

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

SKRIPSI

Diajukan Untuk Skripsi Guna Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Program Studi Sistem Informasi Fakultas Teknik
Universitas Nusantara PGRI Kediri



OLEH :

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1.03.03.0023

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA

UN PGRI KEDIRI

2022

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

SKRIPSI

Diajukan Untuk Skripsi Guna Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer (S.Kom.)

Program Studi Sistem Informasi Fakultas Teknik
Universitas Nusantara PGRI Kediri



OLEH :

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1.03.03.0023

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI**

2022

ii

Skripsi oleh

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1.03.03.0023

Judul:

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

Telah Disetujui untuk diajukan Kepada
Panitia Ujian/Sidang Skripsi Program
Studi Sistem Informasi UN PGRI Kediri

Tanggal: 22 juli 2022

Pembimbing I



Aldina Ristvawan, M.Kom

NIDN. 0721018801

Pembimbing II



M.Nailiuloh Muzaki, M.Cs

NIDN. 0706098902

Skripsi oleh:

MUHAMMAD ZAINAL ARIFIN

NPM : 18.1 03 03 0023




Judul:

**Perbandingan Performa Cluster Dalam Penentuan
Jenis Bunga Iris**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi Program
Studi Sistem Informasi FT UN PGRI Kediri
Pada tanggal: 22 juli 2022

Dan Dinyatakan telah Memenuhi Persyaratan

Panitian Penguji:

- | | | |
|---------------|-----------------------------|---|
| 1. Ketua | : Aidina Ristyawan, M.Kom |  |
| 2. Penguji | : Anita Sari Wardani, M.Kom |  |
| 3. Penguji II | : M.Najibulloh Muzaki, M.Cs |  |

Mengetahui,
Dekan Fakultas Teknik

Dr. Suryo Widodo, M.Pd
NIP. 19640202 199103 1002

PERNYATAAN

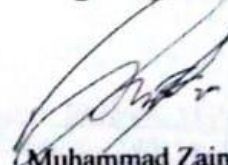
Yang bertanda tangan di bawah ini saya,

Nama : Muhammad Zainal Ariifin
Jenis Kelamin : Laki-Laki
Tempat/tgl.lahir : Nganjuk/ 23 Juli 1999
NPM : 18.1.03.03.0023
Fak./Jur./Prodi. : Teknik / Sistem Informasi

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 22 juli 2022

Yang Menyatakan



Muhammad Zainal Ariifin
NPM: 18.1.03.03.0023

MOTTO

Kertas kosong menjadi bernilai saat tergores oleh tinta

Jiwa yang kosong menjadi berarti saat hati mengukirkan makna kehidupan

Abstrak

Muhammad Zainal Arifin Data Mining Perbandingan Performa Cluster Dalam Penentuan Jenis Bunga Iris, Skripsi, Fakultas Teknik, Universitas Nusantara PGRI Kediri, 2022.

Penelitian ini dilatar belakangi dari Sudah banyak algoritma *clustering* yang dipakai dalam penelitian namun ada beberapa algoritma *clustering* yang populer digunakan dalam analisis komparasi seperti algoritma *K-Means*, *K-Medoids*, *C-Means*, *Hard C-Means*, dan *X-Means*. Namun banyak penelitian yang hanya masih menggunakan satu algoritma *clustering* dalam penentuan jenis bunga *iris*.

Penelitian ini menggunakan 3 algoritma clustering dalam analisis komparasi algoritma yaitu *K-Means*, *X-Means*, dan *K-Medoids*. Dataset menggunakan *data iris* dari *UCI Machine Learning Repository*. Penelitian ini menggunakan nilai *k* yang digunakan adalah tiga kluster untuk diuji menggunakan nilai *Davies Bouldin Index* (DBI) serta menggunakan Validasi *K-Fold Cross Validation* untuk mendapatkan nilai interval dengan membagi data training dan testing dan metode perbandingan uji beda T-test untuk mengetahui ada perbedaan signifikan atau tidak dari perbandingan nilai *Davies Bouldin Index*, sehingga diperoleh algoritma yang memiliki sifat dominan pada tahapan preprosesing data dengan menggunakan data tranformation yaitu berupa metode *Z-score normalization*.

