PENGELOMPOKAN BUKU BERDASARKAN TINGKAT KETERPAKAIAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PERPUSTAKAAN MAN 2 KOTA TANGERANG

SKRIPSI

Oleh:

ACHMAD GHIFFARI FATULLAH
NIM. 210607110003

PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG

2025

PENGELOMPOKAN BUKU BERDASARKAN TINGKAT KETERPAKAIAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PERPUSTAKAAN MAN 2 KOTA TANGERANG

SKRIPSI

Oleh:
ACHMAD GHIFFARI FATULLAH
NIM. 210607110003

Diajukan Kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Sains Informasi (S.S.I)

PROGRAM STUDI PERPUSTAKAAN DAN SAINS INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2025

HALAMAN PERSETUJUAN

PENGELOMPOKAN BUKU BERDASARKAN TINGKAT KETERPAKAIAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PERPUSTAKAAN MAN 2 KOTA TANGERANG

SKRIPSI

Oleh:

ACHMAD GHIFFARI FATULLAH
NIM. 210607110003

Diperiksa dan Disetujui:
Tanggal: 16 Juni 2025

Pembimbing I	Pembimbing II
Fakhris Khusnu Reza Mahfud, M.Kom.	Firma Sahrul Bahtiar, M.Eng.
NIP. 199005062019031007	NIP. 198502012019031009

Mengetahui,
Ketua Program Studi Perpustakaan dan Sains Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Ir. M. Amin Hariyadi, M.T.
NIP. 196700182005011001

HALAMAN PENGESAHAN

PENGELOMPOKAN BUKU BERDASARKAN TINGKAT KETERPAKAIAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA PERPUSTAKAAN MAN 2 KOTA TANGERANG

SKRIPSI

Oleh:

ACHMAD GHIFFARI FATULLAH
NIM. 210607110003

Telah Dipertahankan di Depan Dewan Penguji
Dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Sains Informasi (S.S.I.) Pada 16 Juni 2025

Susunan Dewan Penguji		Tanda Tangan
Ketua Penguji	: Dr. Ir. M. Amin Hariyadi, M.T. NIP. 196700182005011001	( )
Anggota Penguji I	: Ach. Nizam Rifqi, M.A. NIP. 199206092022031002	( )
Anggota Penguji II	: Fakhris Khusnu Reza Mahfud, M.Kom. NIP. 199005062019031007	( )
Anggota Penguji III	: Firma Sahrul Bahtiar, M.Eng. NIP. 198502012019031009	( )

Disahkan Oleh:
Ketua Program Studi Perpustakaan dan Sains Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Ir. M. Amin Hariyadi, M.T.
NIP. 196700182005011001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Achmad Ghiffari Fatullah

NIM : 210607110003

Prodi : Perpustakaan dan Sains Informasi

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasi karya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 16 Juni 2025

Yang membuat pernyataan

Achmad Ghiffari Fatullah

NIM. 210607110003

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Segala puji dan syukur penulis panjatkan ke hadirat ﷻ الله. atas limpahan rahmat, taufik, serta hidayah-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan judul "Pengelompokan Buku Berdasarkan Tingkat Keterpakaian Menggunakan Algoritma K-Means Clustering Pada Perpustakaan Man 2 Kota Tangerang" sebagai salah satu syarat untuk memperoleh gelar Sarjana Sins Informasi (S.S.I) pada Program Studi Perpustakaan dan Sains Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Shalawat serta salam semoga senantiasa tercurah kepada Kanjeng ﷺ محمد, beserta keluarga, sahabat, dan seluruh umatnya hingga akhir zaman.

Penulis menyadari bahwa dalam penyusunan skripsi ini tidak terlepas dari adanya bantuan, bimbingan, serta do'a dan dukungan dari berbagai pihak. Oleh karena itu, melalui kata pengantar ini, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. M. Zainuddin, M.A, selaku Rektor UIN Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Dr. Ir. M. Amin Hariyadi, M.T selaku Ketua Program Studi Perpustakaan dan Sains Informasi, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang.
4. Bapak Fakhris Khusnu Reza Mahfud, M.Kom dan Bapak Firma Sahrul Bahtiar, M.Eng, selaku dosen pembimbing yang telah meluangkan waktu serta pikiran untuk membimbing penulis hingga skripsi ini selesai.
5. Bapak Dr. Ir. M. Amin Hariyadi, M.T serta Bapak Ach. Nizam Rifqi, M.A, selaku dosen penguji yang telah memberikan berbagai saran dan arahan guna menyempurnaan karya ini.

6. Seluruh dosen Program Studi Perpustakaan dan Ilmu Informasi UIN Maulana Malik Ibrahim Malang yang telah memberikan ilmu pengetahuan kepada penulis selama berada di bangku kuliah.
7. Bapak Ali Sutrisno, ST. M.Pd, selaku Kepala Sekolah MAN 2 Kota Tangerang serta Ibu Fatmiarti Kusumaningrum, S.Hum selaku Staff Perpustakaan MAN 2 Kota Tangerang yang telah menyambut baik penulis dan menerima penulis dalam melaksanakan PKL serta melakukan observasi hingga melakukan penelitian.
8. Keluarga besar KH. Abdul Rahman Bin Ba'an, yang selalu memberikan doa, motivasi, serta kepercayaan penuh kepada penulis dalam menempuh pendidikan di Perguruan Tiinggi. Serta Keluarga besar Marhum Bin Mahbub, atas dukungan spiritual, moral, dan semangat yang tak pernah putus.
9. Keluarga penulis yaitu Fatullah Harun S.AR, S.Pd.I, (Ayah), Titin Suprihatun (Umi), Siti Hilyatul Izzah S.Pd (Kakak), Farhan Bin Abdullah As-Syatiri (Kakak Ipar), Farhatun Nisa (Adik), Qafisha Qurratul 'Ain (Adik) yang selalu memberikan dukungan baik secara moral maupun material serta senantiasa mendo'akan penulis selama masa perkuliahan dan sampai penulis bisa menyelesaikan skripsi ini.
10. Keluarga besar Pondok Pesantren Al-Manshuriyah Kota Tangerang, khususnya kepada KH. Najib Syahru Wardi selaku guru yang telah memberikan bimbingan, serta seluruh dewan asatidz yang telah membentuk karakter, kedisiplinan, dan dasar-dasar keilmuan yang sangat berharga bagi penulis.
11. Seluruh keluarga besar penulis yang tidak dapat disebutkan satu per satu, baik dari pihak ayah maupun ibu, yang senantiasa mendoakan dan memberikan semangat di setiap langkah perjuangan ini.
12. Seluruh teman-teman Garyatama Program Studi Perpustakaan dan Sains Informasi angkatan 2021 yang telah memberikan doa serta bantuan untuk penulis dalam menyusun dan menyelesaikan skripsi.
13. Teman-teman seperjuangan Kontrakan The Raid, sahabat, dan semua pihak yang telah membantu secara langsung maupun tidak langsung dalam proses penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, saran dan kritik yang bersifat membangun sangat penulis harapkan demi perbaikan di masa mendatang. Akhir kata, semoga skripsi ini dapat memberikan manfaat dan menjadi kontribusi kecil dalam khazanah keilmuan yang terus berkembang.

Wassalamu'alaikum Wr. Wb.

Malang, 16 Juni 2025
Penulis,
Achmad Ghiffari Fatullah

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Fatullah, Achmad Ghiffari. 2025. **Pengelompokan Buku Berdasarkan Tingkat Keterpakaian Menggunakan Algoritma K-Means Clustering Pada Perpustakaan Man 2 Kota Tangerang. Skripsi. Program Studi Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Firma Sahrul Bahtiar, M.Eng.**

Kata Kunci: *K-Means, Clustering*, Tingkat Keterpakaian Buku, *Data Mining*, Pengelompokan Buku, Perpustakaan

Penurunan angka peminjaman buku di perpustakaan MAN 2 Kota Tangerang menandakan kurang optimalnya pengelolaan koleksi, meskipun banyak judul yang tersedia akan tetapi koleksi belum sepenuhnya memenuhi kebutuhan pengguna. Penelitian ini bertujuan untuk mengukur kinerja pengelompokan buku berdasarkan tingkat keterpakaian menggunakan algoritma *K-Means Clustering* sebagai solusi untuk mengidentifikasi buku yang sangat diminati dan kurang diminati. Penelitian ini menggunakan pendekatan data mining dengan tahapan: studi literatur, pengumpulan dan *pre-processing* data transaksi peminjaman, penerapan algoritma *K-Means*, serta evaluasi hasil pengelompokan menggunakan metode *Silhouette Coefficient*. Data diambil dari sistem informasi perpustakaan SLiMS untuk periode 1 Januari 2023 hingga 1 Januari 2024, dengan 768 data transaksi peminjaman. Hasil nilai *silhouette coefficient* diperoleh sebesar 0,7217 untuk k=2, yang menunjukan bahwa anggota dalam satu *cluster* memiliki tingkat kemiripan yang tinggi, berbeda dengan anggota *cluster* lain, serta adanya kohesi dan separasi yang cukup baik antar *cluster*. Penelitian ini mengelompokan buku menjadi dua kategori, yaitu kurang diminati (*cluster* 0) dan sangat diminati (*cluster* 1). Dari hasil pengelompokan tersebut, diketahui bahwa beberapa buku yang sangat diminati hanya tersedia dalam jumlah eksemplar yang sedikit, sedangkan buku yang kurang diminati justru tersedia dalam jumlah yang terlalu banyak. Secara keseluruhan, penerapan algoritma *k-means clustering* dalam penelitian ini efektif untuk mengelompokkan koleksi buku berdasarkan tingkat keterpakaian. Selain itu, hasil penelitian ini menunjukkan bahwa teknik *data mining* dapat membantu perpustakaan dalam menyusun strategi pengembangan koleksi yang lebih efisien dan sesuai dengan kebutuhan pengguna.

ABSTRACT

Fatullah, Achmad Ghiffari. 2025. ***Book Classification Based on Usage Level Using the K-Means Clustering Algorithm at the Library of MAN 2 Tangerang City. Thesis. Library and Information Science Study Program. Faculty of Science and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors*: (I) Fakhris Khusnu Reza Mahfud, M.Kom. (II) Firma Sahrul Bahtiar, M.Eng.**

***Keywords**: K-Means, Clustering, Book Usage Rate, Data Mining, Book Grouping, Library*

The decline in book borrowing at the MAN 2 Tangerang City library highlights weaknesses in its collection management. Although it offers many titles, the collection does not fully meet user needs. The research aims to evaluate the book classification based on usage levels using the K-Means Clustering algorithm an approach designed to identify both highly popular and less popular books. Using a data mining method, the research consisted of conducting a literature review, collecting and pre-processing borrowing transaction data, applying the K-Means algorithm, and evaluating the clustering results with the Silhouette Coefficient method. The dataset, drawn from the SLiMS library information system, ranged from January 1, 2023, to January 1, 2024, and included 768 borrowing transactions. The clustering produced a Silhouette Coefficient score of 0.7217 for k=2, indicating strong similarity within clusters and clear separation among them. Books were classified into two categories: less popular (Cluster 0) and highly popular (Cluster 1). The results show an imbalance some highly popular books were available in only a few copies, while less popular books were overstocked. In general, the research demonstrates that the K-Means clustering algorithm effectively classifies books by their usage level. The findings also show that the data mining technique can support libraries in creating smarter, user-centered collection development strategies.

Translator,

Rizka Yanuarti
NIPPPK 197801242023212005

Date

مستخلص البحث

فتح الله، أحمد غفاري. 2025. **تصنيف الكتب حسب مستوى الاستخدام باستخدام خوارزمية التجميع K-Means في مكتبة مدرسة MAN 2 بمدينة تانجيرانجرسالة جامعية، برنامج دراسات المكتبات وعلوم المعلومات، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج المشرفان: (أولاً) فخرس خُسن ريزامحفوظ، المجستيرفي الاتصالات. (ثانيًا) فيرما سحر البحتيار، الماجستير في الهندسة.**

الكلمات المفتاحية: K-Means، التجميع، مستوى استخدام الكتب، تنقيب البيانات، تجميع الكتب، المكتبة

إن انخفاض معدل استعارة الكتب في مكتبة MAN 2 بمدينة تانجيرانج يدل على أن إدارة المجموعات المكتبية لم تكن بالمستوى المطلوب من الكفاءة، فرغم توفّر العديد من العناوين، إلا أن المجموعات لم تستطع تلبية حاجات القرّاء بشكل كامل. تهدف هذه الدراسة إلى قياس أداء تصنيف الكتب حسب درجة استخدامها، وذلك عبر استخدام خوارزمية التجميع K-Meansكحلّ لتحديد الكتب ذات الإقبال العالي وتلك ذات الإقبال المنخفض. وقد اتُبع في هذه الدراسة منهج تنقيب البيانات الذي مرَّ بعدة مراحل، منها: مراجعة الدراسات السابقة، جمع البيانات الخاصة بمعاملات الاستعارة ومعالجتها مبدئيا، ثم تطبيق خوارزمية K-Means، وتقييم نتائج التصنيف باستخدام معامل سيلهويت. (Silhouette Coefficient) وقد تم أخذ البيانات من نظام معلومات المكتبة SLiMS، وذلك للفترة الممتدة من 1 يناير 2023 إلى 1 يناير 2024، حيث بلغ عدد معاملات الاستعارة 768 معاملة. وقد بلغ معامل السيلهويت عند القيمة k=2 ما مقداره 0.7217، مما يدل على أن العناصر في كل عنقود تتّسم بدرجة عالية من التشابه فيما بينها، مع اختلاف واضح عن عناصر العناقيد الأخرى، مما يعكس تماسكًا داخليًا وفصلًا جيدًا بين العناقيد. وقد قسّمت الدراسة الكتب إلى فئتين: الكتب قليلة الإقبال (العنقود 0)، والكتب عالية الإقبال (العنقود 1). وأظهرت النتائج أن بعض الكتب ذات الإقبال العالي لم تتوفر إلا بنُسَخٍ قليلة، بينما توفرت الكتب قليلة الإقبال بعددٍ يزيد عن الحاجة. وبصورة عامة، فإن تطبيق خوارزمية K-Means في هذه الدراسة أثبت فعاليته في تصنيف مجموعات الكتب حسب مستوى استخدامها. كما تشير النتائج إلى أن تقنيات تنقيب البيانات يمكن أن تكون أداةً فعالة لمساعدة المكتبات في وضع استراتيجيات أكثر كفاءة لتطوير المجموعات بما يتناسب مع حاجات المستفيدين الفعلية.

BAB I
PENDAHULUAN

1.1 Latar Belakang

Keterpakaian koleksi tercermin dalam statistik peminjaman yang merupakan indikator penting dalam efektivitas dan relevansi pengelolaan koleksi terhadap kebutuhan pengguna (Wahyuntini & Endarti, 2021). Keterpakaian koleksi dapat bervariasi dan penurunannya menjadi perhatian karena mengindikasikan masalah dalam pelayanan atau kurangnya minat pengguna (Muhtadien & Krismayani, 2019). Keterpakaian koleksi berguna untuk mengetahui tingkat pelayanan dan perkembangan minat pemustaka yang kemudian dapat digunakan untuk membantu perpustakaan dalam membuat keputusan yang lebih baik dalam pengembangan koleksi.

Pengelolaan koleksi di perpustakaan sekolah harus dilakukan melalui proses seleksi yang berfokus pada kebutuhan pengguna, yaitu siswa dan guru. Koleksi harus disediakan relevan sesuai dengan kurikulum serta mendukung proses pembelajaran di sekolah mencakup pemilihan buku teks, referensi, dan bahan bacaan untuk meningkatkan minat baca dan pengetahuan siswa (Nabila & Sholihah, 2021). Dengan pengelolaan koleksi yang terorganisir, perpustakaan dapat mengidentifikasi kelompok literatur mana yang paling sering dipinjam dan diminati oleh pengguna (Ulfah & Irtwaty, 2022). Sebagaimana ﷺ berfirman dalam Q.S Al-Hujurat 49: ayat 13 sebagai berikut:

يٰٓاَيُّهَا النَّاسُ اِنَّا خَلَقْنٰكُمْ مِّنْ ذَكَرٍ وَّاُنْثٰى وَجَعَلْنٰكُمْ شُعُوْبًا وَّقَبَاۤىِٕلَ لِتَعَارَفُوْٓا ۚ اِنَّ اَكْرَمَكُمْ عِنْدَ اللّٰهِ اَتْقٰىكُمْ ۗ
اِنَّ اللّٰهَ عَلِيْمٌ خَبِيْرٌ ﴿١٣﴾

Artinya:

"*Wahai manusia, sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan. Kemudian, Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara*

kamu di sisi Allah adalah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui lagi Maha Teliti"(Al-Ḥujurāt [49]:13).

Ayat ini menjelaskan bahwa manusia diciptakan dari keturunan yang sama yaitu Adam dan Hawa, sehingga memiliki derajat kemanusiaan yang setara. Perbedaan bangsa dan suku diciptakan agar manusia saling mengenal dan membantu, bukan untuk saling bermusuhan atau mengolok-olok. Kemuliaan di sisi ﷻ الله terletak pada ketakwaan, bukan pada keturunan, kekayaan, atau jabatan. Oleh karena itu, manusia hendaknya berusaha meningkatkan ketakwaan agar menjadi yang paling mulia di hadapan-Nya. Allah Maha Mengetahui segala sesuatu, sehingga tidak ada perbuatan manusia yang luput dari pengawasan-Nya (quran.kemenag.go.id, 2025). Berdasarkan tafsir ayat ini ﷻ الله telah mengajarkan tentang pengelompokan manusia berdasarkan bangsa dan suku agar saling mengenal. Hal ini sejalan dengan pengelolaan bahan pustaka yang dikelompokkan berdasarkan disiplin ilmu untuk memudahkan pemustaka dalam temu balik informasi (Amanda, 2024).

Perpustakaan MAN 2 Kota Tangerang merupakan fasilitas yang disediakan untuk mendukung pembelajaran bagi pengguna dalam menggali ilmu pengetahuan sesuai dengan salah satu misi dari MAN 2 Kota Tangerang. Berdasarkan observasi yang telah dilakukan, Perpustakaan MAN 2 Kota Tangerang memiliki 3,919 eksemplar dari 473 judul buku yang meliputi buku paket bahan ajar umum dan islam, karya umum, fiksi, serta non fiksi, dan ensiklopedia. Meskipun banyak judul buku yang tersedia, pengelolaan koleksi di perpustakaan ini masih dianggap kurang optimal karena koleksi belum sepenuhnya mewadahi kebutuhan. Statistik peminjaman koleksi di perpustakaan MAN 2 Kota Tangerang menunjukkan penurunan, dengan rata-rata 84 eksemplar per bulan (total 1.228 eksemplar) pada tahun 2023 menjadi 49 eksemplar per bulan (total 589 eksemplar) pada tahun 2024. Hal ini disebabkan karena beberapa judul buku yang sedikit peminatnya terlalu banyak eksemplarnya, sedangkan buku yang lebih diminati justru tidak tersedia dalam jumlah yang cukup.

Dalam hal ini seharusnya perpustakaan MAN 2 Kota Tangerang melakukan musyawarah terlebih dahulu dengan beberapa pihak seperti pustakawan, guru, dan siswa agar keputusan yang diambil lebih bijaksana dan sesuai dengan kebutuhan pengguna (Apriyani et al., 2020). Solusi dalam menyelesaikan masalah yang terjadi berhubungan dengan firman ﷻ الله yang terdapat dalam Q.S Ali Imran :3 ayat 159 sebagai berikut:

فَبِمَا رَحْمَةٍ مِّنَ اللهِ لِنْتَ لَهُمْۚ وَلَوْ كُنْتَ فَظًّا غَلِيْظَ الْقَلْبِ لَانْفَضُّوْا مِنْ حَوْلِكَۖ فَاعْفُ عَنْهُمْ وَاسْتَغْفِرْ لَهُمْ وَشَاوِرْهُمْ فِى الْاَمْرِۚ فَاِذَا عَزَمْتَ فَتَوَكَّلْ عَلَى اللّٰهِۗ اِنَّ اللّٰهَ يُحِبُّ الْمُتَوَكِّلِيْنَ ﴿١٥٩﴾

Artinya:

"*Maka berkat rahmat Allah engkau (Muhammad) berlaku lemah lembut terhadap mereka. Sekiranya engkau bersikap keras dan berhati kasar, tentulah mereka akan menjauhkan diri dari sekitarmu. Karena itu, maafkanlah mereka dan mohonkanlah ampunan untuk mereka. Dan bermusyawarahlah dengan mereka dalam urusan itu. kemudian apabila engkau telah membulatkan tekad, maka bertawakallah kepada Allah. Sungguh, Allah mencintai orang-orang yang bertawakal*" (Al-Imran[3]:159).

Dalam Tafsir Al-Misbah, Quraisy shihab menerangkan ayat ini merupakan perintah untuk melakukan musyawarah. Beliau menekankan bahwa musyawarah merupakan proses kerja sama, kedisiplinan, dan saling menghargai pendapat orang lain, agar menghasilkan keputusan yang terbaik (Shihab, 2002). Selaras dengan tafsir kemenag, perintah musyawarah ini mengandung pesan, sebagai pemimpin tidak boleh meninggalkan musyawarah karena dengan musyawarah dapat memperoleh pandangan dan keinginan dari masyarakat. Esensi musyawarah sendiri adalah pemberian kesempatan kepada orang lain yang memiliki hak untuk berpartisipasi dalam membuat keputusan (Lajnah Pentashihan Mushaf al-Qur`an Kemenag RI, 2009). Tafsir dari ayat ini berhubungan dengan aspek pengelolaan koleksi di perpustakaan. Dimana musyawarah berperan sebagai proses kolaboratif yang dapat memperkuat pengambilan

keputusan agar lebih bijaksana terkait pengadaan koleksi sesuai dengan kebutuhan pengguna.

Berdasarkan permasalahan yang ada pada Perpustakaan MAN 2 Kota Tangerang terkait pengelolaan koleksi, maka perpustakaan tersebut dapat menerapkan teknik *data mining*. *Data mining* merupakan proses penemuan pengetahuan dari basis data *Knowledge Discovery in Databases* (KDD) yang memiliki serangkaian proses untuk menggali nilai tambah berupa informasi tidak diketahui secara manual dari suatu basis data (Marcoulides, 2005). Proses *data mining* menghasilkan informasi berupa angka atau data yang relevan untuk berbagai kebutuhan, dan sering kali dilakukan dengan menggunakan perangkat lunak yang didukung oleh teknik statistika, matematika, atau teknologi terkini seperti *Artificial Intelligence* (AI) (Alkhairi & Windarto, 2019).

Salah satu metode yang dapat digunakan dalam penerapan *data mining* yaitu *clustering*. *Clustering* merupakan salah satu teknik *data mining* untuk melakukan pengelompokan sejumlah data atau objek ke dalam *cluster* (*group*) sehingga setiap *cluster* akan berisi data semirip mungkin dan berbeda dengan objek lainnya (Lestari, 2019). Menurut (Nasir, 2021) dalam pengelompokan buku, metode *clustering* memiliki beberapa keunggulan, Pertama memberikan efisiensi dan skalabilitas tinggi bahkan untuk data bervolume besar. Kedua *clustering* dapat menghasilkan visualisasi data yang mudah diinterpretasikan dan tidak memerlukan label data awal. Terakhir, metode ini memiliki kemampuan reduksi dimensi yang mempermudah visualisasi data kompleks.

Untuk mendukung proses *clustering* maka diperlukan sebuah algoritma, salah satunya adalah *k-means*. Algoritma *k-means* dapat digunakan untuk mengelompokkan data yang besar berdasarkan titik pusatnya (*centroid*) (Huang et al., 2024). Selain itu, data yang akan dikelompokkan diambil secara acak atau secara kebetulan serta jumlah kelompok data akan dibuat dan ditentukan oleh variabel **k**. *K-means* tidak memerlukan label atau kategori pada data set saat pengelompokan (Ahmed et al., 2020). Penggunaan algoritma *k-means* dalam pengelolaan koleksi buku di perpustakaan sangat relevan

karena algoritmanya dapat mengelompokkan data besar berdasarkan titik pusat (*centroid*) kemudian dilakukan identifikasi berdasarkan pola minat pengguna. Penggunaan k-means memiliki kelebihan berupa kesederhanaan karena tidak memiliki banyak parameter yang rumit, mudah diatur, dan proses komputasi yang efisien. Hal ini menyebabkan penggunaan memori yang relatif sedikit untuk menyimpan titik-titik pusat dan tabel *cluster* (Wang et al., 2020).

Selanjutnya pada penggunaan algoritma *k-means* maka diperlukan tahapan evaluasi untuk mengetahui kualitas pengelompokan atau *clustering*. Metode evaluasi tersebut adalah *silhouette coefficient*. *Silhouette coefficient* merupakan metrik yang digunakan untuk menilai sejauh mana *cluster* yang terbentuk saling terpisah dengan baik dengan melihat nilai *koefisien* (Tambunan et al., 2020). Nilai *silhouette coefficient* berkisar antara -1 hingga +1, yang mana semakin tinggi nilai koefisiennya, semakin kohesif *cluster* tersebut. Nilai mendekati +1 mengindikasikan bahwa sampel berada jauh dari *cluster* tetangganya menunjukkan pengelompokan yang baik. Sebaliknya, nilai negatif mengisyaratkan kemungkinan sampel ditempatkan pada *cluster* yang tidak tepat (Shahapure & Nicholas, 2020). *Silhouette coefficient* mengintegrasikan kohesi (keterikatan antar objek dalam *cluster*) dan pemisahan (jarak rata-rata objek ke *cluster* terdekat), dengan perhitungan jarak menggunakan rumus *euclidean* (Hartama & Anjelita, 2022). Kelebihan dari *silhouette coefficient* yaitu, menyediakan analisis *silhouette* menggunakan visualisasi yang jelas melalui plot *silhouette* untuk mempermudah penentuan jumlah *cluster* agar lebih optimal, serta efektivitasnya dalam data berdimensi tinggi (Sumallika et al., 2024). Dengan demikian, *silhouette coefficient* merupakan alat yang sangat berguna untuk memahami struktur data dan meningkatkan pengambilan keputusan terkait metode *clustering* yang digunakan (Shahapure & Nicholas, 2020).

Pada penelitian sebelumnya dengan judul “Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik Lpp Yogyakarta)”. Penelitian tersebut bertujuan untuk mengoptimalkan ketersediaan buku di Perpustakaan Politeknik LPP Yogyakarta

dengan menganalisis data sirkulasi peminjaman buku. Hasil dari penelitian yang dilakukan menghasilkan pengelompokan buku berdasarkan tingkat permintaan menggunakan metode *k-means clustering*, yang membagi buku menjadi tiga kategori: paling diminati, cukup diminati, dan sedikit diminati. Dari 30 judul buku yang diuji, tiga diantaranya diidentifikasi sebagai buku paling diminati, yaitu Buku Pintar Mandor (BPM) Seri Budidaya Tanaman Tebu, Buku Pintar Mandor (BPM) Tanaman Kelapa Sawit, dan Kupas Tuntas Agribisnis Kelapa Sawit. Hasil dari penelitian yang dilakukan diharapkan dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan untuk memperbanyak eksemplar buku yang paling diminati, sehingga rasio antara ketersediaan dan banyaknya peminat dapat seimbang (Hasanah & Purnomo, 2022).

Berdasarkan permasalahan dan solusi yang telah diuraikan, maka diusulkan penelitian dengan judul "Pengelompokan Buku Berdasarkan Tingkat Keterpakaian Menggunakan Algoritma K-Means Clustering Pada Perpustakaan Man 2 Kota Tangerang". Tahapan penelitian ini dimulai dengan studi literatur untuk memahami teori dan aplikasi algoritma tersebut. Selanjutnya, pengumpulan data riwayat peminjaman buku dari Perpustakaan MAN 2 Kota Tangerang dalam jangka waktu 1 tahun terhitung mulai dari 1 Januari 2023 sampai 1 Januari 2024. Kemudian melakukan *preprocessing* data. Setelah itu, Algoritma *k-means clustering* diterapkan kemudian menggunakan *silhouette coefficient* untuk melakukan evaluasi hasil. Terakhir, peneliti menarik kesimpulan dan saran berdasarkan temuan analisis untuk membantu pengelompokan buku yang sangat diminati dan kurang diminati yang ada di Perpustakaan MAN 2 Kota Tangerang. Hasil dari penelitian ini diharapkan dapat membantu pengelola perpustakaan untuk melakukan analisis dan mengambil keputusan.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, identifikasi masalah dalam penelitian ini yaitu berapa hasil kinerja pengelompokan buku berdasarkan tingkat

keterpakaian menggunakan algoritma *k-means clustering* pada Perpustakaan MAN 2 Kota Tangerang?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah mengukur hasil kinerja pengelompokan buku berdasarkan tingkat keterpakaian menggunakan algoritma *k-means clustering* pada Perpustakaan MAN 2 Kota Tangerang.

1.4 Manfaat Penelitian

Adapun manfaat yang akan didapat dari penelitian ini yaitu:

1. Dengan penelitian ini dapat memberikan solusi terhadap masalah pengelolaan koleksi buku di Perpustakaan MAN 2 Kota Tangerang. Dengan menerapkan algoritma *k-means clustering*, pengelola perpustakaan dapat lebih efektif dalam mengidentifikasi buku yang sangat diminati dan kurang diminati.
2. Hasil dari penelitian ini dapat memberikan informasi bagi pihak Perpustakaan MAN 2 Kota Tangerang mengenai kebijakan dalam pengembangan koleksi.
3. Penelitian ini dapat menjadi referensi bagi penelitian selanjutnya terkait pengelolaan perpustakaan dan penggunaan metode *data mining* dalam konteks pendidikan, serta memberikan wawasan baru tentang penerapan teknologi dalam pengelolaan koleksi perpustakaan.

