

Clustering Data Pengunjung UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang Menggunakan Algoritma K-Means

Nanda Karmila Sari¹, Yeka Hendriyani²

¹Program Studi Pendidikan Teknik Informatika, Fakultas Teknik, Universitas Negeri Padang

²Dosen Departemen Teknik Elektronika, Fakultas Teknik, Universitas Negeri Padang

e-mail: nandakarmilasari2806@gmail.com

Abstrak

UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang merupakan salah satu contoh perpustakaan di Universitas Negeri Padang yang memiliki banyak koleksi buku dan peminat pengunjung yang beragam, kelengkapan buku yang tersedia bagi pengunjung dan kenyamanan saat membaca serta pencarian buku sangat mempengaruhi jumlah pengunjung. Banyak pengguna fungsi pencarian yang dapat diakses untuk memahami informasi yang tersembunyi di dalam data tersebut. Atribut data yang akan digunakan adalah bulan kunjungan, fakultas, dan golongan koleksi yang dipinjam. Hasil dan pengolahan data pengunjung ini bertujuan untuk membantu pihak UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang agar memudahkan dalam proses perencanaan untuk meningkatkan mutu perpustakaan. Penggalian data diperlukan untuk mengekstrak informasi yang tersembunyi di dalam database yang ada. Data mining adalah proses interaktif dan fleksibel yang dapat digunakan untuk membuat model baru atau murni yang dapat divalidasi secara menyeluruh terhadap berbagai macam database. Algoritma K-Means dan aplikasi RapidMiner digunakan dalam teknik pengelompokan penelitian ini. Dengan menerapkan algoritma k-means pada proses klasterisasi data UPT Perpustakaan, Penerbitan, dan Percetakan Universitas Negeri Padang, maka dapat diidentifikasi dan ditentukan jumlah cluster yang sesuai serta parameter-parameternya berdasarkan karakteristik masing-masing *cluster*. Dari penelitian dihasilkan 3 cluster yang diantaranya untuk cluster pertama (C1) pada kedelapan fakultas dan dari bulan januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 0 (Karya Umum) dan 100 (Filsafat dan Psikologi). Cluster kedua (C2) pada kedelapan fakultas dan dari bulan januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 800 (Kesusatraan) , 900 (Sejarah, Geografi) , dan FIC (Karya Fiksi). Antara Januari 2022 dan Desember 2022, pengunjung klaster ketiga (C3) dari delapan fakultas meminjam banyak koleksi dalam kategori 500 (Sains dan Matematika), 600 (Teknologi dan Ilmu Terapan), dan 700 (Seni, Hiburan, dan Olahraga). Serta untuk cluster keempat (C4) pada kedelapan fakultas dan dari bulan

januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 200 (Agama), 300 (Ilmu Sosial), dan 400 (Bahasa).

Kata kunci: *Data Pengunjung, Klastering, K-Means, RapidMiner, Data Mining*

Abstract

UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang is one example of a library at Universitas Negeri Padang that has a large collection of books and a variety of visitors, the completeness of the books available to visitors and the convenience when reading and searching for books greatly affects the number of visitors. Many users can access search functions to understand the information hidden in the data. The data attributes that will be used are the month of visit, faculty, and class of books borrowed. The results and processing of visitor data aims to help the UPT Library, Publishing and Printing of Padang State University to facilitate the planning process to improve the quality of the library. Data mining is needed to extract information hidden in existing databases. Data mining is an interactive and flexible process that can be used to create new or pure models that can be thoroughly validated against a wide variety of databases. K-Means algorithm and RapidMiner tool are used in the clustering technique of this research. By applying the k-means algorithm to the data clustering process of UPT Perpustakaan, Penerbitan, dan Percetakan Universitas Negeri Padang, it can be identified and determined the appropriate number of clusters and their parameters based on the characteristics of each cluster. From the research, 3 clusters were produced, including the first cluster (C1) in the eight faculties and from January 2022 to December 2022, many visitors borrowed collections of groups 0 (General Works) and 100 (Philosophy and Psychology). The second cluster (C2) in all eight faculties and from January 2022 to December 2022, visitors borrowed a lot of collections in groups 800 (Literature), 900 (History, Geography), and FIC (Works of Fiction). Between January 2022 and December 2022, visitors to the third cluster (C3) from eight faculties borrowed a lot of collections in categories 500 (Science and Mathematics), 600 (Technology and Applied Sciences), and 700 (Arts, Entertainment, and Sports). And for the fourth cluster (C4) in all eight faculties and from January 2022 to December 2022, visitors borrowed a lot of collections in categories 200 (Religion), 300 (Social Sciences), and 400 (Languages).

