

Fuzzy K-Nearest Neighbor to Predict Rainfall in Padang Pariaman District

Annisa Rizki Amalia, Nonong Amalita*, Yenni Kurniawati, Zamahsary Martha

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: nongmat@fmipa.unp.ac.id

Submitted : 14 Oktober 2023

Revised : 02 Februari 2024

Accepted : 23 Februari 2024

ABSTRACT

Information about rainfall levels at a time and in a region is very important because rainfall influences human activities. Rainfall is the amount of water that falls to the earth in a certain period of time, measured in millimeters. One piece of information related to rainfall is daily rainfall predictions. In this study, an attempt was made to classify daily rainfall at the Padang Pariaman climatology station into 5 categories, namely very light rain, light rain, moderate rain, heavy rain and very heavy rain. There are 4 weather parameters used, namely air temperature, humidity, wind speed and duration of sunlight. One of the methods used to predict rainfall is data mining, a computer learning to analyze data automatically thus obtaining a perfect new model. One of the best prediction algorithms in data mining is Fuzzy K-Nearest Neighbor (FK-NN). FK-NN uses the largest membership degree value of the test data in each class to predict the class. The number of sample classes for rainfall data in Padang Pariaman Regency has an imbalance class. To overcome the imbalance class, Synthetic Minority Over-sampling Technique (SMOTE) method is used to generate minority data as much as majority data. The results of this study by using FK-NN classification with 343 test data, parameters $K = 12$, and euclidean distance is quite good at the accuracy level of 76,38%.

Keywords: Classification, Data Mining, Fuzzy K-Nearest Neighbor, Rainfall



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Besar atau kecilnya curah hujan tidak dapat ditentukan secara pasti, akan tetapi dapat diperkirakan atau diprediksi. Prediksi curah hujan dilakukan untuk melihat dan menganalisa gambaran curah hujan yang akan terjadi beberapa waktu kedepan. Prediksi curah hujan sangat penting dilakukan agar dapat membantu kegiatan para nelayan, petani, masyarakat, pemerintahan, serta pilot. Pada tahun 2023 sudah terjadi beberapa kali banjir bandang di Kabupaten Padang Pariaman, diantaranya pada bulan Juli dan bulan Agustus. Banjir terjadi dikarenakan cuaca ekstrim dan curah hujan yang tinggi. Curah hujan yang tinggi mengakibatkan banjir hingga tanah longsor yang berdampak pada terhambatnya kegiatan masyarakat, rusaknya fasilitas umum dan rumah warga, kerugian di sektor pertanian dan rusaknya pipa penyediaan air minum (Antaraneews, 2023).

Besar kecilnya jumlah curah hujan dipengaruhi oleh beberapa faktor, seperti suhu, kelembapan, lama penyinaran, dan kecepatan angin (Wilson, 1993). Curah hujan tertinggi di Sumatera Barat menurut Badan Pusat Statistik pada tahun 2021 berada di Kabupaten Padang Pariaman. Jumlah curah hujan keseluruhan di Kabupaten Padang Pariaman pada tahun 2021 sebesar 5.332 millimeter. Kabupaten Padang Pariaman memiliki rata-rata hari hujan sebanyak 18 hari per bulannya dan musim kering yang pendek karena daerah lautan yang dipengaruhi oleh angin laut. Kecepatan angin rata-rata Kabupaten Padang Pariaman yaitu 2,14 knot dan kelembaban relatif sebesar 86,75% dengan suhu udara berkisar antara 24,4°C – 25,7°C. Data curah hujan yang diperoleh akan diolah untuk menentukan tingkat curah hujan yang akan terjadi. Namun data curah hujan terlalu besar untuk diolah secara manual. Salah satu cara yang dapat digunakan untuk mengolah data tersebut adalah dengan klasifikasi. Klasifikasi adalah proses mengelompokkan data ke dalam beberapa kelas berdasarkan sifat atau pola yang terdapat dalam data latih. Salah satu metode klasifikasi yang dapat digunakan adalah K-Nearest Neighbor (K-NN).