1.5 Batasan Masalah

Penelitian ini memiliki sejumlah batasan yang membantu memandu dan merinci ruang lingkup serta metode yang akan digunakan sebagai berikut:

1. Penelitian ini tidak mencakup aspek lain selain dari pengelompokan data transaksi peminjaman perpustakaan.
2. Data yang digunakan dalam penelitian ini adalah data transaksi peminjaman buku di Perpustakaan MAN 2 Kota Tangerang dalam jangka waktu 1 tahun mulai 1 Januari 2023 sampai 1 Januari 2024.
3. Variabel yang digunakan dalam penelitian ini terdiri dari empat kategori utama meliputi judul buku, jumlah peminjaman, jumlah anggota yang meminjam dan jumlah eksemplar.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi diperlukan supaya mudah dipahami pembaca dan memudahkan peneliti dalam menyusun skripsi. Penjelasan mengenai isi dari tiap bab adalah sebagai berikut:

BAB I: PENDAHULUAN

Pada bab pendahuluan ini menjelaskan beberapa sub bab yang meliputi latar belakang tentang pengelompokan buku berdasarkan tingkat keterpakaian pada Perpustakaan MAN 2 Kota Tangerang menggunakan algoritma k-means clustering, identifikasi masalah terkait bagaimana hasil *centroid* dari pengelompokan dan pengadaan buku menggunaka algoritma k-means *clustering* pada Perpustakaan MAN 2 Kota Tangerang, tujuan penelitian tentang hasil *centroid* dari pengelompokan buku berdasarkan tingkat keterpakaian menggunakan algoritma *k-means clustering* pada Perpustakaan MAN 2 Kota Tangerang, manfaat penelitian yang dilakukan, batasan masalah dari penelitian, dan sistematika penulisan.

BAB II: TINNJAUAN PUSTAKA

Pada Bab kedua ini akan membahas mengenai tinjauan pustaka dari penelitian terlebih dahulu terkait penerapan *k-means clustering*. Selain itu teori dasar tentang *data mining*, *clustering* dan algoritma *k-means* juga akan dibahas pada bab ini.

BAB III: METODE PENELITIAN

Pada Bab ketiga ini berisikan pola dan rancangan penelitian, bahan penelitian, alat penelitian, alur penelitian, serta analisis hasil penelitian yang digunakan oleh peneliti dalam penelitian ini yang menggunakan metode algoritma *k-means clustering*.

BAB IV: HASIL DAN PEMBAHASAN

Pada bab keempat ini merupakan isi dari hasil penelitian yang dikaji dan dianalisis secara sistematis menggunakan algoritma *k-means clustering* dan *silhouette coefficient* untuk melakukan evaluasi sesuai pembahasan, yakni bagaimana hasil dari penelitian ini dapat membantu pengelompokan buku yang sangat diminati dan kurang diminati yang ada di Perpustakaan MAN 2 Kota Tangerang.

BAB V: KESIMPULAN

Bab kelima ini terdiri dari dua sub bab, yaitu kesimpulan dan saran. Kesimpulan berisi ringkasan dari temuan yang telah dijelaskan dalam hasil dan pembahasan. Sedangkan saran diberikan kepada pihak-pihak yang terkait dengan penelitian yang telah dilakukan dan perbaikan kekurangan penelitian yang dapat dilakukan pada penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Dalam penelitian yang berjudul "Implementasi Data Mining Untuk Pengelompokan Koleksi Buku Di Perpustakaan Perpustakaan Kota Balikpapan Dengan Metode K-Means Clustering". Penelitian ini dilakukan untuk mengelompokan buku di Perpustakaan Kota Balikpapan berdasarkan frekuensi peminjaman dengan menggunakan metode *k-means clustering*. Tujuan dari penelitian ini yaitu untuk mengidentifikasi kelompok judul buku mana yang paling sering dipinjam oleh pengguna. Data yang digunakan dalam jurnal tersebut adalah data peminjaman buku dari tahun 2021 hingga 2023. Penelitian ini mengumpulkan data sebanyak 100 judul buku setiap tahun, dengan total 300 judul buku yang diperiksa. Dari data tersebut, diperoleh 218 judul buku dengan total frekuensi peminjaman sebanyak 2866 kali. Data ini kemudian diolah menggunakan aplikasi *rapidminer* untuk melakukan *clustering* berdasarkan frekuensi peminjaman. Hasil dari penelitian ini berupa tiga *cluster*: *cluster* 0 (135 judul buku jarang dipinjam), *cluster* 1 (35 judul buku paling sering dipinjam), dan *cluster* 2 (48 judul buku sering dipinjam). *Nilai Davies-Bouldin Index* (DBI) untuk **k**=3 adalah 0,541, yang menunjukkan kualitas *clustering* (Ulfah et al., 2024). Dari penelitian diatas, maka diperoleh perbedaan yaitu; penelitian tersebut lebih berfokus pada pemenuhan stok dibandingkan dengan memprioritaskan koleksi berdasarkan frekuensi peminjaman dan *Davies Bouldin Index* sebagai tahapan evaluasi analisis hasil *clustering*. Sedangkan pada penelitian yang akan dilakukan menggunakan *silhouette coefficient* sebagai matrik untuk melakukan evaluasi analisis hasil *clustering*. Sementara itu persamaan terletak pada metode dan algoritma yang digunakan yaitu *k-means clustering*.

Penelitian lain yang membahas tentang penerapan *k-means clustering* untuk manajemen koleksi perpustakaan terdapat pada "Pengelompokkan Buku Dan Rekomendasi Buku Menggunakan K-Means Clustering Pada Dinas Perpustakaan Dan

Kearsipan Kota Medan". Penelitian ini dilakukan karena Dinas Perpustakaan dan Kearsipan Kota Medan belum mempunyai serta mengetahui kelompok buku apa yang paling banyak diminati dan judul buku apa yang dapat direkomendasikan kepada pengunjung. Data yang digunakan terkait penelitian berupa data peminjaman buku dari 2021 sampai 2023. Setelah data terkumpul, kemudian menerapkan metode *k-means clustering* untuk mengelompokkan buku berdasarkan jumlah peminjaman, sehingga dapat mengidentifikasi buku yang paling diminati. Proses tersebut menghasilkan tiga *cluster*: *cluster* 0 (kurang diminati) dengan 362 buku, *cluster* 1 (diminati) dengan 43 buku, dan *cluster* 2 (paling diminati) dengan 8 buku pada penelitian tersebut (Fadlina et al., 2024). Penelitian ini dilakukan hanya menggunakan variable frekuensi peminjaman dan jumlah peminjaman, serta menggunakan aplikasi *Weka* 3.9.6 untuk pengujian lebih lanjut. Sedangkan penelitian yang akan dilakukan menggunakan variabel judul buku, jumlah peminjaman, jumlah anggota yang meminjam dan jumlah eksemplar. Persamaan terletak pada teknik *data mining,* metode *clustering* dan algoritma *k-means*.

Penelitian selanjutnya yang membahas tentang penerapan *clustering* untuk manajemen koleksi perpustakaan berjudul "Sistem Pendukung Keputusan Penentuan Prioritas Pengadaan Buku Perpustakaan Menggunakan Metode K-Means Dan Electre". Penelitian ini berfokus pada pengembangan sistem rekomendasi untuk koleksi perpustakaan STAMI menggunakan metode *k-means clustering* dan *ELECTRE*. Penelitian ini dilakukan untuk mengatasi masalah dalam pengadaan buku di perpustakaan Sekolah Tinggi Akuntansi dan Manajemen Indonesia (STAMI), yang saat ini masih dilakukan secara manual serta mengalami kesulitan dalam melakukan seleksi bahan pustaka, kesulitan tersebut mengakibatkan proses pengadaan koleksi tidak berjalan dengan optimal. Data terkait penelitian didapatkan melalui survei dan wawancara. Data tersebut terdiri dari 163 judul buku yang telah dibersihkan (data *cleaning*) untuk memastikan kualitas dan relevansi informasi yang digunakan dalam analisis. Hasil penelitian menunjukan sebuah sistem pendukung keputusan berbasis web yang terdiri dari 11 kelompok buku, yaitu: Buku-013, Buku-063, Buku-072, Buku-

074, Buku-075, Buku-076, Buku-084, Buku-092, Buku-102, Buku-122, dan Buku-125. Penelitian ini menggunakan variabel berupa jumlah peminjaman, jumlah eksemplar, harga buku, dan jurusan serta menggunakan dua metode yaitu *k-means clustering* dan *ELECTRE* (Purba et al., 2023). Sedangkan pada penelitian yang akan dilakukan hanya menggunakan variabel judul buku, jumlah peminjaman, jumlah anggota yang meminjam dan jumlah eksemplar serta menggunakan satu metode yaitu *k-means clustering*. Persamaan terletak pada teknik *data mining*, metode *clustering* dan algoritma *k-means*.

Penelitian keempat terkait penerapan *clustering* untuk manajemen koleksi perpustakaan berjudul "Penerapan Data Mining Clustering Menggunakan Metode K-Means Dalam Pengelompokan Buku Perpustakaan Politeknik Negeri Balikpapan". Penelitian ini dilakukan untuk mengatasi permasalahan pengadaan koleksi buku di Perpustakaan Politeknik Negeri Balikpapan, karena pengelola perpustakaan tidak mengetahui koleksi apa saja yang harus ditambahkan sesuai kebutuhan pengguna. Data terkait penelitian yaitu data peminjaman buku dari tahun 2019 sampai 2022. Kemudian data yang diperoleh diolah menggunakan *software rapidminer* untuk menerapkan *k-means clustering*, dan terakhir melakukan analisis hasil *clustering* menggunakan *Davies Bouldin Index*. Hasil penelitian didapati melalui aplikasi *rapidminer* dengan **k**=3, didapatkan hasil *cluster* 0 (rendah) terdiri 82 judul buku dengan frekuensi peminjaman buku nya dalam kategori jarang dengan kata lain kurang diminati untuk dipinjam, *cluster* 1 (sedang) terdiri dari 23 judul buku yang merupakan buku-buku dengan frekuensi peminjaman sedang dengan kata lain dikategorikan diminati untuk dipinjam, *cluster* 2 (tinggi) terdiri dari 2 judul buku merupakan kelompok buku yang paling diminati yang terdiri dari 2 judul buku yakni Teknologi Beton dan buku Teori dan Praktik Hotel Front Of Ice. Hasil performansi didapatkan nilai *Index Davies Bouldin* sebesar 0,407 (Ulfah & Irtwaty, 2022). Penelitian ini menggunakan frekuensi peminjaman sebagai variabel dan *Davies Bouldin Index* sebagai tahapan evaluasi analisis hasil *clustering*. Sedangkan pada penelitian yang akan dilakukan menggunakan *silhouette coefficient* sebagai matrik untuk melakukan evaluasi analisis hasil *clustering*.

Penelitian terdahulu terakhir terkait penerapan *clustering* untuk manajemen koleksi perpustakaan berjudul "Analisis Klastering Buku sebagai Evaluasi untuk Peningkatan Minat Baca Perpusatakaan SMAN 1 Grogol". Penelitian ini dilakukan untuk mengoptimalkan penempatan koleksi dan menentukan prioritas pengadaan koleksi berdasarkan minat baca siswa di Perpustakaan SMAN 1 Grogol. Data terkait penelitian ini yaitu data peminjaman buku dari perpustakaan SMAN 1 Grogol pada tahun pelajaran 2018/2019 - 2019/2020, dengan total 8135 records peminjaman. Data tersebut terdiri dari 120 sampel buku dan 2 atribut data, yaitu judul buku dan frekuensi pinjam. Data tersebut kemudian diolah menggunakan tiga algoritma *clustering* yaitu, *k-means*, *k-medoids*, dan *fuzzy c-means*. Selain itu hasil dari perhitungan menggunakan algoritma tersebut, penelitian ini juga menggunakan *silhouette coefficient* untuk menghitung dan mengukur kualitas pengelompokan dalam analisis data. Hasil dari penelitian ini menunjukkan hasil dari indeks *silhouette coefficient*, jumlah *cluster* terbaik dari algoritma *k-means* berada pada *cluster* 9 memiliki nilai indeks sebesar 0,292, algoritma *k-medoids* memiliki nilai indeks sebesar 0,355 terletak pada *cluster* 10, dan algoritma *fuzzy c-means* memiliki nilai indeks sebesar 0,323 terletak pada *cluster* 10 (Karputri & Yustanti, 2022). Perbedaan dengan penelitian yang akan dilakukan adalah penelitian terakhir hanya mempertimbangkan jumlah peminjaman dan judul sebagai variabel *clustering*. Selain itu penelitian ini menggunakan tiga algoritma *clustering* yaitu *k-means*, *k-medoids*, dan *fuzzy c-Means*. Sedangkan pada penelitian yang akan dilakukan menggunakan variabel judul buku, jumlah peminjaman, jumlah anggota yang meminjam dan jumlah eksemplar. Algoritma yang digunakan hanya satu yaitu *k-means clustering*. Selain perbedaan, terdapat persamaan dengan penelitian yang akan dilakukan dengan penelitian terdahulu yaitu menggunakan *silhouette coefficient* untuk untuk menghitung dan mengukur kualitas pengelompokan dalam analisis data.

2.2 Landasan Teori

2.2.1 Pengembangan Koleksi

Pengembangan koleksi dalam konteks perpustakaan sekolah merupakan suatu kegiatan penting yang bertujuan untuk menentukan jenis koleksi yang sesuai dengan kebutuhan pengguna, yaitu siswa dan guru (Wegmann et al., 2021). Pengembangan koleksi adalah serangkaian kegiatan yang mencangkup berbagai aspek terkait dengan peningkatan koleksi di perpustakaan, aspek tersebut meliputi penetapan dan penyesuaian kebijakan seleksi, penilaian kebutuhan pemakai, penelitian pemakaian, evaluasi, identifikasi kebutuhan, pemilihan koleksi, perencanaan kolaborasi dengan sumber bahan pustaka, serta pemeliharaan dan penyiangan koleksi perpustakaan (Yudisman & Rahmi, 2020).

Dalam pengembangan koleksi di perpustakaan, penting untuk memiliki kebijakan secara tertulis yang menjadi dasar untuk perpustakaan dalam meningkatkan kualitas koleksi sesuai dengan kebutuhan pengguna (Levenson & Nichols, 2020). Selain itu dalam pengelolaan koleksi juga dibutuhkan sebuah analisis untuk mengidentifikasi dan mengevaluasi faktor-faktor internal dan eksternal yang mempengaruhi kualitas perpustakaan (Khan & Bhatti, 2021). Menurut Akinola & Opawale (2020) analisis SWOT (*Strengths*, *Weaknesses*, *Opportunities*, *Threats*) dalam pengembangan koleksi perpustakaan adalah alat strategis yang dapat digunakan untuk mengevaluasi posisi perpustakaan dalam konteks pengembangan koleksi. Dengan kata lain menggunakan analisis SWOT (*Strengths*, *Weaknesses*, *Opportunities*, *Threats*) dapat berkontribusi pada perpustakaan dalam pengelolaan koleksi untuk penciptaan keunggulan kompetitif dengan menarik lebih banyak pengguna melalui koleksi yang relevan (Awojobi & Uthman, 2022). Selain itu analisis ini juga dapat membantu perpustakaan untuk menetapkan prioritas dalam pengelolaan koleksi sesuai dengan kebutuhan pengguna (Rochemin & Wicaksono, 2023).

Pengelolaan koleksi merupakan bagian dari pengembangan koleksi dimana perpustakan dapat menambah koleksi yang belum tersedia. Proses ini dilakukan sesuai

dengan jenis perpustakan dan kebutuhan pemustaka (Fatwa, 2020). Pengelolaan koleksi juga merupakan sebuah upaya untuk menyediakan informasi yang dibutuhkan oleh pemustaka melalui berbagai metode, seperti pembelian, pertukaran, hadiah, dan kerjasama. Proses ini harus direncanakan secara jelas dan terukur agar koleksi yang ada dapat memenuhi kebutuhan informasi secara optimal (Ardyawin, 2018). Menurut Golung et al (2023) pengelolaan koleksi yang efektif harus didasarkan pada prinsip-prinsip seperti relevansi, individualisasi, kelengkapan, kemutakhiran, dan kerjasama. Prinsip relevansi memastikan bahwa koleksi yang ada sesuai dengan kebutuhan masyarakat yang dilayani oleh perpustakaan. Individualisasi menekankan pentingnya memenuhi kebutuhan informasi dari berbagai individu, termasuk pelajar, profesional, dan masyarakat umum. Prinsip kelengkapan mengharuskan koleksi mencakup berbagai bidang dan tidak hanya terbatas pada materi pokok, tetapi juga sumber informasi tambahan. Prinsip kemutakhiran menjamin bahwa koleksi selalu diperbarui sesuai dengan perkembangan ilmu pengetahuan dan teknologi. Terakhir, prinsip kerjasama menekankan pentingnya koordinasi antara pihak-pihak yang membutuhkan informasi untuk memastikan pengadaan koleksi berjalan secara efektif dan efisien.

2.2.2 *Data Mining*

Definisi *data mining* adalah serangkaian proses untuk menambah serta mencari informasi baru dari basis data, informasi tersebut sebelumnya tidak diketahui secara manual. Proses ini menghasilkan informasi dengan mengekstraksi dan mengenali pola-pola penting atau menarik dalam data yang tersimpan (Ani et al., 2021). Umumnya, *data mining* digunakan untuk menggali pengetahuan dari basis data besar yang bertujuan untuk mengidentifikasi pola tersembunyi dalam data agar dapat mendukung dalam proses pengambilan keputusan, Sehingga sering disebut sebagai *Knowledge Discovery in Databases* (KDD) (Syahril et al., 2020). Sedangkan menurut Putra & Wadisman (2018) *data mining* merupakan sebuah proses untuk mencari pola atau informasi menarik dalam suatu data dengan menggunakan teknik seperti *clustering* dan klasifikasi yang melibatkan metode dari bagian *Artificial Intelligence* (AI), *Machine*

Learning (ML), dan *Statistic*. Pemilihan metode atau algoritma yang tepat bergantung pada tujuan serta keseluruhan proses *Knowledge Discovery in Databases* (KDD).

Berdasarkan pengertian yang telah dijelaskan, secara garis besar *data mining* merupakan sebuah proses sistematis yang bertujuan untuk mengekstrak informasi berharga dari basis data besar dengan mengidentifikasi pola, tren, dan hubungan yang tersembunyi. Proses ini melibatkan penggunaan teknik dan algoritma dari bidang *Artificial intelligence* (AI), *Machine learning* (ML), dan *Statistic* untuk menemukan pola-pola yang signifikan dalam data. *Data mining* sering kali disebut sebagai *Knowledge Discovery in Databases* (KDD), karena fokus utamanya adalah menggali pengetahuan baru yang dapat digunakan untuk mendukung pengambilan keputusan dalam berbagai konteks, termasuk bisnis, kesehatan, dan penelitian. Pemilihan metode yang tepat dalam data mining sangat bergantung pada tujuan analisis serta karakteristik data yang sedang diproses.

Dalam penggunaan *data mining* untuk pengelompokan buku memberikan sejumlah keuntungan bagi perpustakaan, seperti meningkatkan efisiensi pengelolaan koleksi dan membantu dalam mengidentifikasi minat baca pengunjung. Hal ini memungkinkan perpustakaan untuk mengatur koleksi buku secara lebih sistematis berdasarkan pola peminjaman dan preferensi pengguna, sehingga dapat memenuhi kebutuhan informasi dengan lebih baik (Fitriani et al., 2021). Selain itu, penggunaan *data mining* juga dapat membantu dalam menemukan segmen pasar yang belum terjamah, sehingga perpustakaan dapat menambah koleksi buku sesuai dengan permintaan atau minat dari pengguna (Firmansyah et al., 2022).

2.2.3 *Clustering*

Clustering adalah salah satu teknik yang digunakan dalam *data mining* untuk mengelompokkan data menjadi subset serupa berdasarkan karakteristik atau pola tertentu. Tujuan utama dari proses ini adalah mengidentifikasi struktur tersembunyi dalam data, sehingga dapat memberikan pemahaman yang lebih mendalam tentang kelompok atau kategori yang ada di dalamnya (Hendrastuty, 2024). Sedangkan *clustering* menurut Werdiningsih et al (2020) merupakan sebuah proses pengamatan

dan pengelompokan atribut dalam data untuk membentuk kelas objek yang memiliki kemiripan. Proses ini berbeda karena tidak ada definisi kelas objek dalam pengklasteran yang dikenal sebagai *unsupervised learning* atau pembelajaran tidak terawasi. *Clustering* membagi keseluruhan data menjadi beberapa kelompok yang memiliki kesamaan (homogenitas), di mana tingkat kesamaan dalam satu kelompok akan maksimal, sedangkan kesamaan antar kelompok lainnya akan bernilai minimal (Suhendra et al., 2022).

Dalam buku *data mining* dan penerapan algoritma (Ginantra et al., 2021) penerapan *clustering* memiliki beberapa metode, diantaranya:

1. Metode *Partitioning Clustering*. Metode ini merupakan salah satu pilihan paling populer di kalangan analis untuk melakukan pengelompokan, karena kesederhanaannya dan efektivitasnya dalam menetapkan titik data ke dalam *cluster*. Konsep dasar dari metode ini adalah bahwa *cluster* ditandai dan diwakili oleh vektor pusat, di mana titik data yang dekat dengan vektor tersebut akan dimasukkan ke dalam *cluster* yang sama. Proses pengelompokan ini dilakukan secara berulang dengan mengukur jarak antara *cluster* dan *centroid* menggunakan berbagai matrik jarak (Ginantra et al., 2021).
2. Metode *Hierarchical Clustering*. Metode ini proses pengelompokan *unsupervised learning* yang dimulai dengan struktur cluster dari atas ke bawah, di mana *cluster* telah ditentukan sebelumnya. Proses ini melibatkan dekomposisi objek data berdasarkan hierarki untuk menghasilkan *cluster* yang diinginkan. Metode ini dapat dibagi menjadi dua pendekatan utama yaitu *agglomerative* (*bottom-up*) dan *divisive* (*top-down*). Dalam pendekatan *agglomerative*, setiap objek dimulai sebagai *cluster* tersendiri dan kemudian digabungkan menjadi *cluster* yang lebih besar, sedangkan dalam pendekatan *divisive*, semua objek dimulai dalam satu *cluster* besar yang kemudian dibagi menjadi *cluster* yang lebih kecil (Roy et al., 2019).
3. Metode *Density-Based Clustering*. Metode ini berfokus pada pengelompokan data dengan cara mengidentifikasi area di mana titik-titik data terkonsentrasi dan

memisahkannya dari area yang lebih jarang atau kosong. Metode ini sangat efektif dalam menentukan *cluster* dengan bentuk arbitrer dan dapat menangani noise dengan baik. Contoh terkenal dari metode ini adalah *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), yang hanya memerlukan dua parameter: *epsilon* (ε), yang menentukan radius pencarian, dan minimum jumlah titik (MinPts) yang diperlukan untuk membentuk *cluster* (Campello et al., 2020).

4. Metode *Grid-Based Clustering*. Metode ini merupakan pendekatan yang pertama-tama merangkum dataset menggunakan representasi grid, kemudian menggabungkan sel-sel grid untuk membentuk *cluster*. Pendekatan ini memungkinkan pengolahan data yang lebih cepat dan efisien, terutama ketika berhadapan dengan volume data yang besar. Dengan menggunakan struktur grid, metode ini dapat mengurangi kompleksitas komputasi dan mempercepat proses *clustering* dibandingkan dengan metode yang langsung bekerja pada data mentah (Campello et al., 2020).
5. Metode *Distribution-Based Clustering*. Metode ini mengasumsikan bahwa data dihasilkan dari campuran beberapa distribusi probabilitas, seperti distribusi *Gaussian*. Metode ini memungkinkan fleksibilitas dalam menangani *cluster* dengan bentuk dan ukuran yang berbeda, serta dapat memberikan informasi probabilistik mengenai keanggotaan titik data dalam *cluster* tertentu. Pendekatan ini berbeda dengan metode *centroid-based* yang menggunakan metrik jarak untuk menentukan pengelompokan (Alasali & Ortakci, 2024).
6. Metode *Fuzzy Clustering*. Merupakan metode yang mengelompokan setiap titik data dapat menjadi anggota dari beberapa *cluster* dengan tingkat keanggotaan yang berbeda. Metode ini mengoptimalkan pembagian data dengan mempertimbangkan derajat keanggotaan, sehingga memungkinkan untuk mengelompokkan data yang memiliki karakteristik serupa meskipun ada tumpang tindih antar *cluster*. Pendekatan ini sangat berguna dalam situasi di mana data memiliki ketidakpastian atau ketika batasan antar kategori tidak jelas (Rouza & Fimawahib, 2020).

7. Metode *Constraint-Based Clustering*. Merupakan metode yang memperluas teknik pengelompokan tradisional dengan memasukkan pengetahuan ahli melalui penggunaan batasan yang ditentukan oleh pengguna. Batasan ini dapat berupa hubungan antara data, seperti must-link (data harus berada dalam *cluster* yang sama) dan *cannot-link* (data tidak boleh berada dalam *cluster* yang sama). Pendekatan ini bertujuan untuk meningkatkan relevansi hasil *clustering* dengan mempertimbangkan informasi tambahan yang mengarahkan pembentukan *cluster* (Gançarski et al., 2020).

2.2.4 Algoritma *K-Means*

Algoritma *k-means* ditemukan oleh *Macqueen* pada tahun 1967. Metode ini dikenal sebagai algoritma yang sederhana namun efektif dalam proses pembelajaran. *K-means* merupakan sebuah algoritma *clustering* yang sering digunakan dalam data mining, terutama untuk mengelompokkan data dalam jumlah besar (Izzah et al., 2023). Algoritma ini berfungsi dengan memilih sejumlah objek **k** sebagai pusat awal *cluster*. Jarak antara masing-masing pusat *cluster* dan objek dihitung, kemudian objek tersebut ditetapkan pada jumlah *cluster* terdekat. Proses ini diulang dengan memperbarui rata-rata semua *cluster* hingga konvergensi tercapai (Gantara & Ali, 2023).

K-means merupakan algoritma *clustering* yang termasuk dalam kategori *unsupervised learning*, yang bertujuan untuk menemukan maksimum lokal pada setiap iterasi. Dengan kata lain *k-means* adalah algoritma yang memproses data tanpa memerlukan label kelas sebagai masukan (Lestari & Sumarlinda, 2021). Menurut Ginantra et al., (2021) data *clustering* menggunakan metode *k-means* secara umum dilakukan dengan algoritma dasar sebagai berikut:

1. Menentukan jumlah *cluster* $\boldsymbol{k}$ yang diinginkan, yang biasanya ditentukan oleh pengguna. Misalnya jika kita ingin membagi data menjadi 2 *cluster*, maka $\boldsymbol{k} = 2$.
2. Menghitung jarak *centroid*/rata-rata dari data yang berada pada masing-masing *cluster*. Dalam pengukuran jarak antar *cluster* algoritma yang sering digunakan adalah *euclidean distance*, dengan rumus sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (2.\ 1)$$

Dimana:

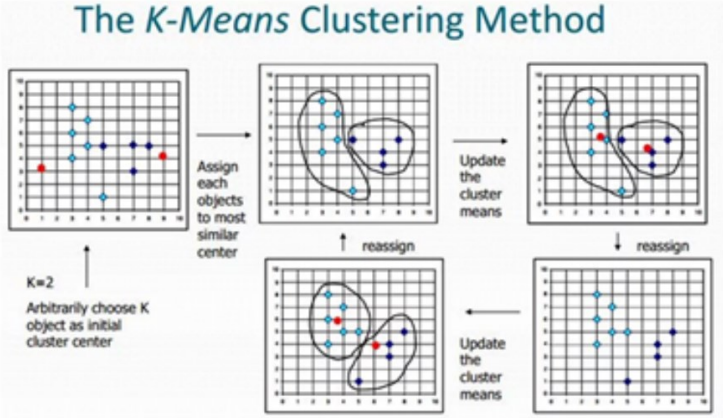
x : Merupakan titik dari data pertama.

y : Titik data kedua

n : Jumlah karakteristik (atribut) dalam terminologi data mining

d(*x*,*y*) : *Euclidean distance* yaitu jarak antara data pada titik *x* dan titik *y* menggunakan kalkulasi matematika

3. Setiap data dialokasikan ke *centroid* atau rata-rata yang terdekat. Dalam proses ini, ditentukan jarak terpendek, di mana data yang memiliki jarak paling dekat akan dimasukkan ke dalam *cluster* tersebut.
4. Melakukan perhitungan nilai rata-rata untuk menentukan *centroid* baru dari data yang berada dalam *cluster* yang sama. Selanjutnya, iterasi dilakukan dengan menggunakan pusat *cluster* yang baru terbentuk, jika hasil yang diperoleh belum menunjukkan konvergensi.



Gambar 2. 1 Metode *Clustering K-Means* (Ginantra et al., 2021)

5. Kembali ke Step 3, apabila masih terdapat data yang berpindah *cluster* atau terjadi perubahan pada nilai *centroid*, terutama jika terdapat perubahan pada nilai diatas

threshold yang ditentukan atau perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang telah ditentukan.

2.2.5 *Silhouette Coefficient*

Silhouette coefficient adalah suatu metode evaluasi yang digunakan untuk mengukur kualitas pengelompokan dalam analisis data, terutama dalam teknik *clustering* seperti *k-means*. Metode ini menggabungkan dua aspek utama yaitu, kohesi dan separasi. Kohesi diukur dengan menghitung jarak rata-rata antara objek dalam satu *cluster*, sedangkan separasi diukur dengan menghitung jarak rata-rata objek dalam *cluster* tersebut dengan objek pada *cluster* terdekat (Paembonan & Abduh, 2021). Nilai *silhouette coefficient* berkisar antara -1 hingga +1, di mana nilai yang mendekati +1 menunjukkan bahwa objek tersebut berada pada *cluster* yang tepat, sedangkan nilai mendekati -1 menunjukkan bahwa objek tersebut mungkin salah penempatan dalam *cluster* (Hasan, 2024).