Keywords : *Visitor Data, Clustering, K-Means, RapidMiner, Data Mining*

PENDAHULUAN

Teknologi informasi modern yang lebih canggih sering kali menghasilkan volume data yang sangat besar dalam berbagai disiplin ilmu. Data dalam jumlah besar dihasilkan di berbagai bidang seperti industri, bisnis, pendidikan, ilmu pengetahuan, teknologi, dan kehidupan sehari-hari. Untuk memahami informasi yang terkandung dalam data, maka perlu melakukan analisis data untuk mengetahui informasi yang

tersembunyi didalam data tersebut, salah satunya informasi yang terdapat dalam data pengunjung perpustakaan.

Salah satu alat yang digunakan oleh Kementerian Pendidikan untuk mendukung siswa dalam pendidikan dan meningkatkan proses belajar mereka adalah perpustakaan. Penambahan atau peningkatan wawasan dan pengetahuan bagi pelajaran, karena perpustakaan sangat membantu. Khususnya pembelajaran yang berkesinambungan dalam pembelajaran sepanjang hayat atau pendidikan seumur hidup bagi masyarakat umum. Dengan meningkatkan fungsi pers secara minimal, diharapkan juga mampu memberikan pendidikan minimal bagi peserta didik. Salah satu metode yang digunakan untuk meningkatkan fungsi perpustakaan itu sendiri adalah dengan menggunakan pegelompokan pengunjung perpustakaan, yang akan memudahkan proses pembesaran mutu perpustakaan. UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang merupakan perpustakaan yang memiliki banyak koleksi mulai dari karya ilmiah, buku teks dan majalah serta peminat pengunjung yang beragam. Kelengkapan buku yang tersedia bagi pengunjung dan kenyamanan saat membaca serta pencarian buku sangat mempengaruhi jumlah pengunjung perpustakaan.

UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang selain memiliki banyak koleksi buku, juga memiliki berbagai fasilitas yang memadai diantaranya tempat membaca buku disetiap lantainya serta layanan digital untuk mencari referensi judul skripsi/tugas akhir/disertasi.

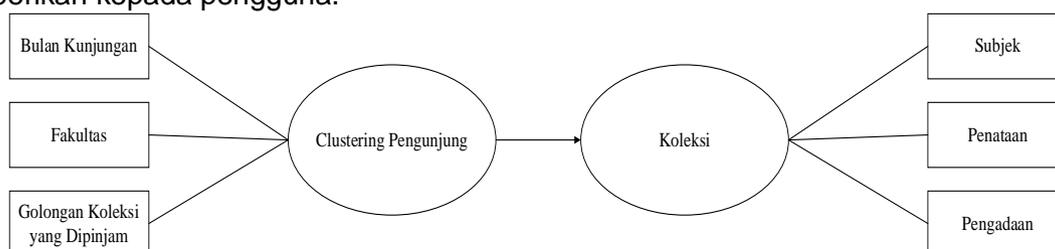
Hal ini memungkinkan untuk mengekstrak informasi tersembunyi dari statistik pengunjung perpustakaan dengan memproses jumlah pengunjung yang sangat banyak. Atribut data yang akan digunakan adalah bulan kunjungan, fakultas, dan golongan koleksi yang dipinjam. Hasil data pengolahan data pengunjung ini bertujuan untuk membantu pihak UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang agar memudahkan proses perencanaan untuk meningkatkan mutu perpustakaan serta memudahkan pihak UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang dalam pengambilan kebijakan dan pengembangan mutu perpustakaan.

