

# Peningkatan Akurasi K-Nearest Neighbor Pada Data Index Standar Pencemaran Udara Kota Pekanbaru

Yuliska<sup>1</sup>, Khairul Umam Syaliman<sup>2</sup>

Teknik Informatika, Politeknik Caltex Riau, Pekanbaru, Riau, Indonesia<sup>1,2</sup>

yuliska@pcr.ac.id<sup>1</sup>, khairul@pcr.ac.id<sup>2</sup>

---

## Article Info

### History :

Dikirim 27 Februari 2020

Direvisi 28 April 2020

Diterima 14 Juli 2020

---

### Kata Kunci :

Akurasi

*Attribute Weighting*

*K-Nearest Neighbor*

*Local Mean*

Peningkatan

---

## Abstrak

kNN adalah salah satu metode yang populer karena mudah dieksploitasi, generalisasi yang baik, mudah dimengerti, kemampuan beradaptasi ke ruang fitur yang rumit, intuitif, atraktif, efektif, flexibility, mudah diterapkan, sederhana dan memiliki hasil akurasi yang cukup baik. Namun kNN memiliki beberapa kelemahan, diantaranya memberikan bobot yang sama pada setiap atribut sehingga atribut yang tidak relevant juga memberikan dampak yang sama dengan atribut yang relevant terhadap kemiripan antar data. Masalah lain dari kNN adalah pemilihan tetangga terdekat dengan system suara terbanyak, dimana system ini mengabaikan kemiripan setiap tetangga terdekat dan kemungkinan munculnya mayoritas ganda serta kemungkinan terpilihnya outlier sebagai tetangga terdekat. Masalah-masalah tersebut tentu saja dapat menimbulkan kesalahan klasifikasi yang mengakibatkan rendahnya akurasi. Pada penelitian kali ini akan dilakukan peningkatan akurasi dari kNN tersebut dalam melakukan klasifikasi terhadap data Index Standar Pencemaran Udara di Pekanbaru dengan menggunakan pembobotan atribut (Attribute Weighting) dan local mean. Adapun hasil dari penelitian ini didapati bahwa metode yang diusulkan mampu untuk meningkatkan akurasi sebesar 2.42% dengan rata-rata tingkat akurasi sebesar 97.09%.

© This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

---

## Koresponden:

Khairul Umam Syaliman

Teknik Informatika

Politeknik Caltex Riau

Jl. Umbansari (Patin), Pekanbaru, Indonesia, 28265

Email : khairul@pcr.ac.id

---

## 1. PENDAHULUAN

K-Nearest Neighbor atau biasa dikenal dengan kNN merupakan suatu metode klasifikasi yang berbasis jarak [1], kNN adalah salah satu metode yang populer dan paling banyak digunakan dalam penyelesaian berbagai kasus klasifikasi, diantaranya text categorization, pengenalan pola, peramalan, image-similarity, data visualization, pengklasifikasian, estimasi posisi, dan lain-lain [2-18]. kNN telah banyak mengalami perubahan dengan tujuan untuk meningkatkan kinerja dari metode ini, hal ini disebabkan karena kNN mudah dieksploitasi di berbagai domain aplikasi, memiliki generalisasi yang baik, mudah dimengerti, memiliki kemampuan beradaptasi ke ruang

fitur yang tidak beraturan, intuitif, atraktif, efektif, flexibility, udah diterapkan, sederhana dan memiliki hasil yang baik [1,3,14-30].

Salah satu modifikasi yang sukses meningkatkan hasil akurasi dari kNN adalah metode yang diusulkan oleh [31] atau biasa dikenal dengan local mean k-Nearest Neighbor (LMkNN). Pada penelitian tersebut disarankan untuk mengganti *system vote majority* (suara terbanyak) dengan *local mean*, hal ini bertujuan untuk mengurangi pengaruh *outlier* [32]. Pada penelitian tersebut local mean terbukti mampu untuk meningkatkan hasil akurasi [31, 20].

Modifikasi selanjutnya diusulkan oleh [33] atau biasa dikenal dengan Distance Weight K-Nearest Neighbor (DWkNN). Pada penelitian ini penulis juga mengusulkan untuk mengganti *system vote majority* dengan metode *distance weight*, hal ini bertujuan untuk memberikan pengaruh yang lebih besar terhadap data yang memiliki tingkat kemiripan yang tinggi, sehingga penentuan kelas bagi data baru menjadi lebih akurat. Pada penelitian ini terbukti bahwa *distance weight* mampu memberikan hasil akurasi yang lebih baik.