Menurut Rousseeuw (1987) dalam Bagirov et al., (2023) nilai *silhouette coefficient* dapat digunakan untuk menilai kualitas hasil *clustering* baik untuk setiap *cluster* maupun keseluruhan *cluster*. Nilai *silhouette* untuk seluruh data dengan jumlah *cluster* $\boldsymbol{k}$ didefinisikan sebagai $s(i)$, yang dihitung dengan menggunakan persamaan 2.2 berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \qquad (2.\ 2)$$

Dimana:

$s(i)$: Merupakan nilai *silhouette coefficient* untuk titik data i.

$a(i)$: Rata-rata jarak dari objek i ke semua objek lain dalam *cluster* yang sama (*intra-cluster distance*).

$b(i)$: Rata-rata jarak dari objek i ke semua objek pada *cluster* terdekat (*inter-cluster distance*).

BAB III
METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan pendekatan *data mining* yaitu, serangkaian proses untuk menambah serta mencari informasi baru dari basis data, informasi tersebut sebelumnya tidak diketahui secara manual. Proses ini menghasilkan informasi dengan mengekstraksi dan mengenali pola-pola penting atau menarik dalam data yang tersimpan (Ani et al., 2021). Penelitian ini menggunakan algoritma *k-means clustering* untuk mengelompokkan buku berdasarkan *cluster* yang sudah ditentukan.

Data riwayat peminjaman diakuisisi dari *database* sistem informasi *SLiMS* Perpustakaan MAN 2 Kota Tangerang berbentuk csv tahun 2023 sampai 2024 terhitung mulai 1 Januari 2023 sampai 1 Januari 2024 dengan jumlah 768 data. Dataset, yang awalnya memiliki atribut id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman, disederhanakan melalui *preprocessing* data dan menghasilkan atribut judul buku, jumlah peminjam, jumlah anggota yang meminjam, dan jumlah eksemplar yang sesuai dengan variabel penelitian. Dari 768 data yang diperoleh, Peneliti mengambil 10% (77 data) sebagai data *testing* dan menggunakan 691 data untuk data *trening*. Peneliti juga menggunakan 15 data dari total data yang diperoleh untuk contoh perhitungan manual algoritma *k-means clustering*.

3.2 Subjek dan Objek Penelitian

Subjek penelitian adalah Perpustakaan MAN 2 Kota Tangerang. Sedangkan untuk objek penelitian ini adalah data riwayat transaksi peminjaman buku Perpustakaan MAN 2 Kota Tangerang dalam jangka waktu 1 tahun mulai dari 1 Januari 2023 sampai 1 Januari 2024.

3.3 Sumber Data

Penelitian ini menggunakan data primer dan sekunder. Data primer berupa data peminjaman buku pada Perpustakaan MAN 2 Kota Tangerang terhitung mulai 1 Januari 2023 sampai 1 Januari 2024. Sedangkan data sekunder yang digunakan adalah

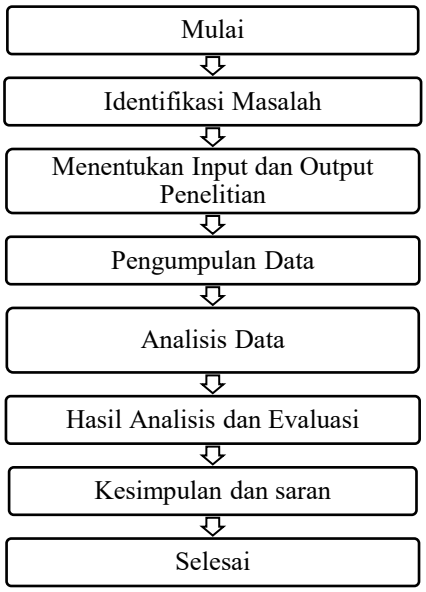
penelitian terdahulu terkait penggunaan *k-means clustering* dalam pengelompokan koleksi di perpustakaan.

3.4 Teknik Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan perizinan pengambilan data terkait peminjaman buku dari sistem otomasi SLiMS berbentuk csv kepada staf perpustakaan bagian sirkulasi Perpustakaan MAN 2 Kota Tangerang dalam jangka waktu 1 tahun, terhitung 1 Januari 2023 sampai 1 Januari 2024. Data yang diperoleh berjumlah 768 data yang terdiri dari beberapa atribut yaitu, id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman. Selain itu peneliti juga menggunakan studi literatur sebagai teknik pengumpulan data dari berbagai sumber ilmiah yang relevan seperti buku dan jurnal ilmiah untuk mendapatkan informasi terkait topik dan permasalahan yang akan diteliti.

3.5 Alur Penelitian

Alur penelitian merupakan rangkaian tahapan penelitian yang tersusun secara sistematis agar pelaksanaan penelitian mendapatkan hasil sesuai dengan tujuan penelitian. Adapun tahapan yang akan dilakukan dalam penelitian ini terlihat pada gambar 3.1 berikut:



Gambar 3. 1 Alur Penelitian

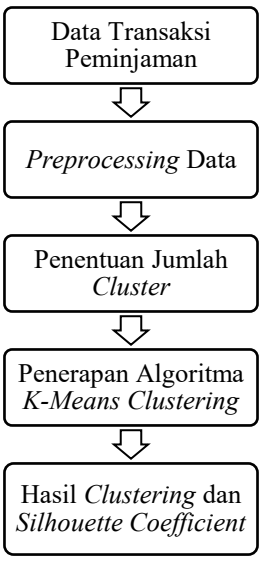
Pada gambar 3.1 terdapat delapan tahapan penelitian yang dilakukan diantaranya:

1. Tahap pertama adalah mengidentifikasi permasalahan yang ada yaitu mengetahui berapa hasil kinerja pengelompokan buku berdasarkan tingkat keterpakaian menggunakan algoritma *k-means clustering* pada Perpustakaan MAN 2 Kota Tangerang.
2. Tahap kedua adalah menentukan input dan output penelitian. Input pada penelitian yang diperlukan adalah data transaksi peminjaman buku Perpustakaan MAN 2 Kota Tangerang meliputi judul buku, jumlah peminjaman, jumlah anggota yang meminjam dan jumlah eksemplar. Sedangkan output penelitian merupakan hasil dari input yang sudah diproses melalui *clustering*. Dari proses tersebut menghasilkan pengelompokkan buku beberapa kategori *cluster* yaitu, *cluster* pertama merupakan buku yang sangat diminati, dan *cluster* kedua buku yang kurang diminati.
3. Tahap ketiga adalah pengumpulan data peminjaman koleksi dari sistem otomasi SLiMS berbentuk csv kepada staf perpustakaan bagian sirkulasi Perpustakaan MAN 2 Kota Tangerang dalam jangka waktu 1 tahun mulai dari 1 Januari 2023 sampai 1 Januari 2024. Data yang diperoleh berjumlah 768 data yang terdiri dari beberapa atribut yaitu, id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman.
4. Tahap keempat adalah melakukan analisis data peminjaman koleksi di Perpustakaan MAN 2 Kota Tangerang untuk mengidentifikasi pola peminjaman buku agar mengetahui hasil kinerja pengelompokan buku berdasarkan tingkat keterpakaian menggunakan algoritma *k-means clustering*. Pada penelitian ini *cluster* yang ditentukan dengan melihat *cluster* terbaik dari matrik *silhouette coefficient*, kemudian setelah itu menentukan titik awal *cluster* secara acak, mengalokasikan setiap data ke *cluster* terdekat berdasarkan jarak *euclidean*, dan menghitung ulang pusat *cluster* berdasarkan rata-rata data dalam setiap *cluster* supaya tidak berubah-ubah untuk mencapai jumlah iterasi yang ditentukan.

5. Tahapan kelima adalah evaluasi hasil analisis data menggunakan *silhouette coefficient* untuk memastikan buku terkelompok sesuai dengan *cluster*nya. Evaluasi hasil dilakukan bertujuan untuk mengetahui nilai *silhouette coefficient* yang berkisar antara -1 hingga +1. Jika nilai yang mendekati +1 akan menunjukkan bahwa objek tersebut berada pada *cluster* yang tepat, sedangkan nilai mendekati -1 menunjukkan bahwa objek tersebut mungkin salah penempatan dalam *cluster*.
6. Tahapan keenam adalah kesimpulan dan saran. Kesimpulan diambil berdasarkan hasil *clustering* dari penggunaan algoritma *k-means clustering* dan evaluasi hasil menggunakan *silhouette coefficient*. Sedangkan saran diberikan untuk pengadaan buku baru selanjutnya berdasarkan hasil yang sudah didapatkan.

3.6 Desain Sistem

Penelitian ini dilaksanakan dengan menggunakan desain sistem yang dijelaskan pada gambar 3.2 berikut:



Gambar 3. 2 Desain Sistem

Dalam gambar 3.2 terdapat lima tahapan desain sistem yang dilakukan, diantaranya:

1. Tahap pertama adalah menyiapkan data transaksi peminjaman koleksi yang diperoleh dari sistem otomasi SLiMS Perpustakaan MAN 2 Kota Tangerang

berbentuk csv dalam jangka waktu 1 tahun mulai dari 1 Januari 2023 sampai 1 Januari 2024. Data berjumlah 768 data yang terdiri dari beberapa atribut yaitu, id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman.

2. Tahap kedua adalah tahap *pre-proses*. Tahap ini bertujuan untuk mengubah data transaksi peminjaman yang awalnya tidak terstruktur menjadi data terstruktur serta menyederhanakan data koleksi menjadi kelas utama (Mahfud et al., 2020). Data tersebut yang awalnya memiliki atribut id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman. Kemudian disederhanakan menjadi judul buku, jumlah peminjam, jumlah anggota yang meminjam dan jumlah eksemplar.
3. Tahap ketiga adalah menentukan jumlah *cluster* **k** yang ingin dibuat untuk memberikan arahan kepada algoritma dalam mengelompokan data. *Cluster* **k** yang ditentukan pada penelitian ini berdasarkan hasil *cluster* terbaik dari matrik *silhouette coefficient*.
4. Tahap keempat adalah mengelompokan buku berdasarkan variabel yang digunakan untuk menghasilkan *centroid* di setiap *cluster* yang sudah dibuat dan menentukan jumlahnya pada proses sebelumnya. Pada dataset dibawah, terdapat 4 kolom variabel yaitu, judul buku, jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar.

Tabel 3. 1 Dataset Sampel Peminjaman Koleksi

Judul Buku	Jumlah Peminjam	Jumlah Anggota yang meminjam	Jumlah Eksemplar
Fikih 1 untuk kelas X Madrasah Aliyah	8	4	10
Fikih 2 untuk kelas X Madrasah Aliyah	5	3	10
Fikih 3 untuk kelas X Madrasah Aliyah	7	4	10
Akidah dan akhlak 1 untuk kelas X Madrasah Aliyah	6	3	10

Akidah dan akhlak 2 untuk kelas X Madrasah Aliyah	4	4	10
Al-Qur'an dan hadis 2 untuk kelas XI Madrasah Aliyah	9	4	10
Bahasa Arab 2 untuk kelas XI Madrasah Aliyah	8	4	10
Bahasa Arab 3 untuk kelas XII Madrasah Aliyah	5	2	10
Al-Qur'an dan hadis 3 untuk kelas XI Madrasah Aliyah	7	5	10
Paris a love journey : kisah pencarian cinta, persahabatan, dan kedamaian jiwa	2	2	5
Paris a miracle of love : cinta, keajaiban, dan keteguhan hati	2	1	5
Biografi para ilmuwan muslim	8	6	10
Belajar sains melalui fenomena di sekitar kita	4	2	10
Raisha dan Trisula Raja	3	3	5
The ultrasmart holiday : ketika liburan mengubah masa depan	3	3	5
Gejolak dalam awan	1	1	5
Laut biru Klara	1	1	5
Hidayah dalam cinta : sebuah novel penggugah nurani	3	1	9
50 rules to be wonderful teenager	6	4	10
125 ilmuwan muslim pengukir sejarah	5	3	10

5. Tahap kelima hasil *clustering* dan *silhouette coefficient*. Hasil *clustering* berupa pengelompokkan buku yang terdiri dari beberapa ketegori cluster yaitu, *cluster* pertama merupakan buku yang sangat diminati, dan *cluster* kedua buku yang

kurang diminati. Selanjutnya melakukan evaluasi hasil *clustering* untuk memastikan buku terkelompok sesuai dengan *cluster*. Evaluasi hasil menggunakan metrik *silhouette coefficient* dengan persamaan 3.1 berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.1)$$

Persamaan tersebut digunakan untuk mengetahui nilai *silhouette coefficient* yang berkisar antara -1 hingga +1. Jika nilai yang mendekati +1 akan menunjukkan bahwa objek tersebut berada pada *cluster* yang tepat, sedangkan nilai mendekati -1 menunjukkan bahwa objek tersebut mungkin salah penempatan dalam *cluster* (Hasan, 2024).

3.7 Analisis Data

Pada proses ini akan dilakukan analisis pengelompokkan data peminjaman koleksi yang diakses melalui database sistem otomasi, menggunakan metode *k-means clustering*. Dari data yang diperoleh sebanyak 768, diambil 15 data untuk dijadikan contoh perhitungan manual penerapan algoritma *k-means clustering*. Jumlah *cluster* pada penelitian menggunakan contoh tiga *cluster* berdasarkan variabel yang telah ditentukan, dengan tujuan untuk menghasilkan *centroid* yang representatif maka $\boldsymbol{k}$=3. *Cluster* ditentukan agar memiliki kesamaan data yang tinggi dan perbedaan yang jelas dengan kelompok lainnya. Selain itu, pemilihan jumlah *cluster* mempertimbangkan evaluasi hasil *clustering* untuk mencapai hasil yang optimal (Hartigan & Wong, 1979). Analisis dilakukan dengan menggunakan perhitungan berikut:

Jumlah data : 15

Jumlah *cluster* : 3

Tabel 3. 2 Contoh Sampel Data

JB	JP	JAM	JE
JB 1	8	4	10
JB 2	5	3	10
JB 3	7	4	10
JB 4	6	3	10
JB 5	4	4	10
JB 6	9	4	10
JB 7	8	4	10
JB 8	5	2	10

JB	JP	JAM	JE
JB 9	7	5	10
JB 10	2	2	5
JB 11	2	1	5
JB 12	8	6	10
JB 13	4	2	10
JB 14	3	3	5
JB 15	3	3	5

Pada tabel 3.3 JB merupakan judul buku, JP (jumlah peminjaman), JAM (jumlah anggota yang meminjam), dan JE (jumlah eksemplar). Proses selanjutnya yaitu menentukan titik *centroid* awal. Penentuan dilakukan secara acak dengan menggunakan data pada tabel dataset. Titik *centroid* awal yang digunakan pada iterasi 1 yaitu data pada JB 12, JB 4 dan JB 11. Penentuan *centroid* awal secara acak dalam algoritma *k-means clustering* bertujuan untuk memastikan inisialisasi yang tidak bias, menghindari *overfitting*, dan memungkinkan eksplorasi variasi dalam hasil *clustering* (Primandana et al., 2019). Sehingga didapatkan titik *centroid* awal adalah:

Cluster 1 : 8.6.10

Cluster 2 : 6.3.10

Cluster 3 : 2.1.5

Untuk mengukur jarak antara data dengan pusat *cluster* menggunakan rumus *euclidean distance* sebagai berikut:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3.\ 2)$$

Dimana:

x : Pusat *cluster*

y : Data

Perhitungan iterasi 1 semua data menggunakan rumus *euclidean distance* ketitik *centroid* pertama.

$$data\ 1{,}1 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_1)^2 + (x_3 - y_1)^2}$$
$$c_1 jb_1 = \sqrt{8-8)^2 + (4-6)^2 + (10-10)^2} = 2$$
$$c_1 jb_2 = \sqrt{(5-8)^2 + (3-6)^2 + (10-10)^2} = 4{,}242641$$
$$c_1 jb_3 = \sqrt{(7-8)^2 + (4-6)^2 + (10-10)^2} = 2{,}236068$$

$$c_1\,jb_4 = \sqrt{(3-8)^2 + (6-6)^2 + (10-10)^2 = 3{,}605551}$$
$$c_1\,jb_5 = \sqrt{(4-8)^2 + (4-6)^2 + (10-10)^2 = 4{,}472136}$$
$$c_1\,jb_6 = \sqrt{(9-8)^2 + (4-6)^2 + (10-10)^2 = 2{,}236068}$$
$$c_1\,jb_7 = \sqrt{(8-8)^2 + (4-6)^2 + (10-10)^2 = 2}$$
$$c_1\,jb_8 = \sqrt{(5-8)^2 + (2-6)^2 + (10-10)^2 = 5}$$
$$c_1\,jb_9 = \sqrt{(7-8)^2 + (5-6)^2 + (10-10)^2 = 1{,}414214}$$
$$c_1\,jb_{10} = \sqrt{(2-8)^2 + (2-6)^2 + (5-10)^2 = 8{,}774964}$$
$$c_1\,jb_{11} = \sqrt{(2-8)^2 + (1-6)^2 + (5-10)^2 = 9{,}273618}$$
$$c_1\,jb_{12} = \sqrt{(8-8)^2 + (6-6)^2 + (10-10)^2 = 0}$$
$$c_1\,jb_{13} = \sqrt{(4-8)^2 + (2-6)^2 + (10-10)^2 = 5{,}656854}$$
$$c_1\,jb_{14} = \sqrt{(3-8)^2 + (3-6)^2 + (5-10)^2 = 7{,}681146}$$
$$c_1\,jb_{15} = \sqrt{(3-8)^2 + (3-6)^2 + (5-10)^2 = 7{,}681146}$$

Perhitungan iterasi 1 semua data menggunakan rumus *euclidean distance* ketitik *centroid* kedua:

$$data\ 1{,}2 = \sqrt{(x_1 - y_2)^2 + (x_2 - y_2)^2 + (x_3 - y_2)^2}$$
$$c_2\,jb_1 = \sqrt{(8-6)^2 + (4-3)^2 + (10-10)^2 = 2{,}236068}$$
$$c_2\,jb_2 = \sqrt{(5-6)^2 + (3-3)^2 + (10-10)^2 = 1}$$
$$c_2\,jb_3 = \sqrt{(7-6)^2 + (4-3)^2 + (10-10)^2 = 1{,}414214}$$
$$c_2\,jb_4 = \sqrt{(6-6)^2 + (3-3)^2 + (10-10)^2 = 0}$$
$$c_2\,jb_5 = \sqrt{(4-6)^2 + (4-3)^2 + (10-10)^2 = 2{,}236068}$$
$$c_2\,jb_6 = \sqrt{(9-6)^2 + (4-3)^2 + (10-10)^2 = 3{,}162278}$$
$$c_2\,jb_7 = \sqrt{(8-6)^2 + (4-3)^2 + (10-10)^2 = 2{,}236068}$$
$$c_2\,jb_8 = \sqrt{(5-6)^2 + (2-3)^2 + (10-10)^2 = 1{,}414214}$$
$$c_2\,jb_9 = \sqrt{(7-6)^2 + (5-3)^2 + (10-10)^2 = 1{,}414214}$$
$$c_2\,jb_{10} = \sqrt{(2-6)^2 + (2-3)^2 + (5-10)^2 = 6{,}480741}$$
$$c_2\,jb_{11} = \sqrt{(2-6)^2 + (1-3)^2 + (5-10)^2 = 6{,}708204}$$
$$c_2\,jb_{12} = \sqrt{(8-6)^2 + (6-3)^2 + (10-10)^2 = 3{,}605551}$$
$$c_2\,jb_{13} = \sqrt{(4-6)^2 + (2-3)^2 + (10-10)^2 = 2{,}236068}$$
$$c_2\,jb_{14} = \sqrt{(3-6)^2 + (3-3)^2 + (5-10)^2 = 5{,}830952}$$
$$c_2\,jb_{15} = \sqrt{(3-6)^2 + (3-3)^2 + (5-10)^2 = 5{,}830952}$$

Perhitungan iterasi 1 semua data menggunakan rumus *euclidean distance* ketitik *centroid* ketiga:

$$data\ 1{,}3 = \sqrt{(x_1 - y_3)^2 + (x_2 - y_3)^2 + (x_3 - y_3)^2}$$

$$c_2\,jb_1 = \sqrt{(8-2+(4-1)^2 + (10-5)^2 = 8{,}36660027}$$
$$c_2\,jb_2 = \sqrt{(5-2)^2 + (3-1)^2 + (10-5)^2 = 6{,}164414}$$
$$c_2\,jb_3 = \sqrt{(7-2)^2 + (4-1)^2 + (10-5)^2 = 7{,}68114575}$$
$$c_2\,jb_4 = \sqrt{(6-2)^2 + (3-1)^2 + (10-5)^2 = 6{,}70820393}$$

$$c_2\,jb_5 = \sqrt{(4-2+(4-1)^2+(10-5)^2} = 6{,}164414$$
$$c_2\,jb_6 = \sqrt{(9-2)^2+(4-1)^2+(10-5)^2} = 9{,}11043358$$
$$c_2\,jb_7 = \sqrt{(8-2)^2+(4-1)^2+(10-5)^2} = 8{,}36660027$$
$$c_2\,jb_8 = \sqrt{(5-2)^2+(2-1)^2+(10-5)^2} = 5{,}91607978$$
$$c_2\,jb_9 = \sqrt{(7-2)^2+(5-1)^2+(10-5)^2} = 8{,}1240384$$
$$c_2\,jb_{10} = \sqrt{(2-2)^2+(2-1)^2+(5-5)^2} = 1$$
$$c_2\,jb_{11} = \sqrt{(2-2)^2+(1-1)^2+(5-5)^2} = 0$$
$$c_2\,jb_{12} = \sqrt{(8-2)^2+(6-1)^2+(10-5)^2} = 9{,}2736185$$
$$c_2\,jb_{13} = \sqrt{(4-2)^2+(2-1)^2+(10-5)^2} = 5{,}47722558$$
$$c_2\,jb_{14} = \sqrt{(3-2)^2+(3-1)^2+(5-5)^2} = 2{,}23606798$$
$$c_2\,jb_{15} = \sqrt{(3-2)^2+(3-1)^2+(5-5)^2} = 2{,}23606798$$

Dari hasil perhitungan *euclidean distance* yang telah dilakukan. Hasil yang didapatkan sebagai berikut:

Tabel 3. 3 Hasil Iterasi I

C1	C2	C3
2	2,236068	8,36660027
4,242641	1	6,164414
2,236068	1,414214	7,68114575
3,605551	0	6,70820393
4,472136	2,236068	6,164414
2,236068	3,162278	9,11043358
2	2,236068	8,36660027
5	1,414214	5,91607978
1,414214	2,236068	8,1240384
8,774964	6,480741	1
9,273618	6,708204	0
0	3,605551	9,2736185
5,656854	2,236068	5,47722558
7,681146	5,830952	2,23606798
7,681146	5,830952	2,23606798

Tabel 3.3 Hasil dari perhitungan iterasi 1 dengan cara menghitung jarak minimum objek terhadap *centroid*. Sehingga didapatkan anggota *cluster* pada iterasi 1.

Tabel 3. 4 Anggota *Cluster* Iterasi I

JB	C1	C2	C3
JB 1	1		
JB 2		1	
JB 3		1	
JB 4		1	
JB 5		1	
JB 6	1		

JB	C1	C2	C3
JB 7	1		
JB 8		1	
JB 9	1		
JB 10			1
JB 11			1
JB 12	1		
JB 13		1	
JB 14			1
JB 15			1

Tabel 3.4. Merupakan anggota *cluster* pada iterasi 1 yang didapatkan dari hasil perhitungan menggunakan rumus *euclidean distance*. Selanjutnya menentukan centroid baru (*Ck*) untuk iterasi 2 dengan cara menghitung rata-rata dari data *centeroid* yang sama sebelumnya, menggunakan persamaan 3.3 berikut:

$$C_k = \left(\frac{1}{n_k}\right)\sum d_1 \tag{3.3}$$

Dimana *nk* adalah jumlah data dalam *cluster* ***k*** dan *d*(*1*) adalah data dalam *cluster* ***k***. Perhitungan *centroid* baru untuk iterasi 2:

Cluster 1

$$C_{k(1),JP} = \left(\frac{40}{5}\right) = 8$$

$$C_{k(1),JAM} = \left(\frac{23}{5}\right) = 4{,}6$$

$$C_{k(1),JE} = \left(\frac{50}{5}\right) = 10$$

Cluster 2

$$C_{k(2),JP} = \left(\frac{31}{6}\right) = 5{,}16$$

$$C_{k(2),JAM} = \left(\frac{18}{6}\right) = 3$$

$$C_{k(2),JE} = \left(\frac{60}{6}\right) = 10$$

Cluster 3

$$C_{k(3),JP} = \left(\frac{10}{4}\right) = 2{,}5$$

$$C_{k(3),JAM} = \left(\frac{9}{4}\right) = 2{,}25$$

$$C_{k(3),JE} = \left(\frac{20}{4}\right) = 5$$

Sehingga didapatkan *centroid* baru iterasi 2 yaitu:

C1 : 8. 4,6. 10

C2 : 5,166666667. 3. 10

C3 : 2,5. 2,25. 5

Hitung kembali jarak objek terhadap *centroid* baru seperti yang telah dilakukan pada tahap iterasi 1 menggunakan persamaan 3.2. Setelah hasil perhitungan, kemudian melakukan perbandingan jika hasil posisi *cluster* pada iterasi ke 2 sama dengan posisi pada iterasi 1, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke 3.

Cluster 1

$$data\ 2{,}1 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_1)^2 + (x_3 - y_1)^2}$$

$$c_1\, jb_1 = \sqrt{8-8)^2 + (4-4{,}6)^2 + (10-10)^2 = 0{,}6}$$

$$c_1 jb_{15} = \sqrt{(3-8)^2 + (3-4{,}6)^2 + (5-10)^2 = 7{,}681146}$$

Cluster 2

$$data\ 2{,}2 = \sqrt{(x_1 - y_2)^2 + (x_2 - y_2)^2 + (x_3 - y_2)^2}$$

$$c_2\, jb_1 = \sqrt{(8-5{,}1)^2 + (4-3)^2 + (10-10)^2 = 3{,}004626063}$$

$$c_2\, jb_{15} = \sqrt{(3-5{,}1)^2 + (3-3)^2 + (5-10)^2 = 7{,}249827584}$$

Cluster 3

$$data\ 2{,}3 = \sqrt{(x_1 - y_3)^2 + (x_2 - y_3)^2 + (x_3 - y_3)^2}$$

$$c_3\, jb_1 = \sqrt{(8-2{,}5 + (4-2{,}25)^2 + (10-5)^2 = 7{,}636262175}$$

$$c_3\, jb_{15} = \sqrt{(3-2,5)^2 + (3-2{,}25)^2 + (5-5)^2 = 0{,}901387819}$$

Dari perhitungan iterasi 2 didapatkan hasil:

Tabel 3. 5 Hasil Iterasi 2

C1	C2	C3
0,6	3,004626	7,636262
1,16619	2,088327	6,950719
1,16619	3,961621	8,385255
0,6	3,004626	7,636262
1,077033	2,713137	7,267221
1,4	4,126473	8,325413
3,4	0,166667	5,640257
2,56125	0,833333	6,149187
4,04475	1,536591	5,505679
3,969887	1,013794	5,595757
4,770744	1,536591	5,226136
8,231646	6,002314	0,559017
8,6	6,247222	1,346291
7,249828	5,449261	0,901388
7,249828	5,449261	0,901388

Tabel 3.5. Hasil dari perhitungan iterasi 2 dengan cara menghitung jarak minimum objek terhadap *centroid*. Sehingga didapatkan anggota *cluster* pada iterasi 2.

Tabel 3. 6 Anggota Cluster Iterasi 2

JB	C1	C2	C3
JB 1	1		
JB 2	1		
JB 3		1	
JB 4		1	
JB 5		1	
JB 6	1		
JB 7	1		
JB 8		1	
JB 9	1		
JB 10			1
JB 11			1
JB 12	1		
JB 13		1	
JB 14			1
JB 15			1

Pada tabel 3.6 terlihat perubahan *cluster* pada JB 2, yang sebelumnya menempati *cluster* 2. Sehingga perlu melakukan perhitungan iterasi kembali dengan

menghitung terlebih dahulu titik *centroid* baru seperti perhitungan iterasi 1 dan 2 menggunakan persamaan 3.4 berikut:

$$C_k = \left(\frac{1}{n_k}\right)\sum d_1 \tag{3.4}$$

Perhitungan *centroid* baru untuk iterasi 3.

Cluster 1

$$C_{k(1),JP} = \left(\frac{47}{6}\right) = 7{,}83$$

$$C_{k(1),JAM} = \left(\frac{27}{6}\right) = 4{,}5$$

$$C_{k(1),JE} = \left(\frac{60}{6}\right) = 10$$

Cluster 2

$$C_{k(2),JP} = \left(\frac{24}{5}\right) = 4{,}8$$

$$C_{k(2),JAM} = \left(\frac{14}{5}\right) = 2{,}8$$

$$C_{k(2),JE} = \left(\frac{50}{5}\right) = 10$$

Cluster 3

$$C_{k(3),JP} = \left(\frac{10}{4}\right) = 2{,}5$$

$$C_{k(3),JAM} = \left(\frac{9}{4}\right) = 2{,}25$$

$$C_{k(3),JE} = \left(\frac{20}{4}\right) = 5$$

Sehingga didapatkan *centroid* baru untuk iterasi 3 yaitu:

C1 : 7,833333. 4,5. 10

C2 : 4,8. 2,8. 10

C3 : 2,5. 2,25. 5

Hitung kembali jarak objek terhadap *centroid* baru seperti yang telah dilakukan pada tahap iterasi 2 menggunakan persamaan 3.2. Setelah hasil perhitungan, kemudian

melakukan perbandingan jika hasil posisi *cluster* pada iterasi ke 2 sama dengan posisi pada iterasi 3, maka proses dihentikan, namun jika tidak maka proses dilanjutkan ke iterasi ke 4.

