



Perbandingan Kinerja k-Nearest Neighbor dan Local Mean Distance k-Nearest Neighbor Pada Data Citra Covid-19

Sapriadi

S1 Farmasi, Fakultas Farmasi dan Kesehatan, Institut Kesehatan Helvetia, Medan, Indonesia
sapriadi@helvet.ac.id

Diserahkan: 19 Juni 2023

Diterima: 21 Juni 2023

Diterbitkan: 28 Juni 2023

ABSTRAK

Corona Virus Disease 2019 (covid-19) merupakan pandemi dunia yang menimbulkan berbagai kerugian di dunia. Diagnosis covid-19 yang valid memerlukan waktu yang cukup lama dan hasil ini tidak sepenuhnya akurat. Salah satu cara untuk meningkatkan hasil akurasi adalah dengan menggunakan image classification. k-Nearest Neighbor (kNN) adalah salah satu Teknik klasifikasi yang paling banyak digunakan untuk melakukan pekerjaan tersebut, hanya saja kNN masih memiliki kelemahan. Untuk mengatasi kelemahan pada kNN, maka dilakukan modifikasi dengan menambahkan local mean dan distance weight, di mana varian kNN ini dikenal dengan nama Local Mean Distance Weight k-Nearest Neighbor (LMDWkNN). Oleh sebab itu, penelitian kali mencoba membandingkan kinerja kedua algoritma ini untuk melakukan image classification pada citra covid-19. Kinerja diukur berdasarkan nilai akurasi, precision, dan recall, di mana dari hasil pengujian terbukti bahwa kinerja LMDWkNN lebih baik dari pada kinerja kNN. Algoritma LMDWkNN mengalami peningkatan rata-rata sebesar 3.5% untuk nilai akurasi, 2.89% untuk precision, dan 3.54% untuk recall. Meskipun begitu kNN tetap mampu menghasilkan kinerja yang sama baiknya yang mana kinerja kNN akan sangat bergantung dari nilai k yang digunakan.

Kata kunci: Corona Virus Disease 2019 (covid-19), Image Classification, Kinerja, k-Nearest Neighbor (kNN), Local Mean Distance Weight k-Nearest Neighbor (LMDWkNN)

ABSTRACT

Corona Virus Disease 2019 (covid-19) is a world pandemic that causes various losses in the world. A valid diagnosis of COVID-19 takes a long time and these results are not completely accurate. One way to improve accuracy is to use image classification. k-Nearest Neighbor (kNN) is one of the most widely used classification techniques to do this work, but kNN still has weaknesses. To overcome the weaknesses in kNN, modifications are made by adding local mean and distance weight, where this kNN variant is known as Local Mean Distance Weight k-Nearest Neighbor (LMDWkNN). Therefore, this research tries to compare the performance of these two algorithms to perform image classification on Covid-19 images. Performance is measured based on the values of accuracy, precision, and recall, where from the test results it is proven that the performance of LMDWkNN is better than kNN performance. The LMDWkNN algorithm has an average increase of 3.5% for accuracy, 2.89% for precision, and 3.54% for recall. Even so, kNN is still able to produce equally good performance, where kNN performance will depend heavily on the value of k used.

Keywords: Corona Virus Disease 2019 (covid-19), Image Classification, k-Nearest Neighbor (kNN), Local Mean Distance Weight k-Nearest Neighbor (LMDWkNN), Performance

1. PENDAHULUAN

Corona Virus Disease 2019 atau biasa disingkat dengan covid-19 merupakan pandemi dunia yang menimbulkan berbagai kerugian di dunia. Menurut data dari World Health Organization (WHO), jumlah kasus virus covid-19 mencapai angka 690 juta infeksi dengan angka kematian 6.8 ribu [1]–[3]. Masa inkubasi yang cukup panjang serta ditambah lagi dengan kemampuan bermutasi, menyebabkan covid-19 sulit untuk dideteksi [4], [5]. Saat ini, diagnosis covid-19 yang *valid* memerlukan waktu yang cukup lama dan hasil ini tidak sepenuhnya akurat serta memungkinkan tes lainnya untuk memastikan diagnosis ini [5], [6]. Menurut [5] salah satu cara untuk meningkatkan hasil akurasi adalah dengan menggunakan Teknik Pengolahan Citra, salah satunya adalah *image classification*.