Dengan menggunakan k-means, perpustakaan UNP dapat mengidentifikasi pengguna yang memiliki preferensi tertentu, seperti jenis buku yang mereka baca atau frekuensi kunjungan mereka. Hal ini meningkatkan pengalaman pengguna dengan memungkinkan rekomendasi buku yang lebih relevan untuk setiap kelompok orang. Selain itu, algoritma k-means dapat membantu proses penyortiran buku sesuai dengan minat dan permintaan pengguna. Dengan cara ini, perpustakaan dapat memaksimalkan tumpukan buku mereka dan memastikan bahwa buku-buku yang paling mereka sukai selalu tersedia.

Algoritma k-means dapat membantu dalam proses penyortiran buku sesuai dengan minat dan permintaan pengguna. Dengan cara ini, perpustakaan dapat memaksimalkan tumpukan buku dan memastikan bahwa buku-buku yang paling mereka sukai selalu tersedia. Perencanaan ulang ruang dapat ditujukan secara ekonomi karena menganalisis pola penggunaan ruang perpustakaan. Jika ada

kelompok pengguna tertentu yang lebih suka bekerja dalam suasana kelompok atau lebih suka suasana yang lebih privat, perpustakaan dapat menyesuaikan fasilitas untuk memenuhi kebutuhan mereka.

Untuk melakukan pencarian menggunakan algoritma k-means, data-data yang relevan harus dikumpulkan, seperti data pemahaman bacaan, data frekuensi kunjungan, data preferensi bacaan, dan lain sebagainya. Setelah itu, data-data tersebut dapat digunakan sebagai input untuk algoritma k-means, yang akan membantu dalam mengklasifikasikan data pengunjung ke dalam kelompok dengan ciri-ciri yang sebanding. Hasil dari upaya ini dapat membantu UPT Perpustakaan, Penerbitan dan Percetakan Universitas Negeri Padang dalam mengembangkan antarmuka pengguna yang lebih efektif dan meningkatkan jumlah dukungan yang diberikan kepada pengguna.



Gambar 1. Kerangka Konseptual

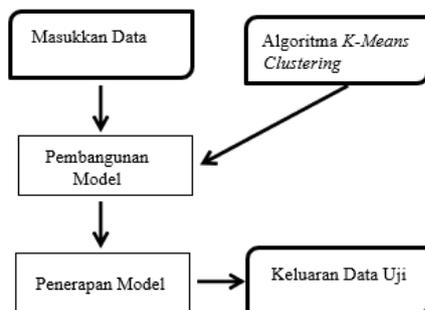
Dari gambar diatas, clustering pengunjung dapat dilakukan dengan menggunakan data beberapa aspek meliputi bulan kunjungan, fakultas dan golongan koleksi yang dipinjam. Clustering pengunjung perpustakaan memiliki relasi dengan koleksi perpustakaan yang dipengaruhi aspek meliputi subjek, penataan dan juga pengadaan yang ada diperpustakaan.

METODE

Penelitian ini dilakukan dengan menggunakan metode sistematis seperti studi literatur, pengumpulan data, pra-pemrosesan, pengelompokan K-Means, dan kesimpulan. Tujuan utamanya adalah untuk mengelompokkan data pengunjung UPT Perpustakaan, Penerbitan, dan Percetakan Univeristas Negeri Padang dengan menggunakan metode tahapan sebagai berikut (1) proses analisis data, (2) perhitungan K-Means Clustering, dan (3) pengujian dengan alat data mining RapidMiner.

1. Proses Analisis Data

Selama proses analisis data, data pengunjung UPT Perpustakaan, Penerbitan, dan Percetakan yang digunakan dari Universitas Negeri Padang. Data tersebut di-cluster menggunakan algoritma K-Means Clustering untuk menghasilkan model. Selanjutnya, model dianalisis menggunakan data uji seperti yang digambarkan pada Gambar 2..



