

Comparison of Distance Function in K-Nearest Neighbor Algorithm to Predict Prospective Customers in Term Deposit Subscriptions

Muhammad Tibri Syofyan, Nonong Amalita*, Dodi Vionanda, Dina Fitria

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: nongaditya@gmail.com

Submitted : 18 Maret 2023

Revised : 10 Mei 2023

Accepted : 12 Mei 2023

ABSTRACT

Data mining is often used to analysis of the big data to obtain new useful information that will be used in the future. One of the best algorithms in data mining is K-Nearest Neighbor (K-NN). K-NN classifier is a distance-based classification algorithm. The distance function is a core component in measuring the distance or similarity between the tested data and the training data. Various measure of distance function exist make this a topic of kind literature problems to determining the best distance function for the performance of the K-NN classifier. This study aims to compare which distance function produces the best K-NN performance. The distance function to be compared is the Manhattan distance and Minkowski distance. The application of K-NN classifier using bank dataset about predict prospective customers in Term Deposit Subscriptions. This study show that Minkowski distance on K-NN algorithm achieved the best result compared to Manhattan distance. Minkowski distance with power $p = 1.5$ produces an accuracy rate of 88.40% when the K value is 7. Thus, performance of K-NN algorithm using Minkowski distance ($p=1,5, K=7$) is best algorithm in predicting prospective costumers in Term Deposit Subscription.

Keywords: *classification, data mining, distance function, k-nearest neighbor, term deposit,*



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

I. PENDAHULUAN

Data mining sekarang ini menarik banyak perhatian dalam berbagai sektor, dan dengan adanya data dalam jumlah yang besar sehingga *data mining* dibutuhkan untuk mendapatkan informasi baru yang mempunyai makna penting dari sekumpulan data tersebut (Eze dkk, 2014). Dengan jumlah data yang besar, diperlukan algoritma yang tepat untuk interpretasi data yang baik (Gheware dkk, 2014). Ada banyak algoritma pada *data mining* yang dikenal dan digunakan pada berbagai penerapan (García dkk, 2015). Namun, salah satu algoritma yang terbaik dalam melakukan *data mining* adalah *K-Nearest Neighbor* (K-NN) (Wu dkk, 2008).

Menurut Prasath dkk (2019) K-NN merupakan algoritma klasifikasi tertua, sederhana, dan dapat menghasilkan akurasi yang hampir sama dengan algoritma klasifikasi lainnya. K-NN dikemukakan pertama kali oleh Fix dan Hodges (1951) dan dikembangkan oleh Cover dan Hart (1967). K-NN telah digunakan pada berbagai permasalahan seperti identifikasi kerusakan jalan (Angreni dkk, 2019), pemasaran langsung (Karim dan Rahman, 2013), identifikasi kualitas air (Tangkelayuk dkk, 2022), dan permasalahan lainnya.

K-NN termasuk klasifikasi berbasis jarak yang menggunakan fungsi jarak untuk menentukan hubungan kedekatan antara data uji dengan data latih (Walters-Williams dan Li, 2010). Sehingga fungsi jarak merupakan komponen inti untuk menghasilkan performa K-NN (Prasath dkk, 2019). Berbagai ukuran fungsi jarak yang ada dari berbagai bidang, seperti statistik, matematika, sains, dan semacamnya (Botchkarev, 2018). Terdapat 65 perhitungan fungsi jarak yang dikelompokkan menjadi sembilan jenis berdasarkan persamaan sintaksis (Cha, 2007). Hal ini menimbulkan suatu topik permasalahan pada berbagai literatur untuk menentukan fungsi jarak mana yang dapat menghasilkan performa K-NN yang terbaik (Hu dkk, 2016).