Dua penelitian diatas yaitu LMkNN dan DWkNN telah berhasil dikombinasikan oleh [34]. Pada penelitian tersebut penulis menyarankan penggabungan beberapa tahapan dari metode LMkNN dengan beberapa tahapan dari metode DWkNN dengan tujuan untuk mengatasi kelemahan pada *system vote majority*. Dimana tahapan dari LMkNN digunakan untuk mengurangi pengaruh *outlier* yang terpilih sebagai *k*-tetangga terdekat, dan tahapan dari DWkNN digunakan untuk menutupi kelemahan *system vote majority* yang mengabaikan kemiripan antar data dan kemungkinan terjadinya kelas mayoritas ganda [33-34].

Namun, penelitian yang disarankan oleh [34] masih memiliki peluang untuk ditingkatkan. Salah satunya dengan memberikan bobot pada setiap atribut (*Feature*). Hal ini dirasa perlu dikarenakan metode yang berbasis jarak selalu memberikan pengaruh yang sama pada setiap atribut dalam proses penentuan kemiripan antar data. Tentu saja hal ini membuat atribut yang tidak relevan terhadap hasil klasifikasi juga akan memberikan pengaruh yang sama besarnya terhadap atribut yang memiliki relevan.

[33] dan [36] sepakat memberikan bobot yang berbeda terhadap setiap atribut, dimana atribut yang memiliki pengaruh yang lebih besar terhadap hasil akurasi diberikan bobot yang lebih besar dibandingkan atribut lainnya. Pada kedua penelitian ini juga disarankan menggunakan Gain Ratio yang dijadikan dasar pembobotan dari setiap atribut. Berdasarkan hasil penelitian gain ratio terbukti mampu untuk meningkatkan nilai akurasi.

Berdasarkan penelitian-penelitian diatas, pada penelitian kali ini penulis berusaha untuk melakukan peningkatan pada metode kNN tersebut, dimana pada penelitian ini penulis akan menggabungkan metode *pembobotan atribut* dengan *local mean* yang telah terbukti berhasil meningkatkan akurasi dari metode kNN konvensional. Diharapkan gabungan dari metode tersebut mampu untuk meningkatkan hasil akurasi dari kNN.

Untuk melihat kinerja dari modifikasi yang dilakukan maka akan digunakan data Index Standar Polusi Udara di Kota Pekanbaru, Riau, Indonesia. Adapun struktur penulisan dari penelitian ini terdiri dari bagian 2 rangkuman mengenai studi literatur. Bagian 3 akan dijelaskan mengenai Metode yang diusulkan, bagian 4 merupakan bagian hasil dan pembahasan serta yang terakhir bagian 5 yang merupakan kesimpulan dari penelitian.

## 2. STUDI LITERATUR

Pada bagian ini akan dijelaskan mengenai studi literature dan metode apa saja yang digunakan pada penelitian ini. Metode yang diusulkan akan menggabungkan beberapa tahapan dari FWkNN dan LMkNN, dimana untuk melakukan pembobotan atributnya akan menggunakan metode Gain Ratio.

### 2.1. Gain Ratio

*Gain Ratio* ditemukan pada algoritma C4.5, dimana *gain ratio* digunakan untuk menghitung pengaruh atribut terhadap target dari suatu data [36]. *Gain Ratio* merupakan pengembangan dari *information gain*, dimana *gain ratio* mengambil informasi intrinsik dan menghilangkan nilai bias

dari setiap atribut [38]. Adapun langkah-langkah dalam perhitungan *Gain Ratio* adalah sebagai berikut :

- Hitung nilai *entropy* dari masing-masing atribut dengan menggunakan persamaan :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

- Hitung nilai *informasi gain* pada masing-masing atribut dengan persamaan :

$$Information\ Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

- Hitung nilai *split information* dengan menggunakan persamaan :

$$SplitInfo_A(D) = -\sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left( \frac{|D_j|}{|D|} \right) \quad (3)$$

- Hitung nilai *Gain Ratio* setiap atribut dengan persamaan :

$$Gain\ Ratio(A) = \frac{Gain(A)}{SplitInfo(A)} \quad (4)$$

## 2.2. Feature Weighting K-Nearest Neighbor

*Feature Weighting K-Nearest Neighbor* (FWKNN) adalah modifikasi dari kNN dengan prinsip kerja memberikan bobot yang berbeda terhadap atribut atau feature data dengan tujuan mengurangi atribut yang tidak atau kurang relevan terhadap kelas data[33,36]. Adapun tahapan dari FWKNN secara sederhana dapat dijelaskan sebagai berikut :

Tahapan 1 : Hitung *Gain Ratio* dari setiap atribut dan jadikan sebagai nilai bobot pada setiap atribut.