Gambar 2. Proses Analisa Data

2. Perhitungan pengelompokan K-Means

Algoritma K-Means melakukan hal berikut secara iteratif hingga kestabilan tercapai atau tidak ada lagi objek yang dapat berubah seperti yang ditampilkan pada Gambar 3.

1. Menentukan koordinat titik tengah untuk setiap cluster.
2. Menentukan jarak antara setiap objek dengan koordinat titik tengah.
3. Mengelompokkan objek berdasarkan jarak minimumnya.



Gambar 3. Flowchart Algoritma Kmeans

3. Menghitung jarak data dengan rumus jarak euclidean

Tahapan yang terlibat dalam perhitungan klasterisasi K-Means meliputi:

- (a) menentukan jumlah cluster,

- (b) mengalokasikan data sesuai dengan jumlah cluster,
- (c) menentukan nilai centroid untuk setiap cluster
- (d) menghitung jarak terdekat dengan rumus jarak Euclidean.

$$d(i, j) = \sqrt{(x1i - x1j)^2 + (x2i - x2j)^2 + \dots + (xki - xkj)^2} \dots \dots \dots (1)$$

Keterangan :

- D(i,j) = Jarak data ke i pusat cluster j
- Xki = Data ke i pada atribut data ke k
- Xkj = Titik Pusat ke j pada atribut ke k

(e) Menampilkan hasil berdasarkan jarak terendah yang diperoleh dari perhitungan hasil

dari langkah 4.

(f) Jika hasil yang sama belum diperoleh, iterasi diulangi dengan menggunakan langkah 3

(g) Iterasi dihentikan jika hasil pengelompokan sama dengan iterasi sebelumnya

4. Menghitung cluster baru

Langkah pertama dari proses klasterisasi adalah menerapkan serta menghitung nilai rata-rata dari setiap kriteria untuk setiap anggota yang termasuk dalam setiap kluster.

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{jsj} x_j$$

dimana : $\mu_j(t + 1)$ = centroid baru di (t+1)
 N_{sj}: banyak data pada klaster S_j

HASIL DAN PEMBAHASAN

1. Hasil Pengelompokan

Terdapat 288 data untuk kelas pertama setelah siklus keenam pemrosesan data selesai, 288 untuk kelas kedua, dan 480 untuk kelas ketiga. Sampel data pengguna yang digunakan dapat dilihat pada Tabel 1-3.

Tabel 1. Hasil Cluster Pertama

No.	Data ke-	Cluster
1	1	Cluster 1
2	2	Cluster 1
3	3	Cluster 1
4	12	Cluster 1
5	13	Cluster 1
6	14	Cluster 1
7	23	Cluster 1
8	24	Cluster 1
9	25	Cluster 1
10	34	Cluster 1
:		
288	1050	Cluster 1

Tabel 1 menunjukkan hasil kluster yang diperoleh setelah iterasi, yang berisi 288 data pada kluster pertama.

Tabel 2. Hasil Cluster Kedua

No.	Data ke-	Cluster
1	4	Cluster 2
2	5	Cluster 2
3	6	Cluster 2
4	15	Cluster 2
5	16	Cluster 2
6	17	Cluster 2
7	26	Cluster 2
8	27	Cluster 2
9	28	Cluster 2
10	37	Cluster 2
:		
288	1053	Cluster 2

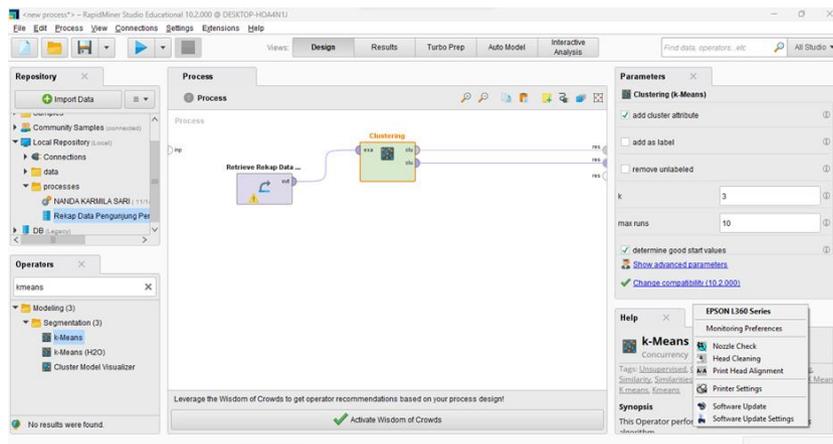
Tabel 2 menunjukkan hasil kluster yang diperoleh setelah iterasi, yang berisi 288 data pada kluster kedua.