Dalam beberapa literatur, fungsi jarak Manhattan dan jarak Minkowski sering digunakan dalam menguji pengaruh fungsi jarak tersebut terhadap kinerja K-NN dalam jenis dan distribusi data yang berbeda. Pada penelitian yang dilakukan oleh Chomboon dkk (2015) menunjukkan bahwa jarak tersebut terhadap 8 dataset buatan menghasilkan akurasi yang tinggi. Lopes dan Ribeiro (2015) menunjukkan jarak Manhattan menghasilkan akurasi

terbaik terhadap beberapa dataset. Hu dkk (2016) menunjukkan jarak Minkowski menghasilkan akurasi yang cukup baik terhadap dataset medis. Wahyono dkk (2020) menunjukkan jarak Minkowski dengan p sebesar 1,5 menghasilkan akurasi terbaik terhadap data tekstual. Namun belum ada membandingkan fungsi jarak pada algoritma K-NN terhadap dataset deposito berjangka.

Berdasarkan permasalahan diatas, tulisan ini menggunakan data nasabah yang tersedia dari Institusi Bank Portugal sebagai objek penelitian dalam menentukan potensi nasabah dalam berlangganan deposito berjangka. Deposito merupakan sumber dana utama dalam keuangan Bank (Zhuang dkk, 2018). Prediksi tanggapan nasabah yang diperoleh dari pemasaran bank merupakan target untuk meningkatkan peluang nasabah dalam berlangganan deposito berjangka (Ruangthong dan Jaiyen, 2015). Hal ini dapat menambah efisiensi kerja bank dalam pemasaran deposito berjangka (Nuraeni, 2021). Prediksi nasabah ini pada dasarnya merupakan suatu permasalahan klasifikasi dalam menentukan pelanggan mana yang memiliki potensi tinggi untuk berlangganan deposito berjangka (Guo dan Hou, 2019). Sehingga pada penelitian ini akan membandingkan jarak Manhattan dan jarak Minkowski pada algoritma K-NN untuk memprediksi potensi nasabah dalam berlangganan deposito berjangka agar dapat meningkatkan strategi pemasaran bank.

II. METODE PENELITIAN

Jenis penelitian ini adalah penelitian studi kasus. Menurut Yin (2014) penelitian studi kasus merupakan suatu penelitian yang memiliki tujuan untuk menguji pertanyaan dan masalah penelitian yang tidak dapat dipisahkan antara fenomena dan institusi dimana fenomena tersebut terjadi, dimana tulisan ini menggunakan institusi bank luar sebagai objek penelitian yang memiliki permasalahan yang sama di Institusi Bank Indonesia. Sehingga, tulisan ini menggunakan data nasabah yang tersedia dari Institusi Bank Portugal dimana pihak bank mencoba memanggil nasabah yang berpotensi berlangganan deposito berjangka (Morani dkk, 2018).

Dataset memiliki data berjumlah besar, yaitu sebanyak 32.950 nasabah dan 13 variabel, diantaranya 12 variabel penjelas dan 1 variabel respon. Berikut penjelas variabel penelitian pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Tipe Data	Keterangan Variabel	Kategori
X1	<i>Age</i>	Numerik	Usia nasabah	
X2	<i>Job</i>	Kategorik	Pekerjaan nasabah	administrasi, jasa, mahasiswa, manajemen, pekerja kasar, pembantu rumah tangga, pengangguran, pensiunan, teknisi, wiraswasta, dan wirausaha
X3	<i>Marital</i>	Kategorik	Status pernikahan nasabah	cerai, menikah, tidak menikah
X4	<i>Education</i>	Kategorik	Pendidikan terakhir nasabah	dasar 4 tahun, dasar 6 tahun, dasar 9 tahun, high school, buta huruf, kursus profesional, lulusan universitas
X5	<i>Housing</i>	Kategorik	Apakah nasabah punya kredit perumahan?	ya, tidak
X6	<i>Loan</i>	Kategorik	Apakah nasabah punya pinjaman untuk pribadi?	ya, tidak
X7	<i>Contact</i>	Kategorik	Jenis komunikasi kontak dengan nasabah	seluler, telepon
X8	<i>Day</i>	Kategorik	Hari kontak terakhir dalam minggu	senin, selasa, rabu, Kamis, jum'at
X9	<i>Month</i>	Kategorik	Bulan kontak terakhir tahun ini	maret, april, mei, juni, juli, agustus, september, oktober, november, desember
X10	<i>Duration</i>	Numerik	Durasi kontak terakhir, dalam detik	
X11	<i>Campaign</i>	Numerik	Jumlah kontak yang dilakukan selama pemasaran ini dan untuk nasabah ini	
X12	<i>Poutcome</i>	Numerik	Hasil dari pemasaran sebelumnya	gagal, tidak ada, sukses
Y	<i>Deposit Term</i>	Kategorik	Apakah nasabah berlangganan deposito?	ya, tidak

