

ANALISIS MINAT BACA MAHASISWA DI PERPUSTAKAAN UNSIKA DENGAN ALGORITMA K-MEANS

Diterima Redaksi: 5 Juli 2023; Revisi Akhir: 27 Oktober 2023; Diterbitkan Online: 18 Mei 2024

Syela Herdina Naibaho¹⁾, Nailufar Farha Afifah²⁾, Yuyun Umaidah³⁾, Nono Heryana⁴⁾

^{1,2,3)}Sistem Informasi, Fakultas Ilmu Komputer dan Universitas Singaperbangsa
^{1,2,3)}Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat, Indonesia, kode pos: 41361
e-mail: 2010631250077@student.unsika.ac.id¹⁾, 2010631250068@student.unsika.ac.id²⁾,
yuyun.umaidah@staff.unsika.ac.id³⁾, nono.heryana.uk@staff.unsika.ac.id⁴⁾

Abstrak: Penelitian ini bertujuan untuk menganalisis minat baca mahasiswa di Perpustakaan Universitas Singaperbangsa Karawang dengan menggunakan algoritma K-Means. Metode K-Means digunakan untuk mengelompokkan buku-buku berdasarkan pola peminjaman, sehingga dapat memberikan wawasan yang lebih baik tentang preferensi baca mahasiswa dan membantu perpustakaan dalam pengelolaan koleksi buku. Data peminjaman buku diperoleh dari database pustakawan universitas, sedangkan studi pustaka dilakukan dengan mempelajari dan mencari referensi pada jurnal dan literatur terkait. Setelah data bersih dan dapat diproses, dilakukan proses klusterisasi menggunakan algoritma K-Means dengan menentukan nilai k. Hasil klusterisasi ditampilkan dalam bentuk tabel yang memperlihatkan kelompok buku-buku yang sering dipinjam, buku-buku dengan frekuensi peminjaman sedang, dan buku-buku yang jarang dipinjam oleh mahasiswa. Dalam pengelompokan data menggunakan algoritma K-means, penting untuk mengukur seberapa efektif hasil kluster yang diperoleh. Metrik evaluasi yang digunakan adalah Davies-Bouldin Index (DBI). Metrik ini membantu mengukur kualitas kluster yang dihasilkan dengan mempertimbangkan jarak antara pusat kluster dan dispersi antar-kluster. Proses penghitungan DBI melibatkan perhitungan rata-rata jarak dalam kluster, serta perbandingan jarak antara pusat-pusat kluster. Temuan penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam mengelompokkan buku-buku berdasarkan pola peminjaman. Dengan pemahaman yang lebih baik tentang minat baca mahasiswa, perpustakaan dapat mengoptimalkan penempatan buku-buku, meningkatkan ketersediaan buku yang diminati, dan mengembangkan strategi untuk meningkatkan minat baca mahasiswa secara keseluruhan. Hasil penelitian ini memberikan kontribusi yang berarti dalam pengembangan perpustakaan universitas dan memberikan panduan yang berguna bagi pengambilan keputusan dalam manajemen perpustakaan.

Kata Kunci— Klusterisasi, K-Means, Minat Baca, Perpustakaan, RapidMiner

Abstract: This study aims to analyze students' reading interest at the Singaperbangsa Karawang University Library using the K-Means algorithm. The K-Means method is used to group books based on borrowing patterns, so as to provide better insight into students' reading preferences and assist libraries in managing book collections. Book borrowing data was obtained from the university librarian database, while literature study was carried out by studying and looking for references in related journals and literature. After the data is clean and can be processed, the clustering process is carried out using the K-Means algorithm by determining the value of k. The clustering results are displayed in the form of a table showing the groups of books that are frequently borrowed, books with a moderate borrowing frequency, and books that are rarely borrowed by students. Cluster quality evaluation was carried out using the Davies Bouldin value. The findings of this study indicate that the K-Means algorithm is effective in classifying books based on borrowing patterns. With a better understanding of students' reading interest, libraries can optimize the placement of books, increase the availability of books of interest, and develop strategies to increase overall student reading interest. The results of this study make a significant contribution to the development of university libraries and provide useful guidelines for decision-making in library management.

Keywords— Clusterization, K-Means, Reading Interest, Library, RapidMiner

I. PENDAHULUAN

Perpustakaan merupakan sebuah lembaga yang memiliki tugas dan tanggung jawab dalam mengelola koleksi berbagai jenis karya tulis, cetak, dan rekam dengan cara yang profesional. Fungsi utama perpustakaan adalah untuk memenuhi berbagai kebutuhan dalam bidang pendidikan, penelitian, pelestarian, informasi, dan rekreasi bagi para pengunjung atau pemustaka [1].

Perpustakaan merupakan komponen penting dalam lingkungan akademik, yang menyediakan akses ke berbagai sumber informasi untuk mendukung kegiatan belajar dan penelitian mahasiswa di universitas. Untuk memastikan keberhasilan perpustakaan dalam memenuhi kebutuhan pengguna, penting untuk memahami minat baca mahasiswa.

Menyelidiki minat baca mahasiswa dapat membantu mengidentifikasi preferensi dan kecenderungan mereka terhadap jenis buku, topik bacaan, dan layanan perpustakaan tertentu. Dengan memahami minat baca mahasiswa secara lebih mendalam, perpustakaan dapat mengoptimalkan koleksi buku, menyediakan layanan yang relevan, dan mempromosikan minat baca yang lebih luas di kalangan mahasiswa. Minat baca sendiri merupakan hal yang fundamental yang harus dimiliki oleh seseorang sebelum mereka mencapai kemampuan membaca [2]. Minat ini menjadi dasar bagi keberhasilan dalam kegiatan membaca. Jika seseorang tidak memiliki minat yang kuat terhadap membaca, maka apapun bahan bacaan yang mereka baca akan menjadi sia-sia. Hal ini disebabkan karena mereka membaca tanpa adanya ketertarikan pribadi atau minat yang mendorong mereka. Sebaliknya, jika kegiatan membaca didasarkan pada keinginan pribadi, kemungkinan besar mereka akan mengalami kegiatan membaca yang efektif. Oleh karena itu, penulis akan menganalisis minat baca di perpustakaan Universitas Singaperbangsa Karawang dengan data mining menggunakan metode k-means untuk clustering atau pengelompokan data judul buku yang dipinjam.

Data mining atau yang disebut juga sebagai *Knowledge Discovery in Database* (KDD) merupakan suatu proses yang dilakukan secara otomatis untuk mencari pola-pola dalam jumlah data yang sangat besar. Proses ini dilakukan dengan menggunakan teknik-teknik seperti Klasifikasi (*Classification*), Regresi (*Regression*), Asosiasi (*Association*), dan Pengelompokan (*Clustering*) [3].

