\frac{f_{cc} + f_{dd}}{f_{cc} + f_{dc} + f_{cd} + f_{dd}}$	Efektivitas keseluruhan prediksi
Presisi	$\frac{f_{cc}}{f_{cc} + f_{dc}}$	Ketepatan prediksi kelas yang dilakukan oleh klasifikasi
Recall (*Sensitivity*)	$\frac{f_{cc}}{f_{cc} + f_{cd}}$	Efektivitas klasifikasi untuk mengidentifikasi kelas prediksi
F-score	$2 \times \frac{Presisi \times Recall}{Presisi + Recall}$	Performa sistem dalam melakukan klasifikasi

4 HASIL DAN PEMBAHASAN

Cara memprediksi cuaca dilakukan dengan menguji data set berdasarkan pada data partisi yang terdapat pada Tabel 3.

4.1 Skenario Eksperimen

Penelitian dilakukan berdasarkan diskritisasi pada kategori curah hujan yang terdapat pada Tabel 2.

4.2 Hasil Eksperimen

Akurasi dan *error rate* pada data uji dan data latih setiap partisi menggunakan *k*={1,3,5,7,9,11,13,15,17,19} yang tercantum pada Tabel 6.

Hasil pengujian menghasilkan nilai akurasi maksimum yang berbeda pada setiap partisi. Pada partisi pertama, kedua dan ketiga memperoleh nilai akurasi masing-masing sebesar 97.22% pada *k*={7,9}, sebesar 97.92% pada *k*={3,5}, sebesar 87.5 pada $\underline{k}$=1. Nilai akurasi yang berbeda disebabkan oleh berbagai faktor, berikut analisis dari faktor-faktor tersebut:

Tabel 6. Hasil Eksperimen

Parameter *k*	Data Latih			Data Uji		
	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3
1	100%	100%	100%	95.83%	95.83%	87.5%
3	95.83%	95.83%	95.83%	94.44%	97.92%	75%
5	95.83%	95.83%	95.83%	95.83%	97.92%	75%
7	95.83%	91.67%	91.67%	97.22%	91.67%	75%
9	95.83%	91.67%	83.33%	97.22%	91.67%	75%
11	90.28%	93.75%	83.33%	91.67%	85.42%	75%
13	91.67%	85.42%	83.33%	91.67%	85.42%	75%
5	91.67%	85.42%	83.33%	91.67%	85.42%	75%
17	90.28%	83.33%	83.33%	91.67%	83.33%	75%
19	81.94%	83.33%	83.33%	83.33%	83.33%	75%

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

a. Pengaruh parameter *k*

Nilai akurasi cenderung mengalami penurunan ketika parameter *k* semakin besar. Semakin kecil parameter *k* berarti semakin sedikit nominasi jumlah tetangga yang digunakan dalam memperoleh proses klasifikasi data baru. Apabila parameter *k* besar, maka semakin banyak tetangga yang digunakan untuk proses klasifikasi, sehingga hasil prediksi kurang maksimal karena data yang label kelasnya berjauhan dengan data yang diprediksi akan diikutsertakan dalam nominasi untuk menentukan label kelas pada data uji tersebut. Selain itu, parameter *k* yang besar menimbulkan biaya komputasi yang tinggi karena diperlukan banyak perhitungan jarak dari setiap ciri latih terhadap ciri uji.

Parameter *k* maksimum berubah pada setiap partisi yang dilakukan. Penentuan nilai *k* maksimum merupakan hal yang sulit untuk dilakukan karena hasil prediksi bergantung berdasarkan sebaran data dengan label kelas tertentu yang bertetanggaan dengan data yang diuji.

Nilai akurasi yang sama dapat terjadi pada parameter *k* yang besarnya berdekatan seperti salah contohnya pada hasil prediksi dengan menggunakan curah hujan pada partisi kedua dengan parameter *k*={3,5} memperoleh akurasi sebesar 95.83% dan parameter *k*={7,9} memperoleh akurasi sebesar 91.67%.

b. Pengaruh jumlah data latih dan data uji

Jumlah data latih berpengaruh terhadap akurasi pengujian. Semakin banyak data latih maka semakin banyak ciri latih yang diikutsertakan dalam proses klasifikasi sehingga memiliki peluang yang lebih besar untuk memprediksi secara akurat. Pada partisi pertama dengan jumlah total data latih dan data uji sebesar 12 tahun pada hasil pengujian curah hujan selalu mendapatkan nilai akurasi yang lebih besar dibandingkan dengan partisi kedua dengan total jumlah data latih dan data uji sebesar 6 tahun.

Pengaruh jumlah data latih dan data uji yang digunakan terhadap akurasi berbanding lurus dengan penentuan parameter *k*. Penggunaan jumlah data uji dan data latih yang semakin besar akan berdampak pada nilai selisih pada nilai akurasi antar parameter *k* yang berdekatan menjadi lebih kecil. Hal tersebut dibuktikan dengan perbandingan antara partisi pertama dan partisi kedua pada hasil pengujian, nilai selisih terbesar pada partisi pertama sebesar 8.34%, sedangkan pada partisi kedua sebesar 8.33%.