Metode *K-Nearest Neighbor* (K-NN) merupakan salah satu algoritma terbaik untuk prediksi dalam penggalian data pada *data mining* (Wu dkk, 2008). Metode K-NN menggunakan konsep perhitungan jarak terdekat dengan sebuah titik menggunakan rumus *euclidean distance*. *Euclidean distance* didapatkan dengan menghitung akar kuadrat dari jumlah selisih kuadrat antara dua titik berbeda, sehingga jarak merupakan komponen inti dalam klasifikasi K-NN (Prasath dkk, 2019). Kelemahan dari algoritma K-NN yaitu sejumlah K tetangga tertentu dianggap sama pentingnya meskipun tidak pasti. Oleh karena itu, Keller mengemukakan suatu metode baru untuk mengatasi permasalahan ini, yaitu dengan menggunakan *Fuzzy K-Nearest Neighbor* (FK-NN) (Keller dkk, 1985). Dataset

diklasifikasikan bertujuan agar analisis data lebih mudah dan mengetahui kondisi secara umum curah hujan di Kabupaten Padang Pariaman.

Penelitian sebelumnya mengenai FK-NN pernah diterapkan oleh Nugraha dkk (2017) dengan menerapkan *Fuzzy K-Nearest Neighbor* dalam penentuan status gizi balita. Pada penelitian tersebut algoritma FK-NN dapat menghasilkan performa klasifikasi yang paling baik. Performa terbaik terjadi pada saat nilai $K=4$ dengan nilai akurasi sebesar 84,37%. Sedangkan penelitian tentang prediksi curah hujan telah dilakukan oleh Nanda dkk (2022) dengan menggunakan metode klasifikasi K-NN. Ada 4 parameter yang digunakan dan 5 kelas output yaitu tingkat curah hujan yang terjadi. Penelitian tersebut memperoleh hasil akurasi sebesar 86,199% dengan hasil uji akurasi menggunakan confusion matrix menghasilkan akurasi 84,38%. Oleh karena itu, berdasarkan permasalahan tersebut maka tujuan dari penelitian ini adalah memprediksi curah hujan di Kabupaten Padang Pariaman menggunakan algoritma *Fuzzy K-Nearest Neighbor*.

II. METODE PENELITIAN

Penelitian ini merupakan penelitian terapan dengan menerapkan algoritma *K-Nearest Neighbor* dan metode *fuzzy* dalam memprediksi curah hujan di Kabupaten Padang Pariaman menggunakan data sekunder. Data bersumber dari *website* Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data tersebut didapatkan berdasarkan pemantauan Stasiun Klimatologi Sicincin, Kabupaten Padang Pariaman pada pengamatan 1 Januari 2021–31 Agustus 2023 dengan jumlah data sebanyak 973 pengamatan dan dengan variabel penelitian seperti Tabel 1.

Tabel 1. Variabel Penelitian

Simbol	Variabel	Kategori	Tipe Data
Y	Curah Hujan (mm)	1: Berawan	Kategorik
		2: Ringan	
		3: Sedang	
		4: Lebat	
		5: Sangat Lebat	
		6: Ekstrem	
X_1	Suhu ($^{\circ}\text{C}$)	-	Numerik
X_2	Rata - rata Kelembapan (%)	-	Numerik
X_3	Lama Penyinaran Matahari (Jam)	-	Numerik
X_4	Kecepatan Angin (m/s)	-	Numerik

Langkah analisis menggunakan algoritma FK-NN pada penelitian ini adalah sebagai berikut:

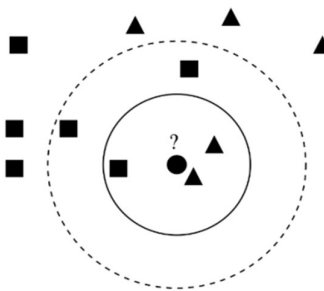
1. Menyiapkan *dataset* yang akan diteliti, yaitu data curah hujan Kabupaten Padang Pariaman periode 1 Januari 2021 sampai 31 Agustus 2023.
2. Melakukan analisis deskriptif untuk melihat gambaran secara umum karakteristik masing-masing variabel pada *dataset*.
3. Sebelum melakukan analisis menggunakan algoritma FK-NN perlu dilakukan *pre-processing* data. *Pre-processing* bertujuan untuk mengatasi data hilang, normalisasi, kesalahan dalam penginputan, dan permasalahan-permasalahan lainnya yang dapat mengganggu proses kerja maupun hasil analisis dari algoritma FK-NN (García dkk, 2015). Proses yang dilakukan pada tahapan ini adalah normalisasi data dan penanganan data tidak seimbang menggunakan SMOTE. Proses normalisasi data merupakan suatu tahapan yang dilakukan dengan penskalaan terhadap data menjadi rentang nilai tertentu. Metode normalisasi yang digunakan adalah *min-max normalization* yang diperoleh menggunakan Persamaan (1).

$$x' = \frac{(x - \min(x))}{\text{Range}(x)} \quad (1)$$

dimana x' merupakan hasil normalisasi, x nilai objek tiap variabel yang akan dinormalisasikan dengan $\min(x)$ adalah nilai minimum dari suatu objek dan $\text{range}(x)$ adalah selisih dari $\max(x) - \min(x)$.