Clustering adalah teknik untuk mengelompokkan rekaman-rekaman dalam sebuah database berdasarkan kondisi tertentu. Prinsip dasar dari clustering adalah mengelompokkan objek-objek ke dalam kelompok, dimana kelompok yang baik yaitu kelompok yang memiliki tingkat kesamaan yang tinggi antara objek-objek di dalamnya dan tingkat perbedaan yang tinggi dengan objek-objek dari kelompok lainnya. [4]

Pengelompokkan minat baca pada perpustakaan unsika ini dilakukan dengan menggunakan metode K-Means *Clustering*. K-Means adalah sebuah prosedur yang digunakan secara iteratif untuk membagi pengelompokan dan mengklasifikasikan sejumlah besar objek. Dalam metode K-Means, objek-objek tersebut dikelompokkan ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik mereka [5]. Metode ini memungkinkan pengelompokan data berdasarkan pola yang terdapat dalam variabel-variabel yang relevan dengan minat baca, seperti jenis buku yang dipinjam, dan topik bacaan yang diminati. Dengan menerapkan metode K-Means, harapannya adalah topik bacaan dari buku yang sering dipinjam akan dikelompokkan bersama dalam kelompok yang terpidah dari topik bacaan dari buku yang jarang dipinjam. Dengan demikian, analisis clustering ini dapat memberikan pemahaman yang lebih mendalam tentang pola minat baca mahasiswa dan memfasilitasi pengelompokan yang efektif dalam konteks minat baca.

Melalui penelitian ini, diharapkan dapat menghasilkan wawasan yang berharga bagi Perpustakaan Universitas Singaperbangsa Karawang. Hasil analisis ini dapat memberikan pemahaman yang lebih baik tentang preferensi dan minat baca mahasiswa, serta membantu perpustakaan dalam pengembangan koleksi yang lebih relevan dan penyediaan layanan yang lebih efektif.

II. TINJAUAN PUSTAKA

Perpustakaan Universitas Singaperbangsa Karawang (Unsika) memiliki tugas utama dalam mengelola dan menyediakan berbagai sumber informasi, baik dalam bentuk buku cetak maupun digital, untuk mendukung aktivitas akademik, penelitian, dan pengembangan ilmu pengetahuan di lingkungan universitas. Koleksi perpustakaan ini mencakup buku teks, referensi, jurnal ilmiah, makalah, tesis, disertasi, serta sumber informasi elektronik lainnya, dan secara berkala diperbarui dan diperluas sesuai dengan perkembangan ilmu pengetahuan dan kebutuhan pengguna. Pengguna perpustakaan dapat dengan mudah mencari dan mengetahui ketersediaan bahan melalui sistem katalog online perpustakaan. Selain itu, akses ke sumber informasi digital juga dapat dilakukan melalui portal perpustakaan atau

basis data berlangganan yang disediakan, sehingga mempermudah pengguna untuk menemukan informasi yang dibutuhkan.

Dalam tahap awal penelitian dilakukan identifikasi masalah yang ada pada perpustakaan Universitas Singaperbangsa Karawang. Kemudian, peneliti mengumpulkan data perpustakaan yang diperoleh dari perpustakaan universitas seperti data peminjaman buku perpustakaan. Data tersebut nantinya akan dilakukan *cluster* dengan menjadikannya subjek penelitian pada perpustakaan Unsika. Dengan 3 kriteria yaitu (1) sering dipinjam, (2) sedang dipinjam dan (3) jarang dipinjam.

Selanjutnya setelah melakukan studi literatur dan penentuan metode, kemudian berdasarkan dengan kriteria-kriteria yang telah ditentukan, metode ini digunakan untuk menentukan pengelompokan pada data peminjaman koleksi buku di perpustakaan universitas tersebut.

A. KDD (*Knowledge Discovery in Databases*)

Metodologi KDD (*Knowledge Discovery in Databases*) adalah suatu pendekatan yang digunakan untuk mendapatkan pengetahuan baru dari data. Metodologi ini melibatkan serangkaian tahapan yang bertujuan untuk mengolah data yang ada dengan maksud memperoleh informasi yang berharga. Tahapan-tahapan dalam metodologi KDD meliputi pemilihan data, pra-pemrosesan data, transformasi data, penambangan data, dan interpretasi/evaluasi. [5] Dengan mengikuti tahapan-tahapan ini, diharapkan informasi yang berharga dapat dihasilkan dari data yang besar tersebut.

1) *Preprocessing*

Sebelum menerapkan algoritma data mining, langkah *preprocessing* dilakukan terhadap data. *Preprocessing* ini mencakup serangkaian proses, seperti pembersihan data, integrasi, seleksi, dan transformasi. Proses ini dilakukan untuk mempersiapkan data agar lebih siap digunakan dalam proses analisis dan penambangan data.

2) *Data Mining*

Proses inti dalam KDD adalah data mining, di mana algoritma-algoritma data mining digunakan untuk menggali pengetahuan dari sumber data. Dalam proses ini, berbagai algoritma diterapkan untuk mengidentifikasi pola, hubungan, dan informasi yang tersembunyi dalam data. Tujuan utamanya adalah mengekstrak pengetahuan yang berharga dan bermanfaat dari data yang ada.

3) *Post processing*

Setelah proses data mining selesai, hasil yang diperoleh akan dievaluasi dalam tahap *post-processing*.

B. *Data Mining*

Data Mining adalah proses yang melibatkan penggunaan satu atau lebih teknik pembelajaran mesin (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Istilah "data mining" digunakan untuk menggambarkan penemuan pengetahuan dalam *database*. Data mining melibatkan penggunaan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi dan mengekstraksi informasi yang berharga dan pengetahuan yang relevan dari berbagai database yang besar. [6] Proses penggalian informasi yang berguna dan menarik dari suatu basis data yang besar dengan menggunakan teknik-teknik tertentu. Tujuan dari data mining sendiri adalah untuk menemukan pola, hubungan, dan informasi yang tersembunyi dalam data yang ada. Data mining dapat membantu dalam mengambil keputusan yang lebih baik dan efektif dalam berbagai bidang, seperti bisnis, Kesehatan, pendidikan, dan lain-lain. Beberapa teknik yang sering digunakan dalam data mining antara lain seperti *association rules*, *decision tree*, *k-means*, dan *naïve bayes*.