Sebelum melakukan analisis data menggunakan algoritma K-NN. Data mentah perlu dilakukan *pre-processing data*, karena data mentah biasanya memiliki data hilang, redundansi data, kesalahan penginputan, dan permasalahan lainnya yang dapat mengganggu proses kerja maupun hasil algoritma K-NN (Garcia dkk, 2015). Proses yang dilakukan pada tahap ini, yaitu dilakukan penanganan data hilang, redundansi data, dan data tak seimbang. Penanganan data hilang dan redundansi data dapat ditangani dengan mengeluarkan nasabah yang mengandung hal tersebut dari dataset, penanganan data tak seimbang menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) dengan membentuk sampel buatan berdasarkan kemiripan antar data (Ruangthong dan Jaiyen,

2015). Sehingga jumlah nasabah yang digunakan pada penelitian ini sebanyak 54.346 nasabah dengan nasabah yang berlangganan deposito berjangka dan tidak berlangganan memiliki jumlah yang sama.

Setelah data dilakukan *pre-processing data*, maka akan dilakukan analisis data menggunakan K-NN. Algoritma K-NN adalah algoritma klasifikasi data dengan menghitung kedekatan antara data baru dengan data lama berdasarkan kemiripan karakteristik pada fitur yang ada (Kusrini dan Luthfi, 2009). K-NN disebut juga sebagai *lazy learning*, karena K-NN melakukan klasifikasi tanpa membentuk model terlebih dahulu namun menggunakan langsung data latih dan data uji untuk melakukan klasifikasi (Leidiyana, 2013). Ada dua komponen penting dalam menggunakan algoritma K-NN, yaitu parameter K untuk menentukan ukuran tetangga terdekat (Angreni dkk, 2019) dan fungsi jarak untuk menghitung kedekatan antara data latih dan data uji (Yang dan Jin, 2006). Menurut Wu dkk (2008) jika nilai parameter K terlalu kecil maka hasil akan sensitif terhadap *noise*, dan sebaliknya jika nilai parameter K terlalu besar akan mengurangi pengaruh terhadap *noise* namun hasil klasifikasi akan mudah menghasilkan *overfitting* karena memasukkan terlalu banyak titik dari kelas lain. Nilai parameter K yang ideal adalah nilai yang terkecil namun terhindar dari *noise* dan *overfitting*. Parameter K yang digunakan pada penelitian adalah bilangan ganjil positif dari 1 hingga 15.

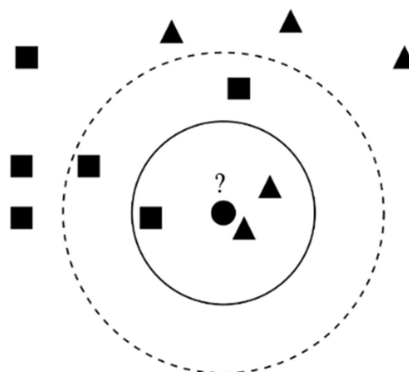
Proses selanjutnya menghitung kedekatan antar data latih dengan data uji menggunakan fungsi jarak, Fungsi jarak yang digunakan adalah jarak Manhattan dan jarak Minkowski. Jarak Manhattan atau *City Block Distance* adalah jumlah dari panjang ruas garis kedua titik terhadap tiap sumbu dalam koordinat Kartesius. Nama jarak ini berasal dari tata letak jalan di Pulau Manhattan yang berbentuk kisi-kisi segiempat. Menurut Prasath dkk (2019) rumus jarak Manhattan dihitung dengan persamaan (1).