[7] *Image classification* atau klasifikasi gambar merupakan proses penentuan kelas gambar. [8], [9] mengatakan salah satu algoritma klasifikasi gambar yang baik adalah algoritma k-Nearest Neighbor (k-NN), hal ini dikarenakan k-NN sederhana, mudah untuk diterapkan dan memiliki hasil yang cukup baik [10]–[14]. [15] dalam bukunya, menjadikan k-NN sebagai salah satu top ten algorithm in data mining. k-NN merupakan algoritma yang banyak digunakan pada berbagai domain penelitian, diantaranya *text categorization*, pengenalan pola, pengklasifikasian, dan lain-lain [8], [11], [16], [17]. [18] dalam penelitiannya menggunakan algoritma berbasis k-NN untuk proses *sampling* data, tepatnya *oversampling*. Hasil dari penelitian ini menunjukkan k-NN yang diterapkan mampu menghasilkan *augmented data* yang baik. Selanjutnya penelitian yang dilakukan oleh [19], [20] dan [21] menggunakan algoritma k-NN untuk melakukan seleksi fitur. Hasil dari kedua penelitian ini algoritma k-NN yang diterapkan mampu memberikan fitur yang paling dominan dalam penentuan kelas data. Akan tetapi, pada dasarnya algoritma k-NN ini merupakan algoritma untuk melakukan proses klasifikasi [22].

[20] dalam penelitiannya menggunakan k-NN untuk melakukan klasifikasi pada gambar medis (*Brain Tumor and Breast Cancer*). Kesimpulan dari penelitian ini mengatakan bahwa, kinerja dari k-NN ini bergantung kepada nilai k yang dipilih. Adapun nilai kinerja dengan akurasi tertinggi pada data Brain tumor adalah sebesar 84%, dan 89% pada data Breast Cancer. Penelitian selanjutnya adalah penelitian yang dilakukan oleh [23] dan [24], pada kedua penelitian ini menggunakan k-NN untuk melakukan klasifikasi pada data *brain image* dan data bunga. Penelitian senada juga telah dilakukan oleh [20], penelitian-penelitian ini menggunakan algoritma k-NN pada proses klasifikasi. Adapun penerapan k-NN pada proses klasifikasi dari penelitian-penelitian menghasilkan kinerja yang baik.

Akan tetapi, dibalik kinerja dan seluruh kelebihan k-NN masih memiliki kelemahan, diantaranya masalah pada sistem *vote majority* atau suara terbanyak. Sistem *vote majority* ini menghasilkan kelas mayoritas ganda yang meningkatkan kesalahan dalam penentuan kelas data baru [25], [26]. [27] melalui penelitiannya menawarkan solusi untuk meningkatkan kinerja dari k-NN, yaitu dengan menggantikan sistem *vote majority* dengan menggunakan gabungan metode *Local Mean* dari [28] dan *Distance Weight* dari [29]. Modifikasi k-NN ini disebut juga dengan nama *Local Mean Distance Weight k-Nearest Neighbor* atau disingkat dengan LMDWk-NN. Kinerja dari LMDWk-NN terbukti lebih baik dari k-NN, akan tetapi pengujian pada penelitian yang dilakukan hanya menggunakan data struktural saja. Oleh sebab itu, pada penelitian kali mencoba membandingkan kinerja k-NN dengan LMDWk-NN menggunakan dataset citra covid-19. Dimana hasil dari penelitian ini nantinya dapat dijadikan sebagai penentuan model klasifikasi yang akan digunakan sehingga menghasilkan kinerja klasifikasi yang lebih baik pada data citra covid-19.