C. *Clustering*

Clustering merupakan sebuah teknik dalam data mining yang bertujuan untuk mengelompokkan objek atau data ke dalam beberapa *cluster* atau kelompok sehingga objek yang memiliki kesamaan dikelompokkan bersama dalam satu *cluster*. Clustering dapat membantu dalam mengungkap pola atau struktur yang tersembunyi dalam data. Salah satu algoritma yang umum digunakan dalam *clustering* adalah algoritma K-Means. Algoritma K-Means bekerja dengan membagi data ke dalam k kelompok atau cluster yang berbeda. Cara kerjanya adalah dengan menghitung jarak antara setiap data dengan *centroid* (titik pusat) dari setiap *cluster*, kemudian mengelompokkan data ke dalam *cluster* yang

memiliki *centroid* terdekat. Algoritma K-Means sering diterapkan dalam berbagai bidang seperti kesehatan, bisnis, dan pendidikan.[7]

D. Teknik Elbow

Teknik *elbow* merupakan metode yang digunakan dalam *clustering* untuk menentukan jumlah *cluster* yang optimal berdasarkan nilai SSE (*Sum of Square Error*) dari setiap *cluster*. Metode ini melibatkan plotting nilai SSE pada sumbu y dan jumlah *cluster* pada sumbu x, kemudian mencari titik pada grafik yang menunjukkan penurunan SSE yang signifikan, dikenal sebagai "*elbow point*". *Elbow point* menunjukkan jumlah *cluster* yang paling optimal untuk data yang sedang dianalisis. Teknik *elbow* sering diterapkan dalam algoritma *clustering* seperti K-Means dan *Partitioning Around Medoids* (PAM). Dalam penelitian yang dilakukan pada CV. HM Putra, teknik *elbow* digunakan untuk menentukan jumlah *cluster* terbaik untuk data transaksi konsumen. Namun, dalam penelitian yang dilakukan pada Dinas Perkebunan Provinsi Kalimantan Timur, metode *elbow* tidak digunakan, melainkan metode *Silhouette Coefficient* untuk mengukur kualitas hasil *clustering*. [8]

E. Rapid Miner

RapidMiner adalah sebuah *platform* perangkat lunak ilmu data yang dikembangkan oleh perusahaan dengan nama yang sama. *Platform* ini menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. RapidMiner dapat digunakan baik untuk tujuan bisnis dan komersial maupun untuk keperluan penelitian, pendidikan, pelatihan, *rapid prototyping*, dan pengembangan aplikasi. *Platform* ini mendukung semua tahapan dalam proses pembelajaran mesin, termasuk persiapan data, visualisasi hasil, validasi model, dan optimasi. [9]

RapidMiner menyediakan lebih dari 500 operator data mining yang mencakup fungsi-fungsi seperti *input*, *output*, *preprocessing data*, dan visualisasi. *Platform* ini merupakan *software* yang mandiri untuk analisis data dan juga dapat diintegrasikan dengan produk lain. RapidMiner ditulis menggunakan bahasa Java, sehingga dapat berjalan di berbagai sistem operasi. Lisensi yang digunakan oleh RapidMiner adalah GPL (*GNU Affero General Public License*) versi 3.

Hingga saat ini, telah banyak aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. Pada polling yang dilakukan oleh *KDnuggets*, sebuah portal data-mining pada tahun 2010-2011, RapidMiner menduduki peringkat pertama sebagai perangkat lunak data mining yang paling populer. Beberapa fitur yang dimiliki oleh RapidMiner mencakup banyaknya algoritma data mining yang tersedia, seperti *decision tree* dan *self-organization map*.

F. Perpustakaan

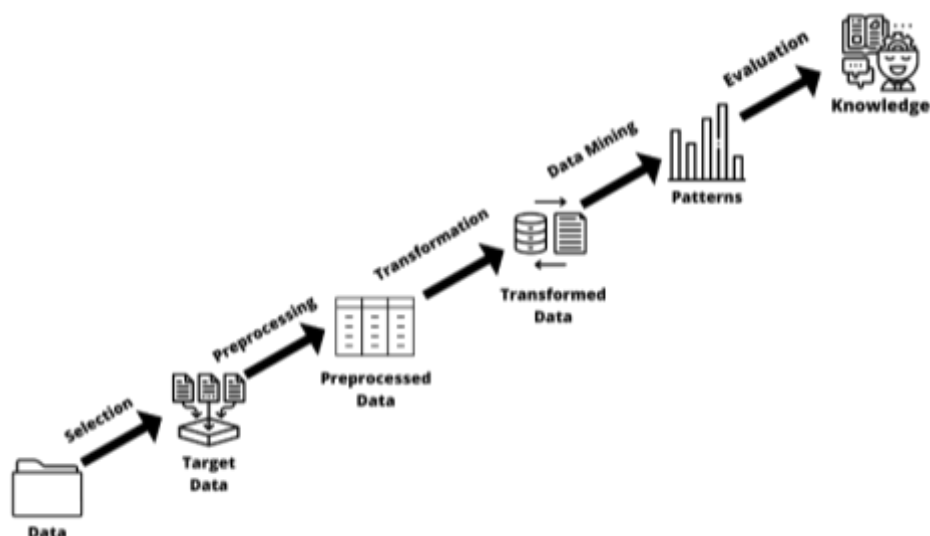
Perpustakaan merupakan lembaga atau tempat yang menyediakan berbagai koleksi bahan bacaan atau sumber informasi dalam berbagai format seperti buku, majalah, jurnal, surat kabar, dan media elektronik. Koleksi tersebut dapat diakses oleh masyarakat umum atau anggota perpustakaan. Fungsi utama perpustakaan adalah melestarikan hasil budaya manusia, terutama dalam bentuk dokumen karya cetak dan karya rekam lainnya, serta menyampaikan gagasan, pemikiran, pengalaman, dan pengetahuan manusia kepada generasi selanjutnya. Selain itu, perpustakaan juga memiliki fungsi rekreasi, khususnya rekreasi kultural, yang bertujuan untuk memberikan hiburan dan edukasi kepada masyarakat. Terdapat berbagai jenis perpustakaan, antara lain perpustakaan umum, perpustakaan sekolah, perpustakaan perguruan tinggi, dan perpustakaan khusus. Setiap jenis perpustakaan memiliki koleksi dan layanan yang disesuaikan dengan kebutuhan pengguna. [10]