4.3 Hasil dan Evaluasi Performa

Data partisi pertama menggunakan *k*=7 memproleh akurasi maksimum. Berikut adalah *confusion matrix* dan hasil evaluasi performa tersebut tertera pada Tabel 7 dan Tabel 8.

Tabel 7. *Confusion Matrix* Pada Partisi Pertama Dengan Menggunakan 7NN

Kategori	**Tidak Hujan**	**Hujan Ringan**	**Hujan Sedang**	**Hujan Lebat**	**Hujan Sangat Lebat**	
Tidak Hujan	0	1	0	0	0	1
Hujan Ringan	0	5	0	0	0	5
Hujan Sedang	0	0	5	0	0	5
Hujan Lebat	0	0	0	9	0	9
Hujan Sangat Lebat	0	0	0	1	51	52
	0	6	5	10	51	

Tabel 8. Evaluasi Performa

Akurasi	**Presisi**	**Recall**	**Fscore**
97.22%	0.76	1	0.864

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

5 SIMPULAN

Berdasarkan penelitian yang telah dilakukan maka didapatkan kesimpulan bahwa metode *k-Nearest Neighbor* dapat d2mplementasikan untuk klasifikasi himpunan data Cuaca di Kabupaten Bandung. Dengan akurasi terbesar 97,92% dengan menggunakan parameter *k*={3,5} pada data partisi kedua. Pada penelitian ini nilai *k* dan jumlah data latih berpengaruh dalam mendapatkan nilai kelas terdekat. Hal ini dibuktikan dengan kenaikan akurasi hampir pada setiap percobaan. Penelitian ini dapat dikembangkan dengan menambahkan himpunan data curah hujan dan membuat kategori curah hujan yang berbeda.

DAFTAR PUSTAKA

Han, J., Pei, J., & Tong, H. (2022). . *Data Mining: Concepts and Techniques, 3rd ed*. The Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers.

Isa, N. E. Z. M., Amir, A., Ilyas, M. Z., & Razalli, M. S. (2017). The performance analysis of K-nearest neighbors (K-NN) algorithm for motor imagery classification based on EEG signal. In *MATEC web of conferences* (Vol. 140, p. 01024). EDP Sciences.

Larose, D. T. (2005). An introduction to data mining. *Traduction et adaptation de Thierry Vallaud*.

Mallika, M., & Nirmala, M. (2016). Chennai annual rainfall prediction using k-nearest neighbour technique. *International Journal of Pure and Applied Mathematics*, *109*(8), 115-120.

Nhita, F. (2013, March). A rainfall forecasting using fuzzy system based on genetic algorithm. In *2013 International Conference of Information and Communication Technology (ICoICT)* (pp. 111-115). IEEE.

Nhita, F., Annisa, S., & Kinasih, S. (2015, May). Comparative study of grammatical evolution and adaptive neuro-fuzzy inference system on rainfall forecasting in Bandung. In *2015 3rd International Conference on Information and Communication Technology (ICoICT)* (pp. 6-10). IEEE.

Nhita, F., Saepudin, D., & Wisesty, U. N. (2015, December). Comparative study of moving average on rainfall time series data for rainfall forecasting based on evolving neural network classifier. In *2015 3rd International symposium on computational and business intelligence (ISCBI)* (pp. 112-116). IEEE.

Nhita, F., Wisesty, U. N., & Ummah, I. (2015). Planting calendar forecasting system using evolving neural network. *Far East Journal of Electronics and Communications*, *14*(2), 81.

Prasetyo, E. (2012). Data mining konsep dan aplikasi menggunakan matlab. *Yogyakarta: Andi*, *1*.

Pratama, S. W., & Nhita, F. (2016, May). Implementation of local regression smoothing and fuzzy-grammatical evolution on rainfall forecasting for rice planting calendar. In *2016 4th International Conference on Information and Communication Technology (ICoICT)* (pp. 1-5). IEEE.

Putra, B. U., Nhita, F., Saepudin, D., & Wisesty, U. N. (2017, September). An implementation of weighted moving average and genetic programming for rainfall forecasting in Bandung Regency. In *2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC)* (pp. 169-173). IEEE.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, *45*(4), 427-437.

E-ISSN: 2655-867X
P-ISSN: 2655-8661
Volume 5 No. 01 Februari 2023

Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, *18*(1), 18.

Yesilbudak, M., Sagiroglu, S., & Colak, I. (2013). A new approach to very short term wind speed prediction using k-nearest neighbor classification. *Energy Conversion and Management*, *69*, 77-86.

Zhang, S., Li, X., Zong, M., Zhu, X., & Cheng, D. (2017). Learning k for knn classification. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *8*(3), 1-19.