2. LANDASAN TEORI

2.1 k-Nearest Neighbor (k-NN)

k-Nearest Neighbor (k-NN) pertama kali diperkenalkan sekitar tahun 1950, k-NN merupakan salah satu metode yang paling banyak digunakan pada permasalahan *text categorization*, *pengenalan pola*, *pengklasifikasian*, dan lain-lain [8], [11], [16]–[20], [22]. Hal ini disebabkan k-NN memiliki karakteristik yang cukup atraktif, mudah untuk diterapkan, intuitif, adaptif, serta sederhana [10], [12], [25], [30].

k-NN merupakan algoritma yang termasuk kedalam kategori distance-based algorithms [10]. Distance-Based Algorithms adalah algoritma yang menentukan kemiripan data atau objek berdasarkan pada kedekatan jarak antar data ke suatu kelas atau label atau kelompok data lainnya [31], [32]. Kemiripan antar data pada k-NN ditentukan dengan menggunakan pengukuran model jarak. Adapun beberapa model jarak yang umum digunakan adalah:

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

k-NN bekerja dengan tujuan menentukan kelas data baru dengan menggunakan data training sebagai acuan. Pada proses training, data yang sudah memiliki label akan diproses dan dipisahkan berdasarkan kemiripannya (gambar 1), lalu data baru akan dihitung jarak terdekat terhadap data training. Setelahnya penentuan kelas data yang didasari pada kelas mayoritas untuk k tetangga terdekat. Alur kerja dari k-NN adalah sebagai berikut [27]:

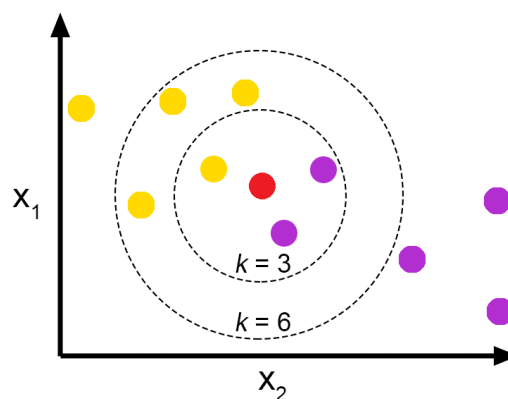
Langkah-1 : Tentukan nilai k

Langkah-2 : Hitung jarak antar data

Langkah-3 : Pilih k tetangga terdekat berdasarkan urutan jarak terdekat

Langkah-4 : Hitung label mayoritas dan jadikan label mayoritas untuk label data baru

[15] dalam bukunya menyatakan dengan segala kekurangan dan kelebihan, k-NN merupakan salah satu dari *top ten algoritma in data mining*, terutama pada proses klasifikasi.



Gambar 1 Simulasi K-Nearest Neighbor

2.2 Local Mean Distance Weigh K-Nearest Neighbor (LMDWk-NN)

Local Mean Distance k-Nearest Neighbor (LMDWk-NN) adalah sebuah metode yang dikembangkan oleh [27]. Dalam papernya LMDWk-NN ini dikembangkan untuk menutupi kelemahan dari k-NN diantaranya pada penentuan kelas data baru dengan menggunakan sistem *vote majority*.

Vote majority memungkinkan munculnya kelas mayoritas ganda dan mengabaikan jarak antar data [25]. Untuk mengatasi masalah pada sistem *vote majority* maka digabungkanlah metode *Local Mean* [28] dan *Distance Weight* [29]. Alur kerja dari LMDWk-NN adalah sebagai berikut [21]:

Langkah-1 : Tentukan nilai k

Langkah-2 : Hitung jarak antar data

Langkah-3 : Pilih k tetangga terdekat berdasarkan jarak untuk setiap kelas

Langkah-4 : Hitung bobot berdasarkan jarak untuk setiap data

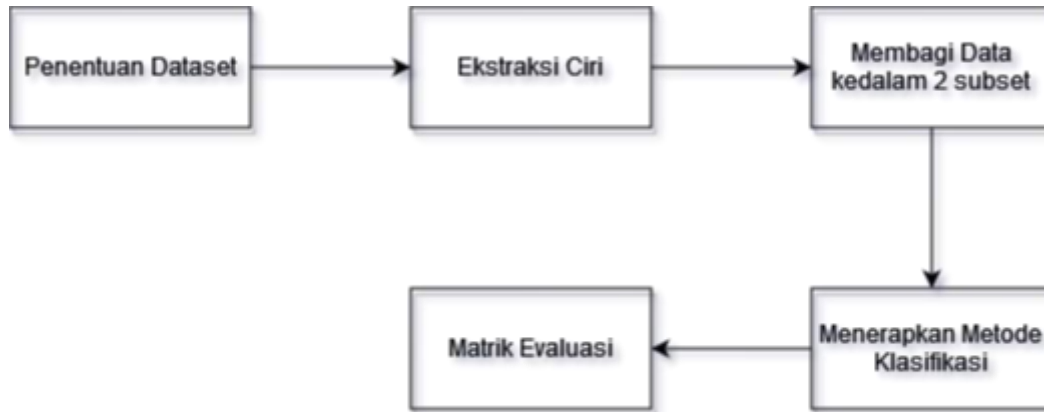
Langkah-5 : Hitung rata-rata bobot untuk setiap kelas