Perpustakaan universitas merupakan perpustakaan yang terletak di lingkungan universitas dan menyediakan berbagai koleksi bahan bacaan atau sumber informasi dalam berbagai format, seperti buku, majalah, jurnal, surat kabar, dan media elektronik. Fasilitas tersebut dapat diakses oleh mahasiswa, dosen, dan masyarakat umum. Perpustakaan universitas memiliki peran penting dalam mendukung kegiatan akademik, penelitian, dan pengabdian masyarakat yang dilakukan oleh universitas. Selain itu, perpustakaan universitas juga berfungsi untuk memberikan layanan informasi dan referensi kepada penggunanya, serta turut membantu meningkatkan minat baca dan literasi masyarakat. Berbagai jenis layanan disediakan oleh perpustakaan universitas, seperti layanan sirkulasi untuk peminjaman

bahan, layanan referensi untuk membantu pencarian informasi, layanan akses elektronik untuk mengakses sumber informasi digital, dan layanan literasi informasi untuk membantu pengguna memanfaatkan informasi secara efektif. Perpustakaan universitas juga terus mengembangkan koleksi bahan bacaan yang beragam sesuai dengan kebutuhan pengguna, serta dilengkapi dengan teknologi informasi dan komunikasi modern guna mempermudah akses informasi. [10]

III. METODE PENELITIAN

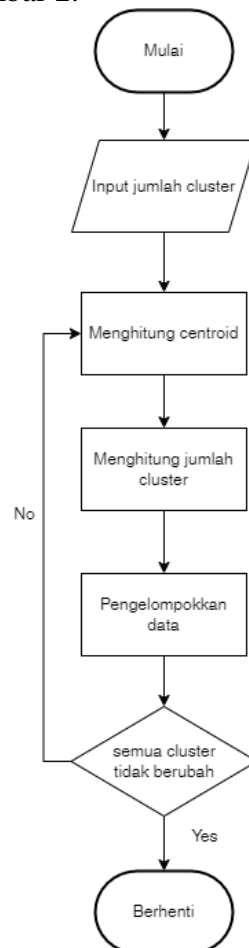
Pada penelitian ini digunakan metodologi KDD karena dilakukan penelitian dan analisis data untuk mengidentifikasi, menggali, dan menghasilkan pengetahuan baru mengenai minat baca mahasiswa di perpustakaan tersebut dengan menggunakan algoritma K-Means. Metodologi ini terdiri dari beberapa tahapan yang dilakukan secara sistematis. Pertama, dilakukan tahap *understanding the domain* untuk memahami konteks dan tujuan penelitian serta mengidentifikasi faktor-faktor yang mempengaruhi minat baca mahasiswa. Pada tahap ini, informasi tentang data-data perpustakaan yaitu data peminjaman buku. Kemudian, pada tahap Seleksi Data, data relevan dan representatif tentang minat baca mahasiswa dipilih. Data tersebut dapat berupa survei, catatan peminjaman buku, atau data terkait lainnya yang berkaitan dengan minat baca mahasiswa di perpustakaan. Pada penelitian ini digunakan data peminjaman koleksi buku perpustakaan. Setelah itu, tahap *Preprocessing Data* dilakukan untuk membersihkan dan mempersiapkan data sebelum analisis dilakukan. Data yang dikumpulkan diperiksa, dikoreksi, dan diubah menjadi format yang sesuai untuk analisis lebih lanjut. Selanjutnya, tahap Transformasi Data dilakukan untuk mengubah data menjadi representasi yang cocok dengan algoritma K-Means. Mungkin perlu dilakukan normalisasi atau pemilihan atribut yang relevan untuk mendapatkan hasil yang akurat. Tahap utama selanjutnya adalah Data Mining, di mana algoritma K-Means diterapkan. Algoritma ini akan mengelompokkan buku-buku ke dalam kategori. Proses ini melibatkan perhitungan jarak antara data dan centroid dari setiap *cluster*, kemudian pengelompokan data ke dalam *cluster* yang memiliki *centroid* terdekat. Setelah melakukan Data Mining, tahap Interpretasi/Evaluasi dilakukan untuk menginterpretasikan hasil analisis data mining. Hasil *cluster* yang ditemukan dievaluasi dan diinterpretasikan untuk mengungkap pola minat baca mahasiswa yang mungkin ada dalam data. Terakhir, penelitian akan menghasilkan penemuan pengetahuan yang relevan mengenai minat baca mahasiswa di perpustakaan universitas. Dengan menerapkan metodologi KDD menggunakan algoritma K-Means, penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang minat baca mahasiswa di perpustakaan Universitas Singaperbangsa Karawang dan membantu perpustakaan dalam meningkatkan layanan dan program literasi yang relevan. Untuk alur metodologinya dapat dilihat seperti pada Gambar 1.



Gambar 1. Metodologi KDD

Algoritma K-Means merupakan salah satu metode dalam data mining yang digunakan untuk melakukan analisis *clustering*. Pada penelitian ini algoritma K-Means digunakan untuk

mengelompokkan peminjaman koleksi buku. Metode K-Means bekerja dengan membagi data menjadi beberapa kelompok atau *cluster* yang berbeda. Dalam penelitian ini, data yang digunakan terkait dengan minat baca mahasiswa di perpustakaan. Algoritma K-Means akan mencari pola atau struktur dalam data untuk mengidentifikasi kelompok mahasiswa dengan minat baca serupa. Proses algoritma K-Means dimulai dengan memilih secara acak k *centroid* awal. Setiap *centroid* mewakili pusat atau rata-rata dari kelompok yang akan terbentuk. Setelah itu, setiap koleksi buku yang dipinjam akan diberi label kelompok berdasarkan *centroid* terdekatnya. *Centroid* kemudian diperbarui dengan menghitung rata-rata dari semua buku dalam kelompok tersebut. Proses ini diulang hingga tidak ada perubahan dalam pengelompokan koleksi buku atau hingga mencapai kriteria berhenti yang ditentukan. Alur dari algoritma K-Means dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Metode K-Means

Perhitungan metode algoritma K-Means pada penelitian ini menggunakan perhitungan jarak dengan *Ecluiden*. Langkah penyelesaian sebuah penelitian menggunakan metode K-means. Untuk menghitung jarak antar *centroid* ke data dari masing-masing *cluster* dengan rumus *Ecluidean*. Untuk menghitung jarak tersebut dapat menggunakan rumus persamaan 1.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Keterangan:

p, q = Dua titik di ruang-n Euklides

q_i, p_i = Vektor Euklides, dimulai dari asal ruang (titik awal)

n = Ruang-n

Hasil akhir dari algoritma K-Means adalah kelompok-kelompok koleksi buku yang memiliki frekuensi peminjaman yang sama. Dalam konteks penelitian ini, informasi ini dapat membantu dalam mengidentifikasi pola minat baca mahasiswa di Perpustakaan Universitas Singaperbangsa Karawang. Contohnya, terdapat kelompok mahasiswa yang tertarik pada bidang akademik tertentu, kelompok

dengan minat baca yang beragam, atau kelompok yang minat bacanya terkait dengan kegiatan kampus tertentu. Dengan hasil penelitian ini dapat membantu perpustakaan untuk tahu koleksi buku mana saja yang kira-kira paling sering diminati oleh pengunjung perpustakaan. Dengan menerapkan algoritma K-Means, penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai minat baca mahasiswa di perpustakaan universitas. Temuan ini dapat digunakan untuk mengembangkan strategi dan program literasi yang lebih efektif, meningkatkan koleksi bahan bacaan yang relevan, serta menyediakan layanan yang sesuai dengan minat baca mahasiswa.