Selanjutnya proses SMOTE merupakan suatu teknik yang dapat diterapkan jika *dataset* memiliki *imbalance class*. Menurut He (2009), *imbalance class* adalah suatu kondisi dimana jumlah data antar kelas yang terdapat dalam *dataset* tidak merata karena salah satu diantara kelasnya memiliki jumlah data yang sangat banyak (kelas mayoritas) dibandingkan dengan kelas yang lainnya (kelas minoritas). Perbedaan besar dalam jumlah data antar kelas ini dapat menyebabkan model tidak mampu memprediksi kelas minoritas secara tepat atau akurat sehingga sejumlah besar data uji yang seharusnya berada pada kelas minoritas tidak dapat diprediksi dengan baik oleh model klasifikasinya (Sun dkk, 2009). Permasalahan data seperti ini dapat diatasi menggunakan metode pengambilan sample atau biasa dikenal dengan *sampling*. Metode pengambilan sampel

- ini bekerja dengan memvariasikan penyebaran dari data antara kelas mayoritas dengan kelas minoritas pada data latih agar jumlah data menjadi seimbang disetiap kelas (Nguyen, 2009). Chawla dkk (2002) mengatakan bahwa metode *sampling* yang sering digunakan salah satunya adalah *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE merupakan metode yang membangkitkan data minoritas sebanyak jumlah data mayoritas (Ruangthong dan Jaiyen, 2015). Pada penelitian ini data mayoritas terdapat pada data dengan sampel kelas berawan dan data minoritas terdapat pada data dengan kelas sangat lebat dan ekstrim. Jumlah sampel yang digunakan pada penelitian ini sebanyak 3.426 dengan jumlah sampel yang sama setiap kelas.
- Membagi *dataset* menjadi dua bagian yaitu 90% data latih untuk membentuk pola data dan 10% data uji untuk menguji performa dari algoritma yang digunakan. Data yang digunakan terdiri dari 3.426 amatan dengan 3.083 amatan untuk data latih dan 343 amatan untuk data uji.
 - Menentukan parameter K yang digunakan.
Menurut Angreni dkk (2019) parameter K menyatakan seberapa banyak jumlah *neighbor* atau data yang terdekat dan nilai K terbaik tergantung pada data. Wu dkk (2008) menyatakan bahwa nilai parameter K yang paling diinginkan adalah nilai yang jaraknya terkecil. Berikut contoh penerapan parameter K pada Gambar 1.



Gambar 1. Ilustrasi Penggunaan Parameter K

Pada ilustrasi Gambar 1 *dataset* memiliki dua label kelas pada variabel respon, yaitu persegi dan segitiga. Sampel uji yang ingin diklasifikasikan disimbolkan dengan lingkaran. Kriteria klasifikasi pada algoritma K-NN yaitu mayoritas kategori pada K tetangga terdekat berdasarkan kemiripan atau jarak terkecil. Berdasarkan Gambar 1 jika parameter $K = 3$ maka hasil klasifikasi pada lingkaran adalah segitiga, sedangkan, jika parameter $K = 5$ maka hasil klasifikasi pada lingkaran adalah persegi.

- Menghitung jarak kedekatan antara data uji dan data latih menggunakan Persamaan (2).
Untuk perhitungan jarak terdekat digunakan fungsi *euclidean distance*. *Euclidean distance* merupakan suatu metode dalam perhitungan jarak. *Euclidean distance* digunakan untuk mengukur jarak dari dua buah titik dalam ruang dua dimensi antara nilai X_1, X_2, X_3 dan X_4 . Adapun perhitungan jarak menggunakan *euclidean distance* menggunakan Persamaan (2) (Barmer, 2007).

$$d_{euc}(A, B) = \sqrt{\sum_{i=1}^p (a_i - b_i)^2} \quad (2)$$

- Melakukan Perhitungan *fuzzy* untuk mencari nilai keanggotaan setiap kelas dengan Persamaan (3).

$$\mu_{ij} = \begin{cases} 0,51 + \left(\frac{n_j}{n}\right) \times 0,49, & \text{jika } j = i \\ \left(\frac{n_j}{n}\right) \times 0,49, & \text{jika } j \neq i \end{cases} \quad (3)$$

dimana μ_{ij} merupakan nilai keanggotaan kelas i pada vektor j , n_j merupakan jumlah anggota kelas j pada suatu data latih n , n adalah jumlah data latih yang digunakan, j merupakan kelas data untuk keanggotaan data latih dan i merupakan kelas data untuk keanggotaan data uji.