Langkah-6 : Jadikan kelas dengan rata-rata bobot tertinggi sebagai kelas untuk data baru

Metode LMDWk-NN ini terbukti mampu memberikan nilai akurasi yang lebih baik dari k-NN, dimana dalam penelitiannya [27] menampilkan rata-rata peningkatan akurasi dari data yang digunakan adalah sebesar 2.8%.

3. METODE PENELITIAN




Pada bagian ini, peneliti akan memberikan gambaran yang komprehensif tentang alur penelitian yang dilakukan dalam rangka mencapai tujuan penelitian. Untuk memberikan gambaran yang lebih jelas, dapat merujuk pada gambar 2.



Gambar 2 Alur Penelitian

Tahapan pertama dari penelitian ini dimulai dengan Penentuan Dataset, dimana pada penelitian ini akan menggunakan dataset berupa citra covid-19 yang bersumber dari Kaggle. Data ini memiliki 3 kelas, yaitu Normal, Covid, dan *Viral Pneumonia*. Adapun jumlah dari dataset ini terdiri dari 317 data dengan 90 untuk kelas normal, 137 data untuk kelas Covid, dan 90 data untuk kelas *Viral Pneumonia*. Adapun sebagian data yang digunakan dapat dilihat pada tabel 1.

Tabel 1 Dataset Covid

No	Gambar	Kelas
1		Normal
2		Viral Pneumonia
:	:	:
317		Covid

Tahapan selanjutnya adalah melakukan ekstraksi ciri pada citra. Adapun untuk ekstraksi ciri dari citra menggunakan algoritma *convolutional neural network* (CNN) dengan arsitektur VGG16. Kelebihan dari arsitektur ini adalah kemampuannya dalam mengekstraksi fitur-fitur yang kaya dan mendalam dari citra, sehingga memberikan citra yang mampu meningkatkan kinerja dari algoritma klasifikasi. Citra akan diubah ukurannya menjadi $224 \times 224 \text{ pixel}$, selanjutnya data yang telah diubah ukurannya tersebut akan diekstraksi cirinya, yang mana dari proses ini akan menghasilkan array satu baris untuk setiap citra.

Setelah tahapan ekstraksi ciri selesai, data akan dilakukan normalisasi dengan menggunakan metode *z-score normalization*. Selanjutnya data tersebut akan digunakan untuk melihat kinerja dari kedua algoritma. Untuk memberikan gambaran yang jelas terkait kinerja algoritma, maka digunakanlah metode *validasi* atau *testing* yaitu *Holdout Validation*. Metode ini membagi data menjadi 2 bagian, yaitu data testing dan data training dengan nilai proporsi 15% sebagai data testing dan 85% sebagai data latih. Kemudian data tersebut akan dilakukan proses klasifikasi dengan menggunakan k-NN dan LMDWk-NN, pada proses klasifikasi ini akan menggunakan *k* bernilai 1 sampai dengan 10.

Tahapan terakhir dari penelitian ini adalah analisis kinerja dari kedua algoritma, di mana pada setiap iterasi saat validasi akan diukur kinerja dari kedua algoritma dengan menggunakan metrik evaluasi seperti akurasi, presisi, dan *recall*. Setelah semua iterasi selesai, dilanjutkan dengan menghitung nilai rata-rata dari metrik evaluasi untuk mendapatkan gambaran yang lebih umum tentang kinerja dari kedua algoritma.

4. HASIL DAN PEMBAHASAN

Dataset berupa covid-19 merupakan dataset citra hasil rontgen paru-paru (dada). Data citra ini akan dilakukan ekstraksi ciri terlebih dahulu menggunakan CNN dengan arsitektur VGG16. Proses ekstraksi ciri dari citra tersebut menghasilkan data dengan jumlah ciri sebanyak 2048 ciri untuk setiap gambarnya. Adapun hasil ekstraksi ciri dapat dilihat pada tabel 2.