IV. HASIL DAN PEMBAHASAN

A. Sumber Data

Dalam penelitian ini, sumber data yang digunakan adalah data peminjaman buku yang berasal dari perpustakaan Universitas Singaperbangsa Karawang. Data ini diperoleh melalui akses ke *database* yang dikelola oleh pustakawan universitas. Proses pengumpulan data dilakukan dengan mengambil informasi tentang peminjaman buku dari staf universitas tersebut. Data yang didapat yaitu data peminjaman buku oleh anggota perpustakaan sejumlah 912 data yang diperoleh dari tahun 2018 sampai pada tanggal 2023. Dari data tersebut sendiri dapat dilihat bahwa peminjaman buku di perpustakaan Unsika pada tahun 2018 sampai 2019 menaik, data tersebut dapat dilihat pada Gambar 3. Kemudian, dari tahun 2020 sampai tahun 2021 mengalami penurunan. Hal ini terjadi akibat pandemi COVID-19. Lalu, dari 2021 sampai tahun 2023 mengalami kenaikan yang signifikan.



Gambar 3. Grafik Data Peminjaman Buku Perpustakaan Unsika

Untuk melengkapi studi pustaka, langkah-langkah tertentu telah diambil. Pertama, dilakukan pemahaman mendalam tentang topik penelitian, yaitu pengelompokan data peminjaman koleksi buku menggunakan algoritma K-Means. Peneliti mempelajari konsep dasar dari algoritma tersebut, termasuk prinsip kerjanya, metode perhitungan jarak, dan langkah-langkah pengelompokan. Selanjutnya, peneliti melakukan pencarian dan seleksi referensi melalui penelusuran jurnal-jurnal terkait serta literatur lain yang relevan dengan topik penelitian ini. Tujuannya adalah untuk memperoleh pemahaman yang komprehensif tentang aplikasi algoritma K-Means dalam pengelompokan data, khususnya dalam konteks peminjaman buku.

Dalam studi pustaka, peneliti mempelajari berbagai artikel jurnal, buku, dan sumber-sumber lainnya yang terkait dengan pengelompokan data peminjaman koleksi buku menggunakan algoritma K-Means. Referensi ini digunakan sebagai dasar teoritis dan panduan dalam merancang analisis dan interpretasi data peminjaman buku yang diperoleh dari perpustakaan Universitas Singaperbangsa Karawang. Dengan demikian, penelitian ini mengandalkan sumber data peminjaman buku dari perpustakaan Universitas Singaperbangsa Karawang sebagai sumber data *primer*, yang diperoleh dari *database* pustakawan universitas. Selain itu, studi pustaka dilakukan melalui pemahaman konsep algoritma K-Means dan pencarian referensi dari jurnal dan literatur terkait untuk mendapatkan wawasan yang mendalam tentang pengelompokan data peminjaman koleksi buku dengan menggunakan algoritma tersebut.

B. Seleksi Data

Setelah mengumpulkan data yaitu yang merupakan tahap pertama yang digunakan pada penelitian ini. Kemudian dilanjutkan ke tahap selanjutnya yaitu seleksi data. Dalam tahap ini peneliti perlu menyeleksi data yang perlu dan sesuai untuk digunakan dalam penelitian ini. Dalam data peminjaman koleksi buku perpustakaan terdapat beberapa kolom seperti ID anggota, nama anggota, kode eksemplar, judul buku, tanggal pinjam, tanggal harus kembali, dan status peminjaman. Untuk data peminjamannya seperti pada Tabel 1.

Tabel 1. Data Peminjaman Buku

Kode Eksemplar	Judul	Tanggal Pinjam	Tanggal Harus Kembali	Status peminjaman
278	Manajemen Sumber Daya Manusia	25/10/2018	30/10/2018	1
13970	NULL	27/03/2019	01/04/2019	1
14950	NULL	27/03/2019	01/04/2019	1
6923	Strategi Dan Metode Pembelajaran	28/03/2019	02/04/2019	1
12695	Hukum Hak Atas Kekayaan Intelektual	28/03/2019	02/04/2019	1

Pada penelitian ini penulis menggunakan kolom judul buku untuk dilakukan clusterisasi pada judul-judul buku yang dipinjam pada anggota perpustakaan. Kemudian hasilnya seperti yang dapat dilihat di Tabel 2.

Tabel 2. Seleksi Data Peminjaman Buku

Judul	Status peminjaman
Manajemen Sumber Daya Manusia	1
NULL	1
NULL	1
Strategi Dan Metode Pembelajaran	1
Hukum Hak Atas Kekayaan Intelektual	1

C. Pemrosesan Data dan Transformasi Data

Setelah dilakukan seleksi data yang dibutuhkan untuk diolah pada penelitian, dilanjutkan dengan pemrosesan data dan transformasi data. Pemrosesan pada data dapat dilakukan dengan *cleaning* (pembersihan) data data yang *missing value* atau redundan serta dilakukan juga integrasi data. Namun, dikarenakan data yang sudah diberikan sudah lolos tahap *cleaning* dan integrasi, maka dilanjut dengan tahap transformasi data, yaitu menggabungkan data yang sudah sesuai dengan data yang digunakan dalam proses data mining nantinya. Dapat dilihat pada Tabel 3 yaitu merupakan data yang sudah dalam tahap transformasi dan sudah sesuai untuk dilakukannya proses data mining.