Kesimpulan hasil dari penelitian ini adalah nilai *Davies Bouldin Index* (DBI) yang didapat adalah sebesar 0,839 untuk algoritma *K-Means*, *X-Means* sebesar 0,841, dan *K-Medoids* sebesar 1,498 , sedangkan dari hasil uji beda t-test yang dilakukan bahwa algoritma *K-Means* dan *X-Means* merupakan algoritma dengan hasil yang paling dominan dari ketiga algoritma yang diterapkan pada dataset iris. Sedangkan untuk algoritma *K-Medoids* merupakan algoritma dengan hasil yang kurang baik dengan nilai *Davies Bouldin Index* (DBI) sebesar 1.498.

Kata Kunci – *Clustering, Algoritma K-Means, Algoritma X-Means, Algoritma K-Medoids*

KATA PENGANTAR

Puji Syukur Kami panjatkan kehadirat Allah Tuhan Yang Maha Kuasa, karena hanya atas perkenannya tugas penyusunan skripsi ini dapat diselesaikan.

Skripsi dengan judul “ **Perbandingan Performa Cluster Dalam Penentuan Jenis Bunga Iris** ” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer, pada Program Studi Sistem Informasi UN PGRI Kediri.

Pada kesempatan ini diucapkan terimakasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak Dr. Zaenal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Bapak Dr. Suryo Widodo, M.Pd. selaku Dekan Fakultas Teknik Universitas Nusantara PGRI Kediri.
3. Ibu Rina Firliana, M.Kom. selaku Ketua Program Studi Sistem Informasi Universitas Nusantara PGRI Kediri.
4. Bapak Aidina Ristyawan, M.Kom selaku Dosen pembimbing I yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
5. Bapak M.Najibulloh Muzaki, M.Cs selaku Dosen pembimbing II yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
6. Ibu Anita Sari Wardani, M.Kom selaku Penguji yang telah memberikan saran dan masukan untuk perbaikan dan penyelesaian skripsi ini.
7. Orangtua tercinta, Ayah Ibunda serta saudara-saudaraku atas ketulusan doa, motivasi dan nasihat-nasehat sepanjang waktu sehingga dapat menyelesaikan skripsi ini.

Kediri, 22 Juli 2022



MUHAMMAD ZAINAL ARIFIN
NPM : 18.1.03.03.0023

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PESETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
MOTTO.....	v
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Batasan Masalah.....	4
1.3. Rumusan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
1.6. Sistematika Penulisan.....	5
BAB II LANDASAN TEORI.....	7
2.1. Kajian Teori.....	7
2.1.1. Data Mining.....	7
2.1.2. <i>Cluster</i>	10
2.1.3. <i>K-Means</i>	11
2.1.4. <i>X-Means</i>	12
2.1.5. <i>K-Medoids</i>	14
2.1.6. Perhitungan Jarak (<i>Measures Distance</i>).....	15
2.1.7. <i>Euclidean Distance</i>	16
2.1.8. <i>Davies Buldin index</i>	16
2.1.9. <i>K-Fold Cross Validation</i>	18
2.1.10. Parametrik uji beda T-test.....	19
2.1.11. Dataset.....	20
2.1.12. Rapidminer.....	21
2.2. Kajian Penelitian terdahulu.....	21
BAB III METODE PENELITIAN.....	24

3.1. Metode Penelitian.....	24
3.2. Metode Pengumpulan Data	24
3.3. Metode Analisis Data	25
3.3.1. Jenis Analisis.....	25
3.4. Alur Penelitian.....	25
BAB IV HASIL DAN PEMBAHASAN	29
4.1. Pengumpulan Data	29
4.2. <i>Preprocessing</i> Data	30
4.2.1. Transformasi Data.....	31
4.3. Implementasi Algoritma Clustering	33
4.3.1. Pemodelan <i>K-Means</i>	33
4.3.2. Pemodelan <i>X-Means</i>	34
4.3.3. Pemodelan K-Medoids.....	36
4.4. Evaluasi	37
4.5. Metode Perbandingan.....	39
BAB V PENUTUP.....	41
4.6. Kesimpulan.....	41
4.7. Saran.....	41
DAFTAR PUSTAKA	43
LAMPIRAN.....	45

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	22
Tabel 4.1 <i>Data sets</i> yang digunakan	29
Tabel 4.2 Hasil Normalisasi Data	32
Tabel 4.3 Perbandingan Performance tiga Algoritma.....	38
Tabel 4.4 Hasil uji beda T-test	39