Cluster 1

$$data\ 3{,}1 = \sqrt{(x_1 - y_1)^2 + (x_2 - y_1)^2 + (x_3 - y_1)^2}$$

$$c_1\, jb_1 = \sqrt{8 - 7{,}83)^2 + (4 - 4{,}5)^2 + (10 - 10)^2 = 0{,}527046277}$$

$$c_1 jb_{15} = \sqrt{(3 - 7{,}83)^2 + (3 - 4{,}5)^2 + (5 - 10)^2 = 7{,}114148657}$$

Cluster 2

$$data\ 3{,}2 = \sqrt{(x_1 - y_2)^2 + (x_2 - y_2)^2 + (x_3 - y_2)^2}$$

$$c_2\, jb_1 = \sqrt{(8 - 4{,}8)^2 + (4 - 2{,}8)^2 + (10 - 10)^2 = 3{,}417601498}$$

$$c_2\, jb_{15} = \sqrt{(3 - 4{,}8)^2 + (3 - 2{,}8)^2 + (5 - 10)^2 = 5{,}31789432}$$

Cluster 3

$$data\ 3{,}3 = \sqrt{(x_1 - y_3)^2 + (x_2 - y_3)^2 + (x_3 - y_3)^2}$$

$$c_3\, jb_1 = \sqrt{(8 - 2{,}5 + (4 - 2{,}25)^2 + (10 - 5)^2 = 7{,}636262175}$$

$$c_3\, jb_{15} = \sqrt{(3 - 2,5)^2 + (3 - 2{,}25)^2 + (5 - 5)^2 = 0{,}901387819}$$

Dari perhitungan iterasi 3 didapatkan hasil:

Tabel 3. 7 Hasil Iterasi 3

C1	C2	C3
0,527046	3,417601	7,636262
0,971825	2,505993	6,950719
1,269296	4,368066	8,385255
0,527046	3,417601	7,636262
0,971825	3,11127	7,267221
1,509231	4,525483	8,325413
3,205897	0,282843	5,640257
2,368778	1,216553	6,149187
3,865805	1,442221	5,505679
3,778595	0,824621	5,595757
4,57651	1,131371	5,226136
8,079466	5,78619	0,559017

C1	C2	C3
8,442617	6,006663	1,346291
7,114149	5,317894	0,901388
7,114149	5,317894	0,901388

Tabel 3.7. Hasil dari perhitungan iterasi 3 dengan cara menghitung jarak minimum objek terhadap *centroid*. Sehingga didapatkan anggota *cluster* pada iterasi 3.

Tabel 3. 8 Anggota *Cluster* Iterasi 3

JB	C1	C2	C3
JB 1	1		
JB 2	1		
JB 3		1	
JB 4		1	
JB 5		1	
JB 6	1		
JB 7	1		
JB 8		1	
JB 9	1		
JB 10			1
JB 11			1
JB 12	1		
JB 13		1	
JB 14			1
JB 15			1

Pada tabel 3.8 terlihat tidak ada perubahan objek *cluster*, sehingga iterasi dianggap telah selesai. Hasil dari perhitungan adalah sebagai berikut.

Cluster 1 : Koleksi dengan kode JB1, JB2, JB6, JB7, JB9, JB12 merupakan buku yang sedikit diminati.

Cluster 2 : Koleksi dengan kode JB3, JB4, JB5, JB8, JB13 merupakan buku yang sangat diminati.

Cluster 3 : Koleksi dengan kode JB10, JB11, JB14, JB15 merupakan koleksi yang cukupt diminati.

Dari hasil yang telah didapatkan melalui proses *clustering* menggunakan algoritma *k-means* maka dapat disimpulkan bahwa koleksi pada *cluster* 2 dan 3 perlu ditambahkan, sedangkan buku pada *cluster* 1 harus dikelola kembali apakah ingin dipertahankan atau dikurangi.

Tahap berikutnya melakukan evaluasi hasil *clustering* untuk memastikan setiap buku terkelompok sesuai dengan *cluster*nya. Adapun tahapan perhitungan menggunakan *silhouette coefficient* adalah sebagai berikut (Handoyo et al., 2014).

1. Menghitung rata-rata jarak dari suatu data, menggunakan persamaan 3.5. Maka didapatkan rata-rata dengan cara memisahkan i terhadap semua data lain yang berbeda dalam satu *cluster* sebagai berikut.

$$a(i) = \frac{1}{|c_i| - 1} \sum_{j \in A, j \neq i} d(i,j) \quad (3.5)$$

   Dimana:

   $a(i)$: Merupakan perbedaan rata-rata objek (i) ke semua objek lain pada Ci.

   $d(i,j)$: Jarak antara i dengan j.

   Ci : Merupakan jumlah data.

2. Mengitung rata-rata jarak data i dengan semua data di *cluster* lain dan mengambil nilai terkecilnya menggunakan persamaan 3.6 berikut:

$$d(i, C) = \frac{1}{|C| - 1} \sum_{j \in C} d(i,j) \quad (3.6)$$

   Dimana:

   $d(i,c)$: Perbedaan rata-rata objek (i) ke semua objek lain pada objek C

   C : *Cluster* lain selain A atau cluster C yang tidak sama dengan *cluster A*.

3. Setelah menghitung $d(i,C)$ untuk semua *cluster C*, maka selanjutnya mengambil nilai minimum $b(i)$ dengan menggunakan persamaan 3.7 berikut:

$$b(i) = min_{C \neq A} d(i, C) \quad (3.7)$$

   Dimana *cluster B* yang sudah mencapai minimum (yaitu, $d(i,B) = b(i)$ sering juga disebut data lain dari objek (i). Objek tersebut merupakan *cluster* terbaik untuk kedua objek (i).

4. Nilai evaluasi dilakukan dengan menggunakan metrik *silhouette coefficient* seperti persamaan 3.8 berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.8)$$

Dimana:

(i) : Merupakan nilai *silhouette coefficient*.

$a(i)$: Rata-rata jarak dari objek i ke semua objek lain dalam *cluster* yang sama (*intra-cluster distance*)

$b(i)$: Rata-raja jarak dari objek i ke semua objek terdekat pada *cluster* (*inter-cluster distance*)

5. Setelah menghitung nilai *silhouette coefficient*, selanjutnya menentukan nilai $s(i)$ pada seluruh dataset dengan menggunakan persamaan 3.9 berikut.

$$SC = \max s\,(k) \quad (3.9)$$

Dimana $s(k)$ merepresentasikan mean $s(i)$ seluruh data pada keseluruhan dataset untuk sejumlah *cluster* tertentu.

Tahap pertama untuk melakukan evaluasi *cluster* menggunakan *silhouette coefficient* yaitu mencari nilai $a(i)$ pada iterasi 1 menggunakan persamaan 3.5, yang dilakukan untuk mencari rata-rata jarak data pada *cluster* 1, 2, dan 3 sebagai berikut:

Cluster 1

$$a(i) = \frac{1}{4}\sqrt{(9-8)^2 + (4-4)^2 + (10-10)^2 + (8-8)^2 + (4-4)^2 + (10-10)^2 + \sqrt{(7-8)^2 + (5-4)^2 + (10-10)^2 + (8-8)^2 + (6-4)^2 + (10-10)^2}}$$

$$= 1$$

Pada *cluster* 1 terdapat lima data yang akan dihitung rata-rata jaraknya dengan data lain dalam satu *cluster*. Sehingga didapatkan jumlah keseluruhan rata-ratanya pada tabel 3.9 berikut.

Tabel 3. 9 Hasil Menghitung *a*(*i*) Pada *Cluster* 1 Iterasi 1

JB	JP	JAM	JE	a(*i*)
JB 1	8	4	10	1
JB 6	9	4	10	2
JB 7	8	4	10	1
JB 9	7	5	10	2
JB 12	8	6	10	2

Tabel 3.9 adalah hasil seluruh perhitungan jarak rata-rata data lain dalam satu *cluster* dengan menggunakan persamaan 3.5. Selanjutnya melakukan perhitungan pada *cluster* 2 dan 3 iterasi 1.

Cluster 2

$$a(i) = \frac{2}{5}\sqrt{(7-5)^2 + (4-3)^2 + (10-10)^2 + (6-5)^2 + (3-3)^2 + (10-10)^2 + \sqrt{(4-5)^2 + (4-3)^2 + (10-10)^2 + (5-5)^2 + (2-3)^2 + (10-10)^2 + \sqrt{(4-5)^2 + 2 - 3)^2 + (10-10)^2}}}$$
$$= 1{,}41289902$$

Cluster 3

$$a(i) = \frac{3}{3}\sqrt{(2-2)^2 + (1-2)^2 + (5-5)^2 + (3-2)^2 + (3-2)^2 + (5-5)^2 + \sqrt{(3-2)^2 + (3-2)^2 + (5-5)^2}}$$
$$= 1{,}276142375$$

Pada *cluster* 2 terdapat enam data dan cluster 3 empat data yang akan yang akan dihitung rata-rata jaraknya dengan data lain dalam satu *cluster*. Sehingga didapatkan jumlah keseluruhan rata-ratanya pada tabel 3.10 berikut.

Tabel 3. 10 Hasil Menghitung *a*(*i*) Pada *Cluster* 2 dan 3 Iterasi 1

JB	JP	JAM	JE	a(*i*)
JB 2	5	3	10	1,41289902
JB 3	7	4	10	2,616851988
JB 4	6	3	10	1,660112616
JB 5	4	4	10	2,177269903
JB 8	5	2	10	1,695741733
JB 13	4	2	10	2,051166563
JB 10	2	2	5	1,276142375
JB 11	2	1	5	1,824045318
JB 14	3	3	5	1,216760513
JB 15	3	3	5	1,216760513

Tabel 3.10 adalah hasil seluruh perhitungan jarak rata-rata data lain dalam satu *cluster* pada *cluster* 2 dan 3 dengan menggunakan persamaan 3.5. Tahap kedua melakukan perhitungan rata-rata jarak data *i* dengan semua data di *cluster* lain pada iterasi 1 yaitu *cluster* 1, 2 dan 3 serta mengambil nilai terkecilnya menggunakan persamaan 3.6.

Cluster 1

$$d(i) = \frac{2}{5}\sqrt{(8-5)^2 + (4-3)^2 + (10-10)^2 + (9-5)^2 + (4-3)^2 + (10-10)^2 + \sqrt{(8-5)^2 + (4-3)^2 + (10-10)^2 + (7-5)^2 + (5-5)^2 + (10-10)^2 + \sqrt{(8-5)^2 + (6-3)^2 + (10-10)^2}}}$$
$$= 3{,}503745752$$

$$d(i) = \frac{3}{5}\sqrt{(8-2)^2 + (4-2)^2 + (10-5)^2 + (9-2)^2 + (4-2)^2 + (10-5)^2 + \sqrt{(8-2)^2 + (4-2)^2 + (10-5)^2 + (7-2)^2 + (5-2)^2 + (10-5)^2 + \sqrt{(8-2)^2 + (6-2)^2 + (10-5)^2}}}$$
$$= 8{,}2824773$$

Tabel 3. 11 Hasil Perhitungan *d(i,1)* Pada *Cluster* 1 Iterasi 1

JB	JP	JAM	JE	d(*i*,1)
JB 1	8	4	10	-
JB 6	9	4	10	-
JB 7	8	4	10	-
JB 9	7	5	10	-
JB 12	8	6	10	-
JB 2	5	3	10	3,503745752
JB 3	7	4	10	3,503745752
JB 4	6	3	10	4,037480598
JB 5	4	4	10	4,121345137
JB 8	5	2	10	4,037480598
JB 13	4	2	10	5,297649682
JB 10	2	2	5	8,2824773
JB 11	2	1	5	8,2824773
JB 14	3	3	5	8,2824773
JB 15	3	3	5	8,2824773

Tabel 3.11 merupakan hasil dari seluruh perhitungan *d*(*i*) pada *cluster* 1. *cluster* 1 tidak dihitung karena dijadikan objek untuk dibandingkan dengan data selain didalam *cluster*nya. Dengan kata lain, jarak dari titik data pada *cluster* pertama ke *centroid* atau titik lain di *cluster* 1 atau lainnya tidak dihitung sebagai bagian dari perbandingan untuk *b*(*i*) (Triandini et al., 2021). Selanjutnya menghitung *d*(*i*) pada *cluster* 2 dan 3 menggunakan persamaan 3.6 sebagai berikut.

Cluster 2

$$d(i) = \frac{1}{6}\sqrt{(5-8)^2 + (3-4)^2 + (10-10)^2 + (7-8)^2 + (4-4)^2 + (10-10)^2 + \sqrt{(6-8)^2 + (3-4)^2 + (10-10)^2 + (5-8)^2 + (2-4)^2 + (10-10)^2 + \sqrt{(4-8)^2 + (2-4)^2 + (10-10)^2}}}$$

$$= 3{,}099856416$$

$$d(i) = \frac{3}{6}\sqrt{(2-5)^2 + (2-3)^2 + (5-10)^2 + (2-5)^2 + (1-3)^2 + (5-10)^2 + \sqrt{(3-5)^2 + (3-3)^2 + (5-10)^2 + (3-5)^2 + (3-3)^2 + (5-10)^2}}$$

$$= 3{,}808470567$$

Cluster 3

$$d(i) = \frac{1}{4}\sqrt{(2-8)^2 + (2-4)^2 + (5-10)^2 + (2-8)^2 + (1-4)^2 + (5-10)^2 + \sqrt{(3-8)^2 + (3-4)^2 + (5-10)^2 + (3-8)^2 + (3-4)^2 + (5-10)^2 +}}$$

$$= 7{,}677928718$$

$$d(i) = \frac{3}{4}\sqrt{(2-5)^2 + (2-3)^2 + (5-10)^2 + (2-5)^2 + (1-3)^2 + (5-10)^2 + \sqrt{(3-5)^2 + (3-3)^2 + (5-10)^2 + (3-5)^2 + (3-3)^2 + (5-10)^2}}$$

$$= 5{,}71270585$$

Tabel 3. 12 Hasil Perhitungan *d*(*i*,2) dan *d*(*i*,3) Pada *Cluster* 2 dan 3 Iterasi 1

JB	JP	JAM	JE	d(*i*,2)	d(*i*,3)
JB 1	8	4	10	3,099856416	7,677928718
JB 6	9	4	10	3,079338811	7,677928718
JB 7	8	4	10	3,079338811	7,677928718
JB 9	7	5	10	3,079338811	7,677928718
JB 12	8	6	10	3,079338811	7,677928718
JB 2	5	3	10	-	5,71270585
JB 3	7	4	10	-	5,71270585
JB 4	6	3	10	-	5,71270585
JB 5	4	4	10	-	5,71270585
JB 8	5	2	10	-	5,71270585
JB 13	4	2	10	-	5,602992298
JB 10	2	2	5	3,808470567	-
JB 11	2	1	5	3,808470567	-
JB 14	3	3	5	3,808470567	-
JB 15	3	3	5	3,808470567	-

Tabel 3.12 hasil seluruh perhitungan seluruh perhitungan *d*(*i*) pada *cluster* 2 dan 3. Dimana *cluster* yang akan dijadikan objek tidak dihitung karena akan dibandingkan dengan data selain didalam clusternya. Dengan kata lain, jarak dari titik data dalam *cluster* yang dijadikan objek tidak dihitung sebagai bagian dari perbandingan untuk *b*(*i*). Tahap berikutnya yaitu menghitung nilai *b*(*i*) menggunakan persamaan 3.7 berikut.

Cluster 1

$$b(i) = \min(3{,}099856416 : 7{,}67792871) = 3{,}099856416$$

Cluster 2

$$b(i) = \min(3{,}503745752 : 5{,}71270585) = 3{,}503745752$$

Cluster 3

$$b(i) = \min(8{,}2824773 : 3{,}808470567) = 3{,}808470567$$

Hasil perhitungan $b(i)$ terlihat pada tabel 3.13 sebagai berikut:

Tabel 3. 13 Hasil Seluruh Perhitungan *b*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 1

JB	d(*i*,1)	d(*i*,2)	d(*i*,3)	b(*i*)
JB 1	-	3,099856416	7,677928718	3,099856416
JB 6	-	3,079338811	7,677928718	3,079338811
JB 7	-	3,079338811	7,677928718	3,079338811
JB 9	-	3,079338811	7,677928718	3,079338811
JB 12	-	3,079338811	7,677928718	3,079338811
JB 2	3,503745752	-	5,71270585	3,503745752
JB 3	3,503745752	-	5,71270585	3,503745752
JB 4	4,037480598	-	5,71270585	4,037480598
JB 5	4,121345137	-	5,71270585	4,121345137
JB 8	4,037480598	-	5,71270585	4,037480598
JB 13	5,297649682	-	5,602992298	5,297649682
JB 10	8,2824773	3,808470567	-	3,808470567
JB 11	8,2824773	3,808470567	-	3,808470567
JB 14	8,2824773	3,808470567	-	3,808470567
JB 15	8,2824773	3,808470567	-	3,808470567

Tabel 3.13. Hasil seluruh perhitungan $b(i)$ iterasi 1 pada *cluster* 1, 2, dan 3. Perhitungan dilakukan dengan mencari nilai minimum pada setiap *cluster* yang memiliki data, apabila dalam *cluster* tidak terdapat data maka tidak dihitung. Karena nilai $b(i)$ yang merupakan nilai minimal dari rata-rata jarak data ke-*i* antara semua data pada setiap *cluster* (Dewi & Pramita, 2019).

Tahap terakhir yaitu menghitung nilai $s(i)$ atau *silhouette cluster* 1, 2, dan 3 pada iterasi 1 menggunakan persamaan 3.8 dan 3.9 berikut.

Cluster 1

$$s\,(jb_1) = \frac{3{,}09 - 1}{\max\big(a(1), b(3{,}09)\big)} = 0{,}64$$

$$s\ (jb_6) = \frac{3{,}07 - 2}{\max\big(a(2), b(3{,}07)\big)} = 0{,}47$$

$$s\ (jb_7) = \frac{3{,}07 - 1}{\max\big(a(1), b(3{,}07)\big)} = 0{,}64$$

$$s\ (jb_9) = \frac{3{,}07 - 2}{\max\big(a(2), b(3{,}07)\big)} = 0{,}47$$

$$s\ (jb_{12}) = \frac{3{,}07 - 2}{\max\big(a(2), b(3{,}07)\big)} = 0{,}37$$

Cluster 2

$$s\ (jb_2) = \frac{3{,}50 - 1{,}41}{\max\big(a(1{,}41), b(3{,}50)\big)} = 0{,}59$$

$$s\ (jb_3) = \frac{3{,}50 - 2{,}61}{\max\big(a(2{,}61), b(3{,}50)\big)} = 0{,}25$$

$$s\ (jb_4) = \frac{4{,}03 - 1{,}66}{\max\big(a(1{,}66), b(4{,}03)\big)} = 0{,}58$$

$$s\ (jb_5) = \frac{4{,}12 - 2{,}17}{\max\big(a(2{,}17), b(4{,}12)\big)} = 0{,}47$$

$$s\ (jb_8) = \frac{4{,}03 - 1{,}69}{\max\big(a(1{,}69), b(4{,}03)\big)} = 0{,}58$$

$$s\ (jb_{13}) = \frac{5{,}29 - 2{,}05}{\max\big(a(2{,}05), b(5{,}29)\big)} = 0{,}61$$

Cluster 3

$$s\ (jb_{10}) = \frac{3{,}80 - 1{,}27}{\max\big(a(1{,}27), b(3{,}80)\big)} = 0{,}66$$

$$s\ (jb_{11}) = \frac{3{,}80 - 1{,}82}{\max\big(a(1{,}82), b(3{,}80)\big)} = 0{,}52$$

$$s\ (jb_{14}) = \frac{3{,}80 - 1{,}21}{\max\big(a(1{,}21), b(3{,}80)\big)} = 0{,}68$$

$$s\ (jb_{15}) = \frac{3{,}80 - 1{,}21}{\max\big(a(1{,}21), b(3{,}80)\big)} = 0{,}68$$

Selanjutnya menentukan nilai $s(i)$ pada seluruh dataset dengan menggunakan persamaan 3.9 berikut.

$$SC = max\ 0{,}643998546 : 0{,}680512034 = 0{,}550887945$$

Hasil dari perhitungan *silhouette coefficient* pada iterasi 1 dapat dilihat pada tabel 3.14 berikut.

Tabel 3. 14 Hasil Seluruh Perhitungan *s*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 1

JB	a(*i*)	b(*i*)	s(*i*)
JB 1	1	3,099856416	0,643998546
JB 6	2	3,079338811	0,47455149
JB 7	1	3,079338811	0,641626512
JB 9	2	3,079338811	0,474017877
JB 12	2	3,079338811	0,378902257
JB 2	1,41289902	3,503745752	0,596746134
JB 3	2,616851988	3,503745752	0,253127318
JB 4	1,660112616	4,037480598	0,588824621
JB 5	2,177269903	4,121345137	0,471708913
JB 8	1,695741733	4,037480598	0,580000029
JB 13	2,051166563	5,297649682	0,612815742
JB 10	1,276142375	3,808470567	0,664919985
JB 11	1,824045318	3,808470567	0,521055687
JB 14	1,216760513	3,808470567	0,680512034
JB 15	1,216760513	3,808470567	0,680512034
			SC(0,550887945)

Tabel 3.14. Hasil iterasi 1 didapatkan nilai *silhouette coefficient* yaitu 0,55, untuk mengetahui apakah nilai tersebut sudah mendekati +1 yang akan menunjukkan bahwa objek tersebut berada pada *cluster* yang tepat, sedangkan jika nilai mendekati -1 menunjukkan bahwa objek tersebut mungkin salah penempatan dalam *cluster*.

Berikutnya melakukan perhitungan kembali nilai *silhouette* untuk iterasi 2 pada *cluster* 1, 2, dan 3 menggunakan persamaan 3.5, 3.6, 3.7, 3.8, 3.9. Pertama yaitu mencari nilai $a(i)$ menggunakan persamaan 3.5, hasil dari seluruh perhitungan $a(i)$ pada iterasi 2 dapat dilihat pada tabel 3.15 berikut.

Cluster 1

$$a(i) = \frac{1}{5}\sqrt{(7-8)^2 + (4-4)^2 + (10-10)^2 + (9-8)^2 + (4-4)^2 + (10-10)^2 + \sqrt{(8-8)^2 + (4-4)^2 + (10-10)^2 + (7-8)^2 + (5-4)^2 + (10-10)^2} \sqrt{(8-8)^2 + (6-4)^2 + (10-10)^2}} = 1{,}082842712$$

Cluster 2

$$a(i) = \frac{2}{4}\sqrt{(6-5)^2 + (3-3)^2 + (10-10)^2 + (4-5)^2 + (3-3)^2 + (10-10)^2 + \sqrt{(5-5)^2 + (2-3)^2 + (10-10)^2 + (4-5)^2 + (2-3)^2 + (10-10)^2}} = 1{,}207106781$$

Cluster 3

$$a(i) = \frac{3}{3}\sqrt{(2-2)^2 + (1-5)^2 + (5-5)^2 + (3-2)^2 + (3-2)^2 + (5-5)^2 + \sqrt{(3-2)^2 + (3-2)^2 + (5-5)^2}} = 1{,}276142375$$

Hasil perhitungan $a(i)$ terlihat pada tabel 3.15 sebagai berikut:

Tabel 3. 15 Hasil Seluruh Perhitungan *a*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 2

JB	JP	JAM	JE	a(*i*)
JB 1	8	4	10	1,082843
JB 3	7	4	10	1,447214
JB 6	9	4	10	1,694427
JB 7	8	4	10	1,082843
JB 9	7	5	10	1,495742
JB 12	8	6	10	1,97727
JB 2	5	3	10	1,207107
JB 4	6	3	10	1,721587
JB 5	4	4	10	1,971587
JB 8	5	2	10	2,690671
JB 13	4	2	10	1,66257
JB 10	2	2	5	1,276142
JB 11	2	1	5	1,824045
JB 14	3	3	5	1,216761
JB 15	3	3	5	1,216761

Tabel 3.15 adalah hasil seluruh perhitungan jarak rata-rata data lain dalam satu *cluster* pada *cluster* 1, 2, dan 3 iterasi ke 2. Setelah melakukan perhitungan nilai $a(i)$ pada iterasi 2, selanjutnya melakukan perhitungan rata-rata jarak data *i* atau $d(i)$ dengan semua data di *cluster* lain pada iterasi 2 yaitu *cluster* 1, 2 dan 3 serta mengambil nilai terkecilnya menggunakan persamaan 3.6.

Cluster 1

$$d(i) = \frac{2}{6}\sqrt{(8-5)^2 + (4-3)^2 + (10-10)^2 + (9-5)^2 + (4-3)^2 + (10-10)^2 + \sqrt{(8-5)^2 + (4-3)^2 + (10-10)^2 + (7-5)^2 + (5-5)^2 + (10-10)^2 + \sqrt{(8-5)^2 + (6-3)^2 + (10-10)^2}} = 3{,}292466123$$

$$d(i) = \frac{3}{6}\sqrt{(2-2)^2 + (1-2)^2 + (5-5)^2 + (3-2)^2 + (3-2)^2 + (5-5)^2 + } = 8{,}126809288$$

Cluster 2

$$d(i) = \frac{1}{5}\sqrt{(5-8)^2 + (3-4)^2 + (10-10)^2 + (7-8)^2 + (4-4)^2 + (10-10)^2 + \sqrt{(6-8)^2 + (3-4)^2 + (10-10)^2 + (5-8)^2 + (2-4)^2 + (10-10)^2 + \sqrt{(4-8)^2 + (2-4)^2 + (10-10)^2}} = 3{,}495206574$$

$$d(i) = \frac{3}{5}\sqrt{(2-5)^2 + (2-3)^2 + (5-10)^2 + (2-5)^2 + (1-3)^2 + (5-10)^2 + \sqrt{(3-5)^2 + (3-3)^2 + (5-10)^2 + (3-5)^2 + (3-3)^2 + (5-10)^2}} = 4{,}57016468$$

Cluster 3

$$d(i) = \frac{1}{4}\sqrt{(8-2)^2 + (4-2)^2 + (10-5)^2 + (7-2)^2 + (4-2)^2 + (10-5)^2 + \sqrt{(9-2)^2 + (4-2)^2 + (10-5)^2 + (8-2)^2 + (4-2)^2 + (10-5)^2 + \sqrt{(7-2)^2 + (5-2)^2 + (10-5)^2 + (8-2)^2 + (6-2)^2 + (10-5)^2 + } = 12{,}19021393$$

$$d(i) = \frac{2}{4}\sqrt{(2-5)^2 + (2-3)^2 + (5-10)^2 + (2-5)^2 + (1-3)^2 + (5-10)^2 + \sqrt{(3-5)^2 + (3-3)^2 + (5-10)^2 + (3-5)^2 + (3-3)^2 + (5-10)^2} = 5{,}71270585$$

Hasil perhitungan $d(i)$ terlihat pada tabel 3.16 sebagai berikut:

Tabel 3. 16 Hasil Perhitungan *d*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 2

JB	JP	JAM	JE	d(*i*,1)	d(*i*,2)	d(*i*,3)
JB 1	8	4	10	-	3,495206574	12,19021393
JB 3	7	4	10	-	3,495206574	12,19021393
JB 6	9	4	10	-	3,495206574	12,14315146
JB 7	8	4	10	-	3,495206574	12,19021393
JB 9	7	5	10	-	4,137427084	12,19021393
JB 12	8	6	10	-	3,495206574	12,19021393
JB 2	5	3	10	3,292466	-	5,71270585
JB 4	6	3	10	3,292466	-	5,71270585
JB 5	4	4	10	3,452604	-	5,71270585
JB 8	5	2	10	3,292466	-	5,71270585
JB 13	4	2	10	3,292466	-	5,71270585

JB	JP	JAM	JE	d(*i*,1)	d(*i*,2)	d(*i*,3)
JB 10	2	2	5	8,126809	4,57016468	-
JB 11	2	1	5	8,235988	4,57016468	-
JB 14	3	3	5	8,126809	4,57016468	-
JB 15	3	3	5	8,126809	4,57016468	-

Tabel 3.16, hasil seluruh perhitungan seluruh perhitungan $d(i)$ pada *cluster* 1, 2, dan 3. Dimana *cluster* yang akan dijadikan objek tidak dihitung karena akan dibandingkan dengan data selain didalam *cluster*nya. Dengan kata lain, jarak dari titik data dalam *cluster* yang dijadikan objek tidak dihitung sebagai bagian dari perbandingan untuk $b(i)$. Langkah berikutnya yaitu, menghitung nilai $b(i)$ pada iterasi 2 menggunakan persamaan 3.7.