Tabel 3. Hasil Cluster Ketiga

No.	Data ke-	Cluster
1	7	Cluster 3
2	8	Cluster 3
3	9	Cluster 3
4	10	Cluster 3
5	11	Cluster 3
6	18	Cluster 3
7	19	Cluster 3
8	20	Cluster 3
9	21	Cluster 3
10	22	Cluster 3
:		
480	1056	Cluster 3

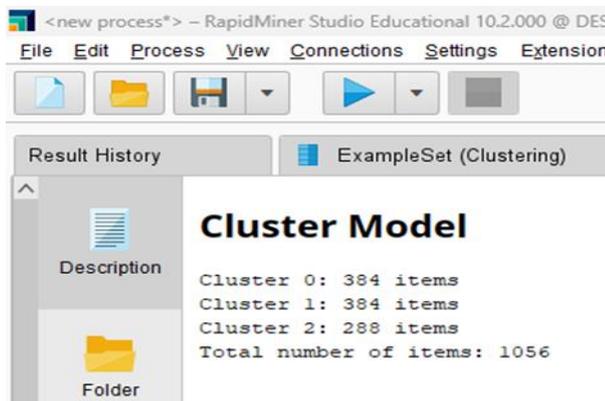
Tabel 3 menunjukkan hasil kluster yang diperoleh setelah iterasi, yang berisi 480 data pada kluster ketiga.

2. Pengujian dengan Tool Rapidminer



Gambar 4. Proses Data

Gambar di atas (Gambar 4) menunjukkan mengolah data dengan teknik clustering K-Means, yang menghasilkan dan menampilkan data keluaran sebagai berikut.



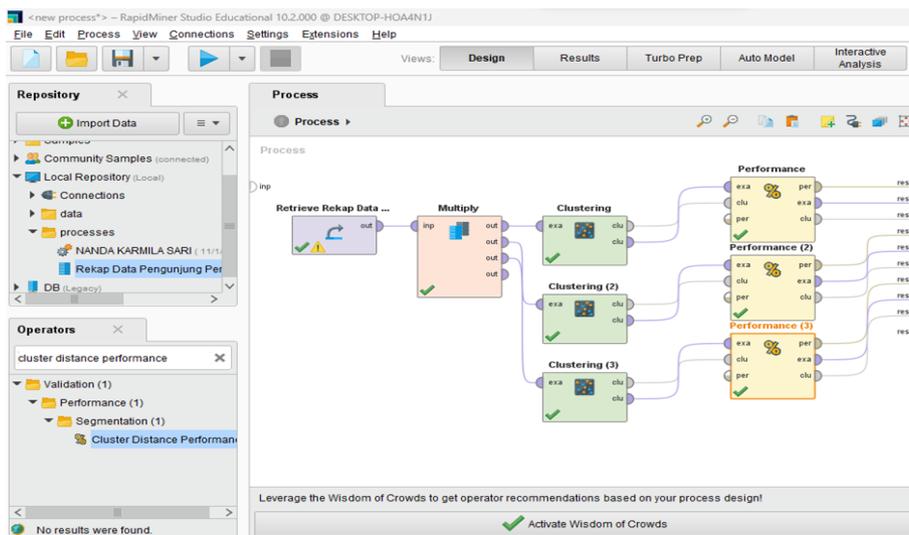
Gambar 5. Hasil Cluster Model

Berdasarkan gambar di atas (Gambar 5), dapat diamati bahwa model cluster mengelompokkan total 1056 data ke dalam 3 cluster. Klaster 1 yaitu sebanyak 384 data, klaster 2 yaitu sebanyak 384 data dan klaster 3 yaitu sebanyak 288 data.