DAFTAR GAMBAR

Gambar 1.1 Bunga Iris	1
Gambar 2.1 Bidang Ilmu <i>Data Mining</i>	8
Gambar 2.2 Tahapan KDD	9
Gambar 2.3 Skenario Pengujian <i>10-Fold Cross Validation</i>	19
Gambar 3.1 Alur Penelitian.....	26
Gambar 4.1 visualisasi data iris	30
Gambar 4.2 Rancangan Model <i>K-Means</i>	33
Gambar 4.3 Desain Bagian Cross Validation	34
Gambar 4.4 Hasil <i>description Performance Vector</i>	34
Gambar 4.5 Rancangan Model <i>X-Means</i>	35
Gambar 4.6 Desain Bagian Cross Validation	35
Gambar 4.7 Hasil <i>description Performance Vector</i>	36
Gambar 4.8 Rancangan Model <i>K-Medoids</i>	36
Gambar 4.9 Desain Bagian Cross Validation	37
Gambar 4.10 Hasil <i>description Performance Vector</i>	37
Gambar 4.11 Desain Perbandingan uji beda T-test.....	39

DAFTAR LAMPIRAN

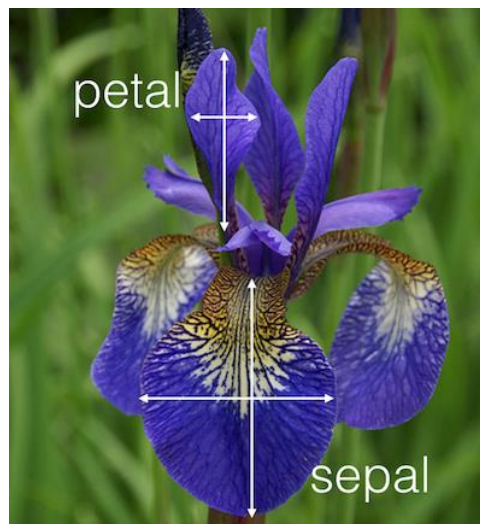
Lampiran 1 <i>Data Iris</i>	1
Lampiran 2 <i>Data Iris Z-Score Normalization</i>	8

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Data iris merupakan data dari 150 bunga, Gambar 1.1 menunjukkan bahwa bunga iris diidentifikasi berdasarkan *Sepal Length*, *Sepal Width*, *Petal Length*, *Petal Width*. Dari 150 data tersebut pada umumnya peneliti-peneliti sebelumnya mengelompokkan menjadi tiga kelompok bunga, yaitu *iris setosa*, *iris virginica* dan *iris versi color*. Dalam pengujian metode cluster banyak dari peneliti sebelumnya yang menggunakan *data iris*, karena data iris merupakan data yang gampang untuk didapatkan dan sederhana (Candrasari Hermanto, 2017).



Gambar 1.1 Bunga iris (Taha Chicho et al, 2021)

Data mining adalah ekstraksi informasi dari sejumlah data yang besar untuk melihat pengetahuan yang tersembunyi dan memfasilitasi pemakainya secara *real time*. *Data mining* memiliki algoritma yang berbagai macam untuk analisis data.

Beberapa algoritma *data mining* sering digunakan untuk analisis data adalah *Clustering, Association, Classification* dan lain-lain. *Cluster* adalah teknik untuk analisis data eksplorasi secara efektif (Herviany et al, 2021).

Analisis *cluster* adalah teknik multivariat yang memiliki tujuan primer buat mengelompokkan objek-objek menurut karakteristik yang dimiliki (Awalludin & Taufik, 2017). Analisis cluster mengklasifikasi objek sehingga objek-objek yang paling dekat kesamaanya menggunakan objek lain berada pada cluster yang sama (Sari & Sukestiyarno, 2021).

Di antara banyak algoritma *clustering*, algoritma *clustering K-means* banyak digunakan karena algoritmanya yang sederhana dan konvergensi yang cepat. Namun, nilai K dari pengelompokan perlu diberikan terlebih dahulu dan pilihan nilai K secara langsung mempengaruhi hasil konvergensi (Yuan & Yang, 2019).

K-Medoids adalah salah satu metode partisi, karena menggunakan objek yang paling terpusat (*medoids*) di *cluster* menjadi pusat *cluster* dari nilai rata-rata objek dalam sebuah *cluster*. Metode *K-medoids* lebih cocok untuk mengelompokkan data dibandingkan metode *K-Means* (Marlina et al, 2018).