Cluster 1

$$b(i) = \min(3{,}495206574 : 12{,}19021393) = 3{,}495206574$$

Cluster 2

$$b(i) = \min(3{,}292466123 : 5{,}71270585) = 3{,}292466123$$

Cluster 3

$$b(i) = \min(8{,}126809288 : 4{,}57016468) = 4{,}57016468$$

Hasil dari perhitungan $b(i)$ terlihat pada tabel 3.17 sebagai berikut:

Tabel 3. 17 Hasil Seluruh Perhitungan *b*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 2

JB	d(*i*,1)	d(*i*,2)	d(*i*,3)	b(*i*)
JB 1	-	3,495206574	12,19021393	3,495206574
JB 3	-	3,495206574	12,19021393	3,495206574
JB 6	-	3,495206574	12,14315146	3,495206574
JB 7	-	3,495206574	12,19021393	3,495206574
JB 9	-	4,137427084	12,19021393	4,137427084
JB 12	-	3,495206574	12,19021393	3,495206574
JB 2	3,292466	-	5,71270585	3,292466123
JB 4	3,292466	-	5,71270585	3,292466123
JB 5	3,452604	-	5,71270585	3,452604117
JB 8	3,292466	-	5,71270585	3,292466123
JB 13	3,292466	-	5,71270585	3,292466123
JB 10	8,126809	4,57016468	-	4,57016468
JB 11	8,235988	4,57016468	-	4,57016468
JB 14	8,126809	4,57016468	-	4,57016468
JB 15	8,126809	4,57016468	-	4,57016468

Tabel 3.17, Hasil seluruh perhitungan $b(i)$ iterasi 2 pada *cluster* 1, 2, dan 3. Perhitungan dilakukan dengan mencari nilai minimum pada setiap *cluster* yang memiliki data, apabila dalam *cluster* tidak terdapat data maka tidak dihitung. Tahap terakhir yaitu menghitung nilai $s(i)$ atau *silhouette cluster* 1, 2, dan 3 pada iterasi 2 menggunakan persamaan 3.8 dan 3.9.

Cluster 1

$$s\ (jb_1) = \frac{3{,}49 - 1{,}08}{\max\big(a(1{,}08), b(3{,}49)\big)} = 0{,}69$$

$$s\ (jb_{12}) = \frac{3{,}49 - 1{,}97}{\max\big(a(1{,}97), b(3{,}49)\big)} = 0{,}43$$

Cluster 2

$$s\ (jb_2) = \frac{3{,}29 - 1{,}20}{\max\big(a(1{,}20), b(3{,}29)\big)} = 0{,}63$$

$$s\ (jb_{13}) = \frac{3{,}29 - 1{,}66}{\max\big(a(1{,}66), b(3{,}29)\big)} = 0{,}49$$

Cluster 3

$$s\ (jb_{10}) = \frac{4{,}57 - 1{,}27}{\max\big(a(1{,}27), b(4{,}57)\big)} = 0{,}72$$

$$s\ (jb_{15}) = \frac{4{,}57 - 1{,}21}{\max\big(a(1{,}21), b(4{,}57)\big)} = 0{,}73$$

$$\text{SC} = max\ 0{,}69019207 : 0{,}733760029 = 0{,}57071627$$

Hasil perhitungan $s(i)$ atau s*ilhouette score* dapat dilihat pada tabel 3.18 sebagai berikut:

Tabel 3. 18 Hasil Seluruh Perhitungan *s*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 2

JB	a(*i*)	b(*i*)	s(*i*)
JB 1	1,082843	3,495206574	0,69019207
JB 3	1,447214	3,495206574	0,58594333
JB 6	1,694427	3,495206574	0,515214007
JB 7	1,082843	3,495206574	0,69019207
JB 9	1,495742	4,137427084	0,638485053
JB 12	1,97727	3,495206574	0,434290975
JB 2	1,207107	3,292466123	0,63337306
JB 4	1,721587	3,292466123	0,477113108
JB 5	1,971587	3,452604117	0,42895643
JB 8	2,690671	3,292466123	0,182779534

JB	a(*i*)	b(*i*)	s(*i*)
JB 13	1,66257	3,292466123	0,495037968
JB 10	1,276142	4,57016468	0,720766654
JB 11	1,824045	4,57016468	0,600879739
JB 14	1,216761	4,57016468	0,733760029
JB 15	1,216761	4,57016468	0,733760029
			SC(0,57071627)

Tabel 3.18, hasil seluruh perhitungan $s(i)$ pada iterasi 2 didapatkan nilai *silhouette coefficient* yaitu 0,57, untuk mengetahui apakah nilai tersebut mengalami perubahan pada iterasi 3. Maka selanjutnya melukan kembali perhitungan di iterasi 3 menggunakan persamaan 3.5, 3.6, 3.7, 3.8, 3.9 yang sudah dilakukan pada iterasi 1 dan 2. Menghitung nilai $a(i)$ pada *cluster* 1, 2, dan 3 menggunakan persamaan 3.5. Hasil perhitungan dapat dilihat pada tabel 3.19 sebagai berikut:

Cluster 1

$$a(i) = \frac{1}{5}\sqrt{(7-8)^2 + (4-4)^2 + (10-10)^2 + (9-8)^2 + (4-4)^2 + (10-10)^2 + \sqrt{(8-8)^2 + (4-4)^2 + (10-10)^2 + (7-8)^2 + (5-4)^2 + (10-10)^2}\sqrt{(8-8)^2 + (6-4)^2 + (10-10)^2}}$$
$$= 1{,}082842712$$

Cluster 2

$$a(i) = \frac{2}{4}\sqrt{(6-5)^2 + (3-3)^2 + (10-10)^2 + (4-5)^2 + (3-3)^2 + (10-10)^2 + \sqrt{(5-5)^2 + (2-3)^2 + (10-10)^2 + (4-5)^2 + (2-3)^2 + (10-10)^2}}$$
$$= 1{,}207106781$$

Cluster 3

$$a(i) = \frac{3}{3}\sqrt{(2-2)^2 + (1-5)^2 + (5-5)^2 + (3-2)^2 + (3-2)^2 + (5-5)^2 + \sqrt{(3-2)^2 + (3-2)^2 + (5-5)^2}}$$
$$= 1{,}276142375$$

Tabel 3. 19 Hasil Perhitungan *a*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 3

JB	JP	JAM	JE	a(*i*)
JB 1	8	4	10	1,082843
JB 3	7	4	10	1,447214
JB 6	9	4	10	1,694427
JB 7	8	4	10	1,082843
JB 9	7	5	10	1,495742
JB 12	8	6	10	1,97727
JB 2	5	3	10	1,207107

JB	JP	JAM	JE	a(*i*)
JB 4	6	3	10	1,721587
JB 5	4	4	10	1,971587
JB 8	5	2	10	2,690671
JB 13	4	2	10	1,66257
JB 10	2	2	5	1,276142
JB 11	2	1	5	1,824045
JB 14	3	3	5	1,216761
JB 15	3	3	5	1,216761

Tabel 3.19 adalah hasil seluruh perhitungan jarak rata-rata data lain dalam satu *cluster* pada *cluster* 1, 2, dan 3 iterasi ke 3. Setelah melakukan perhitungan nilai $a(i)$, selanjutnya melakukan perhitungan rata-rata jarak data i atau $d(i)$ dengan semua data di *cluster* lain pada iterasi 3 yaitu *cluster* 1, 2 dan 3 serta mengambil nilai terkecilnya menggunakan persamaan 3.6.

Cluster 1

$$d(i) = \frac{2}{6}\sqrt{(8-5)^2+(4-3)^2+(10-10)^2+(9-5)^2+(4-3)^2+(10-10)^2+}$$
$$\sqrt{(8-5)^2+(4-3)^2+(10-10)^2+(7-5)^2+(5-5)^2+(10-10)^2+}$$
$$\sqrt{(8-5)^2+(6-3)^2+(10-10)^2}$$
$$= 3{,}292466123$$

$$d(i) = \frac{3}{6}\sqrt{(2-2)^2+(1-2)^2+(5-5)^2+(3-2)^2+(3-2)^2+(5-5)^2+}$$
$$\sqrt{(3-2)^2+(3-2)^2+(5-5)^2+(3-2)^2+(3-2)^2+(5-5)^2}$$
$$= 8{,}126809288$$

Cluster 2

$$d(i) = \frac{1}{5}\sqrt{(5-8)^2+(3-4)^2+(10-10)^2+(7-8)^2+(4-4)^2+(10-10)^2+}$$
$$\sqrt{(6-8)^2+(3-4)^2+(10-10)^2+(5-8)^2+(2-4)^2+(10-10)^2+}$$
$$\sqrt{(4-8)^2+(2-4)^2+(10-10)^2}$$
$$= 3{,}495206574$$

$$d(i) = \frac{3}{5}\sqrt{(2-5)^2+(2-3)^2+(5-10)^2+(2-5)^2+(1-3)^2+(5-10)^2+}$$
$$\sqrt{(3-5)^2+(3-3)^2+(5-10)^2+(3-5)^2+(3-3)^2+(5-10)^2}$$
$$= 4{,}57016468$$

Cluster 3

$$d(i) = \frac{1}{4}\sqrt{(8-2)^2+(4-2)^2+(10-5)^2+(7-2)^2+(4-2)^2+(10-5)^2+}$$
$$\sqrt{(9-2)^2+(4-2)^2+(10-5)^2+(8-2)^2+(4-2)^2+(10-5)^2+}$$
$$\sqrt{(7-2)^2+(5-2)^2+(10-5)^2+(8-2)^2+(6-2)^2+(10-5)^2+}$$
$$= 12{,}19021393$$

$$d(i) = \frac{2}{4}\sqrt{(2-5)^2+(2-3)^2+(5-10)^2+(2-5)^2+(1-3)^2+(5-10)^2+} \\ \sqrt{(3-5)^2+(3-3)^2+(5-10)^2+(3-5)^2+(3-3)^2+(5-10)^2} \\ = 5{,}71270585$$

Tabel 3. 20 Hasil Perhitungan *d*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 3

JB	JP	JAM	JE	d(*i*,1)	d(*i*,2)	d(*i*,3)
JB 1	8	4	10	-	3,495206574	12,19021393
JB 3	7	4	10	-	3,495206574	12,19021393
JB 6	9	4	10	-	3,495206574	12,14315146
JB 7	8	4	10	-	3,495206574	12,19021393
JB 9	7	5	10	-	4,137427084	12,19021393
JB 12	8	6	10	-	3,495206574	12,19021393
JB 2	5	3	10	3,292466	-	5,71270585
JB 4	6	3	10	3,292466	-	5,71270585
JB 5	4	4	10	3,452604	-	5,71270585
JB 8	5	2	10	3,292466	-	5,71270585
JB 13	4	2	10	3,292466	-	5,71270585
JB 10	2	2	5	8,126809	4,57016468	-
JB 11	2	1	5	8,235988	4,57016468	-
JB 14	3	3	5	8,126809	4,57016468	-
JB 15	3	3	5	8,126809	4,57016468	-

Tabel 3.20 hasil seluruh perhitungan seluruh perhitungan *d*(*i*) pada *cluster* 1, 2, dan 3. Dimana *cluster* yang akan dijadikan objek tidak dihitung karena akan dibandingkan dengan data selain didalam *cluster*nya. Dengan kata lain, jarak dari titik data dalam *cluster* yang dijadikan objek tidak dihitung sebagai bagian dari perbandingan untuk *b*(*i*). Langkah berikutnya yaitu, menghitung nilai *b*(*i*) pada iterasi 3 menggunakan persamaan 3.7.

Cluster 1

$$b(i) = \min(3{,}495206574 : 12{,}19021393) = 3{,}495206574$$

Cluster 2

$$b(i) = \min(3{,}292466123 : 5{,}71270585) = 3{,}292466123$$

Cluster 3

$$b(i) = \min(8{,}126809288 : 4{,}57016468) = 4{,}57016468$$

Tabel 3. 21 Hasil Seluruh Perhitungan *b*(*i*) Pada *Cluster* 1, 2, dan 3 iterasi 3

JB	d(*i*,1)	d(*i*,2)	d(*i*,3)	b(*i*)
JB 1	-	3,495206574	12,19021393	3,495206574
JB 3	-	3,495206574	12,19021393	3,495206574

JB	d(*i*,1)	d(*i*,2)	d(*i*,3)	b(*i*)
JB 6	-	3,495206574	12,14315146	3,495206574
JB 7	-	3,495206574	12,19021393	3,495206574
JB 9	-	4,137427084	12,19021393	4,137427084
JB 12	-	3,495206574	12,19021393	3,495206574
JB 2	3,292466	-	5,71270585	3,292466123
JB 4	3,292466	-	5,71270585	3,292466123
JB 5	3,452604	-	5,71270585	3,452604117
JB 8	3,292466	-	5,71270585	3,292466123
JB 13	3,292466	-	5,71270585	3,292466123
JB 10	8,126809	4,57016468	-	4,57016468
JB 11	8,235988	4,57016468	-	4,57016468
JB 14	8,126809	4,57016468	-	4,57016468
JB 15	8,126809	4,57016468	-	4,57016468

Tabel 3.21, hasil seluruh perhitungan $b(i)$ iterasi 3 pada *cluster* 1, 2, dan 3. Perhitungan dilakukan dengan mencari nilai minimum pada setiap *cluster* yang memiliki data, apabila dalam *cluster* tidak terdapat data maka tidak dihitung. Tahap terakhir yaitu menghitung nilai $s(i)$ atau *silhouette score* pada *cluster* 1, 2, dan 3 pada iterasi 3 menggunakan persamaan 3.8 dan 3.9.

Cluster 1

$$s\ (jb_1) = \frac{3{,}49 - 1{,}08}{\max\left(a(1{,}08), b(3{,}49)\right)} = 0{,}69$$

$$s\ (jb_{12}) = \frac{3{,}49 - 1{,}97}{\max\left(a(1{,}97), b(3{,}49)\right)} = 0{,}43$$

Cluster 2

$$s\ (jb_2) = \frac{3{,}29 - 1{,}20}{\max\left(a(1{,}20), b(3{,}29)\right)} = 0{,}63$$

$$s\ (jb_{13}) = \frac{3{,}29 - 1{,}66}{\max\left(a(1{,}66), b(3{,}29)\right)} = 0{,}49$$

Cluster 3

$$s\ (jb_{10}) = \frac{4{,}57 - 1{,}27}{\max\left(a(1{,}27), b(4{,}57)\right)} = 0{,}72$$

$$s\ (jb_{15}) = \frac{4{,}57 - 1{,}21}{\max\left(a(1{,}21), b(4{,}57)\right)} = 0{,}73$$

$$\text{SC} = max\ 0{,}69019207 : 0{,}733760029 = 0{,}57071627$$

Tabel 3. 22 Hasil Seluruh Perhitungan *s*(*i*) Pada *Cluster* 1, 2, dan 3 Iterasi 3

JB	a(*i*)	b(*i*)	s(*i*)
JB 1	1,082843	3,495206574	0,69019207
JB 3	1,447214	3,495206574	0,58594333
JB 6	1,694427	3,495206574	0,515214007
JB 7	1,082843	3,495206574	0,69019207
JB 9	1,495742	4,137427084	0,638485053
JB 12	1,97727	3,495206574	0,434290975
JB 2	1,207107	3,292466123	0,63337306
JB 4	1,721587	3,292466123	0,477113108
JB 5	1,971587	3,452604117	0,42895643
JB 8	2,690671	3,292466123	0,182779534
JB 13	1,66257	3,292466123	0,495037968
JB 10	1,276142	4,57016468	0,720766654
JB 11	1,824045	4,57016468	0,600879739
JB 14	1,216761	4,57016468	0,733760029
JB 15	1,216761	4,57016468	0,733760029
			SC(0,57071627)

Tabel 3.22. Hasil seluruh perhitungan *s*(*i*) pada iterasi 3 didapatkan nilai *silhouette coefficient* yaitu 0,57, dari perhitungan evaluasi hasil *clustering* iterasi 1, 2, dan 3, dapat diketahui bahwa nilai *silhouette* pada iterasi 2 mengalami peningkatan dari iterasi 1. Sedangkan nilai *silhouette* pada iterasi 2 dan 3 memiliki jumlah yang sama. Dari perhitungan tersebut dapat disimpulkan titik data telah dialokasikan ke dalam *cluster* yang sesuai, hal ini dibuktikan dengan nilai *silhouette* pada iterasi 2 dan 3 mendekati +1.

BAB IV
HASIL DAN PEMBAHASAN

4.1 Hasil

Pada tahap ini, dilakukan proses pengelompokkan (*clustering*) buku berdasarkan tingkat keterpakaian koleksi menggunakan algoritma *k-means*. Penelitian ini memiliki tahapan yang meliputi pengumpulan data, *prepocessing* data, penentuan jumlah *cluster*, penerapan algoritma *k-means*, dan evaluasi hasil *clustering* menggunakan *silhouette coefficient*.

4.1.1 Data Transaksi Peminjaman

Tahap pertama yaitu menyiapkan data transaksi peminjaman koleksi yang diperoleh dari sistem otomasi SLiMS Perpustakaan MAN 2 Kota berbentuk csv dalam jangka waktu 1 tahun mulai dari 1 Januari 2023 sampai 1 Januari 2024. Data berjumlah 768 data yang terdiri dari beberapa atribut yaitu, id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman.

4.1.2 *Pre-Processing* Data

Pada tahap *preprocessing*, dilakukan transformasi data untuk mendapatkan atribut data yang relevan dengan tingkat keterpakaian buku. Proses transformasi meliputi:

1. Pembersihan data dari nilai yang hilang atau tidak valid.
2. Penghapusan data duplikat pada tabel judul, nama anggota dan kode eksemplar.
3. Penyederhanaan data untuk memperoleh tabel baru yaitu:
    a. Jumlah peminjaman per judul buku.
    b. Jumlah anggota yang meminjam per judul.
    c. Jumlah eksemplar perjudul.

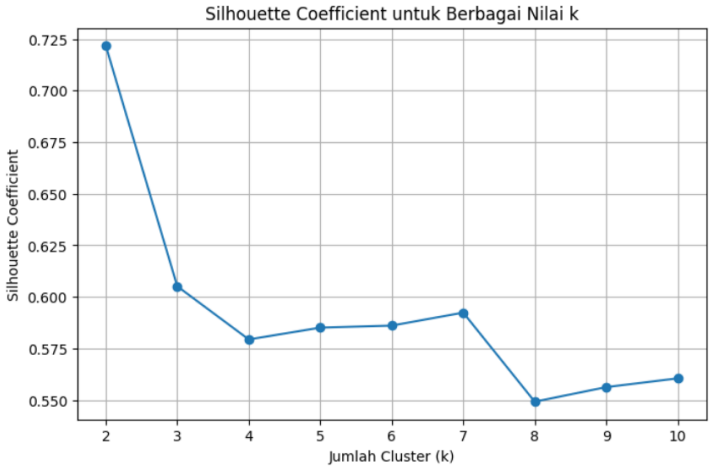
Sebelum memulai transformasi data terlebih dulu memuat data kedalam python dengan format csv. Data awal berjumlah 768 dan memiliki atribut id anggota, nama anggota, kode eksemplar, judul, tanggal pinjam, tanggal kembali dan status

peminjaman. Selanjutnya melakukan transformasi data menggunakan *library pandas* di *jupyter notebook*. Transformasi data dilakukan karena pada data awal terdapat beberapa data duplikat, jika langsung digunakan tanpa diolah, hasil analisis bisa jadi tidak akurat. Dengan melakukan transformasi, data diubah ke dalam format yang seragam dan mudah diproses.

Hasil dari transformasi dan penyederhaan data yang sebelumnya berjumlah 768 serta memiliki atribut id anggota, nama anggota, kode eksemplar, judul buku, tanggal peminjaman, tanggal pengembalian dan status peminjaman. Kemudian setalah dilakukan transformasi, menghasilkan 193 data dan memiliki atribut Judul, Jumlah Peminjaman, Jumlah Anggota, dan Jumlah Eksemplar sesuai dengan varibel yang digunakan dalam penelitian. Selanjutnya data disimpan kedalam file csv.

4.1.3 Penentuan Jumlah *Cluster*

Dalam menentukan jumlah *cluster* (***k***) yang paling sesuai merupakan langkah penting agar hasil pengelompokan dapat mencerminkan pola alami dalam data. Salah satu cara yang umum digunakan untuk memilih nilai ***k*** yang tepat adalah dengan menggunakan matrik *silhouette coefficient*. Matrik ini menilai seberapa baik objek-objek dalam suatu *cluster* saling menyerupai (*kohesi*), serta seberapa jauh mereka berbeda dengan objek di *cluster* lain (*separasi*). Nilai koefisien ini berada pada rentang -1 hingga +1, di mana nilai yang lebih tinggi menunjukkan bahwa struktur *cluster* tersebut semakin baik. Adapun hasil untuk melihat jumlah *cluster* terbaik menggunakan matrik *silhouette coefficient* dengan rentang jumlah *cluster* 2 sampai 10 dapat dilihat pada gambar 4.1 sebagai berikut:



Gambar 4. 1 Grafik *Cluster Silhouette Coefficient*

Berdasarkan grafik *silhouette coefficient* pada gambar 4.1 yang dianalisis untuk berbagai nilai ***k***, diketahui bahwa nilai tertinggi terlihat pada jumlah *cluster* adalah 2, dengan angka sekitar 0,7217. Hal ini menunjukkan bahwa pemisahan antar *cluster* paling efektif ketika data dibagi menjadi dua kelompok. Semakin tinggi nilai koefisien ini, semakin baik objek-objek dalam satu *cluster* saling berkaitan dan jelas perbedaannya dengan objek di *cluster* lain. Dari hasil tersebut, dapat disimpulkan bahwa jumlah *cluster* yang digunakan pada penelitian ini berdasarkan hasil dari matrik *silhouette coefficient* adalah ***k***=2.

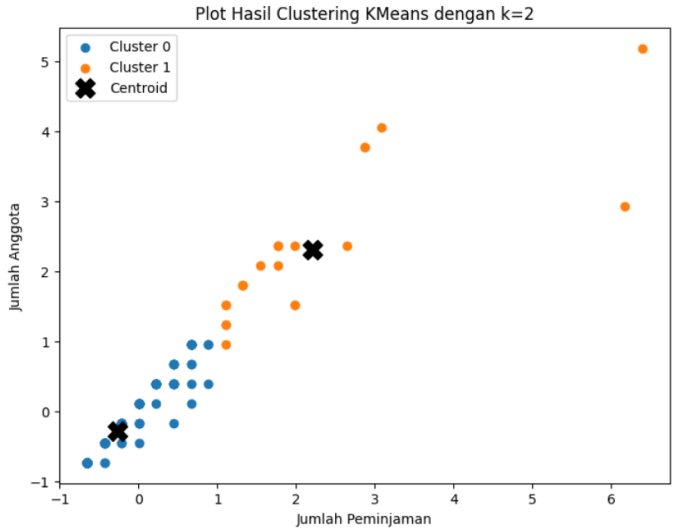
4.1.4 Penerapan Algoritma *K-Means*

Setelah jumlah *cluster* ditentukan, proses selanjutnya yaitu *clustering* dengan algoritma *k-means* menggunakan variabel judul buku, jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar. Hasil dari proses ini menghasilkan pembagian data ke dalam *cluster* yang masing-masing merepresentasikan pola keterpakaian buku berdasarkan karakteristik yang serupa. Setiap entri data kemudian diberi label dalam kolom baru bernama "*cluster*" dalam dataframe hasil. Penambahan kolom ini bertujuan untuk mempermudah interpretasi dan analisis lanjutan, khususnya dalam memahami distribusi dan kecenderungan pemanfaatan koleksi buku di perpustakaan. Selanjutnya, seluruh hasil *clustering* tersebut disimpan ke dalam sebuah

file csv dengan nama ‘Hasil_clustering.csv’ sebagai dokumentasi dan bahan analisis lebih lanjut.

Adapun untuk anggota kelompok *cluster* berdasarkan karakteristiknya yaitu, *cluster* 0 terdiri dari 172 judul buku yang terlihat dari jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar yang tersedia. Sementara itu, *cluster* 1 mencakup 21 judul buku yang juga dikelompokkan berdasarkan karakteristik yang sama seperti *cluster* 0.

Setelah melakukan proses *clustering*, kemudian melihat plot hasil *clustering* algoritma *k-means*. Plot berfungsi untuk menampilkan hasil *clustering* menggunakan metode k-means dalam bentuk grafik sebar. Masing-masing *cluster* digambarkan dengan warna dan label yang berbeda sehingga memudahkan dalam membedakan antar kelompok data. Titik pusat dari setiap *centroid*, ditandai dengan simbol khusus berupa tanda ‘**X’** berwarna hitam. Hasil dari plot dapat dilihat pada gambar 4.2 sebagai berikut:

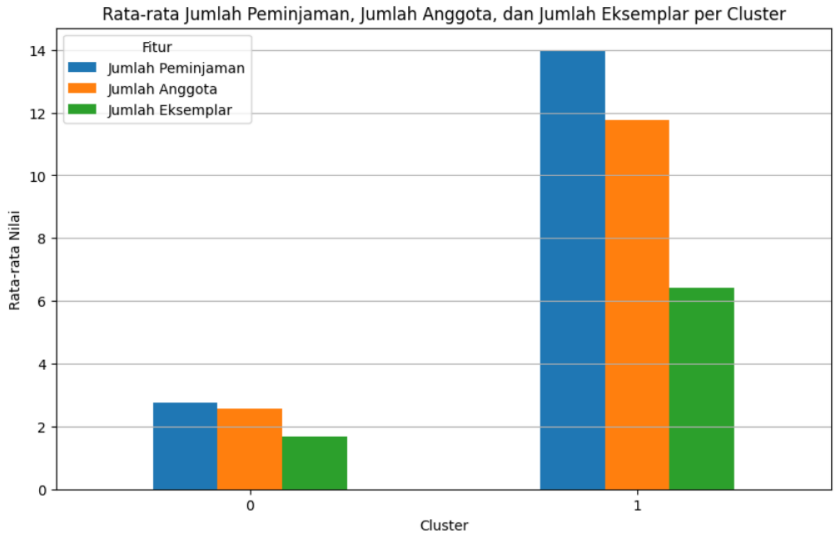


Gambar 4. 2 Hasil Plot *Clustering*

Gambar 4.2 menunjukkan visualisasi hasil *clustering* dalam bentuk grafik sebar yang memudahkan dalam melihat pemisahan antar *cluster* berdasarkan karakteristik data. Setiap *cluster* diberi warna berbeda, sehingga pola-pola dalam data bisa terlihat lebih jelas. Titik-titik pada grafik mewakili judul buku dan dikelompokkan berdasarkan

kemiripan atribut, seperti frekuensi peminjaman dan jumlah eksemplar. Di tengah tiap *cluster* terdapat tanda '**X**' berwarna hitam yang menunjukkan titik pusat (*centroid*), yaitu rata-rata posisi data dalam *cluster* tersebut.

Langkah berikutnya adalah menghitung rata-rata dari setiap fitur yaitu jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar untuk masing-masing *cluster*. Adapun hasil menghitung rata-rata setiap *cluster* divisualisasikan dalam bentuk grafik batang, di mana setiap grafik mewakili satu buku dan diberi warna sesuai dengan *cluster* yang terbentuk. Gambar 4.3 merupakan tampilan visualisasi rata-rata setiap *cluster*.

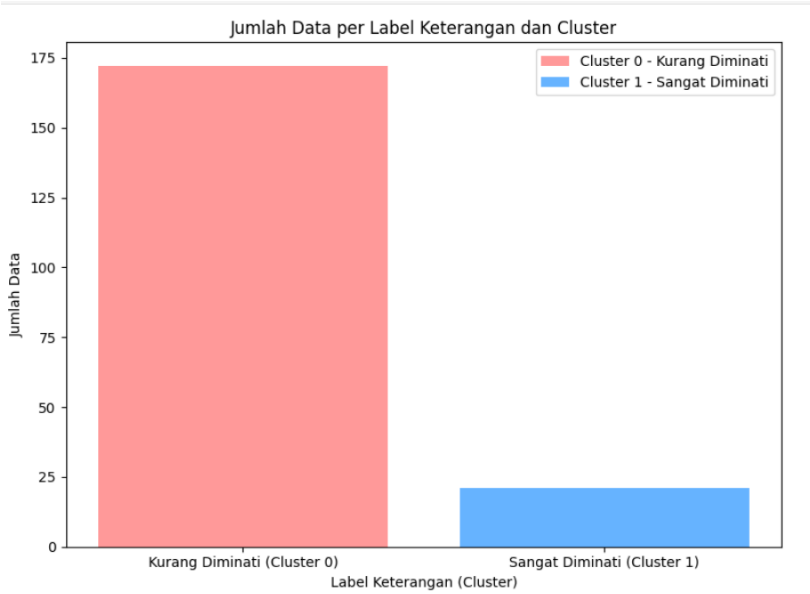


Gambar 4. 3 Grafik Rata-Rata Per *Cluster*

Gambar 4.3 menunjukkan visualisasi rata-rata dari setiap *cluster* yang dihasilkan berdasarkan empat variabel utama, yaitu judul, jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar koleksi. Hasil visualisasi menunjukan *cluster* 1 memiliki nilai rata-rata tertinggi pada ketiga fitur yang dianalisis, yaitu jumlah peminjaman sebesar 14,00, jumlah anggota 11,76, dan jumlah eksemplar 6,43. Sebaliknya, *cluster* 0 menunjukkan rata-rata jumlah peminjaman sebesar 2,76, jumlah anggota 2,58, dan jumlah eksemplar 1,69. Dari perbandingan ini, dapat diketahui bahwa *cluster* 1 mewakili kelompok buku yang paling sering digunakan di

perpustakaan. Sementara itu, *cluster* 0 menggambarkan kelompok buku dengan tingkat penggunaan yang lebih rendah.

Setelah melihat grafik rata-rata per-*cluster*, selanjutnya memberikan label keterangan pada *cluster*. Pemberian label keteranga terhadap hasil *clustering* data berdasarkan *cluster* yang telah diperoleh sebelumnya. Setiap *cluster* diberi label deskriptif, yaitu "sangat diminati", dan "kurang diminati" sesuai dengan anggota *cluster*. Label ini ditambahkan ke dalam kolom baru bernama (label_keterangan) pada dataframe dan disimpan kedalam file csv bernama Keterangan.csv. Visualisasi hasil pemberian label keterangan dapat dilihat pada gambar 4.4 sebagai berikut:



Gambar 4. 4 Grafik Hasil Label Keterangan

Visualisasi pada Gambar 4.4 menunjukkan bahwa mayoritas data berada pada *cluster* 0 yang diberi label "kurang diminati", dengan jumlah data mencapai sekitar 172 entri. Sementara itu, *cluster* 1 yang diberi label "sangat diminati" hanya terdiri dari 21 entri. Pemberian label keterangan pada setiap anggota *cluster* bertujuan untuk mempermudah dalam memahami hasil *clustering* yang telah dilakukan. Setiap *cluster* yang terbentuk melalui metode *k-means* pada dasarnya hanya dibedakan berdasarkan nilai angka yang belum tentu langsung dimengerti maknanya.