- Menghitung derajat keanggotaan data uji terhadap masing-masing kelas dengan menggunakan Persamaan (4).

$$\mu_i(x) = \frac{\sum_{j=1}^k \mu_{ij} \left(\|x - x_j\|^{-\frac{2}{m-1}} \right)}{\sum_{j=1}^k \left(\|x - x_j\|^{-\frac{2}{m-1}} \right)} \quad (4)$$

- dimana $\mu_i(x)$ merupakan nilai keanggotaan suatu data x kedalam kelas i , $x - x_j$ merupakan selisih jarak data uji x ke data latih x_j dalam k tetangga terdekat dan m merupakan bobot pangkat yang besarnya $m > 1$
9. Memilih kelas dengan derajat keanggotaan terbesar sebagai hasil akhir.
 10. Menghitung tingkat akurasi prediksi curah hujan menggunakan Persamaan (5).

Nilai akurasi pada penelitian ini dihitung menggunakan *confusion matrix*. Pada *confusion matrix* terdapat informasi nyata dari suatu objek beserta hasil prediksi klasifikasinya (Deng dkk, 2015). Akurasi digunakan untuk menilai performa dari suatu sistem klasifikasi dalam mengklasifikasikan data uji. Semakin besar akurasi maka kemampuan prediksi metode klasifikasi bekerja semakin baik. Tabel 2 menunjukkan bentuk dari *confusion matrix* untuk klasifikasi multi-kelas dengan label A_n dan N_{ij} merupakan jumlah sampel aktual yang termasuk ke label A_i tetapi diklasifikasikan ke dalam label A_j .

Tabel 2. Confusion Matrix

		Prediksi				
		A_1	...	A_j	...	A_n
Aktual	A_1	N_{11}		N_{1j}	N_{1n}	
	\vdots			\vdots		
	A_i	N_{i1}	...	N_{ij}	...	N_{in}
	\vdots			\vdots		
	A_n	N_{n1}		N_{nj}	N_{nn}	

(Sumber: Deng dkk, 2015)

Akurasi merupakan rasio hasil prediksi sesuai dengan hasil aktual terhadap keseluruhan data. Rumus perhitungan akurasi menggunakan Persamaan (4) (Deng, dkk 2015).

$$Akurasi = \frac{\sum_{i=1}^n N_{ii}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}} \quad (5)$$

dimana N_{ii} merupakan data hasil prediksi yang sesuai dengan hasil aktual dan N_{ij} merupakan jumlah data uji.

III. HASIL DAN PEMBAHASAN

Dataset yang digunakan pada penelitian ini bersumber dari *website* resmi BMKG. *Dataset* yang digunakan merupakan data curah hujan di Kabupaten Padang Pariaman periode 1 Januari 2021 hingga 31 Agustus 2023. Jumlah amatan pada setiap kelas dalam penelitian ini ditunjukkan pada Tabel 3.

Tabel 3. Jumlah Data per Kelas

Kelas	Jumlah Amatan
Berawan	571
Ringan	174
Sedang	147
Lebat	64
Sangat Lebat	9
Ekstrem	8

Tabel 3 menampilkan hasil dari jumlah amatan tiap kelas tidak seimbang, dimana kelas berawan memiliki jumlah amatan yang sangat banyak dibandingkan kelas lainnya. Perbedaan banyaknya amatan antar kelas mengakibatkan algoritma tidak mampu memprediksi kelas dengan tepat dan akurat sehingga menyebabkan kesalahan dalam klasifikasi. Untuk mengatasi permasalahan tersebut dibutuhkan tambahan teknik khusus untuk menyeimbangkan jumlah data dalam setiap kelas, yaitu menggunakan teknik SMOTE. Hasil dari teknik SMOTE dapat dilihat pada Tabel 4.