Tahapan 2 : Tentukan nilai k

Tahapan 3 : Hitung kemiripan dengan menggunakan persamaan :

$$d^w(X' - X) = \sqrt{\sum_{i=1}^y W_y \times (x'_i - x_i)^2} \quad (5)$$

Tahapan 4 : Pilih k tetangga terdekat berdasarkan kemiripan.

Tahapan 5 : Tentukan kelas mayoritas dari k tetangga terdekat dan jadikan sebagai kelas bagi data yang baru.

Banyak metode yang dapat digunakan dalam pembobotan atribut, salah satunya dengan menggunakan *gain ratio*. [33] dan [36] membuktikan bahwa pembobotan atribut dengan menggunakan *Gain Ratio* mampu memberikan hasil yang lebih baik dari metode kNN konvensional.

## 2.3. Local Mean K-Nearest Neighbor

*Local mean K-Nearest Neighbor* atau biasa disingkat dengan LMkNN adalah metode yang pertama kali dikenalkan oleh [31]. Pada penelitian tersebut *vote system majority* digantikan dengan *local mean* dalam proses penentuan kelas bagi data baru. Adapun tahapan dari LMkNN adalah sebagai berikut :

Tahapan 1 : Tentukan nilai k.

Tahapan 2 : Hitung kemiripan data training dan data testing dengan menggunakan persamaan :

$$d(X' - X) = \sqrt{\sum_{i=1}^y (x'_i - x_i)^2} \quad (6)$$

Tahapan 3 : Pilih k tetangga terdekat dari setiap kelas.

Tahapan 4 : Hitung *local mean* dengan persamaan :

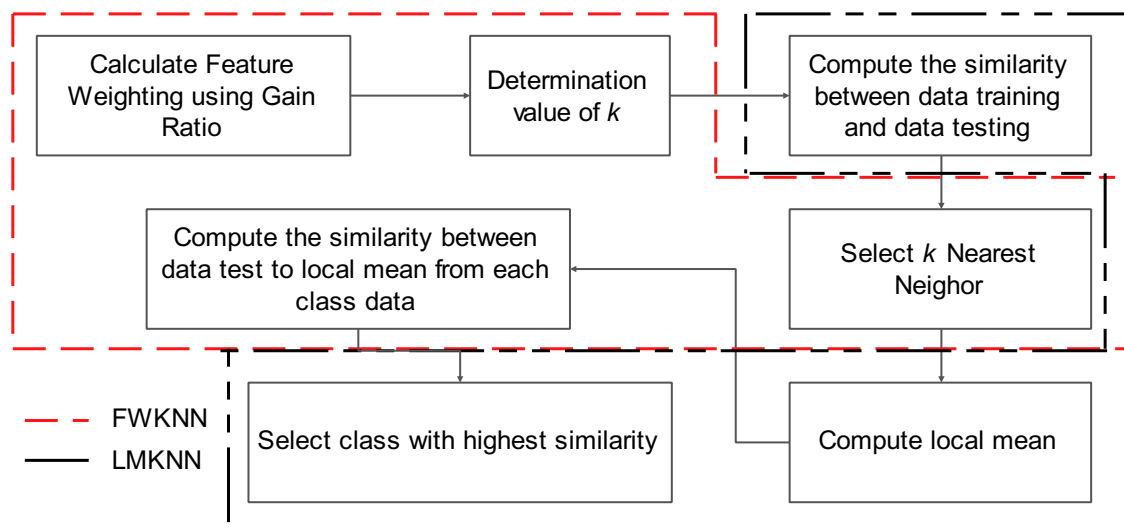
$$m_{w_j}^k = \frac{1}{k} \sum_{i=1}^k y_{i,j}^{NN} \quad (7)$$

Tahapan 5 : Hitung kemiripan data testing terhadap *local mean* dari setiap kelas data dengan persamaan (6).