4.1.5 Hasil *Clustering* dan Evaluasi Dengan *Silhoutte Coefficient*

Hasil *clustering* menggunakan algoritma *k-means* dengan ***k***=2 menghasilkan dua *cluster* yaitu *cluster* 0, dan *cluster* 1. Masing-masing *cluster* memiliki anggota yang berbeda berdasarkan tingkat keterpakaian koleksi, kemudian setiap *cluster* diberikan label keterangan berbeda yaitu, sangat diminati (*cluster* 0) dan kurang iminati (*cluster* 1). Untuk memberikan gambaran yang lebih jelas mengenai karakteristik masing-masing *cluster*, berikut disajikan tabel rekapitulasi hasil *clustering*:

Tabel 4. 1 Tabel Rekapitulasi Keterangan *Cluster*

Cluster	Rata-Rata Jumlah Peminjaman	Rata-Rata Jumlah Anggota meminjam	Rata-Rata Jumlah Eksemplar	Keterangan	Jumlah
0	2.755.814	2.575.581	1.686.047	Kurang Diminati	172
1	14.000.000	11.761.905	6.428.571	Sangat Diminati	21

Tabel 4.1 merupakan rekapitulasi hasil pengelompokan berdasarkan rata-rata dari masing-masing variabel, disertai dengan interpretasi keterangan setiap *cluster*. *Cluster* 0 memiliki rata-rata jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar yang relatif rendah dibandingkan *cluster* 1, sehingga dikelompokkan sebagai "kurang diminati" dengan jumlah anggota sebanyak 172. Sebaliknya, *cluster* 1 menunjukkan nilai rata-rata yang jauh lebih tinggi pada ketiga variabel tersebut, sehingga dikategorikan sebagai "sangat diminati" dengan jumlah anggota sebanyak 21

Setelah mendapatkan *cluster* dan pemberian label keterangan, selanjutnya dilakukan evaluasi hasil *clustering* menggunakan *silhouette coefficient*. Matrik ini digunakan untuk mengukur seberapa baik setiap objek (buku) cocok dengan *cluster*nya sendiri (*cohesion*) dibandingkan cluster terdekat (*separation*). Hasil *score silhouette* dapat dilihat pada tabel 4.2 sebagai berikut.

Tabel 4. 2 Hasil Score Silhouette Coefficient

Silhouette Coefficient untuk ***k***=2: 0.7217

Nilai *silhouette coefficient* yang diperoleh sebesar 0,7217 menunjukkan kualitas pengelompokan yang kuat dan optimal. Berdasarkan klasifikasi Rousseeuw (1987) nilai di atas 0,5 termasuk kategori "kuat" berarti pemisahan antar-*cluster* sangat jelas dan objek dalam satu *cluster* memiliki kemiripan yang tinggi. Dengan demikian, sebagian besar buku telah dikelompokkan ke dalam *cluster* yang tepat dan optimal.

4.2 Pembahasan

Hasil analisis *cluster* pada data transaksi peminjaman koleksi di perpustakaan MAN 2 Kota Tangerang menggunakan algoritma *k-means clustering* dengan empat variabel utama yaitu, judul, jumlah peminjaman, jumlah anggota yang meminjam, dan jumlah eksemplar memperoleh dua kelompok utama yang menggambarkan tingkat keterpakaian koleksi yang tersedia. *Cluster* pertama dalam hasil pengelompokan ditandai sebagai *cluster* 0 yang menandakan koleksi dengan tingkat keterpakaian kurang diminati. Sementara itu, *cluster* kedua atau *cluster* 1 menunjukan koleksi yang sangat diminati oleh pemustaka.

Berdasarkan tabel 4.1 dapat dilihat bahwa *cluster* 1 memiliki rata-rata jumlah peminjaman, anggota peminjam, dan eksemplar yang jauh lebih tinggi dibandingkan *cluster* 0, meskipun jumlah koleksi dalam *cluster* ini lebih sedikit. Pada *cluster* 1, beberapa koleksi dalam kategori ini memiliki jumlah eksemplar yang relatif sedikit, namun tingkat peminjamannya cukup tinggi. Adapun koleksi-koleksi yang masuk dalam *cluster* ini di antaranya adalah Akidah dan Akhlak 3 untuk Kelas XII Madrasah Aliyah, Belajar Bahasa Arab Otodidak, Tafsir Ayat-Ayat Ahkam, Ensiklopedi Akhir Zaman, dan Psikologi Pendidikan Islam. Hasil pengelompokan ini menunjukan bahwa keberadaan koleksi tersebut sesuai dengan kebutuhan pemustaka. Sejalan dengan pendapat Yuliani (2020) dalam pengembangan koleksi perpustakaan, aspek kesesuaian antara bahan pustaka dan kebutuhan pengguna memiliki peran yang lebih penting dibandingkan dengan jumlah koleksi semata. Dengan kata lain koleksi yang relevan dan sesuai dengan kebutuhan pemustaka cenderung memiliki tingkat pemanfaatan yang lebih tinggi, meskipun jumlah eksemplarnya sangat terbatas.

Sebaliknya, pada *cluster* 0 meskipun rata-rata jumlah eksemplarnya relatif lebih banyak, tingkat keterpakaiannya justru rendah, dengan jumlah pemustaka yang meminjam pun terbatas. Beberapa koleksi yang masuk dalam *cluster* 0 ini adalah 100 Wanita yang Membentuk Sejarah Dunia, 50 Rules to be Wonderful Teenager, 7 Prajurit Bapak, A Brief History of East Asia, dan Akidah dan Akhlak 2 untuk Kelas XI Madrasah Aliyah. Hasil pengelompokan ini menjadi catatan penting bahwa banyaknya koleksi tidak menjamin koleksi tersebut akan dimanfaatkan secara optimal. Dalam kasus *cluster* 0, rendahnya tingkat peminjaman meskipun jumlah eksemplar melimpah akan tetapi menunjukkan adanya ketidaksesuaian antara koleksi yang disediakan dengan kebutuhan pengguna. Evans & Saponaro (2005) menekankan bahwa proses pengembangan koleksi seharusnya tidak hanya didasarkan pada asumsi kebutuhan atau sekadar menambah jumlah judul dan eksemplar, melainkan harus selalu dievaluasi secara berkala berdasarkan data transaksi peminjaman koleksi. Berdasarkan temuan pada *cluster* 0, dapat dilihat bahwa banyaknya jumlah koleksi yang tersedia tidak selalu membuat koleksi tersebut lebih sering digunakan. Oleh karena itu, pengembangan koleksi sebaiknya dilakukan dengan pendekatan yang terencana, yaitu dengan menganalisis pola peminjaman secara rutin untuk menentukan jenis dan jumlah koleksi yang benar-benar dibutuhkan oleh pengguna.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan berbasis data seperti analisis *clustering* dengan algoritma *k-means* dapat digunakan sebagai alat evaluasi untuk mengetahui tingkat keterpakaian koleksi di perpustakaan. Pendekatan ini tidak hanya membantu dalam mengidentifikasi koleksi yang efektif, tetapi juga memberikan pemahaman dasar untuk menyusun strategi pengembangan koleksi yang lebih sesuai dengan kebutuhan pemustaka. Sejalan dengan panduan yang dikemukakan oleh (IFLA, 2001) pengembangan koleksi hendaknya berorientasi pada prinsip relevansi, aksesibilitas, dan aktualitas informasi. Ketiga prinsip ini menjadi landasan utama dalam memastikan bahwa perpustakaan mampu berfungsi secara optimal sebagai penyedia informasi sesuai yang dibutuhkan oleh pemustaka.

Nilai *silhouette coefficient* untuk jumlah *cluster* **k**=2 yang diperoleh pada penelitian ini adalah 0,7217, menunjukkan bahwa anggota dalam satu *cluster* memiliki tingkat kemiripan yang tinggi serta berbeda dari anggota *cluster* lain, serta menunjukkan adanya kohesi dan separasi yang cukup baik. Hal ini mengindikasikan bahwa hasil pengelompokan data menggunakan algoritma kmeans sudah berada pada kategori baik. Sebagai mana yang dinyatakan oleh Bagirov et al (2023) bahwa nilai *silhouette coefficient* dalam kisaran 0,5 hingga 0,7 menunjukkan struktur *cluster* yang cukup baik dan jelas.

Berdasarkan hasil analisis ini, terlihat bahwa metode *clustering* menggunakan algoritma *k-means* mampu mengelompokkan koleksi buku perpustakaan berdasarkan tingkat keterpakaian memiliki kinerja yang cukup baik. Pengelompokan ini memberikan gambaran mengenai pola peminjaman buku oleh anggota perpustakaan yang terbagi menjadi dua kategori *cluster,* yaitu kurang diminati dan sangat diminati.

Penelitian ini bertujuan untuk mengukur kinerja pengelompokan buku berdasarkan tingkat keterpakaian dengan menerapkan algoritma *k-means clustering* pada Perpustakaan MAN 2 Kota Tangerang. Pendekatan ini tidak hanya memberikan manfaat teknis dalam pengelolaan koleksi, tetapi juga memiliki relevansi yang kuat dengan nilai-nilai islam, khususnya dalam aspek pengelolaan ilmu pengetahuan secara efisien dan maslahat (Iswanto, 2017). Dalam islam, pemanfaatan sumber daya secara optimal sangat dianjurkan, sebagaimana firman ﷻ الله dalam Q.S Al-Isra:17 ayat 27 sebagai berikut:

اِنَّ الْمُبَذِّرِيْنَ كَانُوْٓا اِخْوَانَ الشَّيٰطِيْنِۗ وَكَانَ الشَّيْطٰنُ لِرَبِّهٖ كَفُوْرًا ﴿٢٧﴾

Artinya:

"*Sesungguhnya para pemboros itu adalah saudara-saudara setan dan setan itu sangat ingkar kepada Tuhannya*" (Al-Isrā' [17]:27).

Tafsir ayat ini memberikan penjelasan tentang sebuah peringatan dari ﷻ الله kepada manusia untuk tidak menggunakan sesuatu secara berlebihan tanpa manfaat yang jelas. Pemborosan dikecam karena bertentangan dengan prinsip keseimbangan

dan tanggung jawab terhadap nikmat yang diberikan ﷻ الله (quran.kemenag.go.id, 2025). Selain itu Imam Al-Qurtubi (1964) dalam tafsirnya menjelaskan bahwa pemborosan (tabdzir) adalah menggunakan sesuatu di luar kepatutan atau tidak sesuai kebutuhan yang sebenarnya. Dalam konteks perpustakaan, buku-buku yang tidak dimanfaatkan secara maksimal merupakan bentuk pemborosan aset pengetahuan. Oleh karena itu, pengelompokan buku berdasarkan keterpakaian menggunakan algoritma *k-means clustering* dapat dianggap sebagai upaya untuk menghindari pemborosan dan meningkatkan efisiensi pemanfaatan koleksi (Yulia & Silalahi, 2021).

Dengan mengelompokan buku berdasarkan tingkat keterpakaian, maka perpustakaan berperan aktif dalam menjaga dan memanfaatkan aset koleksi secara optimal untuk para pengguna (Lolytasari et al., 2023). Menurut Fang, (2023) dalam Ranganathan (1931) dari sudut pandang ilmu perpustakaan, prinsip ini sangat relevan dengan hukum kedua dari buku "*The Five Laws of Library Science*" yaitu "*Books are for use*". Artinya, koleksi yang dimiliki perpustakaan harus digunakan secara optimal oleh pemustaka. Maka, koleksi yang kurang diminati harus dievaluasi, sementara koleksi yang diminati harus diprioritaskan pengembangannya agar kebermanfaatannya meningkat. Dengan demikian, tafsir dari ayat ini menjelaskan bahwa pengelompokan koleksi berdasarkan tingkat keterpakaian di perpustakaan menggunakan algoritma *k-means clustering* tidak hanya meningkatkan efisiensi dan efektivitas pengelolaan sumber daya, tetapi juga merupakan manifestasi dari prinsip-prinsip islam dalam menjaga dan menyebarkan ilmu pengetahuan.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai pengelompokan buku berdasarkan tingkat keterpakaian menggunakan algoritma *k-means clustering* pada Perpustakaan MAN 2 Kota Tangerang memperoleh dua kategori *cluster* yaitu *cluster* 0 dan *cluster* 1. Perolehan jumlah *cluster* terbaik dilakukan menggunakan metode evaluasi *silhouette coefficient*, berdasarkan *silhouette coefficient* diperoleh nilai sebesar 0.7217 untuk k=2. Nilai tersebut menunjukkan bahwa anggota dalam satu *cluster* memiliki tingkat kemiripan yang tinggi, berbeda dari anggota *cluster* lain, serta menunjukkan adanya kohesi dan separasi yang cukup baik. Hasil penelitian ini mengelompokan buku menjadi dua kategori, yaitu kurang diminati (*cluster* 0) dan sangat diminati (*cluster* 1). Berdasarkan hasil pengelompokan tersebut diketahui bahwa beberapa buku yang sangat diminati hanya tersedia dalam jumlah eksemplar yang sedikit, sedangkan untuk buku yang kurang diminati eksemplarnya tersedia dalam jumlah yang terlalu banyak. Secara keseluruhan, penerapan algoritma *k-means clustering* dalam penelitian ini efektif digunakan untuk mengelompokkan koleksi buku berdasarkan tingkat keterpakaian serta memberikan bukti bahwa teknik *data mining* dapat membantu perpustakaan dalam menyusun strategi dalam pengembangan koleksi yang lebih efisien dan tepat sasaran sesuai dengan kebutuhan pengguna.

5.2 Saran

Berdasarkan hasil penelitian dan kesimpulan yang telah diperoleh, berikut beberapa saran yang dapat diberikan:

1. Bagi pihak Perpustakaan MAN 2 Kota Tangerang, hasil pengelompokan buku yang diperoleh dari penerapan algoritma *k-means clustering* dapat dimanfaatkan sebagai dasar dalam pengembangan koleksi. Koleksi yang tergolong dalam kategori “sangat diminati” sebaiknya ditambah jumlah eksemplarnya agar dapat memenuhi kebutuhan pengguna, sedangkan koleksi yang masuk kategori “kurang

diminati" dapat dievaluasi lebih lanjut baik melalui promosi untuk meningkatkan keterpakaian, atau dipertimbangkan untuk penyiangan apabila sudah tidak relevan atau tidak lagi sesuai kebutuhan pengguna.

2. Untuk penelitian selanjutnya, disarankan menggunakan data peminjaman dalam jangka waktu yang lebih panjang, karena akan memberikan hasil yang lebih akurat dalam melihat pola keterpakaian buku.

DAFTAR PUSTAKA

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*, *9*(8), 1295.

Akinola, S. A., & Opawale, F. Opeyemi. (2020). SWOT analysis (Strengths, Weaknesses, Opportunities and Threats) of the Joseph A eats) of the Joseph Ayo Babalola Univ o Babalola University Libr ersity Library, Ikeji - Arakeji, Nigeria. *Library Philosophy and Practice (e-Journal)*, *8153*.

Alasali, T., & Ortakci, Y. (2024). Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications. *Computer Science*, *9*(1), 32–50.

Alkhairi, P., & Windarto, A. Perdana. (2019). Penerapan K-Means Cluster Pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, *1*(1), 762–767.

Al-Qurtubi, S. (1964). *Tafsir Al-Jami' Li Ahkami Al-Qur'an* (XVII). Darul Kutub Mishriyah.

Amanda, D. Y. (2024). Analisis Penerapan Sistem Klasifikasi Islam Dalam Pengolahan Bahan Pustaka di Perpustakaan Pondok Pesantren Assalam. *PARADIGM: Journal Of Multidisciplinary Research and Innovation*, *2*(02), 112–123.

Ani, H., Nofriansyah, D., & Mariami, I. (2021). Implementasi Data Mining Untuk Pengelempokan Buku Di Perpustakaan Yayasan Nurul Islam Indonesia Baru Dengan Metode K-Means Clustering. *Jurnal Cyber Tech*, *1*(1), 1–12.

Apriyani, D., Harapan, E., & Houtman, H. (2020). Manajemen Perpustakaan Sekolah Dasar. *JMKSP (Jurnal Manajemen, Kepemimpinan, Dan Supervisi Pendidikan)*, *6*(1), 132–139.

Ardyawin, I. (2018). Urgensi Pengembangan Koleksi Sebagai Upaya Menyediakan Koleksi yang Berkualitas di Perpustakaan. *Jurnal Adabiya*, *20*(1), 86–108.

Awojobi, E. A., & Uthman, K. O. (2022). Evaluation of collection development and management in selected university libraries in Ogun State, Nigeria. *Ghana Library Journal*, *27*(2), 259–272.

Bagirov, A. M., Aliguliyev, R. M., & Sultanova, N. (2023). Finding compact and well-separated clusters: Clustering using silhouette coefficients. *Pattern Recognition*, *135*, 109144.

Campello, R. J. G. B., Kröger, P., Sander, J., & Zimek, A. (2020). Density-based clustering. *WIREs Data Mining and Knowledge Discovery*, *10*(2).

Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, *9*(3), 102–109.

Evans, G. E., & Saponaro, M. Z. (2005). *Developing Library and Information Center Collections* (5th ed.). Libraries Unlimited.

Fadlina, J., Utami, R., & Sitinjak, M. N. (2024). Pengelompokkan Buku Dan Rekomendasi Buku Menggunakan K-Means Clustering Pada Dinas Perpustakaan Dan Kearsipan Kota Medan. *JURNAL WIDYA*, *5*(2), 1045–1058.

Fang, M. (2023). A Book Review of "The Five Laws of Library Science." *Information and Knowledge Management*, *4*(2), 1–5.

Fatwa, A. N. (2020). Proses Pengembangan Koleksi Di Perpustakaan Smpit Bina Anak Sholeh (Bias) Yogyakarta. *JIPER: Jurnal Ilmu Perpustakaan*, *2*(2), 45–54.

Firmansyah, T., Poningsih, P., & Andani, S. R. (2022). Analisis Clustering Algoritma K-Means Sebagai Rekomendasi Penambahan Koleksi Buku Di Perpustakaan Madrasah Tsanawiyah Negeri 2 Simalungun. *Zahra: Bulletin Big Data, Data Science and Artificial Intelligence*, *1*(1), 44–48.

Fitriani, S., Marsono, & Azlan. (2021). Implementasi Data Mining Dalam Pengelompokkan Minat Baca Pengunjung Pada Perpustakaan STMIK Triguna Dharma MedanMenggunakan Metode K-Means. *Jurnal CyberTech*, *1*(2).

Gançarski, P., Dao, T. H., Crémilleux, B., Forestier, G., & Lampert, T. (2020). Constrained Clustering: Current and New Trends. *Artificial Intelligence Research*, *2*, 447–484.

Gantara, N. P., & Ali, I. (2023). Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station. *E-Link: Jurnal Teknik Elektro Dan Informatika*, *18*(1), 28.

Ginantra, N. L. R. W. S., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P., Effendy, F., Hazriani, A. I., Sari, I. Y., Prianto, Z. C. G., Gustian, D., & Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma* (R. Watrianthos & J. Simarmata, Eds.; 1st ed.). Yayasan Kita Menulis.

Golung, K. T., Golung, A. M., & Londa, J. W. (2023). Prinsip-Prinsip Pengadaan Bahan Pustaka Di Perpustakaan Melalui Pemilihan Untuk Memenuhi Koleksi Yang Relevan Dengan Kebutuhan Siswa Sma Negeri 2 Tomohon. *Jurnal Acta Diurna Komunikasi*, *5*(2), 5.

Handoyo, R., Mangkudjaja, R., & Nasution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K - Means pada Pengelompokan Dokumen. *Jurnal SIFO Mikroskil*, *15*(2), 73–82.

Hartama, D., & Anjelita, M. (2022). Analysis of Silhouette Coefficient Evaluation with Euclidean Distance in the Clustering Method (Case Study: Number of Public Schools in Indonesia). *Journal Mantik*, *6*(3), 3667–3677.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1), 100–108.

Hasan, Y. (2024). Pengukuran Silhouette Score Dan Davies-Bouldin Index Pada Hasil Cluster K-Means Dan DBSCAN. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3S1).

Hasanah, N. N., & Purnomo, A. S. (2022). Implementasi Data Mining Untuk Pengelompokan Buku Menggunakan Algoritma K-Means Clustering (Studi Kasus: Perpustakaan Politeknik LPP Yogyakarta). *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *4*(2), 300–311.

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56.

Huang, Z., Zheng, H., Li, C., & Che, C. (2024). Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection. *Academic Journal of Science and Technology*, *10*(1), 33–39.

IFLA. (2001). Guidelines For A Collection Development Policy Using The Conspectus Model. International Federation Of Library Associations And Institutions. *International Federation of Library Associations and Institutions (IFLA)*.

Iswanto, R. (2017). Kebijakan Pengembangan Koleksi dan Pemanfaatannya di Perpustakaan Perguruan Tinggi (Analisis Penerapan Kebijakan Pengembangan Koleksi Perpustakaan Utama Universitas Islam Negeri Syarif Hidayatullah Jakarta). *Tik Ilmeu: Jurnal Ilmu Perpustakaan Dan Informasi*, *1*(1), 1.

Izzah, T., Arifiyanti, A. A., & Najaf, A. E. R. (2023). Implementasi Segmentasi Pelanggan E-Commerce Menggunakan Algoritma K-Means Pada Website. *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi*, *3*(1), 217–226.

Karputri, D. L., & Yustanti, W. (2022). Analisis Klastering Buku sebagai Evaluasi untuk Peningkatan Minat Baca Perpusatakaan SMAN 1 Grogol. *Journal of Emerging Information System and Business Intelligence (JEISBI)*, *3*(3), 94–101.

Khan, G., & Bhatti, R. (2021). An Argument on Collection Development and Collection Management. *Library Philosophy and Practice (e-Journal)*, *5383*.

Lajnah Pentashihan Mushaf al-Qur`an Kemenag RI. (2009). *Tafsir al-Qur'an Tematik, Etika berkeluarga, Bermasyarakat dan Berpolitik.* (1st ed.). Lajnah Pentashihan Mushaf Al-Qur′an.

Lestari, W. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari). *Jurnal Sistem Informasi Dan Sistem Komputer*, *4*(2), 35–48.

Lestari, W., & Sumarlinda, S. (2021). Clustering Model Of Lecturers Performa In Publication Using K-Means For Decision Support Data. *International Journal of Multi Science*, *1*(10), 88–95.

Levenson, H. N., & Nichols, H. A. (2020). Collaborative collection development: current perspectives leading to future initiatives. *The Journal of Academic Librarianship*, *46*(5), 102201.

Lolytasari, Hayati, N., & Nuratikha, V. (2023). Keterpakaian Koleksi Perpustakaan Fakultas Kedokteran Uin Syarif Hidayatullah Jakarta Berdasarkan Analisis Sitiran Pada Skripsi. *Al-Maktabah: Jurnal Komunikasi Dan Informasi Perpustakaan*, *22*(2), 53–64.

Mahfud, F. K. R., Mudawwamah, N. S., & Hariyanto, W. (2020). Sentiment Analysis of Perpustakaan Nasional Republik Indonesia Through Social Media Twitter. *MATICS*, *12*(1), 90.

Marcoulides, G. A. (2005). Discovering Knowledge in Data: an Introduction to Data Mining. *Journal of the American Statistical Association*, *100*(472), 1465–1465.

Muhtadien, S., & Krismayani, I. (2019). Faktor-Faktor Penyebab Rendahnya Minat Kunjung Siswa Ke Perpustakaan SMAN 2 Mranggen. *Jurnal Ilmu Perpustakaan*, *6*(4), 341–350.

Nabila, H. N. H., & Sholihah, B. (2021). Optimalisasi Pengelolaan Koleksi Perpustakaan Di SMP Negeri 1 Bawen. *Management of Education: Jurnal Manajemen Pendidikan Islam*, *7*(2), 1–25.

Nasir, J. (2021). Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, *11*(2), 690–703.

Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48.

Primandana, A., Adinugroho, S., & Dewi, C. (2019). Optimasi Penentuan Centroid pada Algoritme K-Means Menggunakan Algoritme Pillar (Studi Kasus:

Penyandang Masalah Kesejahteraan Sosial di Provinsi Jawa Timur). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(11), 10678–10683.

Purba, A. T., Sugara, H., Simarmata, H. P. M., Saragih, D. Y., & Damanik, E. (2023). Sistem Pendukung Keputusan Penentuan Prioritas Pengadaan Buku Perpustakaan Menggunakan Metode K-Means Dan Electre. *Jurnal TEKINKOM*, *6*(1), 196–203.

Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, *1*(1), 72–77.

quran.kemenag.go.id. (2025). *qur'an kemenag*. Quran.Kemenag.Go.Id. https://quran.kemenag.go.id/quran/per-ayat/surah/6?from=67&to=165

Ranganathan, S. R. (1931). *The Five Laws of Library Science.* Madras Library Association.

Rochemin, & Wicaksono, H. (2023). Rancangan Strategi Pemanfaatan Koleksi Perpustakaan Berdasarkan Analisis Swot Di Perpustakaan Desa Bantarsari. *JIPKA*, *3*(2), 91–106.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Rouza, E., & Fimawahib, L. (2020). Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu. *Techno.Com*, *19*(4), 481–495.

Roy, S., Mondal, S., Ekbal, A., & Desarkar, M. S. (2019). Dispersion Ratio based Decision Tree Model for Classification. *Expert Systems with Applications*, *116*, 1–9.

Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 747–748.

Shihab, Q. (2002). *Tafsir Al-Mishbah Jilid 02* (2nd ed., Vol. 2). Lentera Hati.

Sigit, R. (2024). Penerapan Algoritma K-Means Clustering dalam Menganalisis Pola Peminjaman Buku di Perpustakaan. *The Indonesian Journal of Computer Science*, *13*(5).

Suhendra, O., Tiro, M. A., & Ruliana. (2022). K-Means Cluster Analysis for Grouping Districts in South Sulawesi Province Based on Village Potential. *ARRUS Journal of Mathematics and Applied Science*, *2*(1), 49–59.

Sumallika, T., Alekya, V., Raju, P. V. M., Rao, R., Shiney, D. E. G., & Sudha, M. V. (2024). Exploring Optimal Cluster Quality in Health Care Data (HCD):

Comparative Analysis utilizing k-means Elbow and Silhouette Analysis. *International Journal of Chemical and Biochemical Sciences (IJCBS)*, *25*(16), 48–60.

Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, *3*(1), 118–136.

Tambunan, H. B., Barus, D. H., Hartono, J., Alam, A. S., Nugraha, D. A., & Usman, H. H. H. (2020). Electrical Peak Load Clustering Analysis Using K-Means Algorithm and Silhouette Coefficient. *2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, 258–262.

Tan, P. N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Introduction to Data Mining (2nd ed.). *Pearson Education Limited*.

Triandini, M., Defit, S., & Nurcahyo, G. W. (2021). Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia. *Jurnal Informasi Dan Teknologi*, *3*(2), 167–172.

Ulfah, M., Irtawaty, A. S., & Abrar, A. (2024). Implementasi Data Mining Untuk Pengelompokan Koleksi Buku Di Perpustakaan Perpustakaan Kota Balikpapan Dengan Metode K-Means Clustering. *Jurnal Fokus Elektroda: Energi Listrik, Telekomunikasi, Komputer, Elektronika Dan Kendali)*, *9*(3), 178–184.

Ulfah, M., & Irtwaty, A. S. (2022). Penerapan Data Mining Clustering Menggunakan Metode K-Means Dalam Pengelompokan Buku Perpustakaan Politeknik Negeri Balikpapan. *Fidelity: Jurnal Teknik Elektro, 4*(3), 62–68.

Wahyuntini, S., & Endarti, S. (2021). Tantangan Digital dan Dinamisasi Koleksi Dalam Pemanfaatan Koleksi Perpustakaan Bagi Prestasi Belajar Mahasiswa. *ABDI PUSTAKA: Jurnal Perpustakaan Dan Kearsipan*, *1*(1), 1–6.

Wang, S., Sun, Y., & Bao, Z. (2020). On the E iciency of K-Means Clustering: Evaluation, Optimization, and Algorithm Selection. *Proceedings of the VLDB Endowment*, *14*(2), 163–175.

Wegmann, M., LaDuke, A., Lear, M., & Henry Casesa, R. (2021). Assessing the Role of a Children's Collection in an Academic Library: A Case Study of Collaborative Collection Management. *Collection Management*, *46*(3–4), 257–272.

Werdiningsih, I., Nuqoba, B., & Muhammadun. (2020). *Data Mining Menggunakan Android, Weka, dan SPSS* (1st ed.). Airlangga University Press.

Yudisman, S. N., & Rahmi, L. (2020). Kebijakan Pengembangan Koleksi Di Perpustakaan Sekolah Tinggi Pembangunan Masyarakat Desa (STPMD) Yogyakarta. *UNILIB: Jurnal Perpustakaan*, *11*(2).

Yulia, & Silalahi, M. (2021). Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means. *Indonesian Journal of Computer Science*, *10*(1).

Yuliani, T. (2020). Analisis kebutuhan pemustaka pada kegiatan layanan pengembangan koleksi buku Perpustakaan IAIN Batusangkar. *Al-Kuttab: Jurnal Kajian Perpustakaan, Informasi Dan Kearsipan*, *2*(1), 41–52.