Tabel 3. Dataset Peminjaman Buku

Judul	Status peminjaman
Manajemen Sumber Daya Manusia	1
Kamus Inggris Indonesia	1
Asas-Asas Hukum Pidana	1
Strategi Dan Metode Pembelajaran	1
Hukum Hak Atas Kekayaan Intelektual	1
Pokok-pokok Hukum Perdana	1
Metodologi penelitian kualitatif	1
Metode Penelitian Kualitatif	1
Pengantar Material Teknik	1
Metodik Khusus pengajaran Agama Islam	1

Berdasarkan data yang ada, terdapat data sebanyak 841 data yang nantinya akan diolah menggunakan metode data mining dengan algoritma K-Means *Clustering* dengan cara mengelompokkan judul-judul buku tersebut ke dalam *cluster* untuk buku-buku yang sering dipinjam sampai dengan buku yang jarang dipinjam. Sehingga nantinya dapat dilakukan analisis pola pembentukan *cluster*.

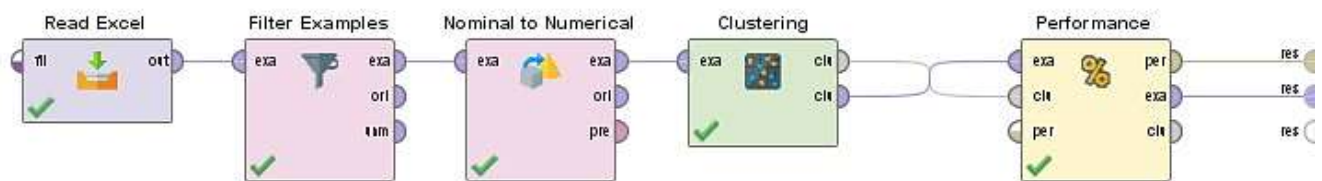
D. Metode K-Means Clustering

Setelah data telah dibersihkan dan siap untuk diproses, langkah selanjutnya adalah melakukan tahap data mining dengan menggunakan algoritma *clustering*. Dalam hal ini, tujuannya adalah untuk mengelompokkan judul-judul buku yang telah dipinjam ke dalam beberapa kelompok atau *cluster*, sehingga dapat dilakukan analisis terhadap pola pembentukan *cluster* tersebut. Proses *clustering* dimulai dengan menetapkan nilai *centroid* secara acak. Selanjutnya, estimasi jarak antara setiap data dengan setiap *centroid* dilakukan menggunakan korelasi antara dua objek dalam rumus *Euclidean Distance* pada Persamaan. Langkah-langkah yang diikuti dalam proses ini melibatkan penentuan jumlah *k* kelompok, inisiasi *centroid*, dan kemudian menghitung jarak pusat *cluster* menggunakan *Euclidean distance* menggunakan rumus yang telah ditentukan.

Dengan demikian, proses ini melibatkan penggunaan algoritma *clustering* untuk mengelompokkan judul-judul buku berdasarkan pola yang ada. Langkah-langkah yang diambil termasuk menentukan jumlah kelompok, inisiasi *centroid*, dan menghitung jarak pusat *cluster* dengan menggunakan *Euclidean distance* sesuai dengan rumus yang telah ditentukan.

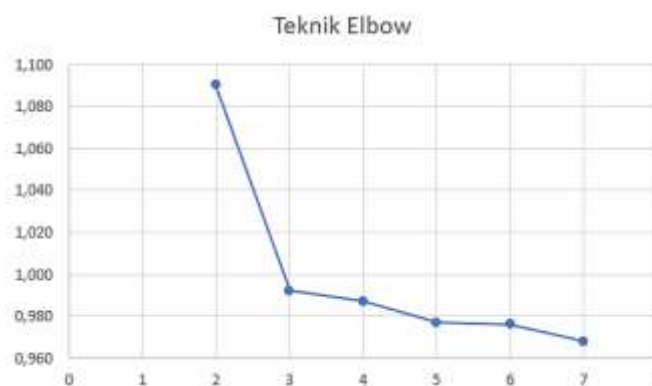
E. Implementasi Algoritma K-Means Clustering Menggunakan RapidMiner

Data inputan yang digunakan berupa data excel dengan langkah-langkah pengerjaannya dapat dilihat pada Gambar 4.



Gambar 4. Proses Pengolahan K-Means

Proses ini pertama dilakukan *read excel* untuk membaca dataset peminjaman koleksi buku dan dilakukan pemilihan kolom untuk memilih kolom mana saja yang akan digunakan. Kemudian dilakukan filter untuk memfilter data yang kosong atau NULL. Selanjutnya dilakukan pengubahan format data judul buku dari *nominal* ke *numerical*. Setelah itu dimasukkan algoritma K-Means *Clustering* untuk mengolah datanya setelah itu dimasukkan *operator performance* untuk penghitungannya. Setelah itu dilakukan pencarian nilai *k* dengan menggunakan metode *Elbow*. Dalam metode *Elbow*, nilai *cluster* yang diambil sebagai *cluster* optimal adalah titik di mana terbentuknya siku dalam grafik. [11]



Gambar 5. Grafik Teknik Elbow

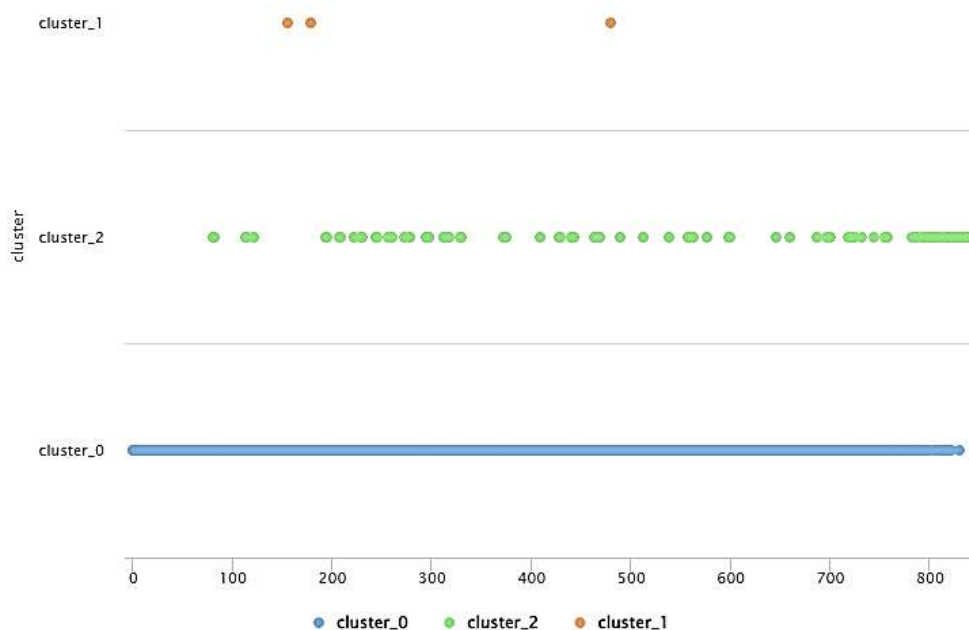
Setelah tahapan sebelumnya selesai, langkah selanjutnya adalah melakukan proses seperti yang ditunjukkan pada Gambar 5 dengan menggunakan nilai *k* yang telah ditentukan sebelumnya, yaitu 3. Proses ini melibatkan penghubungan antara *connector* satu dengan yang lainnya untuk memastikan tidak terjadi kesalahan (*error*) dalam jalannya program. Setelah semua koneksi telah diatur dengan benar, peneliti menjalankan kembali program dengan menggunakan data hasil proses yang telah dipersiapkan sebelumnya. Hasil dari proses tersebut kemudian ditampilkan pada Gambar 6.