Tahapan 6 : Jadikan kelas data dari *local mean* terdekat sebagai kelas bagi data testing.

## 3. METODE YANG DIUSULKAN

Pada bagian ini akan dijelaskan mengenai metode yang diusulkan. Dimana pada penelitian kali ini penulis akan menggabungkan beberapa tahapan dari FWkNN dan LMkNN. Untuk melihat lebih jelas langkah-langkah dari metode yang diusulkan dapat dilihat pada gambar 1.



Gambar 1. Metode Yang Diusulkan.

Berdasarkan gambar 1 dapat dijelaskan langkah-langkah penggabungan antara *feature weighting* dan *local mean* dalam penelitian ini adalah sebagai berikut :

Tahapan 1 : Hitung bobot setiap atribut dengan menggunakan *Gain Ratio*.

Tahapan 2 : Tentukan nilai  $k$ .

Tahapan 3 : Hitung kemiripan antar data latih dan data uji dengan persamaan (5).

Tahapan 4 : Pilih  $k$  tetangga terdekat dari setiap kelas data.

Tahapan 5 : Tentukan *local mean* dengan persamaan (7).

Tahapan 6 : Hitung kemiripan antara data uji dengan *local mean* dari setiap kelas menggunakan persamaan (5).

Tahapan 7 : Jadikan kelas data dari *local mean* terdekat sebagai kelas bagi data testing.

Tahapan 1 s/d 3 merupakan proses yang diadopsi dari metode FWkNN, sedangkan tahapan 4 s/d 7 merupakan bagian proses dari metode LMKNN. Dimana pembobotan atribut dilakukan untuk mengurangi pengaruh atribut yang tidak atau kurang relevan terhadap kelas data, sedangkan *local mean* digunakan untuk mengatasi kelemahan pada sistem suara terbanyak.

#### 4. HASIL DAN PEMBAHASAN

Untuk mengetahui apakah metode yang diusulkan mampu menghasilkan akurasi yang lebih baik, maka pada penelitian kali ini akan dilakukan pengujian dengan menggunakan data Index Standart Pencemaran Udara (ISPU) Kota Pekanbaru, Riau, Indonesia. Data ISPU ini didapat dari lab Udara Pemerintah Kota Pekanbaru. Data ini terdiri dari 5 atribut dengan 4 kategori kelas data (Bagus, Sehat, Tidak Sehat, Sangat Tidak Sehat), dimana jumlah data sebanyak 992 data. Data ini juga merupakan data yang *unbalance*, artinya jumlah data dari setiap kelas-nya tidak seimbang. Adapun rincian dari data dapat dilihat pada table 1.

Tabel 1. Rincian Data Yang digunakan

Data	Nilai Atribut Ke					Jumlah Data Per-Kelas				Jumlah Data
	1	2	3	4	5	B	S	STS	TS	
Index Pencemaran Udara	4...870	0...57	0...150	2...211	0...105	796	150	15	31	992

Pada penelitian ini akan menggunakan metode Hold-Out pada proses validasinya dimana sebanyak 80% dari jumlah data akan dijadikan sebagai data latih dan 20% sisanya akan dijadikan

sebagai data uji. Pada penelitian ini hanya akan menggunakan  $k$  bernilai 1 s/d  $K$  bernilai 5 dikarenakan data bersifat *unbalance*. Adapun rincian data uji dan data latih dapat dilihat pada tabel 2.

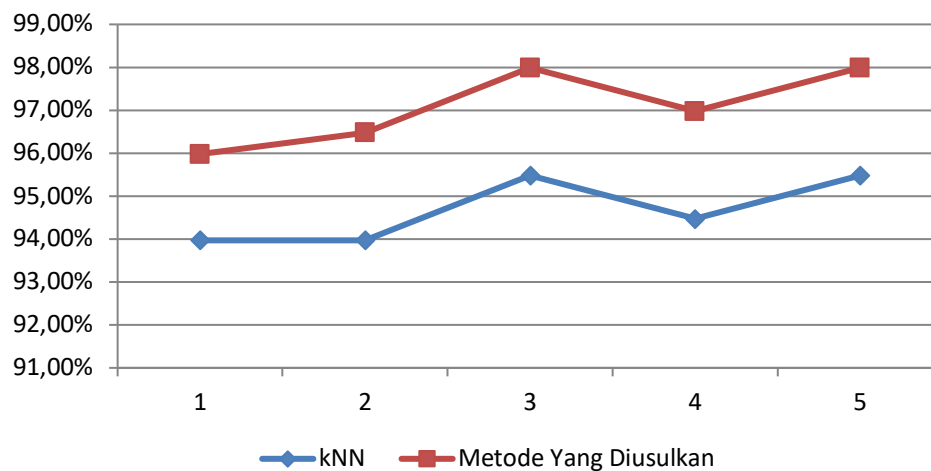
Tabel 2. Rincian Data Latih dan Data Uji