Tabel 2 Data Hasil Ekstraksi Ciri

No	X1	...	X512	...	X1024	X2048	Kelas
1	0.00	...	0.02	...	0.00	...	0.00	Normal
2	3.83	...	0.00	...	2.33	...	0.00	Normal
...
317	0.00	...	38.56	...	32.04	...	9.21	Covid

Sebelum dilakukan ekstraksi ciri, dilakukan normalisasi data dengan menggunakan *metode z-score normalization* (tabel 3). Data yang telah dinormalisasi ini selanjutnya dibagi menjadi 2 subset, yaitu data training dan data testing dengan jumlah data training sebanyak 269, dan jumlah data testing 48. Untuk rincian pembagian data, dapat dilihat pada tabel 4.

Tabel 3 Data Hasil Normalisasi

No	X1	...	X512	...	X1024	X2048	Kelas
1	-0.62	...	-0.79	...	-0.75	...	-0.42	Normal
2	-0.44	...	-0.80	...	-0.53	...	-0.42	Normal
3	-0.62	...	2.82	...	2.04	...	-0.42	Normal
...
317	-0.62	...	1.35	...	-0.75	...	-0.42	Covid

Tabel 4 Pembagian Data

Kelas	Training	Testing	Kelas	Training	Teting
Normal	80	10	Viral Pneumonia	75	15
Covid	114	23			

Proses selanjutnya adalah melakukan klasifikasi dengan k-NN dan LMDWk-NN menggunakan k bernilai 1 sampai 10. Untuk perbandingan kinerja dari k-NN dan LMDWk-NN dapat dilihat pada tabel 5.

Tabel 5 Kinerja Algoritma

k	Akurasi		Precision		Recall	
	M1	M2	M1	M2	M1	M2
1	89.58%	89.58%	90.01%	90.01%	89.58%	89.58%
2	85.42%	91.67%	85.89%	92.00%	85.42%	91.67%
3	89.58%	93.75%	90.30%	93.93%	89.58%	93.75%
4	89.58%	93.75%	91.44%	93.93%	89.58%	93.75%
5	87.50%	91.67%	87.98%	92.07%	87.50%	91.67%
6	93.75%	91.67%	94.47%	92.07%	93.75%	91.67%
7	91.67%	91.67%	91.68%	92.07%	91.67%	91.67%
8	85.42%	91.67%	86.24%	92.07%	85.42%	91.67%
9	85.42%	91.67%	87.68%	92.07%	85.42%	91.67%
10	85.42%	91.67%	87.68%	92.07%	85.42%	91.67%
Rata-rata	88.33%	91.88%	89.34%	92.23%	88.33%	91.88%
STD	0.028	0.011	0.026	0.010	0.028	0.011

Keterangan:

M1: k-NN

M2: LMDWk-NN

Berdasarkan tabel 5, dapat dilihat bahwa metode LMDWk-NN dan k-NN sama-sama mampu menghasilkan nilai akurasi tertinggi sebesar 93.75%, di mana pada k-NN nilai akurasi tertinggi ini dicapai saat k bernilai 6, dan LMDWk-NN saat k bernilai 3 dan 4. Sedangkan untuk *precision*, k-NN mampu mencapai nilai tertinggi sebesar 94.47% saat k bernilai 6, sementara itu LMDWk-NN hanya mampu mencapai nilai tertinggi sebesar 93.39% saat k bernilai 3 dan 4. Untuk nilai *recall*, k-NN mampu mencapai nilai tertinggi sebesar 93.75% saat k bernilai 6, dan LMDWk-NN sebesar 93.75% untuk k bernilai 3 dan 4.