$$d_{\text{Manhattan}}(A,B) = \sum_{i=1}^f |a_i - b_i| \tag{1}$$

Jarak Minkowski adalah jarak antar dua titik data dalam vektor bernorma. Perhitungan jarak menggunakan parameter p yang nilainya lebih besar dari nol ($p > 0$). Jarak Minkowski disebut juga sebagai generalisasi jarak Manhattan. Jika nilai parameter p bernilai 1 ($p = 1$) maka akan membentuk jarak Manhattan (Yean dkk, 2018). Azevedo dan Machado (2013) menunjukkan nilai parameter p terbaik bernilai 1,5. Menurut Prasath dkk (2019) rumus jarak Minkowski dihitung dengan persamaan (2).

$$d_{\text{Minkowski}}(A,B) = \sqrt[p]{\sum_{i=1}^f |a_i - b_i|^p} \tag{2}$$

Pada tahap klasifikasi ditentukan berdasarkan mayoritas label pada K tetangga terdekat. Tetangga terdekat ditentukan berdasarkan jarak yang terkecil. Berikut contoh penerapan parameter K pada Gambar 1.



Gambar 1. Ilustrasi penggunaan parameter K pada algoritma K-NN
(Sumber: Abu Alfeilat dkk, 2019)

Gambar 1 merupakan ilustrasi kedekatan sampel uji dengan data latih setelah dihitung kedekatan atau kemiripan data. Pada Gambar 1 terdapat tiga simbol yaitu persegi, segitiga, dan lingkaran. Persegi dan segitiga merupakan hasil klasifikasi pada data latih, sedangkan lingkaran merupakan sampel uji yang ingin diklasifikasi. Penggunaan parameter K berdasarkan Gambar 1, jika parameter $K = 3$ maka lingkaran diklasifikasikan menjadi kategori segitiga dan jika $K = 5$ maka lingkaran diklasifikasikan menjadi kategori persegi.

Perhitungan kinerja algoritma K-NN menggunakan *confusion matrix* yang mana memuat informasi tentang klasifikasi aktual dan prediksi. *Confusion matrix* memiliki dua dimensi, satu dimensi berindeks sebagai kelas aktual sebuah objek, dimensi lainnya berindeks sebagai kelas yang diprediksi oleh klasifikasi (Deng dkk, 2016). Tabel 2 memperlihatkan bentuk dasar *confusion matrix* untuk klasifikasi multi-kelas, dengan label A_1, A_2, \dots, A_n , dan N_{ij} merupakan jumlah sampel aktual yang termasuk ke label A_i tetapi terklasifikasikan sebagai label A_j .

Tabel 2. *Confusion Matrix*

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

(Sumber: Deng dkk, 2016).

Pengukuran kinerja klasifikasi yang digunakan pada algoritma K-NN antara lain akurasi, presisi, dan *recall*. Akurasi merupakan rasio hasil prediksi sesuai dengan hasil aktual terhadap keseluruhan data. Rumus perhitungan akurasi menggunakan Persamaan (3).

$$\text{Accuracy} = \frac{\sum_{i=1}^n N_{ii}}{\sum_{i=1}^n \sum_{j=1}^n N_{ij}} \tag{3}$$

Presisi merupakan rasio hasil prediksi sesuai dengan hasil aktual yang bersifat positif atau negatif terhadap keseluruhan data yang bersifat sama. Rumus perhitungan presisi untuk setiap variabel menggunakan Persamaan (4).

$$\text{Precision}_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ki}}, k = \text{banyaknya kategori} \tag{4}$$

Sensitifitas merupakan rasio hasil prediksi sesuai dengan hasil aktual yang bersifat positif atau negatif terhadap keseluruhan data yang hasil aktualnya bersifat sama. Rumus perhitungan sensitifitas untuk setiap variabel menggunakan Persamaan (5).