3. Evaluasi Cluster

Untuk menentukan jumlah cluster yang ideal, gunakan teknik Davies Bouldin Index (DBI). Teknik DBI, yang pertama kali seperti yang dinyatakan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979, berfungsi sebagai metrik untuk menilai kinerja algoritma optimasi. Berdasarkan kriteria ini, algoritma yang menghasilkan jumlah pengklasifikasi terbanyak dengan indeks Davies-Bouldin dianggap sebagai yang terbaik. Percobaan dari 3 hingga 6 cluster dilakukan

dengan menggunakan aplikasi RapidMiner, seperti yang digambarkan pada Gambar 6.



Gambar 6. Proses Hitung Nilai DBI

Setelah di proses di tool RapidMiner di dapatkan nilai DBI yang terdapat pada tabel 4 berikut.

Tabel 4. Nilai DBI

Kluster Set	DBI
3	0,494
4	0,463
5	0,490

Berdasarkan hasil perhitungan pada tabel di atas yaitu cluster set 4 yang menghasilkan 4 cluster memiliki DBI terendah yaitu 0.463. Berikut ini adalah hasil untuk setiap klaster:

1. Cluster pertama (C1) pada kedelapan fakultas dan dari bulan januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 0 (Karya Umum) dan 100 (Filsafat dan Psikologi).
2. Cluster kedua (C2) pada kedelapan fakultas dan dari bulan januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 800 (Kesusastaan), 900 (Sejarah, Geografi) dan FIC (Karya Fiksi).
3. Cluster ketiga (C3) pada kedelapan fakultas dan dari bulan januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 500 (IPA dan Matematika), 600 (Teknologi dan Ilmu Terapan), dan 700 (Kesenian, Hiburan, dan Olahraga).

4. Cluster keempat (C4) pada kedelapan fakultas dan dari bulan januari 2022 hingga desember 2022, pengunjung banyak meminjam koleksi golongan 200 (Agama), 300 (Ilmu Sosial), dan 400 (Bahasa).

SIMPULAN

Hasil percobaan dengan menggunakan 1056 set data dan centroid (titik pusat) yang ditentukan di awal Algoritma K-Means memiliki dampak yang signifikan terhadap hasil klaster; nilai centroid yang bervariasi juga menghasilkan hasil klaster yang bervariasi. Tiga cluster - cluster satu dengan 384 item, cluster dua dengan 384 item, dan cluster tiga dengan 288 item digambarkan pada gambar setelah pengumpulan data selesai menggunakan algoritma K-Means. Salah satu tujuan dari proses data mining yang telah disebutkan di atas adalah untuk mendapatkan informasi baru mengenai teknik untuk mengumpulkan dan mengevaluasi data sesuai dengan standar yang telah ditetapkan. Hal ini dapat menjadi tantangan bagi UPT Perpustakaan, Penerbitan, dan Percetakan Universitas Negeri Padang untuk melakukan penelitian dan pengembangan berdasarkan koleksi yang paling populer dan diminati oleh pengunjung untuk meningkatkan jumlah pengunjung setiap tahunnya.