Jika dilihat berdasarkan rata-rata nilai akurasi, *precision*, dan *recall* yang dihasilkan oleh kedua metode, maka LMDWk-NN terbukti memiliki kinerja yang lebih baik dari k-NN pada klasifikasi data citra ini, dimana peningkatan rata-rata LMDWk-NN terhadap k-NN pada akurasi adalah sebesar 3.5%, *precision* sebesar 2.89%, dan *recall* sebesar 3.54%. Berdasarkan nilai *standar deviasi* dari nilai akurasi, *precision*, dan *recall* juga menunjukkan bahwa nilai yang dihasilkan oleh LMDWk-NN cenderung konsisten. Namun k-NN tetap mampu menghasilkan nilai akurasi yang sama baiknya jika menemukan nilai k yang tepat. Hanya saja penentuan nilai k pada k-NN ini tidaklah mudah, penentuan nilai k pada k-NN akan tetap akan menjadi tantangan tersendiri dari algoritma ini.

5. KESIMPULAN

Berdasarkan pengujian yang dilakukan pada bagian 4, maka dapat disimpulkan bahwa kinerja LMDWk-NN lebih baik dari pada kinerja k-NN. Dimana dari parameter pengujian kinerja yang digunakan, LMDWk-NN memiliki nilai rata-rata lebih tinggi jika dibandingkan dengan k-NN. Dimana pada nilai akurasi LMDWk-NN mengalami peningkatan kinerja sebesar 3.5%, *precision* sebesar 2.89%, dan *recall* sebesar 3.54%. Meskipun begitu k-NN tetap mampu menghasilkan kinerja yang sama baiknya, yang mana kinerja k-NN akan sangat bergantung dari nilai k yang digunakan.

DAFTAR PUSTAKA

- [1] Kementerian Kesehatan Republik Indonesia, "Peran Ditjen Kesmas Daman Pandemi COVID 19 2020-2021," Jakarta, 2021.
- [2] A. Syauqi, "Jalan panjang covid19 (Sebuah Refleksi Dikala Wabah Merajalela Berdampak Pada

- Perekonomian),” *JKUBS J. Chem. Inf. Model.*, vol. 1, no. 1, pp. 1–19, 2020.
- [3] N. Mamoon and G. Rasskin, “COVID-19,” 2019. [Online]. Available: <https://www.covidvisualizer.com/>
- [4] M. F. Asriansyah, “Pandemi Covid 19 dan Upaya Pencegahan,” 2022. <https://www.djkn.kemenu.go.id/artikel/baca/15799/Pandemi-Covid-19-dan-Upaya-Pencegahan.html>
- [5] E. Martínez Chamorro, A. Díez Tascón, L. Ibáñez Sanz, S. Ossaba Vélez, and S. Borrueal Nacenta, “Radiologic diagnosis of patients with COVID-19,” *Radiologia*, vol. 63, no. 1, pp. 56–73, 2021, doi: 10.1016/j.rx.2020.11.001.
- [6] S. Ebrahimzadeh *et al.*, “Thoracic imaging tests for the diagnosis of COVID-19,” 2022. doi: 10.1002/14651858.CD013639.pub5.
- [7] S. Kamthan, H. Singh, and T. Meitzler, “Hierarchical fuzzy deep learning for image classification,” *Memories - Mater. Devices, Circuits Syst.*, vol. 2, no. June, p. 100016, 2022, doi: 10.1016/j.memori.2022.100016.
- [8] X. Zhang, H. Xiao, R. Gao, H. Zhang, and Y. Wang, “K-nearest neighbors rule combining prototype selection and local feature weighting for classification,” *Knowledge-Based Syst.*, vol. 243, 2022, doi: 10.1016/j.knosys.2022.108451.
- [9] Y. Guo, S. Han, Y. Li, C. Zhang, and Y. Bai, “K-Nearest Neighbor combined with guided filter for hyperspectral image classification,” *Procedia Comput. Sci.*, vol. 129, pp. 159–165, 2018, doi: 10.1016/j.procs.2018.03.066.
- [10] 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, doi: 10.1016/j.patrec.2006.07.002.
- [11] J. A. Romero-del-Castillo, M. Mendoza-Hurtado, D. Ortiz-Boyer, and N. García-Pedrajas, “Local-based k values for multi-label k-nearest neighbors rule,” *Eng. Appl. Artif. Intell.*, vol. 116, no. June, p. 105487, 2022, doi: 10.1016/j.engappai.2022.105487.
- [12] 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, doi: 10.1109/PCI.2012.69.
- [13] L. Cai, Y. Song, T. Liu, and K. Zhang, “A Hybrid BERT Model That Incorporates Label Semantics via Adjustive Attention for Multi-Label Text Classification,” *IEEE Access*, vol. 8, pp. 152183–152192, 2020, doi: 10.1109/ACCESS.2020.3017382.
- [14] K. Li, H. Wang, W. Wang, F. Wang, and Z. Cui, “Improving artificial bee colony algorithm using modified nearest neighbor sequence,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 8807–8824, 2022, doi: 10.1016/j.jksuci.2021.10.012.
- [15] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008. doi: 10.1007/s10115-007-0114-2.
- [16] G. I. Okolo, S. Katsigiannis, and N. Ramzan, “IEViT: An enhanced vision transformer architecture for chest X-ray image classification,” *Comput. Methods Programs Biomed.*, vol. 226, p. 107141, 2022, doi: 10.1016/j.cmpb.2022.107141.
- [17] S. Suyanto, P. E. Yunanto, T. Wahyuningrum, and S. Khomsah, “A multi-voter multi-commission nearest neighbor classifier,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6292–6302, 2022, doi: 10.1016/j.jksuci.2022.01.018.
- [18] A. Islam, S. B. Belhaouari, A. U. Rehman, and H. Bensmail, “K Nearest Neighbor OverSampling approach: An open source python package for data augmentation,” *Softw. Impacts*, vol. 12, no. February, p. 100272, 2022, doi: 10.1016/j.simpa.2022.100272.
- [19] M. Kumar, N. K. Rath, A. Swain, and S. K. Rath, “Feature Selection and Classification of Microarray Data using MapReduce based ANOVA and K-Nearest Neighbor,” *Procedia Comput. Sci.*, vol. 54, pp. 301–310, 2015, doi: 10.1016/j.procs.2015.06.035.
- [20] P. Nair *et al.*, “A new COVID-19 intubation prediction strategy using an intelligent feature selection and K-NN method,” *Int. J. Sci. Technol. Res.*, vol. 28, no. 3, pp. 221–226, 2020, doi: 10.1016/j.imu.2021.100825.
- [21] K. U. Syaliman, Yuliska, and N. F. Najwa, “Seleksi Fitur Menggunakan Pendekatan k-Nearest Neighbor,” *J. Sist. Inf. dan Teknol. Jar.*, vol. 3, no. 1, pp. 8–13, 2022.