$$\text{Sensitifitas}_i = \frac{N_{ii}}{\sum_{k=1}^n N_{ik}}, k = \text{banyaknya kategori} \tag{5}$$

III. HASIL DAN PEMBAHASAN

A. *K-Nearest Neighbor*

Terdapat beberapa tahapan dalam menjalankan algoritma K-NN, yaitu penentuan parameter K, perhitungan kedekatan atau kemiripan antar data, klasifikasi berdasarkan mayoritas kategori pada K buah tetangga terdekat. Parameter K yang digunakan adalah bilangan ganjil dengan awalan satu hingga lima belas. Perhitungan kedekatan antar data menggunakan jarak Manhattan dan jarak Minkowski dengan parameter $p=1,5$.

Hasil perhitungan jarak Manhattan terhadap kinerja algoritma K-NN pada Tabel 3.

Tabel 3. Kinerja Jarak Manhattan

Parameter K	Akurasi	Presisi	Sensitifitas
1	0,8735	0,8616	0,8927
3	0,8784	0,8595	0,9073
5	0,8824	0,8642	0,9100
7	0,8816	0,8635	0,9092
9	0,8828	0,8648	0,9101
11	0,8820	0,8640	0,9093
13	0,8795	0,8624	0,9058
15	0,8787	0,8628	0,9034

Pada Tabel 3 menunjukkan bahwa tiap parameter K yang digunakan menghasilkan akurasi yang berbeda. Nilai akurasi yang tertinggi dihasilkan pada parameter K = 9. Hal ini menunjukkan kinerja algoritma K-NN menggunakan jarak Manhattan dengan parameter K = 9 menghasilkan akurasi sebesar 88,28%. Presisi menunjukkan bahwa sebesar

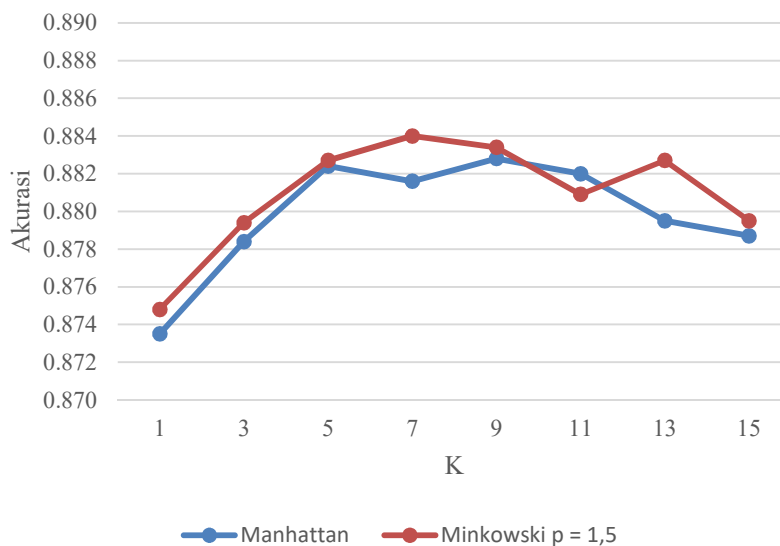
86,4% nasabah yang diprediksi benar-benar berlangganan deposito berjangka dari keseluruhan nasabah yang diprediksi berlangganan deposito berjangka. Sensitifitas menunjukkan bahwa sebesar 91,01% nasabah yang benar diprediksi berlangganan deposito berjangka dari keseluruhan nasabah yang berlangganan deposito berjangka.

Hasil perhitungan jarak Minkowski dengan parameter $p = 1,5$ terhadap kinerja algoritma K-NN pada Tabel 4.