Tabel 4. Jumlah Data per Kelas Setelah SMOTE

Kelas	Jumlah Amatan
1	571
2	571
3	571
4	571
5	571
6	571
Total	3.426

Tabel 4 merupakan jumlah amatan tiap kelas setelah teknik SMOTE diterapkan. Jumlah amatan pada tiap kelas setelah penerapan teknik SMOTE menjadi sama, yaitu 571 amatan sehingga jumlah amatan total pada data juga bertambah menjadi 3426. Pada tahapan selanjutnya, akan dilihat hasil prediksi dari data uji dengan rasio data uji dan data latih sebesar 90:10. Sehingga, jumlah data uji pada penelitian ini adalah 343 data. Pada Tabel 5 akan ditampilkan perbandingan kelas data uji asli dengan data uji prediksi menggunakan algoritma FK-NN dengan nilai $K=12$.

Tabel 5. Perbandingan Kelas Data Uji Asli dengan Data Uji Prediksi

Amatan	Kelas Data Uji		Keterangan
	Asli	Prediksi	
1	Berawan	Sangat Lebat	Tidak Sesuai
2	Lebat	Lebat	Sesuai
3	Berawan	Berawan	Sesuai
4	Berawan	Berawan	Sesuai
5	Sedang	Sedang	Sesuai
6	Ringan	Ringan	Sesuai
7	Berawan	Ringan	Tidak Sesuai
8	Berawan	Berawan	Sesuai
9	Berawan	Lebat	Tidak Sesuai
10	Berawan	Berawan	Sesuai
:	:	:	:
334	Ekstream	Ekstream	Sesuai
335	Ekstream	Ekstream	Sesuai
336	Ekstream	Ekstream	Sesuai
337	Ekstream	Ekstream	Sesuai
338	Ekstream	Ekstream	Sesuai
339	Ekstream	Ekstream	Sesuai
340	Ekstream	Ekstream	Sesuai
341	Ekstream	Ekstream	Sesuai
342	Ekstream	Ekstream	Sesuai
343	Ekstream	Ekstream	Sesuai

Berdasarkan Tabel 5 terlihat bahwa ada beberapa amatan yang memiliki hasil prediksi yang tidak sama dengan data aslinya. Contohnya pada amatan 1 kelas data uji yang asli berada di kelas berawan dan kelas data uji hasil prediksi menggunakan FK-NN juga menunjukkan hasil sangat lebat. Sedangkan pada amatan 2 kelas data uji hasil prediksi sama dengan kelas data uji aslinya. Selanjutnya, dilakukan perhitungan akurasi dengan menggunakan Persamaan 4 dan diperoleh hasil sebagai berikut.

$$Akurasi = \frac{\sum_{i=1}^n N_{ii}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}} = \frac{262}{343} = 0,7638$$

Nilai akurasi yang dihasilkan melalui perhitungan dengan algoritma FK-NN dan parameter $K = 12$ adalah sebesar 0,7638. Hal ini menunjukkan kinerja algoritma FK-NN menggunakan jarak *euclidean* dengan parameter $K = 12$ menghasilkan akurasi sebesar 76,38%.

IV. KESIMPULAN

Dataset yang bersumber dari *website* BMKG memiliki *imbalance class*. Oleh karena itu diterapkan metode *Synthetic Minority OverSampling Technique* (SMOTE) untuk mengatasi masalah *imbalance class*. Penelitian ini melakukan pemodelan terhadap *dataset* curah hujan di Kabupaten Padang Pariaman dengan menggunakan algoritma *Fuzzy K-Nearest Neighbor*. Jumlah data yang digunakan dalam penelitian sebanyak 343 data, pembagian data latih dan data uji dengan perbandingan 90:10. Nilai K yang digunakan untuk menguji kinerja algoritma FK-NN dalam memprediksi curah hujan yaitu $K = 12$, dimana didapatkan nilai akurasi sebesar 76,38%. Berdasarkan hasil tersebut algoritma FK-NN memiliki tingkat akurasi yang cukup baik dalam melakukan proses prediksi curah hujan di Kabupaten Padang Pariaman. Untuk penelitian selanjutnya dalam memprediksi tingkat curah hujan menggunakan algoritma FKNN dapat mengubah nilai parameter k dan juga dapat menggunakan fungsi jarak yang lainnya. Selain itu, curah hujan juga dapat diprediksi menggunakan algoritma klasifikasi yang lain, seperti *Naïve Bayes Classifier* atau *Backpropagation Neural Network*.