- [22] K. U. Syaliman, A. Labellapansa, and A. Yulianti, "Improving the Accuracy of Features Weighted k-Nearest Neighbor using Distance Weight," no. ICoSET 2019, pp. 326–330, 2020, doi: 10.5220/0009390903260330.
- [23] P. A. Charde and S. D. Lokhande, "Classification Using K Nearest Neighbor for Brain Image Retrieval," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 8, pp. 760–765, 2013, doi: 10.25126/jtiik.2020722608.
- [24] L. Farokhah, "Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1129, 2020, doi: 10.25126/jtiik.2020722608.
- [25] 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, doi: 10.1016/j.eswa.2016.09.031.
- [26] 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.*, vol. 115, pp. 356–372, 2019, doi: 10.1016/j.eswa.2018.08.021.
- [27] 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, pp. 1–6, 2018, doi: 10.1088/1742-6596/978/1/012047.
- [28] Y. Mitani and Y. Hamamoto, "A local mean-based nonparametric classifier," *Pattern Recognit. Lett.*, vol. 27, no. 10, pp. 1151–1159, 2006, doi: 10.1016/j.patrec.2005.12.016.
- [29] S. A. Dudani, "The Distance-Weighted k-Nearest-Neighbor Rule," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. SMC-8, no. 4, pp. 311–313, 1978, doi: 10.1109/tsmc.1978.4309958.
- [30] N. García-Pedrajas and D. Ortiz-Boyer, "Boosting k-nearest neighbor classifier by means of input space projection," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10570–10582, 2009, doi: 10.1016/j.eswa.2009.02.065.
- [31] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 6, pp. 354–360, 2013.
- [32] Z. Lei, S. Wang, and D. Xu, "Protein sub-cellular localization based on noise-intensity-weighted linear discriminant analysis and an improved k-nearest-neighbor classifier," *Proc. - 2016 9th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2016*, no. 3, pp. 1871–1876, 2017, doi: 10.1109/CISP-BMEI.2016.7853022.