	Jumlah Data Per-Kelas				Jumlah Data
	B	S	STS	TS	
Latih	640	119	10	24	793
Uji	156	31	5	7	199
Total Data					992

Untuk mengetahui dengan pasti apakah metode yang diusulkan mampu memberikan hasil akurasi yang terbaik, maka selanjutnya akan dilakukan pengujian dengan menggunakan data yang telah dibagi menjadi 2 bagian, adapun hasil dari akurasi dapat dilihat pada tabel 3 dan gambar 2.

Tabel 3. Result Accuracy from APSI data

$k$	kNN <sup>(1)</sup>	Accuracy		Metode Terbaik
		Metode Yang Diusulkan <sup>(2)</sup>		
1	93.97%	95.98%		2
2	93.97%	96.48%		2
3	95.48%	97.99%		2
4	94.47%	96.98%		2
5	95.48%	97.99%		2
Rata-rata	94.67%	97.09%		2



Gambar 2. Hasil Akurasi Klasifikasi Data ISPU.

Berdasarkan tabel 3 dan gambar 2 terlihat bahwa metode yang diusulkan selalu memberikan hasil akurasi yang lebih baik dari kNN konvensional, dimana nilai akurasi terendah yang didapati kNN sebesar 93.97% dan yang tertinggi hanya sebesar 95.48% saat  $k$  bernilai 3 dan 5, sedangkan metode yang diusulkan mampu menghasilkan nilai akurasi terendah sebesar 95.98% dan tertinggi sebesar 97.99% saat  $k$  bernilai 3 dan 5. Rata-rata peningkatan yang dicapai pada metode yang diusulkan adalah sebesar 2.42%.

## 5. KESIMPULAN

Berdasarkan bagian dari hasil dan pembahasan terlihat bahwa pembobotan attribut dan *local mean* mampu meningkatkan hasil akurasi dari kNN konvensional. Akurasi tertinggi yang mampu dicapai dengan menggunakan kNN konvensional hanya sebesar 95.48% saat  $k$  bernilai 3 dan 5,

sedangkan metode yang diusulkan mampu mencapai nilai akurasi sebesar 97.99% saat  $k$  bernilai 3 dan 5 dengan rata-rata peningkatan akurasi sebesar 2.42%.

### UCAPAN TERIMA KASIH



Penulis mengucapkan terima kasih kepada Sri Indah Darlianti, S.T untuk sumbangan data ISPU Kota Pekanbaru, Lembaga Penelitian Politeknik Caltex Riau (PCR) untuk dukungan fasilitas serta financial dan kepada Raudhatul Jannah untuk dukungan moral maupun materil yang tak bias diungkapkan dengan kata-kata pada bagian ini.