DAFTAR PUSTAKA

- D. Aprillia, D. A. Baskoro, L. Ambarwati dan I. W. S. Wicaksana, “ *Belajar Data Mining dengan RapidMiner* ”, Jakarta, 2013.
- Dony Prisma. 2012. "Karakteristik Pemustaka Perpustakaan" <https://donyprisma.wordpress.com/2012/07/24/karakteristik-pemakai-perpustakaan>, diakses 20 april 2018
- Fina Nasari dan Charles Jhony Maoto Sianturi, 2016. "Penerapan *Algoritma K- Means Clustering* untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat". Jurnal. Medan: Program Studi Sistem Informasi, Universitas Potensi Utama.
- Ginangjar Abdurrahman. 2016. "*Clustering* Data Ujian Tengah Semester (UTS) Data Mining Menggunakan *Algoritma K-Means*". *Jurnal Jember: Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember*.
- Ge, Z., Song, Z., Ding, S.X. and Huang, B., (2017). Data mining and analytics in the process industry: The role of machine learning. *IEEE Access*, 5, pp.20590-20616.
- Gojun Shi, Bingkun Gao, Li Zhang. "The Optimized K-means Algorithms for Improving Randomly-initialed Midpoints". 2013 2nd International Conference on Measurement, Information and Control. 2013:1212-1216.
- Lhorend Mutiara Pratiwi 2017 "*Penerapan K-means Clustering* untuk memprediksi minat nasabah pada PT Asuransi Jiwa Bersama 1912 Bumiputera Prabumulih Jurnal Palembang.
- Novaliendry, D., Hendriyani, Y., Yang, C. H., & Hamimi, H. (2015). The optimized K-means *clustering* algorithms to analyzed the budget revenue expenditure in Padang. *International Conference on Electrical Engineering, Computer Science*

- and Informatics (EECSI), 2(August), 61–66.
<https://doi.org/10.11591/eecsi.v2i1.771>
- N.N, "WhyRapidMiner," RapidMiner, [Online]. Available: <https://rapidminer.com/why-rapidminer/>. [Diakses 30 Mei 2023].
- Prayoga, Y., Tambunan, H. S., & Parlina, I. (2019). Penerapan *Clustering* Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 1(1), 24-30.
- Ramadhani, R., Hendriyani, Y., & Pendahuluan, I. (2021). *Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus : SMKN 2 Padang) (Studi Kasus : SMKN 2 Padang)*
- R. Gustrinda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-means Dan K-Medoids," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 27, 2022, doi: 10.30865/mib.v6i1.3294.
- S. Angra dan S. Ahuja, "Analysis of Student's Data Using Rapid Miner," *Journal on Today's Ideas – tomorrow's Technologies*, vol.4, pp. 49-58, 2016.
- Sardi, H. Y., Budayawan, K., New, T., Pendidikan, P., Informatika, T., Teknik, F., & Negeri, U. (2020). Klasifikasi Tingkat Kelulusan Mahasiswa Elektronika Menggunakan Algoritma Naïve Bayes Classifier (Studi Kasus : Pendidikan Teknik Informatika FT-UNP) P - ISSN : 2302-3295. 8(4).
- Santoso. 2007, "*Data Mining Teknik Pemanfaatan Dara untuk Keperluan Bisnis*". Yogyakarta: Graha Ilmu Sugiyono 2017, "*Metode Penelitian Kombinasi (Mixed Methods)*", Bandung Alfabeta
- Saputro, D. T., & Sucihermayanti, W. P. (2021). Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara. *Jurnal Buana Informatika*, 12(2), 146–155. <https://doi.org/10.24002/jbi.v12i2.4861>
- Sugiyono. 2017. "*Metode Penelitian Kombinasi (Mixed Methods)*". Bandung: Alfabeta
- Sumadi Suryabata. 2012. "*Metodologi Penelitian*". Jakarta: Rajawali Pers.
- Suyanto, 2017. "*Data Mining untuk Klasifikasi dan Klasterisasi Data*", Bandung Informatika.
- Wibowo, T., Novaliendry, D., (2023). *Optimizing Patient Medical Records Grouping through Data Mining and K-Means Clustering Algorithm (A Case Study at RSUD Mohammad Natsir Solok)*.
- Yi-feng XU, Chun-ming Chen, Yun-qing XU. "An improved K-means clustering algorithm[J]". *Computer applications and software*. March 2008, Vol.25 No.3:275-277
- Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru. *Sistemasi*, 7(3), 238. <https://doi.org/10.32520/stmsi.v7i3.388>
- Universitas Negeri Padang. 2014. "*Buku Pedoman Penulisan Tugas Akhir/Skripsi*" Padang: Universitas Negeri Padang.