Dalam Adhitama et al. (2020), Pelleg dan Moore melakukan penelitian pada tahun 2000 untuk melakukan ekstensi kepada metode *K-Means* untuk menghasilkan proses komputasi yang lebih cepat dan melakukan estimasi jumlah klaster yang paling efisien dalam pemrosesan data. Hasil dari penelitian tersebut adalah metode *X-Means* (*extended K-Means*) dapat digunakan pada dataset yang kecil dan proses komputasi yang lebih cepat dibandingkan dengan *K-Means* .

Sudah banyak algoritma *clustering* yang dipakai dalam penelitian namun ada beberapa algoritma *clustering* yang populer digunakan dalam analisis komparasi seperti algoritma *K-Means*, *K-Medoids*, *C-Means*, *Hard C-Means*, dan *X-Means*. Namun banyak penelitian yang hanya masih menggunakan satu algoritma *clustering* dalam penentuan jenis bunga *iris* (Candrasari Hermanto, 2017).

Data yang digunakan didalam penelitian ini adalah *data set* iris, data tersebut dipublikasikan oleh UCI Repository yang mana data tersebut sudah diakui oleh penelitian yang bergerak dalam data science, Metode yang dipakai dalam penelitian ini adalah Algoritma *X-Means*, *K-Means*, *K-Medoids*. Penelitian ini membandingkan performa dari ketiga algoritma Dengan menghitung nilai *Davies Bouldin Index* (DBI) serta menggunakan Validasi *K-Fold Cross Validation* untuk mendapatkan nilai interval dengan membagi data *training* dan *testing* dan metode perbandingan uji beda T-test untuk mengetahui ada perbedaan signifikan atau tidak dari perbandingan nilai *davies bouldin index*, sehingga diperoleh algoritma yang optimal dari ketiga algoritma dengan *data set iris*. Pada tahapan preprosesing data dengan menggunakan *data tranformation* yaitu berupa metode *Z-score normalization*.

Clustering dipilih dalam pemrosesan data karena mampu menentukan K terbaik pada pengelompokan dalam *data set*. Penganalisaan k suatu cara efektif karena hasil *cluster* bergantung pada jumlah k yang digunakan.

Berdasarkan latar belakang yang diuraikan maka dilakukan komparasi algoritma yang paling optimal dalam clustering bunga iris.

1.2. Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut :

1. *Dataset* yang digunakan dalam penelitian ini berasal dari data sekuler yang diperoleh dari *UCI Repository* berupa data iris.
2. Algoritma *X-Means*, *K-Means* dan *K-Medoids* menggunakan Metode Validasi *K-Fold Cross Validation* dan Evaluasi DBI.
3. Untuk metode perbandingan menggunakan metode uji beda T-test

1.3. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana menetapkan performa *clustering* yang optimal dari komparasi Algoritma *X-Means*, *K-Means*, dan *K-Medoids*?
2. Bagaimana membandingkan hasil *cluster* dari 3 algoritma menggunakan metode Evaluasi DBI dan uji beda T-test?

1.4. Tujuan Penelitian

Sesuai dengan rumusan masalah yang telah dibahas diatas maka tujuan dari penelitian ini adalah:

1. Membandingkan performa dari ketiga algoritma dengan menghitung nilai *Davies Bouldin Index* (DBI) yang menggunakan metode DBI.
2. Mengetahui algoritma *Cluster* yang optimal untuk penelitian data iris.

1.5. Manfaat Penelitian

Dalam penelitian ini diharapkan dapat memberikan manfaat berupa, menjadi perbandingan peneliti lain dalam menganalisis menggunakan metode *clustering*.

1.6. Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri dari lima bagian, yaitu sebagai berikut:

Bab 1: Pendahuluan

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

Bab 2: Landasan Teori

Bab ini berisi tentang teori-teori yang digunakan untuk memahami permasalahan yang dibahas pada penelitian ini yaitu teori-teori yang berhubungan dengan *data mining*, *clustering*, *K-Means*, *K-Medoids*, *X-Means*, dan beberapa penelitian terdahulu.

Bab 3: Metodologi Penelitian

Bab ini berisi tentang tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja.

Bab 4: Hasil Dan Pembahasan

Bab ini berisi tentang proses yang terdapat pada langkah penelitian yang telah direncanakan.