LAMPIRAN

Lampiran.1 Dataset awal

No	ID Anggota	Nama Anggota	Kode Eksemplar	Judul	Tanggal Pinjam	Tanggal Kembali	Status peminjaman
1	77796189	VILZA LIANA	P01006S	Dikta dan hukum	08/08/2023	22/08/2023	1
2	63301218	ALIFIANDA APRIANDI	P00961S	Bumi	11/08/2023	25/08/2023	1
3	71411237	BUNGA NATIA PUTRI DEWI	P01058S	Great expectations	11/08/2023	25/08/2023	1
4	68876615	ALIFAH JULIA RAHMA	P01007S	Dikta dan hukum	18/08/2023	01/09/2023	1
5	67341493	ALYKA RAMADHANI	P01010S	What's so wrong about your self healing	22/08/2023	05/09/2023	1
6	67341493	ALYKA RAMADHANI	P00456S	Kami (bukan) sarjana kertas	22/08/2023	05/09/2023	1
7	65427309	ANGGI DEALOVA	P01029S	Berani tidak disukai	22/08/2023	05/09/2023	1
8	52743781	KHERN AHGIAT	P01059S	Atomic habits : perubahan kecil yang memberikan hasil luar biasa	22/08/2023	05/09/2023	1
9	65308661	SABRINA ANANDITA	P01060S	Atomic habits : perubahan kecil yang memberikan hasil luar biasa	23/08/2023	06/09/2023	1
10	64971398	INNA LUTFIAH FATIH	P01129S	Penuntun penyelesaian kimia kelas X-XI-XII SMA/MA	23/08/2023	06/09/2023	1
11	68012020	ALSA ALSIFANI	P01065S	Kenali bakatmu lejitkan potensimu	24/08/2023	07/09/2023	1
12	51326711	NATASYA FITRIA RAHMADANI	P01153S	Top rank tes skolastik SNBT 2023	25/08/2023	08/09/2023	1
13	51326711	NATASYA FITRIA RAHMADANI	P00338S	Super handbook for (super) teens	25/08/2023	08/09/2023	1
14	56837028	NAJWA AWALIA	P01149S	Top spoiler tes skolastik SNBT 2023	25/08/2023	08/09/2023	1
15	61063174	NAYLA ZAKIATUZZAHRA	P00195S	Bulan	28/08/2023	11/09/2023	1
16	64754570	NABIA USWATUN HASANAH	P00909S	Ekonomi : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas XI	29/08/2023	12/09/2023	1
17	56355772	NADYA FARHATI	P00912S	Ekonomi : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas XI	29/08/2023	12/09/2023	1
18	76570666	SISKA FITRIANI	P01458S	Rich dad poor dad : apa yang diajarkan orang kaya pada anak-anak mereka tentang uang yang tidak diajarkan oleh orang miskin dan kelas menengah!	30/08/2023	13/09/2023	1
19	74444684	ANISA KHOLISHOH NURROHMAH	P01015S	Loneliness is my best friend	30/08/2023	13/09/2023	1

20	74684814	NISWATUL AZIZAH	P01461S	Mariposa	30/08/2023	13/09/2023	1
21	66491800	OKTAVIA PUTRI RAHMADHANI	P01009S	What's so wrong about your self healing	30/08/2023	13/09/2023	1
22	74444684	ANISA KHOLISHOH NURROHMAH	P01028S	Berani tidak disukai	31/08/2023	14/09/2023	1
23	65427309	ANGGI DEALOVA	P01467S	Pengantar filsafat	01/09/2023	15/09/2023	1
24	79983494	ANNISA AULIA RAHMA	P01032S	Menginstall nyali baru	04/09/2023	18/09/2023	1
25	64320341	PERMATA INDAH PRATAMA	P01062S	Atomic habits : perubahan kecil yang memberikan hasil luar biasa	04/09/2023	18/09/2023	1
26	64320341	PERMATA INDAH PRATAMA	P01155S	Top rank tes skolastik SNBT 2023	04/09/2023	18/09/2023	1
27	61063174	NAYLA ZAKIATUZZAHRA	P00037S	Komet	05/09/2023	19/09/2023	1
28	68873636	FITRI HERIYANTI	P01007S	Dikta dan hukum	06/09/2023	20/09/2023	1
29	52743781	KHERN AHGIAT	P00806S	Tak kasat mata	07/09/2023	21/09/2023	1
30	68876615	ALIFAH JULIA RAHMA	P00663S	Bumi	08/09/2023	22/09/2023	1
31	66853065	WARDAH SA'ADATUL HAFAZAH	P00034S	Negeri 5 Menara	08/09/2023	22/09/2023	1
32	74444684	ANISA KHOLISHOH N.	P01013S	Insecurity is my middle name	11/09/2023	25/09/2023	1
33	52375379	NADIA ZAHRA	P00572S	Sukseskan salat anda : menengok kembali salat Rasulullah Saw.	12/09/2023	26/09/2023	1
34	66160814	MAGDALENA KINO SAPUTRI	P01087S	Kartun Aljabar	12/09/2023	26/09/2023	1
35	63007375	NASYWA AULIA MIDAFI	P00575S	Sukseskan salat anda : menengok kembali salat Rasulullah Saw.	12/09/2023	26/09/2023	1
36	65820733	ZUNEVA DWINA KUSNADI	P00745S	Kata	12/09/2023	26/09/2023	1
37	64320341	PERMATA INDAH PRATAMA	P01151S	Top spoiler tes skolastik SNBT 2023	12/09/2023	26/09/2023	1
38	72306651	SUCI RAHMAWATI	P00709S	Senior	14/09/2023	28/09/2023	1
39	64513853	ALYSSA CAHYA DARMAWAN	P01006S	Dikta dan hukum	15/09/2023	29/09/2023	1
40	62289221	SITI NURKHALIZA	P01007S	Dikta dan hukum	18/09/2023	02/10/2023	1
41	68873636	FITRI HERIYANTI	P01014S	Insecurity is my middle name	18/09/2023	02/10/2023	1
42	68873636	FITRI HERIYANTI	P01011S	What's so wrong about your self healing	18/09/2023	02/10/2023	1
43	64513853	ALYSSA CAHYA DARMAWAN	P01022S	Berjuta rasanya	18/09/2023	02/10/2023	1

Lampiran.2 Hasil Transformasi Data

No	Judul	Jumlah Peminjaman	Jumlah Anggota	Jumlah Eksemplar
1	100 wanita yang membentuk sejarah dunia	1	1	1
2	50 rules to be wonderful teenager	6	6	5
3`	7 Prajurit Bapak	7	7	1
4	A Brief history of East Asia : sejarah lengkap Korea, Jepang, dan Tiongkok dari zaman kuno hingga era modern	2	2	2
5	Akidah dan akhlak 2 untuk kelas XI Madrasah Aliyah	1	1	1
6	Akidah dan akhlak 3 untuk kelas XII Madrasah Aliyah	17	17	9
7	Al-Qur'an dan hadis 2 untuk kelas XI Madrasah Aliyah	4	4	3
8	Al-Qur'an dan hadis 1 untuk kelas X Madrasah Aliyah	2	2	2
9	Al-Qur'an dan hadis 3 untuk kelas XII Madrasah Aliyah	5	4	4
10	Ana uhibbuka oppa!	2	2	1
11	Anak sejuta bintang	1	1	1
12	Andalkan Allah aja!	1	1	1
13	Antara aku dan mamaku	3	3	2
14	Anti salah jurusan : panduan memilih jurusan yang tepat untuk masa depan yang hebat	1	1	1
15	Atlas Indonesia dan dunia. Edisi 34 provinsi di Indonesia.	1	1	1
16	Atomic habits : perubahan kecil yang memberikan hasil luar biasa	17	17	6
17	Azzamine	6	6	2
18	Badai matahari Andalusia	1	1	1
19	Bahasa Arab 1 untuk kelas X Madrasah Aliyah	1	1	1
20	Bahasa Arab 2 untuk kelas XI Madrasah Aliyah	1	1	1
21	Bahasa Arab 3 untuk kelas XII Madrasah Aliyah	2	2	2
22	Bahasa Indonesia 1 untuk SMA/MA/SMK/MAK kelas X	1	1	1
23	Bahasa Inggris 1 untuk SMA/MA/SMK/MAK kelas X	1	1	1
24	Bahasa Inggris 2 untuk SMA/MA/SMK/MAK kelas XI	2	2	2
25	Be calm, be strong, be grateful	2	2	1

26	Be strong, be brave, be happy with Allah	1	1	1
27	Belajar sains melalui fenomena di sekitar kita	1	1	1
28	Berani tidak disukai	9	8	3
29	Berjuta rasanya	11	11	3
30	Bidadari berbisik	3	3	1
31	Bingkai kenangan yang terlupakan	3	2	1
32	Biografi empat imam mazhab : Abu Hanifah, Malik, Al-Syafi'I, Ahmad	1	1	1
33	Biografi para ilmuwan muslim	1	1	1
34	Blue heaven	1	1	1
35	Boot, poem, and a piece of cupcake	1	1	1
36	Bukan pelajar biasa	1	1	1
37	Buku pintar video editing di android	1	1	1
38	Bulan	6	6	1
39	Bumi	9	9	4
40	Cameo Revenge	1	1	1
41	Ceros dan Batozar	3	3	1
42	Cinta di ujung sajadah	9	8	5
43	Cool boy vs cool girl	7	7	1
44	Dan perang pun usai	1	1	1
45	Debat Islam vs non-Islam : argumen cerdas Zakir Naik yang membuat orang tercengang bahkan masuk Islam	1	1	1
46	Dikta dan hukum	13	12	4
47	Ekonomi 3 : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas XII	3	3	3
48	Ekonomi : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas X	1	1	1
49	Ekonomi : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas XI	2	2	2
50	Elang Menoreh : perjalanan Purwa Kala	2	2	2
51	Emotional intelligence = kecerdasan emosional	1	1	1
52	Fikih 1 untuk kelas X Madrasah Aliyah	10	10	6

53	Fikih 3 untuk kelas XII Madrasah Aliyah	33	22	25
54	Gadis-gadis misterius	2	2	1
55	Galaksi	6	5	1
56	Garis waktu	1	1	1
57	Gejolak dalam awan	3	3	3
58	Geografi 1 untuk kelas X SMA dan MA	12	12	11
59	Geografi 3 untuk SMA/MA kelas XII	2	2	2
60	Going offline : menemukan jati diri di dunia penuh distraksi	1	1	1
61	Great expectations	1	1	1
62	Halo, mantan!	8	7	1
63	Hanya ingin memilikimu	2	2	2
64	Hello	7	7	1
65	Hidayah dalam cinta : sebuah novel penggugah nurani	1	1	1
66	Hot lecturer and me	3	3	1
67	Hujan	4	4	2
68	Hujan bulan Juni	1	1	1
69	Hujan untuk pelangi	2	2	1
70	IPA & IPS	7	7	2
71	IPS : Ilmu Pengetahuan Sosial untuk SMA/MA kelas X	8	5	6
72	IPS Geografi 1	4	4	4
73	IPS Sosiologi 2 untuk SMA/MA kelas XI	1	1	1
74	IPS Sosiologi untuk SMA/MA kelas X	1	1	1
75	Ilmu hadis 2 untuk kelas XI madrasah aliyah program keagamaan	16	12	11
76	Ilmu hadis 3 untuk kelas XI madrasah aliyah program keagamaan	13	9	9
77	Ilmu pengetahuan alam untuk SMA/MA kelas X	7	5	7
78	Ilmu tafsir 2 untuk kelas XI Madrasah Aliyah program keagamaan	5	5	4
79	Ilmu tafsir 3 untuk kelas XI Madrasah Aliyah program keagamaan	9	7	7

80	Insecurity is my middle name	10	10	4
81	Jilbab traveler : love sparks in Korea	1	1	1
82	Jurai : kisah anak-anak emak di setapak impian	1	1	1
83	Kami (bukan) sarjana kertas	1	1	1
84	Kartun Aljabar	2	2	2
85	Kartun Genetika	2	2	2
86	Kartun fisika	1	1	1
87	Kartun kimia	2	2	1
88	Kata	6	5	1
89	Kenali bakatmu lejitkan potensimu	4	4	3
90	Kerudung dan Kaukasia	1	1	1
91	Ketua kelas	2	2	1
92	Khazanah peradaban Islam	1	1	1
93	Kisah murid paling bahagia dan sebelas kisah lainnya	3	3	2
94	Komet	7	6	3
95	Komet Minor	6	5	3
96	Laskar Pelangi	1	1	1
97	Laut biru Klara	4	3	2
98	Lima sekawan : dalam lorong pencoleng	1	1	1
99	Lingkar	1	1	1
100	Loneliness is my best friend	10	10	3
101	Lovasket 4 : your heart	1	1	1
102	Loving you	4	3	1
103	Lusi Lindri	1	1	1
104	Mariposa	5	5	2
105	Matahari	6	6	4
106	Matematika 1 untuk SMA/MA/SMK/MAK kelas X	1	1	1

107	Matematika 2 untuk SMA/MA/SMK/MAK kelas XI	1	1	1
108	Matematika untuk SMA/MA/SMK/MAK Kelas X	1	1	1
109	Me and my broken heart	3	3	1
110	Melodylan	1	1	1
111	Membaca emosi orang	2	2	1
112	Menginstall nyali baru	1	1	1
113	Mengkaji ilmu geografi 2 untuk kelas XI SMA dan MA	4	4	2
114	Mengkaji ilmu geografi untuk kelas XII SMA dan MA	6	5	3
115	Menjelajah dunia biologi 2 untuk kelas XI SMA dan MA	4	4	4
116	Menjelajah dunia biologi 3 untuk kelas XII SMA dan MA	5	5	4
117	Miss complicated designer	1	1	1
118	Misteri terakhir	6	3	5
119	Misteri terakhir. Buku kesatu	1	1	1
120	Moga bunda disayang Allah	4	4	2
121	Nebula	5	5	2
122	Nebula : aku akan menjadi bintangmu	8	7	3
123	Negeri 5 Menara	3	3	2
124	Negeri para bedebah	3	3	1
125	Orang-orang biasa	1	1	1
126	PJOK 1 : Pendidikan Jasmani, Olahraga, dan Kesehatan untuk SMA/MA/SMK/MAK Kelas X	4	4	4
127	Panggil aku,Salma	1	1	1
128	Paris a love journey : kisah pencarian cinta, persahabatan, dan kedamaian jiwa	6	5	4
129	Paris a miracle of love : cinta, keajaiban, dan keteguhan hati	3	2	3
130	Pendidikan Pancasila 1 untuk SMA/MA/SMK/MAK kelas X	2	2	2
131	Pengantar filsafat	1	1	1
132	Penuntun penyelesaian kimia kelas X-XI-XII SMA/MA	1	1	1
133	Perjalanan hati	1	1	1

134	Petak umpet Minako	5	5	2
135	Petualangan di Puri Rajawali	1	1	1
136	Petualangan di lembah maut	4	2	3
137	Pulang	3	3	1
138	Qishashul Anbiya' : kisah para nabi	2	2	1
139	Queen of hearts. Vol. 2 : the wonder	2	2	2
140	Rain in Paris	2	2	1
141	Raja untuk Ratu	5	5	1
142	Rasa	18	18	5
143	Reano	5	5	1
144	Rich dad poor dad : apa yang diajarkan orang kaya pada anak-anak mereka tentang uang yang tidak diajarkan oleh orang miskin dan kelas menengah!	4	4	2
145	Rindu	1	1	1
146	Rumah mati di siberia	2	2	1
147	Sahabat kok romantis?	7	7	2
148	Sahabat tanpa cinta	2	2	2
149	Segi tiga	3	3	3
150	Sejarah 2 untuk SMA/MA/SMK/MAK kelas XI	1	1	1
151	Sejarah 2 : Untuk Kelas XI SMA dan MA Program Minat Ilmu-Ilmu Sosial	3	3	3
152	Sejarah 3 : Untuk Kelas XII SMA dan MA Program Minat Ilmu-Ilmu Sosial	5	5	5
153	Sejarah Indonesia 3 untuk SMA/MA kelas XII (kelompok wajib)	1	1	1
154	Sejarah Indonesia untuk SMA/MA kelas XI (kelompok wajib)	3	3	3
155	Sejarah kebudayaan Islam 1 untuk kelas X Madrasah Aliyah dan Madrasah Aliyah Program Keagamaan	2	2	2
156	Sejarah kebudayaan Islam 3 untuk kelas XII Madrasah Aliyah dan Madrasah Aliyah Program Keagamaan	4	4	3
157	Selena	5	5	2
158	Senior	5	5	2
159	Sepatu Dahlan	2	1	1
160	Seperti hujan yang jatuh ke bumi	2	2	1

161	September wish : harapan selalu berbanding terbalik dengan kenyataan	2	2	1
162	Sherlock Holmes Series 1 : a study in Scarlet. Buku 1	2	2	1
163	Si anak badai	4	4	1
164	Si anak spesial	9	9	1
165	Simple quiz about me seri 2	3	3	1
166	Sirah Nabawiyah : sejarah lengkap kehidupan Nabi Muhammad Saw.	4	4	2
167	Song of soul : penantian bidadari kesunyian	1	1	1
168	Sosiologi 1: Kelompok Peminatan ilmu Pengetahuan Sosial SMA/MA Kelas X	1	1	1
169	Sosiologi 2: Kelompok Peminatan Ilmu Pengetahuan Sosial SMA/MA Kelas XI	1	1	1
170	Sosiologi 3: Kelompok Peminatan Ilmu Pengetahuan Sosial SMA/MA Kelas XII	1	1	1
171	Sukseskan salat anda : menengok kembali salat Rasulullah Saw.	3	3	2
172	Super handbook for (super) teens	2	2	2
173	Supernova 1 : kesatria, putri, dan bintang jatuh	1	1	1
174	Tak di Ka'bah, di Vatikan, atau di Tembok Ratapan, Tuhan ada di hatim	5	5	2
175	Tak kasat mata	2	1	1
176	Tapak jejak	1	1	1
177	Teluk Alaska	4	3	2
178	Tentang kamu	7	7	1
179	The Queen alone : perjalanan melintasi waktu	1	1	1
180	The adventures of Tom Sawyer	1	1	1
181	The best guide book of TOEFL preparation	2	2	1
182	The ultrasmart holiday : ketika liburan mengubah masa depan	2	2	2
183	Top rank tes skolastik SNBT 2023	13	9	5
184	Top spoiler tes skolastik SNBT 2023	7	4	5
185	Tulisan Sastra	12	11	1
186	Usul fikih 1 untuk kelas X Madrasah Aliyah program keagamaan	2	2	1
187	Usul fikih 2 untuk kelas XI Madrasah Aliyah program keagamaan	6	5	4

188	Usul fikih 3 untuk kelas XII Madrasah Aliyah program keagamaan	32	14	10
189	Wajah baru Perang Salib : fakta-fakta konspirasi dunia Kristen untuk menghancurkan Islam	1	1	1
190	What's so wrong about your self healing	12	12	3
191	Yang telah lama pergi	7	6	1
192	Yoo Shi-Jin?s scandal	2	2	1
193	You(th)	1	1	1

Lampiran.3 Hasil *Clustering* dan Label Keterangan

No	Judul	Jumlah Peminjaman	Jumlah Anggota	Jumlah Eksemplar	cluster	label_keterangan
1	100 wanita yang membentuk sejarah dunia	1	1	1	0	Kurang Diminati
2	50 rules to be wonderful teenager	6	6	5	0	Kurang Diminati
3	7 Prajurit Bapak	7	7	1	0	Kurang Diminati
4	A Brief history of East Asia : sejarah lengkap Korea, Jepang, dan Tiongkok dari zaman kuno hingga era modern	2	2	2	0	Kurang Diminati
5	Akidah dan akhlak 2 untuk kelas XI Madrasah Aliyah	1	1	1	0	Kurang Diminati
6	Akidah dan akhlak 3 untuk kelas XII Madrasah Aliyah	17	17	9	1	Sangat Diminati
7	Al-Qur'an dan hadis 2 untuk kelas XI Madrasah Aliyah	4	4	3	0	Kurang Diminati
8	Al-Qur'an dan hadis 1 untuk kelas X Madrasah Aliyah	2	2	2	0	Kurang Diminati
9	Al-Qur'an dan hadis 3 untuk kelas XII Madrasah Aliyah	5	4	4	0	Kurang Diminati
10	Ana uhibbuka oppa!	2	2	1	0	Kurang Diminati
11	Anak sejuta bintang	1	1	1	0	Kurang Diminati
12	Andalkan Allah aja!	1	1	1	0	Kurang Diminati
13	Antara aku dan mamaku	3	3	2	0	Kurang Diminati
14	Anti salah jurusan : panduan memilih jurusan yang tepat untuk masa depan yang hebat	1	1	1	0	Kurang Diminati
15	Atlas Indonesia dan dunia. Edisi 34 provinsi di Indonesia.	1	1	1	0	Kurang Diminati
16	Atomic habits : perubahan kecil yang memberikan hasil luar biasa	17	17	6	1	Sangat Diminati
17	Azzamine	6	6	2	0	Kurang Diminati
18	Badai matahari Andalusia	1	1	1	0	Kurang Diminati

19	Bahasa Arab 1 untuk kelas X Madrasah Aliyah	1	1	1	0	Kurang Diminati
20	Bahasa Arab 2 untuk kelas XI Madrasah Aliyah	1	1	1	0	Kurang Diminati
21	Bahasa Arab 3 untuk kelas XII Madrasah Aliyah	2	2	2	0	Kurang Diminati
22	Bahasa Indonesia 1 untuk SMA/MA/SMK/MAK kelas X	1	1	1	0	Kurang Diminati
23	Bahasa Inggris 1 untuk SMA/MA/SMK/MAK kelas X	1	1	1	0	Kurang Diminati
24	Bahasa Inggris 2 untuk SMA/MA/SMK/MAK kelas XI	2	2	2	0	Kurang Diminati
25	Be calm, be strong, be grateful	2	2	1	0	Kurang Diminati
26	Be strong, be brave, be happy with Allah	1	1	1	0	Kurang Diminati
27	Belajar sains melalui fenomena di sekitar kita	1	1	1	0	Kurang Diminati
28	Berani tidak disukai	9	8	3	1	Sangat Diminati
29	Berjuta rasanya	11	11	3	1	Sangat Diminati
30	Bidadari berbisik	3	3	1	0	Kurang Diminati
31	Bingkai kenangan yang terlupakan	3	2	1	0	Kurang Diminati
32	Biografi empat imam mazhab : Abu Hanifah, Malik, Al-Syafi'I, Ahmad	1	1	1	0	Kurang Diminati
33	Biografi para ilmuwan muslim	1	1	1	0	Kurang Diminati
34	Blue heaven	1	1	1	0	Kurang Diminati
35	Boot, poem, and a piece of cupcake	1	1	1	0	Kurang Diminati
36	Bukan pelajar biasa	1	1	1	0	Kurang Diminati
37	Buku pintar video editing di android	1	1	1	0	Kurang Diminati
38	Bulan	6	6	1	0	Kurang Diminati
39	Bumi	9	9	4	1	Sangat Diminati
40	Cameo Revenge	1	1	1	0	Kurang Diminati
41	Ceros dan Batozar	3	3	1	0	Kurang Diminati
42	Cinta di ujung sajadah	9	8	5	1	Sangat Diminati
43	Cool boy vs cool girl	7	7	1	0	Kurang Diminati
44	Dan perang pun usai	1	1	1	0	Kurang Diminati
45	Debat Islam vs non-Islam : argumen cerdas Zakir Naik yang membuat orang tercengang bahkan masuk Islam	1	1	1	0	Kurang Diminati

46	Dikta dan hukum	13	12	4	1	Sangat Diminati
47	Ekonomi 3 : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas XII	3	3	3	0	Kurang Diminati
48	Ekonomi : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas X	1	1	1	0	Kurang Diminati
49	Ekonomi : kelompok peminatan ilmu pengetahuan sosial untuk kelas SMA/MA Kelas XI	2	2	2	0	Kurang Diminati
50	Elang Menoreh : perjalanan Purwa Kala	2	2	2	0	Kurang Diminati
51	Emotional intelligence = kecerdasan emosional	1	1	1	0	Kurang Diminati
52	Fikih 1 untuk kelas X Madrasah Aliyah	10	10	6	1	Sangat Diminati
53	Fikih 3 untuk kelas XII Madrasah Aliyah	33	22	25	1	Sangat Diminati
54	Gadis-gadis misterius	2	2	1	0	Kurang Diminati
55	Galaksi	6	5	1	0	Kurang Diminati
56	Garis waktu	1	1	1	0	Kurang Diminati
57	Gejolak dalam awan	3	3	3	0	Kurang Diminati
58	Geografi 1 untuk kelas X SMA dan MA	12	12	11	1	Sangat Diminati
59	Geografi 3 untuk SMA/MA kelas XII	2	2	2	0	Kurang Diminati
60	Going offline : menemukan jati diri di dunia penuh distraksi	1	1	1	0	Kurang Diminati
61	Great expectations	1	1	1	0	Kurang Diminati
62	Halo, mantan!	8	7	1	0	Kurang Diminati
63	Hanya ingin memilikimu	2	2	2	0	Kurang Diminati
64	Hello	7	7	1	0	Kurang Diminati
65	Hidayah dalam cinta : sebuah novel penggugah nurani	1	1	1	0	Kurang Diminati
66	Hot lecturer and me	3	3	1	0	Kurang Diminati
67	Hujan	4	4	2	0	Kurang Diminati
68	Hujan bulan Juni	1	1	1	0	Kurang Diminati
69	Hujan untuk pelangi	2	2	1	0	Kurang Diminati
70	IPA & IPS	7	7	2	0	Kurang Diminati
71	IPS : Ilmu Pengetahuan Sosial untuk SMA/MA kelas X	8	5	6	0	Kurang Diminati
72	IPS Geografi 1	4	4	4	0	Kurang Diminati

73	IPS Sosiologi 2 untuk SMA/MA kelas XI	1	1	1	0	Kurang Diminati
74	IPS Sosiologi untuk SMA/MA kelas X	1	1	1	0	Kurang Diminati
75	Ilmu hadis 2 untuk kelas XI madrasah aliyah program keagamaan	16	12	11	1	Sangat Diminati
76	Ilmu hadis 3 untuk kelas XI madrasah aliyah program keagamaan	13	9	9	1	Sangat Diminati
77	Ilmu pengetahuan alam untuk SMA/MA kelas X	7	5	7	0	Kurang Diminati
78	Ilmu tafsir 2 untuk kelas XI Madrasah Aliyah program keagamaan	5	5	4	0	Kurang Diminati
79	Ilmu tafsir 3 untuk kelas XI Madrasah Aliyah program keagamaan	9	7	7	1	Sangat Diminati
80	Insecurity is my middle name	10	10	4	1	Sangat Diminati
81	Jilbab traveler : love sparks in Korea	1	1	1	0	Kurang Diminati
82	Jurai : kisah anak-anak emak di setapak impian	1	1	1	0	Kurang Diminati
83	Kami (bukan) sarjana kertas	1	1	1	0	Kurang Diminati
84	Kartun Aljabar	2	2	2	0	Kurang Diminati
85	Kartun Genetika	2	2	2	0	Kurang Diminati
86	Kartun fisika	1	1	1	0	Kurang Diminati
87	Kartun kimia	2	2	1	0	Kurang Diminati
88	Kata	6	5	1	0	Kurang Diminati
89	Kenali bakatmu lejitkan potensimu	4	4	3	0	Kurang Diminati
90	Kerudung dan Kaukasia	1	1	1	0	Kurang Diminati
91	Ketua kelas	2	2	1	0	Kurang Diminati
92	Khazanah peradaban Islam	1	1	1	0	Kurang Diminati
93	Kisah murid paling bahagia dan sebelas kisah lainnya	3	3	2	0	Kurang Diminati
94	Komet	7	6	3	0	Kurang Diminati
95	Komet Minor	6	5	3	0	Kurang Diminati
96	Laskar Pelangi	1	1	1	0	Kurang Diminati
97	Laut biru Klara	4	3	2	0	Kurang Diminati
98	Lima sekawan : dalam lorong pencoleng	1	1	1	0	Kurang Diminati
99	Lingkar	1	1	1	0	Kurang Diminati

100	Loneliness is my best friend	10	10	3	1	Sangat Diminati
101	Lovasket 4 : your heart	1	1	1	0	Kurang Diminati
102	Loving you	4	3	1	0	Kurang Diminati
103	Lusi Lindri	1	1	1	0	Kurang Diminati
104	Mariposa	5	5	2	0	Kurang Diminati
105	Matahari	6	6	4	0	Kurang Diminati
106	Matematika 1 untuk SMA/MA/SMK/MAK kelas X	1	1	1	0	Kurang Diminati
107	Matematika 2 untuk SMA/MA/SMK/MAK kelas XI	1	1	1	0	Kurang Diminati
108	Matematika untuk SMA/MA/SMK/MAK Kelas X	1	1	1	0	Kurang Diminati
109	Me and my broken heart	3	3	1	0	Kurang Diminati
110	Melodylan	1	1	1	0	Kurang Diminati
111	Membaca emosi orang	2	2	1	0	Kurang Diminati
112	Menginstall nyali baru	1	1	1	0	Kurang Diminati
113	Mengkaji ilmu geografi 2 untuk kelas XI SMA dan MA	4	4	2	0	Kurang Diminati
114	Mengkaji ilmu geografi untuk kelas XII SMA dan MA	6	5	3	0	Kurang Diminati
115	Menjelajah dunia biologi 2 untuk kelas XI SMA dan MA	4	4	4	0	Kurang Diminati
116	Menjelajah dunia biologi 3 untuk kelas XII SMA dan MA	5	5	4	0	Kurang Diminati
117	Miss complicated designer	1	1	1	0	Kurang Diminati
118	Misteri terakhir	6	3	5	0	Kurang Diminati
119	Misteri terakhir. Buku kesatu	1	1	1	0	Kurang Diminati
120	Moga bunda disayang Allah	4	4	2	0	Kurang Diminati
121	Nebula	5	5	2	0	Kurang Diminati
122	Nebula : aku akan menjadi bintangmu	8	7	3	0	Kurang Diminati
123	Negeri 5 Menara	3	3	2	0	Kurang Diminati
124	Negeri para bedebah	3	3	1	0	Kurang Diminati
125	Orang-orang biasa	1	1	1	0	Kurang Diminati
126	PJOK 1 : Pendidikan Jasmani, Olahraga, dan Kesehatan untuk SMA/MA/SMK/MAK Kelas X	4	4	4	0	Kurang Diminati