DAFTAR PUSTAKA

- Ahmad, dkk. (2018). "Using Fuzzy K-Nearest Neighbor for Predicting University Students Graduation In Teknokrat". *Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)*, 1(1), 47-52.
- Angreni, I. A., Adisasmita, S. A., Ramli, M. I., & Hamid, S. (2019). "Pengaruh Nilai K pada Metode K-Nearest Neighbor (K-NN) Terhadap Tingkat Akurasi Identifikasi Kerusakan Jalan", *Rekayasa Sipil*, 7(2), 63-70.
- Anggraini, Nicky. (2019). Penerapan Algoritma Artificial Immune System Untuk Menentukan Optimasi Batasan Fungsi Keanggotaan Fuzzy Tsukamoto Pada Kasus Prediksi Curah Hujan. *Skripsi*, 120 Hal, Universitas Islam Negeri Syarif Kasim Riau, Pekanbaru, Indonesia, Desember 2019.
- Anugrah, Andhika. S. P., Indriati., & Candra Dewi. (2018). Implementasi Algoritme Fuzzy K-Nearest Neighbor untuk Penentuan Lulus Tepat Waktu (Studi Kasus: Fakultas Ilmu Komputer Universitas Brawijaya). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(4), 1726-1732.
- Bramer, Max. 2007. Principles of Data Mining. London: Springer.
- Chawla, N. V, K. W. Bowyer, and L. O. Hall. (2017). "Handling Imbalance Data Prediksi Churn menggunakan metode SMOTE dan KNN Based on Kernel", *e-Proceeding of Engineering*, 4(117), 1-15.
- Chawla, N. V, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. (2002). "SMOTE: Synthetic Minority Over-sampling Technique", *J. Artif. Intell. Res*, 16, 321–357.
- Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An Improved Method to Construct Basic Probability Assignment Based on the Confusion Matrix for Classification Problem. *Information Sciences*, 340–341, 250–261.
- García, S., Luengo, J., & Herrera, F. (2015). "Data Preprocessing in Data Mining", *Springer International Publishing*. 72.
- He, H, & E. A. Garcia. (2009). "Learning From Imbalanced Data", *IEEE Trans. Knowl. Data Eng*, 21(9), 1263–1284.
- Jin, Weiqing. (2004). *Fuzzy Classification Based on Fuzzy Association Rule Mining*. North Carolina State University.
- Nanda, D. M., Pudjiantoro, T. H., & Sabrina, C. W. (2022). Metode K-Nearest Neighbor (KNN) dalam Memprediksi Curah Hujan di Kota Bandung. *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika (SNESTIK)*, 387-393.
- Nguyen, E. W., & K. Kamei. (2011). "Borderline Over-Sampling for Imbalanced Data Classification", *Int. J. Knowl. Eng. Soft Data Paradig*, 3(1), 4–21.
- Nugraha, S. D., Rakyant, R. M. D., & Randy, C. W. (2017). Penerapan Fuzzy K-Nearest Neighbor (FK-NN) Dalam Menentukan Status Gizi Balita. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(9), 925-932.
- Prasath, V. B. S., Alfeilat, H. A. A., Hassanat, A. B. A., Lasasmeh, O., Tarawneh, A. S., Alhasanah, M. B., & Salman, H. S. E. (2019). "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier—A Review". *Big Data*, 7(4), 221–248.
- Ruangthong, P., & Jaiyen, S. (2015). "Bank Direct Marketing Analysis of Asymmetric Information Based on Machine Learning", *12th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 93–96.
- Sun, Y. A. K. C. Wong, and M. S. Kamel. (2009). "Classification of imbalanced data: A review", *Int J Pattern Recognit Artif Intell*, 23(4), 687–719.
- Triatmodjo, Bambang. (2008). *Hidrologi Terapan*. Yogyakarta: Beta Offset.

- Wilson. E. M. (1973). *Hidrologi Teknik*. Edisi ke-4. Bandung: ITB.
- Wirawan, I Made Agus. (2017). *Metode Penalaran dalam Kecerdasan Buatan*. Depok: RajaGrafindo Persada.
- Wisdario, A. (2013). Penerapan Metode Fuzzy K-Nearest Neighbor (FK-NN) Untuk Pengklasifikasian Spam Email. *Skripsi*. 110 Hal. Program Studi Ilmu Komputer Universitas Brawijaya. Malang.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). "Top 10 Algorithms in Data Mining. Knowledge and Information Systems", *Knowl Inf Syst*, 14(1), 1–37.