Tabel 4. Kinerja Jarak Minkowski ($p = 1,5$)

Parameter K	Akurasi	Presisi	Sensitifitas
1	0,8748	0,8637	0,8928
3	0,8794	0,8606	0,9082
5	0,8827	0,8638	0,9112
7	0,8840	0,8648	0,9127
9	0,8834	0,8649	0,9114
11	0,8809	0,8630	0,9083
13	0,8827	0,8650	0,9093
15	0,8795	0,8621	0,9062

Pada Tabel 4 menunjukkan bahwa tiap parameter K yang digunakan menghasilkan akurasi yang berbeda. Nilai akurasi yang tertinggi dihasilkan pada parameter $K = 7$. Hal ini menunjukkan kinerja algoritma K-NN menggunakan jarak Manhattan dengan parameter $K = 7$ menghasilkan akurasi sebesar 88,40%. Presisi menunjukkan bahwa sebesar 86,48% nasabah yang diprediksi benar-benar berlangganan deposito berjangka dari keseluruhan nasabah yang diprediksi berlangganan deposito berjangka. Sensitifitas menunjukkan bahwa sebesar 91,27% nasabah yang benar diprediksi berlangganan deposito berjangka dari keseluruhan nasabah yang berlangganan deposito berjangka. Untuk memperoleh fungsi jarak yang terbaik akan dibandingkan hasil kinerja setiap jarak yang ditampilkan pada Gambar 2.



Gambar 2. Hasil Akurasi Klasifikasi K-NN

Gambar 2 merupakan grafik hasil akurasi klasifikasi algoritma K-NN pada dataset pemasaran deposito berjangka untuk jarak Manhattan dan jarak Minkowski ($p = 1,5$). Grafik ini membandingkan kinerja algoritma K-NN terhadap fungsi jarak yang digunakan. Perbandingan ini berdasarkan pada hasil akurasi yang diperoleh untuk setiap parameter K dan fungsi jarak yang digunakan. Pada Gambar 2 menunjukkan bahwa jarak Minkowski menghasilkan akurasi yang signifikan lebih baik dari jarak Manhattan. Hal ini ditunjukkan akurasi tertinggi pada jarak Minkowski dengan parameter $K = 7$ sebesar 88,40%.

IV. KESIMPULAN

Pemilihan fungsi jarak pada algoritma K-NN merupakan hal yang penting untuk dikaji guna meningkatkan kinerja algoritma K-NN terhadap permasalahan, khususnya prediksi potensi nasabah dalam berlangganan deposito

berjangka. Jarak Minkowski ($p = 1,5$) menghasilkan kinerja algoritma K-NN yang lebih baik dari jarak Manhattan. Akurasi yang diperoleh tertinggi terdapat pada penggunaan parameter $K = 7$ sebesar 88,40%. Sehingga kinerja algoritma K-NN dalam memprediksi potensi nasabah dalam berlangganan deposito berjangka yang terbaik adalah menggunakan jarak Minkowski dengan parameter $p = 1,5$ dan parameter $K = 7$. Namun tidak menutup kemungkinan bahwa menggunakan beberapa fungsi jarak yang lain dapat berpeluang menghasilkan kinerja algoritma K-NN lebih baik terutama dalam memprediksi potensi nasabah dalam berlangganan deposito berjangka. Maka dapat disimpulkan bahwa jarak Minkowski lebih baik digunakan pada algoritma K-NN daripada jarak Manhattan untuk memprediksi potensi nasabah dalam berlangganan deposito berjangka.