### DAFTAR PUSTAKA

- [1] J. Wang, P. Neskovic, and L. N. Cooper, "Improving nearest neighbor rule with a simple adaptive distance measure," *Pattern Recognit. Lett.*, vol. 28, no. 2, pp. 207–213, 2007.
- [2] N. Bhatia, and Vandana, "Survey of nearest neighbor techniques" *Int. J. of Computer Science and Information Security*, vol 8, no. 2, pp. 302-305, 2010.
- [3] J. Gou, Z. Yi, L. Du, and T. Xiong, "A Local Mean-Based  $k$ -Nearest Centroid Neighbor Classifier," vol. 55, no. 9, 2012.
- [4] A. Suárez Sánchez, F. J. Iglesias-Rodríguez, P. Riesgo Fernández, and F. J. de Cos Juez, "Applying the  $K$ -nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders," *Int. J. Ind. Ergon.*, vol. 52, pp. 92–99, 2014.
- [5] H. B. Jaafar, N. B. Mukahar, and D. A. B. Ramli, "A methodology of nearest neighbor: Design and comparison of biometric image database," *Proc. - 14th IEEE Student Conf. Res. Dev. Adv. Technol. Humanit. SCOREd 2016*, 2017.
- [6] K. Zheng, G. Si, L. Diao, Z. Zhou, J. Chen, and W. Yue, "Applications of support vector machine and improved  $k$ -Nearest neighbor algorithm in fault diagnosis and fault degree evaluation of gas insulated switchgear," *ICEMPE 2017 - 1st Int. Conf. Electr. Mater. Power Equip.*, pp. 364–368, 2017.
- [7] Y. Cai, H. Huang, H. Cai, and Y. Qi, "-Nearest Neighbor Locally Search Regression Algorithm for Short-Term Traffic Flow Forecasting," no. Icmic, pp. 624–629, 2017.
- [8] I. Gazalba, N. Gayatri, and I. Reza, "Comparative Analysis of  $K$ -Nearest Neighbor and Modified  $K$ -Nearest Neighbor Algorithm for Data Classification," pp. 294–298, 2017.
- [9] F. Chen, Z. Ye, C. Wang, L. Yan, and R. Wang, "A Feature Selection Approach for Network Intrusion Detection Based on Tree-Seed Algorithm and  $K$ -Nearest Neighbor," 2018 IEEE 4th Int. Symp. Wirel. Syst. within Int. Conf. Intell. Data Acquis. Adv. Comput. Syst., pp. 68–72, 2018.
- [10] S. Han and Y. Li, "ScienceDirect ScienceDirect  $K$ -Nearest Neighbor combined with guided filter for hyperspectral  $K$ -Nearest Neighbor combined with guided filter for hyperspectral image classification image classification," vol. 00, 2018.
- [11] L. Le, "Deep Similarity-Enhanced  $K$  Nearest Neighbors," 2018 IEEE Int. Conf. Big Data (Big Data), pp. 2643–2650, 2018.
- [12] J. Kim, "Adapt tive  $K$ -Neare est Ne ighbo our Alg gorith m for WiFi Finge rprint t Posit ioning g," *ICT Express*, pp. 4–7, 2018.
- [13] H. Kaneko, "SC," *Chemom. Intell. Lab. Syst.*, 2018.
- [14] A. Swetapadma and A. Yadav, "A novel single-ended fault location scheme for parallel transmission lines using  $k$ -nearest neighbor algorithm ☆," *Comput. Electr. Eng.*, vol. 69, no. May, pp. 41–53, 2018.
- [15] A. R. Winnersyah, "Identification and Position Estimation Method with  $K$ -Nearest Neighbour and Home Occupants Activity Pattern," 2018 6th Int. Conf. Cyber IT Serv. Manag., no. Citsm, pp. 1–4, 2018.
- [16] F. Borghesan, M. Chioua, and N. F. Thornhill, "Forecasting of process disturbances using  $k$ -nearest neighbours , with an application in process control R," *Comput. Chem. Eng.*, vol. 128, no. 675215, pp. 188–200, 2020.
- [17] M. Cao, L. I. N. Li, W. Xie, W. E. I. Jia, M. Ieee, and Z. Lv, "Parallel  $K$  Nearest Neighbor Matching for 3D Reconstruction," *IEEE Access*, vol. 7, pp. 55248–55260, 2019.Test
- [18] J. Gou, H. Ma, W. Ou, S. Zeng, Y. Rao, and H. Yang, "A generalized mean distance-based  $k$ -nearest neighbor classifier," *Expert Syst. Appl.*, 2018.
- [19] N. Garcia-Pedrajas, J. A. Romero Del Castillo, and G. Cerruela-Garcia, "A Proposal for Local  $k$  Values for  $k$ -Nearest Neighbor Rule," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 28, no. 2, pp. 470–475, 2017.