Cluster Model

```
Cluster 0: 741 items  
Cluster 1: 3 items  
Cluster 2: 97 items  
Total number of items: 841
```

Gambar 6. Hasil Data Proses

Dengan demikian, langkah-langkah yang dilakukan adalah mengikuti proses yang ditunjukkan pada Gambar 3 dengan menggunakan nilai *k* yang telah ditentukan sebelumnya. Selanjutnya, peneliti menghubungkan setiap *connector* agar tidak terjadi kesalahan dalam program, dan program dijalankan kembali dengan menggunakan data hasil proses yang telah disiapkan. Hasil dari proses tersebut ditampilkan pada Gambar 6 sebagai output dari analisis yang dilakukan.



Gambar 7. Hasil Visualisasi Cluster

Dapat dilihat pada Gambar 7 terdapat hasil visualisasi Cluster. Diperoleh bahwa pada Cluster 0 terdapat banyak koleksi buku yang sering dipinjam oleh pengunjung perpustakaan Unsika. Setelah itu, pada cluster 1 ada 3 koleksi buku yang frekuensinya sedang yang dipinjam oleh pengunjung perpustakaan Unsika. Terakhir, pada cluster 2 terdapat beberapa koleksi buku yang jarang dipinjam oleh pengunjung perpustakaan Unsika.

F. Pengujian Hasil (Evaluasi)

Setelah hasil perhitungan K-Means *Clustering* melalui RapidMiner dihasilkan, maka akan dilakukan pengujian hasil dari perhitungan melalui aplikasi. Dapat dilihat hasil *performance vector*nya seperti pada Gambar 8.