127	Panggil aku,Salma	1	1	1	0	Kurang Diminati
128	Paris a love journey : kisah pencarian cinta, persahabatan, dan kedamaian jiwa	6	5	4	0	Kurang Diminati
129	Paris a miracle of love : cinta, keajaiban, dan keteguhan hati	3	2	3	0	Kurang Diminati
130	Pendidikan Pancasila 1 untuk SMA/MA/SMK/MAK kelas X	2	2	2	0	Kurang Diminati
131	Pengantar filsafat	1	1	1	0	Kurang Diminati
132	Penuntun penyelesaian kimia kelas X-XI-XII SMA/MA	1	1	1	0	Kurang Diminati
133	Perjalanan hati	1	1	1	0	Kurang Diminati
134	Petak umpet Minako	5	5	2	0	Kurang Diminati
135	Petualangan di Puri Rajawali	1	1	1	0	Kurang Diminati
136	Petualangan di lembah maut	4	2	3	0	Kurang Diminati
137	Pulang	3	3	1	0	Kurang Diminati
138	Qishashul Anbiya' : kisah para nabi	2	2	1	0	Kurang Diminati
139	Queen of hearts. Vol. 2 : the wonder	2	2	2	0	Kurang Diminati
140	Rain in Paris	2	2	1	0	Kurang Diminati
141	Raja untuk Ratu	5	5	1	0	Kurang Diminati
142	Rasa	18	18	5	1	Sangat Diminati
143	Reano	5	5	1	0	Kurang Diminati
144	Rich dad poor dad : apa yang diajarkan orang kaya pada anak-anak mereka tentang uang yang tidak diajarkan oleh orang miskin dan kelas menengah!	4	4	2	0	Kurang Diminati
145	Rindu	1	1	1	0	Kurang Diminati
146	Rumah mati di siberia	2	2	1	0	Kurang Diminati
147	Sahabat kok romantis?	7	7	2	0	Kurang Diminati
148	Sahabat tanpa cinta	2	2	2	0	Kurang Diminati
149	Segi tiga	3	3	3	0	Kurang Diminati
150	Sejarah 2 untuk SMA/MA/SMK/MAK kelas XI	1	1	1	0	Kurang Diminati
151	Sejarah 2 : Untuk Kelas XI SMA dan MA Program Minat Ilmu-Ilmu Sosial	3	3	3	0	Kurang Diminati
152	Sejarah 3 : Untuk Kelas XII SMA dan MA Program Minat Ilmu-Ilmu Sosial	5	5	5	0	Kurang Diminati
153	Sejarah Indonesia 3 untuk SMA/MA kelas XII (kelompok wajib)	1	1	1	0	Kurang Diminati

154	Sejarah Indonesia untuk SMA/MA kelas XI (kelompok wajib)	3	3	3	0	Kurang Diminati
155	Sejarah kebudayaan Islam 1 untuk kelas X Madrasah Aliyah dan Madrasah Aliyah Program Keagamaan	2	2	2	0	Kurang Diminati
156	Sejarah kebudayaan Islam 3 untuk kelas XII Madrasah Aliyah dan Madrasah Aliyah Program Keagamaan	4	4	3	0	Kurang Diminati
157	Selena	5	5	2	0	Kurang Diminati
158	Senior	5	5	2	0	Kurang Diminati
159	Sepatu Dahlan	2	1	1	0	Kurang Diminati
160	Seperti hujan yang jatuh ke bumi	2	2	1	0	Kurang Diminati
161	September wish : harapan selalu berbanding terbalik dengan kenyataan	2	2	1	0	Kurang Diminati
162	Sherlock Holmes Series 1 : a study in Scarlet. Buku 1	2	2	1	0	Kurang Diminati
163	Si anak badai	4	4	1	0	Kurang Diminati
164	Si anak spesial	9	9	1	1	Sangat Diminati
165	Simple quiz about me seri 2	3	3	1	0	Kurang Diminati
166	Sirah Nabawiyah : sejarah lengkap kehidupan Nabi Muhammad Saw.	4	4	2	0	Kurang Diminati
167	Song of soul : penantian bidadari kesunyian	1	1	1	0	Kurang Diminati
168	Sosiologi 1: Kelompok Peminatan ilmu Pengetahuan Sosial SMA/MA Kelas X	1	1	1	0	Kurang Diminati
169	Sosiologi 2: Kelompok Peminatan Ilmu Pengetahuan Sosial SMA/MA Kelas XI	1	1	1	0	Kurang Diminati
170	Sosiologi 3: Kelompok Peminatan Ilmu Pengetahuan Sosial SMA/MA Kelas XII	1	1	1	0	Kurang Diminati
171	Sukseskan salat anda : menengok kembali salat Rasulullah Saw.	3	3	2	0	Kurang Diminati
172	Super handbook for (super) teens	2	2	2	0	Kurang Diminati
173	Supernova 1 : kesatria, putri, dan bintang jatuh	1	1	1	0	Kurang Diminati
174	Tak di Ka'bah, di Vatikan, atau di Tembok Ratapan, Tuhan ada di hatim	5	5	2	0	Kurang Diminati
175	Tak kasat mata	2	1	1	0	Kurang Diminati
176	Tapak jejak	1	1	1	0	Kurang Diminati
177	Teluk Alaska	4	3	2	0	Kurang Diminati
178	Tentang kamu	7	7	1	0	Kurang Diminati
179	The Queen alone : perjalanan melintasi waktu	1	1	1	0	Kurang Diminati
180	The adventures of Tom Sawyer	1	1	1	0	Kurang Diminati

181	The best guide book of TOEFL preparation	2	2	1	0	Kurang Diminati
182	The ultrasmart holiday : ketika liburan mengubah masa depan	2	2	2	0	Kurang Diminati
183	Top rank tes skolastik SNBT 2023	13	9	5	1	Sangat Diminati
184	Top spoiler tes skolastik SNBT 2023	7	4	5	0	Kurang Diminati
185	Tulisan Sastra	12	11	1	1	Sangat Diminati
186	Usul fikih 1 untuk kelas X Madrasah Aliyah program keagamaan	2	2	1	0	Kurang Diminati
187	Usul fikih 2 untuk kelas XI Madrasah Aliyah program keagamaan	6	5	4	0	Kurang Diminati
188	Usul fikih 3 untuk kelas XII Madrasah Aliyah program keagamaan	32	14	10	1	Sangat Diminati
189	Wajah baru Perang Salib : fakta-fakta konspirasi dunia Kristen untuk menghancurkan Islam	1	1	1	0	Kurang Diminati
190	What's so wrong about your self healing	12	12	3	1	Sangat Diminati
191	Yang telah lama pergi	7	6	1	0	Kurang Diminati
192	Yoo Shi-Jin?s scandal	2	2	1	0	Kurang Diminati
193	You(th)	1	1	1	0	Kurang Diminati

Lampiran.4 Hasil *Score Sillhoute Coffecient*

Silhouette Coefficient untuk **k**=2: 0.7217

Lampiran.5 Surat Izin Penelitian

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jalan Gajayana 50 Malang 65144 Telepon/Faksimile (0341) 558933
Website: http://saintek.uin-malang.ac.id, email: saintek@uin-malang.ac.id

Nomor : B-179.O/FST.01/TL.00/12/2024
Lampiran : -
Hal : Permohonan Penelitian

Yth. Pimpinan MAN 2 KOTA TANGERANG
Jl. Panglima Polim No.6, RT.005/RW.003, Poris Plawad Utara, Kec. Cipondoh, Kota Tangerang, Banten 15141

Dengan hormat,
Sehubungan dengan penelitian mahasiswa Jurusan Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang atas nama:

Nama	: Achmad Ghiffari Fatullah
NIM	: 210607110003
Judul Penelitian	: PENGELOMPOKAN BUKU MENGGUNAKAN METODE ALGORITMA K-MEANS CLUSTERING UNTUK PENGADAAN BUKU PADA PERPUSTAKAAN MAN 2 KOTA TANGERANG
Dosen Pembimbing	: FAKHRIS KHUSNU REZA MAHFUD,M.Kom

Maka kami mohon Bapak/Ibu berkenan memberikan izin pada mahasiswa tersebut untuk melakukan penelitian di MAN 2 KOTA TANGERANG dengan waktu pelaksanaan pada tanggal 23 Desember 2024 sampai dengan 07 Januari 2025.

Malang, 24 Desember 2024
a.n Dekan

Scan QRCode ini

Untuk verifikasi keaslian surat

Wakil Dekan Bidang Akademik,

Dr. Anton Prasetyo, M.Si
NIP. 19770925 200604 1 003

\

Lampiran.6 Surat Izin Pengambilan Data

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jalan Gajayana 50 Malang 65144 Telepon/Faksimile (0341) 558933
Website: http://saintek.uin-malang.ac.id, email: saintek@uin-malang.ac.id

Nomor : B-95.O/FST.01/TL.00/12/2024
Lampiran : -
Hal : Permohonan Data

Yth. Pimpinan MAN 2 KOTA TANGERANG
Jl. Panglima Polim No.6, RT.005/RW.003, Poris Plawad Utara, Kec. Cipondoh, Kota Tangerang, Banten 15141

Dengan hormat,
Sehubungan dengan penelitian mahasiswa Jurusan Perpustakaan dan Sains Informasi Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang atas nama:

Nama : Achmad Ghiffari Fatullah
NIM : 210607110003
Judul : PENGELOMPOKAN BUKU MENGGUNAKAN METODE ALGORITMA K-MEANS CLUSTERING UNTUK PENGADAAN BUKU PADA PERPUSTAKAAN MAN 2 KOTA TANGERANG
Dosen Pembimbing : FAKHRIS KHUSNU REZA MAHFUD,M.Kom

Maka kami mohon Bapak/Ibu berkenan memberikan izin pada mahasiswa tersebut untuk melakukan penelitian dan mendapatkan data Data Peminjaman Koleksi di MAN 2 KOTA TANGERANG dengan waktu pelaksanaan pada tanggal 07 Januari 2025.

Demikian permohonan ini, atas perhatian dan kerjasamanya disampaikan terimakasih.

Malang, 16 Mei 2025

Wakil Dekan Bidang Akademik,

Dr. Anton Prasetyo, M.Si
NIP. 19770925 200604 1 003

Scan QRCode ini

Untuk verifikasi keaslian surat

Lampiran.7 Surat Balasan Penelitian dan Pengambilan Data

KEMENTERIAN AGAMA REPUBLIK INDONESIA
KANTOR KEMENTERIAN AGAMA KOTA TANGERANG
MADRASAH ALIYAH NEGERI 2
Jalan Panglima Polim Nomor 6 Poris Plawad Utara Kec. Cipondoh Kota Tangerang Kode Pos : 15141
NSM : 131136710002, NPSN : 20623289 Telp/Fax : 02155754525 Website : www.man2kotatangerang.sch.id

Nomor : 403 /Ma.28.05.03.02/PP.00.10/05/2025 26 Mei 2025
Lampiran : -
Hal : Izin Penelitian

Yth. Dekan Fakultas Sains dan Teknologi
UIN Maulana Malik Ibrahim Malang
Di
Malang

Assalamu'alaikum Wr. Wb.

Berdasarkan surat dari Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang Nomor : B-176.O/FST.01/TL.00/12/2024, tanggal 24 Desember 2024, Perihal Izin Penelitian, maka Kepala Madrasah Aliyah Negeri (MAN) 2 Kota Tangerang, menyampaikan hal-hal sebagai berikut :

1. Pada prinsipnya kami mengizinkan Mahasiswa yang bernama Achmad Ghiffari Fatullah, NIM : 210607110003, Jurusan Perpustakaan dan Ilmu Informasi, Semester VIII (Delapan), untuk mengadakan Penelitian dan Pengumpulan data pada MAN 2 Kota Tangerang dalam rangka penyusunan Tugas Akhir/Skripsi dengan ketentuan :
   a. sebatas bahan untuk memenuhi syarat akademis penulisan atau penyelesaian Studi Strata Satu dan tidak mengganggu kegiatan dan proses belajar dan mengajar;
   b. Metode Penelitian disesuaikan dengan sistem belajar online dan tatap muka yang sedang berlaku pada saat penelitian;
2. Waktu penelitian dan pengumpulan data dilakukan terhitung sejak tanggal 23 Desember 2024 s.d. 07 Januari 2025;
3. Agar 1 (satu) salinan/rangkap skripsi yang telah selesai wajib diserahkan pada MAN 2 Kota Tangerang sebagai bukti penelitian dan bahan pembelajaran untuk dapat dimanfaatkan sesuai dengan kebutuhan;

Demikian Izin Penelitian ini disampaikan untuk dapat dipergunakan sebagaimana mestinya.

Wassalamu'alaikum Wr. Wb.

Kepala,

ALI SUTRISNO

Lampiran.8 Surat Keterangan Selesai Melakukan Penelitian dan Pengambilan Data

KEMENTERIAN AGAMA REPUBLIK INDONESIA
KANTOR KEMENTERIAN AGAMA KOTA TANGERANG
MADRASAH ALIYAH NEGERI 2
Jalan Panglima Polim Nomor 6 Poris Plawad Utara Kec. Cipondoh Kota Tangerang Kode Pos : 15141
NSM : 131136710002, NPSN : 20623289 Telp/Fax : 02155754525 Website :
www.man2kotatangerang.sch.id

SURAT KETERANGAN
Nomor : 904 /Ma.28.05.03.02/PP.00.6/05/2025

Berdasarkan surat Universitas Islam Negeri Maulana Malik Ibrahim Malang Fakultas Sains dan Teknologi Nomor : B-179.O/FST.01/TL.00/12/2024 tanggal 24 Desember 2024, perihal Permohonan Izin Penelitian, dengan ini menerangkan bahwa :

Nama	: **Achmad Ghiffari Fatullah**
NIM	: 210607110003
Program Studi	: Perpustakaan dan Ilmu Informasi
Semester	: VIII (Delapan)

telah selesai mengadakan Penelitian dan Pengambilan Data di MAN 2 Kota Tangerang yang dilaksanakan pada tanggal 23 Desember 2024 s.d. tanggal 07 Januari 2025, dalam rangka kebutuhan siding skripsi dengan Tema "*Pengelompokan Buku Menggunakan Metode Algoritma K- Means Clustering untuk Pengadaan Buku pada Perpustakaan MAN 2 Kota Tangerang*"

Demikian Surat ini dibuat, agar dapat dipergunakan sebagaimana mestinya.

Tangerang, 26 Mei 2025
Kepala,

ALI SUTRISNO

Lampiran.9 *Screenshot* Tahap *Pre-Processing* Menggunakan *Python*

```python
import pandas as pd

# Membaca data dari file CSV
df = pd.read_csv('loan_history.csv', sep=';')

# 1. Pembersihan data
# Menghapus baris dengan nilai kosong di kolom penting
df_clean = df.dropna(subset=['Judul', 'Nama Anggota', 'Kode Eksemplar'])

# 2. Penghapusan duplikat berdasarkan judul, nama anggota, dan kode eksemplar
df_clean = df_clean.drop_duplicates(subset=['Judul', 'Nama Anggota', 'Kode Eksemplar'])

# 3. Penyederhanaan data
# a. Jumlah peminjaman per judul buku
peminjaman_per_judul = df_clean.groupby('Judul').size().reset_index(name='Jumlah Peminjaman')

# b. Jumlah anggota unik per judul
anggota_per_judul = df_clean.groupby('Judul')['Nama Anggota'].nunique().reset_index(name='Jumlah Anggota')

# c. Jumlah eksemplar unik per judul
eksemplar_per_judul = df_clean.groupby('Judul')['Kode Eksemplar'].nunique().reset_index(name='Jumlah Eksemplar')

# Menggabungkan ketiga tabel
final_table = peminjaman_per_judul.merge(
    anggota_per_judul, on='Judul'
).merge(
    eksemplar_per_judul, on='Judul'
)

# Menampilkan hasil
print(final_table)
```

```
                                          Judul  Jumlah Peminjaman
0             100 wanita yang membentuk sejarah dunia                  1
1                  50 rules to be wonderful teenager                   6
2                                  7 Prajurit Bapak                    7
3    A Brief history of East Asia : sejarah lengkap...                 2
4    Akidah dan akhlak 2 untuk kelas XI Madrasah Al...                 1
..                                                 ...               ...
188  Wajah baru Perang Salib : fakta-fakta konspira...                 1
189           What's so wrong about your self healing                 12
190                            Yang telah lama pergi                   6
191                            Yoo Shi-Jin?s scandal                   2
192                                        You(th)                     1

     Jumlah Anggota  Jumlah Eksemplar
0                 1                 1
1                 6                 5
2                 7                 1
3                 2                 2
4                 1                 1
..              ...               ...
188               1                 1
189              12                 3
190               6                 1
191               2                 1
192               1                 1

[193 rows x 4 columns]
```

```python
# Simpan ke CSV dan Excel
final_table.to_csv('Hasil_Transformasi_Data_Revisi_1.csv', index=False)
final_table.to_excel('Hasil_Transformasi_Data_Revisi_1.xlsx', index=False)
```

Lampiran.10 Screenshot Tahap Penentuan Jumlah *Cluster* Dengan *Silhouette Coefficient* Menggunakan *Python*

```python
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt

# Membaca dataset
df = pd.read_csv('Hasil_transformasi_data.csv')

# Pilih hanya kolom numerik selain 'Judul'
fitur = ['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar']
df_fitur = df[fitur].fillna(df[fitur].mean())

# Standarisasi fitur
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df_fitur)

# Range jumlah cluster yang akan diuji
range_n_clusters = range(2, 11)
silhouette_scores = []

for k in range_n_clusters:
    kmeans = KMeans(n_clusters=k, random_state=42)
    cluster_labels = kmeans.fit_predict(X_scaled)
    score = silhouette_score(X_scaled, cluster_labels)
    silhouette_scores.append(score)
    print(f"Untuk k = {k}, Silhouette Coefficient = {score:.4f}")

# Plot hasil silhouette score
plt.figure(figsize=(8,5))
plt.plot(list(range_n_clusters), silhouette_scores, marker='o')
plt.title('Silhouette Coefficient untuk Berbagai Nilai k')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Silhouette Coefficient')
plt.grid(True)
plt.show()
```

```
Untuk k = 2, Silhouette Coefficient = 0.7217
Untuk k = 3, Silhouette Coefficient = 0.6051
Untuk k = 4, Silhouette Coefficient = 0.5793
Untuk k = 5, Silhouette Coefficient = 0.5851
Untuk k = 6, Silhouette Coefficient = 0.5861
Untuk k = 7, Silhouette Coefficient = 0.5923
Untuk k = 8, Silhouette Coefficient = 0.5492
Untuk k = 9, Silhouette Coefficient = 0.5562
Untuk k = 10, Silhouette Coefficient = 0.5605
```

Lampiran.11 Screenshot Tahap *Clustering* Menggunakan *Python*

```python
import pandas as pd
from sklearn.cluster import KMeans

# Membaca dataset
df = pd.read_csv('Hasil_transformasi_data.csv')

# Pilih hanya kolom numerik selain 'Judul'
fitur = ['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar']
df_fitur = df[fitur].fillna(df[fitur].mean())

# Tentukan centroid awal secara manual
initial_centroids = [[10, 5, 2], [30, 15, 4]]

# Jalankan K-Means dengan k=2 dan centroid manual
k = 2
kmeans = KMeans(n_clusters=k, init=initial_centroids, n_init=1, random_state=42)
kmeans.fit(df_fitur)

# Ambil label cluster dan centroid
labels = kmeans.labels_
centroids = kmeans.cluster_centers_

# Menambahkan hasil cluster ke dataframe asli
df['cluster'] = labels

# Melihat hasil cluster
print(df[['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar', 'cluster']].head(200))

# Simpan hasil clustering ke file csv
df.to_csv('hasil_clustering_Revisi_II.csv', index=False)
```

```
     Jumlah Peminjaman  Jumlah Anggota  Jumlah Eksemplar  cluster
0                    1               1                 1        0
1                    6               6                 5        0
2                    7               7                 1        0
3                    2               2                 2        0
4                    1               1                 1        0
..                 ...             ...               ...      ...
188                  1               1                 1        0
189                 12              12                 3        1
190                  7               6                 1        0
191                  2               2                 1        0
192                  1               1                 1        0

[193 rows x 4 columns]
```

Lampiran.12 *Screenshot* Tahap Melihat Plot Hasil *Clustering* Menggunakan *Python*

```python
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Membaca dataset
df = pd.read_csv('hasil_clustering_Revisi_II.csv')

# Pilih dua fitur untuk visualisasi
fitur = ['Jumlah Peminjaman', 'Jumlah Anggota']
df_fitur = df[fitur].fillna(df[fitur].mean())

# Standarisasi fitur
scaler = StandardScaler()
X_scaled = scaler.fit_transform(df_fitur)

# KMeans clustering dengan k=2
k = 2
kmeans = KMeans(n_clusters=k, random_state=42)
kmeans.fit(X_scaled)
df['cluster'] = kmeans.labels_

# Plot hasil cluster
plt.figure(figsize=(8,6))
for i in range(k):
    plt.scatter(
        X_scaled[df['cluster']==i, 0],
        X_scaled[df['cluster']==i, 1],
        label=f'Cluster {i}'
    )
```

```python
# Plot centroid
plt.scatter(
    kmeans.cluster_centers_[:, 0],
    kmeans.cluster_centers_[:, 1],
    s=200, c='black', marker='X', label='Centroid'
)

plt.xlabel('Jumlah Peminjaman')
plt.ylabel('Jumlah Anggota')
plt.title('Plot Hasil Clustering KMeans dengan k=2')
plt.legend()
plt.show()
```

Lampiran.13 *Screenshot* Tahap Memberikan Rata-Rata Per-*Cluster* Menggunakan *Python*

```python
# Menghitung rata-rata tiap fitur per cluster
rata_rata_per_cluster = df.groupby('cluster')[['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar']].mean()

print("Rata-rata Jumlah Peminjaman, Jumlah Anggota, dan Jumlah Eksemplar per cluster:")
print(rata_rata_per_cluster)

# Mencari cluster dengan nilai rata-rata tertinggi untuk setiap fitur
cluster_terbanyak_peminjaman = rata_rata_per_cluster['Jumlah Peminjaman'].idxmax()
cluster_terbanyak_anggota = rata_rata_per_cluster['Jumlah Anggota'].idxmax()
cluster_terbanyak_eksemplar = rata_rata_per_cluster['Jumlah Eksemplar'].idxmax()

print(f"\nCluster dengan jumlah peminjaman terbanyak: Cluster {cluster_terbanyak_peminjaman}")
print(f"Cluster dengan jumlah anggota terbanyak: Cluster {cluster_terbanyak_anggota}")
print(f"Cluster dengan jumlah eksemplar terbanyak: Cluster {cluster_terbanyak_eksemplar}")
```

```
Rata-rata Jumlah Peminjaman, Jumlah Anggota, dan Jumlah Eksemplar per cluster:
         Jumlah Peminjaman  Jumlah Anggota  Jumlah Eksemplar
cluster
0                 2.755814        2.575581          1.686047
1                14.000000       11.761905          6.428571

Cluster dengan jumlah peminjaman terbanyak: Cluster 1
Cluster dengan jumlah anggota terbanyak: Cluster 1
Cluster dengan jumlah eksemplar terbanyak: Cluster 1
```

Lampiran.14 Screenshot Tahap Melihat Grafik Rata-Rata Per-Cluster Menggunakan *Python*

```python
import matplotlib.pyplot as plt

# Misal df sudah memiliki kolom 'cluster' dan fitur numerik
rata_rata_per_cluster = df.groupby('cluster')[['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar']].mean()

# Membuat plot bar chart
rata_rata_per_cluster.plot(kind='bar', figsize=(10,6))

plt.title('Rata-rata Jumlah Peminjaman, Jumlah Anggota, dan Jumlah Eksemplar per Cluster')
plt.xlabel('Cluster')
plt.ylabel('Rata-rata Nilai')
plt.xticks(rotation=0)
plt.legend(title='Fitur')
plt.grid(axis='y')

plt.show()
```

Lampiran.15 Screenshot Tahap Memberikan Label Keterangan Menggunakan *Python*

```python
# Hitung rata-rata fitur per cluster
rata_rata_per_cluster = df.groupby('cluster')[['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar']].mean()

# Hitung total rata-rata ketiga fitur untuk setiap cluster sebagai ukuran minat
rata_rata_per_cluster['total_rata_rata'] = rata_rata_per_cluster.sum(axis=1)

# Tentukan threshold, misal median total rata-rata untuk membedakan diminati dan kurang diminati
threshold = rata_rata_per_cluster['total_rata_rata'].median()

# Buat mapping label berdasarkan threshold
def label_cluster(total):
    if total >= threshold:
        return 'Sangat Diminati'
    else:
        return 'Kurang Diminati'

# Buat dictionary mapping cluster ke label
label_mapping = {cluster: label_cluster(total) for cluster, total in rata_rata_per_cluster['total_rata_rata'].items()}

# Tambahkan kolom label keterangan ke dataframe asli
df['label_keterangan'] = df['cluster'].map(label_mapping)

# Tampilkan hasil
print(rata_rata_per_cluster)
print("\nMapping cluster ke label keterangan:")
print(label_mapping)
print("\nContoh data dengan label keterangan:")
print(df[['cluster', 'label_keterangan']].head(200))

# Simpan hasil Label Keterangan ke file Excel
df.to_excel('Label_Keterangan_Revisi_II.xlsx', index=False)
```

```
         Jumlah Peminjaman  Jumlah Anggota  Jumlah Eksemplar  total_rata_rata
cluster
0                 2.755814        2.575581          1.686047         7.017442
1                14.000000       11.761905          6.428571        32.190476

Mapping cluster ke label keterangan:
{0: 'Kurang Diminati', 1: 'Sangat Diminati'}

Contoh data dengan label keterangan:
     cluster label_keterangan
0          0  Kurang Diminati
1          0  Kurang Diminati
2          0  Kurang Diminati
3          0  Kurang Diminati
4          0  Kurang Diminati
..       ...              ...
188        0  Kurang Diminati
189        1  Sangat Diminati
190        0  Kurang Diminati
191        0  Kurang Diminati
192        0  Kurang Diminati

[193 rows x 2 columns]
```

Lampiran.16 Screenshot Tahap Melihat Grafik Label Keterangan Menggunakan *Python*

```python
import pandas as pd
import matplotlib.pyplot as plt

# Membaca data dari file Excel
df = pd.read_excel('Label_Keterangan_Revisi_II.xlsx')

# Membuat tabel jumlah data per cluster dan label_keterangan
summary = df.groupby(['cluster', 'label_keterangan']).size().reset_index(name='Jumlah Data')

# Plot grafik batang
plt.figure(figsize=(8,6))
colors = ['#ff9999','#66b3ff']  # warna untuk tiap label_keterangan

for idx, cluster in enumerate(sorted(df['cluster'].unique())):
    data = summary[summary['cluster'] == cluster]
    plt.bar(
        data['label_keterangan'] + f' (Cluster {cluster})',
        data['Jumlah Data'],
        color=colors[idx % len(colors)],
        label=f'Cluster {cluster} - {data["label_keterangan"].values[0]}'
    )

plt.title('Jumlah Data per Label Keterangan dan Cluster')
plt.xlabel('Label Keterangan (Cluster)')
plt.ylabel('Jumlah Data')
plt.legend()
plt.tight_layout()
plt.show()
```

Lampiran.17 Screenshot Tahap Mendapatkan Nilai *Score Sillhoute Cofficient* Menggunakan *Python*

```python
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_samples, silhouette_score
import matplotlib.pyplot as plt
import numpy as np

# Membaca dataset
df = pd.read_csv('hasil_clustering_Revisi_II.csv')

# Pilih fitur numerik
fitur = ['Jumlah Peminjaman', 'Jumlah Anggota', 'Jumlah Eksemplar']
X = df[fitur].fillna(df[fitur].mean())

# Standarisasi fitur
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# KMeans dengan k=2
k = 2
kmeans = KMeans(n_clusters=k, random_state=42)
labels = kmeans.fit_predict(X_scaled)

# Hitung silhouette score dan nilai silhouette tiap sample
silhouette_avg = silhouette_score(X_scaled, labels)
sample_silhouette_values = silhouette_samples(X_scaled, labels)

print(f"Silhouette Coefficient rata-rata untuk k=2: {silhouette_avg:.4f}")
```

```
Silhouette Coefficient rata-rata untuk k=2: 0.7217
```

Lampiran. 18 Screenshot Hasil Turnitin

Skripsi_Komplit_Saya_Revisi_18.06.2025_Versi IV_FIXS_Bismillah.docx

ORIGINALITY REPORT

22%	20%	12%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	etheses.uin-malang.ac.id Internet Source	1%
2	123dok.com Internet Source	1%
3	ejournal.unesa.ac.id Internet Source	1%
4	doczz.fr Internet Source	1%
5	fidelity.nusaputra.ac.id Internet Source	1%
6	media.neliti.com Internet Source	1%
7	repository.its.ac.id Internet Source	<1%
8	repository.uin-suska.ac.id Internet Source	<1%
9	repository.unimus.ac.id Internet Source	<1%
10	repository.ub.ac.id Internet Source	<1%
11	Submitted to Konsorsium Perguruan Tinggi Swasta Indonesia II Student Paper	<1%
12	repositori.unsil.ac.id Internet Source	