- [20] Z. Pan, Y. Wang, and W. Ku, "A new k-harmonic nearest neighbor classifier based on the multi-local means," *Expert Syst. Appl.*, vol. 67, pp. 115–125, 2017.
- [21] S. Ougiaroglou and G. Evangelidis, "Fast and accurate k-nearest neighbor classification using prototype selection by clustering," *Proc. 2012 16th Panhellenic Conf. Informatics, PCI 2012*, no. i, pp. 168–173, 2012.
- [22] F. Yu, J. C. Liu, and D. M. Liu, "An approach for fault diagnosis based on an improved k-nearest neighbor algorithm," *Chinese Control Conf. CCC*, vol. 2016-August, no. 1, pp. 6521–6525, 2016.
- [23] S. K. Shukla and E. Koley, "Detection and classification of open conductor faults in six-phase transmission system using k-nearest neighbour algorithm," *2017 7th Int. Conf. Power Syst. ICPS 2017*, pp. 157–161, 2018.
- [24] K. Fathoni, M. Zikky, A. S. Nurhayati, and I. Prasetyaningrum, "Application of K-Nearest Neighbor Algorithm for Puzzle Game of Human Body's System Learning on Virtual Mannequin," *Proc. - 2018 Int. Conf. Appl. Sci. Technol. iCAST 2018*, pp. 530–535, 2018.
- [25] S. S. Mullick, S. Datta, and S. Das, "Adaptive learning-based k-nearest neighbor classifiers with resilience to class imbalance," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 11, pp. 5713–5725, 2018.
- [26] K. Nyodu and K. Sambyo, "Automatic Identification of Arunachal language Using K-Nearest Neighbor Algorithm," *Proc. - IEEE 2018 Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2018*, pp. 213–216, 2018.
- [27] M. Pujari, C. Awati, and S. Kharade, "Efficient Classification with an Improved Nearest Neighbor Algorithm," *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–5, 2018.
- [28] G. A. Sandag, N. E. Tedry, and S. Lolong, "Classification of Lower Back Pain Using K-Nearest Neighbor Algorithm," *2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018*, no. Citsm, pp. 1–5, 2019.
- [29] M. Marzouq, H. El Fadili, Z. Lakhliai, A. Mechaqrane, and K. Zenkouar, "New distance weighted k Nearest Neighbor model for hourly global solar irradiation estimation," *2019 Int. Conf. Wirel. Technol. Embed. Intell. Syst. WITS 2019*, pp. 1–5, 2019.
- [30] Y. Wang, Z. Pan, and Y. Pan, "A Training Data Set Cleaning Method by Classification Ability Ranking for the k-Nearest Neighbor Classifier," *IEEE Trans. Neural Networks Learn. Syst.*, no. 1, pp. 1–13, 2019.
- [31] Y. Mitani and Y. Hamamoto, "A local mean-based nonparametric classifier," *Pattern Recognit. Lett.*, vol. 27, no. 10, pp. 1151–1159, 2006.
- [32] Z. Pan, Y. Wang, and W. Ku, "A new general nearest neighbor classification based on the mutual neighborhood information," *Knowledge-Based Syst.*, vol. 121, pp. 142–152, 2017.
- [33] A. Duneja and T. Puyalnithi, "Enhancing Classification Accuracy of K-Nearest Neighbours Algorithm Using Gain Ratio," *Int. Res. J. Eng. Technol.*, vol. 4, no. 9, pp. 1385–1388, 2017.
- [34] K. U. Syaliman, E. B. Nababan, and O. S. Sitompul, "Improving the accuracy of k-nearest neighbor using local mean based and distance weight," *J. Phys. Conf. Ser.*, vol. 978, no. 1, 2018.
- [35] Y. Chen and Y. Hao, "A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction," *Expert Syst. Appl.*, vol. 80, pp. 340–355, 2017.
- [36] A. A. Nababan, O. S. Sitompul, and Tulus, "Attribute Weighting Based K-Nearest Neighbor Using Gain Ratio," *J. Phys. Conf. Ser.*, vol. 1007, no. 1, 2018.
- [37] Thomas M. Mitchell. 1997. *Machine Learning* (1 ed.). McGraw-Hill, Inc., New York, NY, USA.
- [38] P. P. R., V. M.L., and S. S., "Gain Ratio Based Feature Selection Method for Privacy Preservation," *ICTACT J. Soft Comput.*, vol. 01, no. 04, pp. 201–205, 2011.

**BIOGRAFI PENULIS**

	<p><b>Yuliska, S.T., M.Eng</b> obtained her bachelor degree in informatics engineering from State Islamic University of Suska Riau, in 2009, obtained her master degree in Computer Science and Communications Engineering from Waseda University, in 2017. She is now a lecturer in Informatics Engineering major, Politeknik Caltex Riau. Her current research interests are Natural Language Processing, Text Mining, Deep Learning and Human Computer Interaction</p>
	<p><b>Khairul Umam Syaliman</b> is now a lecturer in Informatics Engineering major, Politeknik Caltex Riau. His current research interests are Artificial Intelligence, Data Mining, Machine Learning, Deep Learning and Natural Language Processing.</p>