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: 0.992  
Avg. within centroid distance_cluster_0: 0.997  
Avg. within centroid distance_cluster_1: 0.000  
Avg. within centroid distance_cluster_2: 0.988  
Davies Bouldin: 1.653
```

Gambar 8. Performance Vector

Hasil analisis menunjukkan bahwa kelompok buku dengan jarak terdekat terhadap *Avg. within centroid distance* adalah C2, yang menandakan bahwa kelompok ini memiliki buku-buku yang jarang dipinjam oleh mahasiswa. Sebaliknya, kelompok buku dengan jarak terjauh terhadap *Avg. within centroid distance* adalah C1, yang menunjukkan bahwa buku-buku dalam kelompok ini memiliki frekuensi peminjaman yang lebih tinggi atau sering dipinjam oleh mahasiswa. Selain itu, penilaian menggunakan nilai *Davies Bouldin* memberikan informasi tentang kualitas *cluster* yang dihasilkan oleh metode K-Means. Semakin kecil nilai *Davies Bouldin*, semakin baik kualitas *cluster* yang diperoleh. Dalam penelitian ini, nilai *Davies Bouldin* sebesar 1.653 menunjukkan bahwa masing-masing objek dalam *cluster* memiliki tingkat kesamaan yang cukup baik, mendekati angka 0.

Dengan demikian, hasil analisis menunjukkan bahwa terdapat perbedaan dalam pola peminjaman buku antara kelompok C1 dan C2. Kelompok C1 memiliki buku-buku yang lebih sering dipinjam, sementara kelompok C2 memiliki buku-buku yang jarang dipinjam. Selain itu, hasil evaluasi menggunakan nilai *Davies Bouldin* menunjukkan kualitas yang baik dalam pembentukan *cluster* menggunakan metode K-Means pada penelitian ini.

Tabel 4. Judul Buku Setelah Clustering

Cluster 0	Cluster 1	Cluster 2
Manajemen Sumber Daya Manusia	Penelitian Pendidikan Metode & Paradigma Baru	Praxis pembelajaran pesantren
Kamus Inggris Indonesia		Media Pembelajaran
Asas-Asas Hukum Pidana		Manajemen Mutu Terpadu
Strategi Dan Metode Pembelajaran		Hukum Tentang Merger
Hukum Hak Atas Kekayaan Intelektual		Pengantar Kriminologi Teori, Metode, Dan Perilaku Kriminal
Pokok-pokok Hukum Perdana		Perancangan Dan Implementasi Sistem Basis Data
Metodologi penelitian kualitatif		Kesuburan Tanah Dan Nutrisi Tanaman
Metode Penelitian Kualitatif		Studi Kelayakan Bisnis (Manajemen, Metode & Kasus)
Pengantar Material Teknik		Perilaku Konsumsi
Metodik Khusus pengajaran Agama Islam		Perlindungan Hukum Bagi Konsumen Terhadap Iklan Yang Menyesatkan

Informasi yang tertera pada Tabel 4 memberikan gambaran tentang hasil klasterisasi buku-buku dalam penelitian ini. Klaster 0 menunjukkan buku-buku yang memiliki frekuensi peminjaman yang tinggi, yang berarti buku-buku ini sering dipinjam oleh anggota perpustakaan. Klaster 1 mencakup buku-buku dengan frekuensi peminjaman yang sedang, menunjukkan bahwa buku-buku ini dipinjam dengan frekuensi yang lebih moderat. Sedangkan, klaster 2 terdiri dari buku-buku yang jarang dipinjam oleh anggota perpustakaan.

Dengan menggunakan analisis klasterisasi, penelitian ini telah mengelompokkan buku-buku berdasarkan pola peminjaman mereka. Hasil klasterisasi ini memberikan wawasan yang berguna dalam memahami tingkat minat baca mahasiswa terhadap berbagai buku yang tersedia di perpustakaan. Dengan mengetahui klaster buku-buku yang sering dipinjam, sedang dipinjam, dan jarang dipinjam, perpustakaan dapat mengoptimalkan penempatan buku-buku tersebut dan mengembangkan strategi untuk meningkatkan minat baca mahasiswa. Hasil akhir ini memberikan pandangan yang jelas dan terstruktur tentang hasil klasterisasi, yang membantu peneliti dan pembaca jurnal ini dalam melihat pola-pola yang muncul dari analisis data. Klasterisasi buku-buku ini memberikan pemahaman yang lebih mendalam tentang preferensi dan minat baca mahasiswa, sehingga dapat digunakan sebagai dasar untuk mengambil keputusan yang lebih baik dalam manajemen perpustakaan.

Dalam kesimpulannya, penelitian ini telah berhasil mengklasterisasi buku-buku dalam perpustakaan Universitas Singaperbangsa Karawang berdasarkan pola peminjaman mereka. Dengan mengidentifikasi buku-buku yang sering dipinjam, sedang dipinjam, dan jarang dipinjam, penelitian ini memberikan wawasan yang berharga untuk mengoptimalkan layanan perpustakaan dan meningkatkan

minat baca mahasiswa. Tabel 5 menyajikan dengan jelas hasil klasterisasi tersebut, memberikan panduan yang berguna bagi pembaca jurnal ini dalam memahami temuan penelitian ini.

V. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan mengenai Analisis Minat Baca Mahasiswa di Perpustakaan Universitas Singaperbangsa Karawang dengan menggunakan algoritma K-Means, dapat diambil beberapa kesimpulan penting yaitu algoritma K-Means efektif digunakan untuk mengelompokkan buku-buku dalam perpustakaan berdasarkan pola peminjaman. Hasil klasterisasi memungkinkan pengidentifikasian buku-buku yang sering dipinjam, buku-buku dengan frekuensi peminjaman sedang, dan buku-buku yang jarang dipinjam oleh mahasiswa. Klasterisasi buku-buku memungkinkan perpustakaan untuk memahami minat baca mahasiswa dengan lebih baik. Dengan mengetahui preferensi baca mahasiswa dan pola peminjaman buku, perpustakaan dapat mengoptimalkan penempatan buku-buku, meningkatkan ketersediaan buku yang diminati, serta mengembangkan strategi untuk meningkatkan minat baca mahasiswa secara keseluruhan.

Analisis *Davies Bouldin* yang dilakukan dalam penelitian ini menunjukkan bahwa kualitas *cluster* yang dihasilkan oleh algoritma K-Means cukup baik. Nilai *Davies Bouldin* yang mendekati 0 menunjukkan bahwa objek-objek dalam setiap klaster memiliki kesamaan yang signifikan. Hasil penelitian ini memberikan sumbangan penting dalam pengembangan perpustakaan universitas. Dengan pemahaman yang lebih baik tentang minat baca mahasiswa dan pola peminjaman buku, perpustakaan dapat meningkatkan layanan, memperluas koleksi buku yang diminati, dan meningkatkan partisipasi mahasiswa dalam kegiatan literasi.

Dalam kesimpulannya, penelitian ini menunjukkan bahwa analisis menggunakan algoritma K-Means pada minat baca mahasiswa di perpustakaan Universitas Singaperbangsa Karawang memberikan hasil yang berharga. Klasterisasi buku-buku memungkinkan perpustakaan untuk mengoptimalkan sumber daya dan meningkatkan layanan kepada mahasiswa. Dengan peningkatan minat baca mahasiswa, diharapkan akan terjadi peningkatan pemahaman, keterampilan, dan pengetahuan yang berdampak positif pada perkembangan akademik dan profesional mereka.

UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih yang tulus kepada Dosen Pembimbing, Staf Perpustakaan, dan semua pihak yang terlibat dalam penelitian ini. Kontribusi dan dukungan semuanya telah memberikan nilai tambah yang besar dalam penulisan jurnal "Analisis Minat Baca Mahasiswa di Perpustakaan Universitas Singaperbangsa Karawang dengan Algoritma K-Means." Terima kasih atas bimbingan, bantuan, dan kerjasama yang berharga.

DAFTAR PUSTAKA

- [1] M. Afdal, M. Rosadi, P. Studi Sistem Informasi, F. H. Sains dan Teknologi UIN Suska Riau Jl Soebrantas KM, and P. Pekanbaru -Riau, "PENERAPAN ASSOCIATION RULE MINING UNTUK ANALISIS PENEMPATAN TATA LETAK BUKU DI PERPUSTAKAAN MENGGUNAKAN ALGORITMA APRIORI," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, 2019.
- [2] R. Taylor, "Impact of Text Interest on Reading Comprehension," 2020.
- [3] R. Muliono and Z. Sembiring, "DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASTERISASI TINGKAT TRIDARMA PENGAJARAN DOSEN," 2019.
- [4] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient," 2021. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [5] N. Salsabila, N. Sulistiyowati, and T. N. Padilah, "Pencarian Pola Pemakaian Obat Menggunakan Algoritma FP-Growth," 2022. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [6] M. Afdal, M. Rosadi, P. Studi Sistem Informasi, F. H. Sains dan Teknologi UIN Suska Riau Jl Soebrantas KM, and P. Pekanbaru -Riau, "PENERAPAN ASSOCIATION RULE MINING

UNTUK ANALISIS PENEMPATAN TATA LETAK BUKU DI PERPUSTAKAAN MENGGUNAKAN ALGORITMA APRIORI,” *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, 2019.

- [7] A. Kurnia, “Perbandingan Algoritma K-Means dan Fuzzy C-Means Untuk Clustering Puskesmas Berdasarkan Gizi Balita Surabaya,” *Jurnal PROCESSOR*, vol. 18, no. 1, Apr. 2023, doi: 10.33998/processor.2023.18.1.696.
- [8] D. D. Abdurrahman, F. Agus, and G. M. Putra, “Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Mengelompokkan Hasil Produksi Komoditi Perkebunan (Studi Kasus: Dinas Perkebunan Provinsi Kalimantan Timur),” *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, vol. 16, no. 2, p. 130, Oct. 2021, doi: 10.30872/jim.v16i2.6520.
- [9] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER,” *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.
- [10] N. Chakimah Mutharoh, “Perspektif Mahasiswa S-1 Ilmu Perpustakaan Universitas Diponegoro dalam Menanggapi Judul Clickbait pada Portal Berita Online di Indonesia,” *ANUVA*, vol. 6, no. 4, pp. 455–464, 2022.
- [11] R. Y. Sari, H. Oktavianto, and H. W. Sulistyono, “ALGORITMA K-MEANS DENGAN METODE ELBOW UNTUK MENGELOMPOKKAN KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN KOMPONEN PEMBENTUK INDEKS PEMBANGUNAN MANUSIA.”