Bab 5: Penutup

Bab ini berisi tentang kesimpulan dari semua uraian pada bab-bab sebelumnya dan saran-saran yang diharapkan berguna untuk pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA

- Adhitama, R., Burhanuddin, A., & Ananda, R. (2020). Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clusterin. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 1–5. <https://doi.org/10.33387/jiko.v3i1.1635>
- Awalludin, A. S., & Taufik, I. (2017). *Analisis Cluster Data Longitudinal pada Pengelompokan Daerah Berdasarkan Indikator IPM di Jawa Barat*. 978, 187–194.
- Candrasari Hermanto, D. M. (2017). ANALISIS ALGORITMA CLUSTERING DALAM KASUS PENENTUAN JENIS BUNGA IRIS. *Jurnal Pendidikan Teknologi Dan Kejuruan*.
- Emzir. (2010). *Metodologi Penelitian Pendidikan Kuantitatif dan Kualitatif*. Jakarta: Raja Grafindo Persada.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*.
- Fitriyadi, akhmad upi, & ana kurniawati. (2021). Algoritma K-Means dan K-Medoids Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional. *Kilat*, 10(1), 157–168. <https://doi.org/10.33322/kilat.v10i1.1174>
- G.Soni, K., & Patel, D. A. (2017). Comparative analysis of FCM and HCM algorithm on Iris data set. *International Journal of Computer Applications*, 5(2), 33–37. <https://doi.org/10.5120/888-1261>
- Garcia, S., Luengo, J., & Herrera, F. (2015). *Data Preprocessing in Data Mining*.
- Hasan, F. N., Hikmah, N., & Utami, D. Y. (2018). Perbandingan Algoritma C4.5, KNN, dan Naive Bayes untuk Penentuan Model Klasifikasi Penanggung jawab BSI Entrepreneur Center. *Jurnal Pilar Nusa Mandiri*, 14(2), 169. <https://doi.org/10.33480/pilar.v14i2.908>
- Hasibuan, Z. A. (2007). Metodologi Penelitian Pada Bidang Ilmu Komputer Dan Teknologi Informasi. *Konsep, Teknik, Dan Aplikasi*, (Universitas Indonesia), 194.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*.
- Herviany, M., Delima, S. P., Nurhidayah, T., & Kasini. (2021). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 34–40.
- James and others, M. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1(14), 281–297.
- Marlina, D., Lina, N., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 4(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- Muhamad, H., Prasojo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle

- Swarm Optimization Pada Data Iris. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 180. <https://doi.org/10.25126/jtiik.201743251>
- Muzaki, M. (2019). Penerapan Triangular Kernel Nearest Neighbor Sebagai Metode Clustering Dasar Pada Metode Bagging. *Jurnal Teknologi Dan Riset Terapan (JATRA)*, 1(1 SE-Research Articles). Retrieved from <https://jurnal.polibatam.ac.id/index.php/JATRA/article/view/1313>
- Pitria, P. (2019). Analisis Sentimen Pengguna Twitter Pada Akun Resmi Samsung Indonesia Dengan Menggunakan Naïve Bayes. *Undergraduate Theses from JBPTUNIKOMPP*.
- Pramudiono, Iko. (2006). *Apa itu data mining?*
- Prasetyo, eko. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*.
- Ren, Q., Li, M., & Han, S. (2019). Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives. *Big Earth Data*, 3(1), 8–25. <https://doi.org/10.1080/20964471.2019.1572452>
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sari, D. N. P., & Sukestiyarno, Y. L. (2021). Analisis Cluster dengan Metode K-Means pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 602–610. Retrieved from <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Taha Chicho, B., Mohsin Abdulazeez, A., Qader Zeebaree, D., & Assad Zebari, D. (2021). Machine Learning Classifiers Based Classification For IRIS Recognition. *Qubahan Academic Journal*, 1(2), 106–118. <https://doi.org/10.48161/qaj.v1n2a48>
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. Milan: WILEY.
- Wahidin, A. J., & Sensuse, D. I. (2021). Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air. *Jurnal ICT: Information Communication & Technology*, 20(2), 298–302. <https://doi.org/10.36054/jict-ikmi.v20i2.387>
- Wurdianarto, S., Novianto, S., & Rosyidah, U. (2014). Perbandingan Euclidean Distance Dengan Canberra Distance Pada Face Recognition. *Techno.Com*, 13(1), 31–37. Retrieved from <https://publikasi.dinus.ac.id/index.php/technoc/article/view/539>
- Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, 2(2), 226–235. <https://doi.org/10.3390/j2020016>
- Zizwan, P. A., Zarlis, M., & Nababan, E. B. (2020). *Bigdata Clustering using X-means method with Euclidean Distance Bigdata Clustering using X-means method with Euclidean Distance*. <https://doi.org/10.1088/1742-6596/1566/1/012103>