DAFTAR PUSTAKA

- Abu Alfeilat, H. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, M. B., Eyal Salman, H. S., & Prasath, V. B. S. (2019). Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data*, 7(4), 221–248.
- 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.
- Azevedo, F., & Machado, J. T. (2013). Multidimensional Scaling Analysis of Electricity Market Prices. In A. Madureira, C. Reis, & V. Marques (Eds.), *Computational Intelligence and Decision Making* (Vol. 61, pp. 345–354). Springer Netherlands.
- Botchkarev, A. (2018). *Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology*.
- Cha, S.-H. (2007). Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. *International Journal of Mathematical Models and Methods in Applied Sciences*, 1(4), 300–307.
- Chomboon, K., Chujai, P., Teerassammee, P., Kerdprasop, K., & Kerdprasop, N. (2015). *An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm*. 280–285.
- Cover, T., & Hart, P. E. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1), 21–27.
- 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.
- Eze, U. F., Adeoye, O. S., & Ikemelu, C. R. (2014). Industry Wide Applications of Data Mining. *International Journal of Advanced Studies in Computer Science and Engineering*, 3(2), 10.
- Fix, E., & Hodges, J. L. (1951). *Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties* (No. 4). USAF School of Aviation Medicine.
- García, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining* (Vol. 72). Springer International Publishing.
- Gheware, S. D., Kejkar, A. S., & Tondare, S. M. (2014). Data Mining Task Tools Techniques and Applications. *International Journal of Advanced Research in Computer and Communication Engineering*, 3(10), 8095–8098.
- Guo, J., & Hou, H. (2019). Statistical Decision Research of Long-Term Deposit Subscription in Banks Based on Decision Tree. *2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, 614–617.
- Hu, L.-Y., Huang, M.-W., Ke, S.-W., & Tsai, C.-F. (2016). The Distance Function Effect on K-Nearest Neighbor Classification for Medical Datasets. *SpringerPlus*, 5(1), 1304.

- Karim, M., & Rahman, R. M. (2013). Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. *Journal of Software Engineering and Applications*, 6(4), 196–206.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining* (1st ed.). Penerbit ANDI.
- Leidiyana, H. (2013). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, 1(1), 65–76.
- Lopes, N., & Ribeiro, B. (2015). *On the Impact of Distance Metrics in Instance-Based Learning Algorithms BT - Pattern Recognition and Image Analysis* (R. Paredes, J. S. Cardoso, & X. M. Pardo, Eds.; pp. 48–56). Springer International Publishing.
- Morani, K., Ayana, E. K., & Engin, S. N. (2018). Development of Prediction in Clients' Consent to a Bank Term Deposit Using Feature Selection. *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, 1–5.
- Nuraeni, N. (2021). Klasifikasi Data Mining untuk Prediksi Potensi Nasabah dalam Membuat Deposito Berjangka. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 3(1), 65–75.
- Prasath, V. B. S., Alfeilat, H. A. A., Hassanat, A. B. A., Lasassmeh, O., Tarawneh, A. S., Alhasanat, 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. *2015 12th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 93–96.
- Tangkelayuk, A., Mailoa, E., & Diponegoro, J. (2022). Klasifikasi Kualitas Air Menggunakan Metode K-NN, Naïve Bayes Dan Decision Tree. *Jurnal Teknik Informatika dan Sistem Informasi*, 9(2), 1109–1119.
- Wahyono, W., Trisna, I. N. P., Sariwening, S. L., Fajar, M., & Wijayanto, D. (2020). Comparison of Distance Measurement on K-Nearest Neighbour in Textual Data Classification. *Jurnal Teknologi Dan Sistem Komputer*, 8(1), 54–58.
- Walters-Williams, J., & Li, Y. (2010). Comparative Study of Distance Functions for Nearest Neighbors. In K. Elleithy (Ed.), *Advanced Techniques in Computing Sciences and Software Engineering* (pp. 79–84). Springer Netherlands.
- 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*, 14(1), 1–37.
- Yang, L., & Jin, R. (2006). Distance Metric Learning: A Comprehensive Survey. *Michigan State University*, 2.
- Yean, C. W., Khairunizam, W., Omar, M. I., Murugappan, M., Zheng, B. S., Bakar, S. A., Razlan, Z. M., & Ibrahim, Z. (2018). Analysis of The Distance Metrics of KNN Classifier for EEG Signal in Stroke Patients. *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, 1–4.
- Yin, R. K. (2014). *Case Study Research: Design and Methods* (Fifth edition). SAGE.
- Zhuang, Q. R., Yao, Y. W., & Liu, O. (2018). Application of Data Mining in Term Deposit Marketing. *Proceedings of The International MultiConference of Engineers and Computer Scientists 2018*, 2, 707–710.