

PENERAPAN ALGORITMA *HYBRID K-MEANS* DAN *FUZZY C-MEANS* UNTUK MENGANALISIS POLA MINAT BACA DI PERPUSTAKAAN PROVINSI JAWA TENGAH

LAPORAN TUGAS AKHIR

Laporan Ini Disusun Untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Strata 1 (S1) Pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



DISUSUN OLEH:

DWIKY CHANDRA MULYO UTOMO

32602200058

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG**

2026

FINAL PROJECT

***APPLICATION OF HYBRID K-MEANS AND FUZZY C-MEANS
ALGORITHMS TO ANALYZE READING INTEREST PATTERNS
IN CENTRAL JAVA PROVINCIAL LIBRARIES***

*Proposed to complete the requirement to obtain a bachelor's degree
(S-1) at Informatics Engineering Departement of Industrial
Technology Faculty Sultan Agung Islamic University*



DWIKY CHANDRA MULYO UTOMO

32602200058

**MAJORING OF INFORMATICS ENGINEERING
INDUSTRIAL TECHNOLOGY FACULTY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG**

2026

**LEMBAR PENGESAHAN
LAPORAN TUGAS AKHIR**

**PENERAPAN ALGORITMA *HYBRID K-MEANS* DAN *FUZZY C-MEANS*
UNTUK MENGANALISIS POLA MINAT BACA DI PERPUSTAKAAN
PROVINSI JAWA TENGAH**

**DWIKY CHANDRA MULYO UTOMO
32602200058**

Telah dipertahankan di depan tim penguji laporan tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 26 Februari 2026

TIM PENGUJI SEMINAR TUGAS AKHIR

| | | |
|--|----------------|--------------------------------|
| <u>Dedy Kurniadi, ST, M.Kom</u> NIK. 210615048 (Ketua Penguji) | (tanda tangan) | (tanggal) <u>26-02-2026</u> |
| <u>Sam Farisa CH, ST, M.Kom</u> NIK. 210615046 (Anggota Penguji) | (tanda tangan) | (tanggal) <u>26-02-2026</u> |
| <u>Andi Riansyah, ST, M.Kom</u> NIK. 210616053 (Pembimbing) | (tanda tangan) | (tanggal) <u>26-02-2026</u> |

Semarang, 26 Februari 2026

Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch. Taufik, ST., MIT
NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Dwiky Chandra Mulyo Utomo
NIM : 32602200058
Judul Tugas Akhir : Penerapan Algoritma Hybrid K-Means dan Fuzzy C-Means untuk Menganalisis Pola Minat Baca di Perpustakaan Provinsi Jawa Tengah

Dengan ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka. Apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 26 Februari 2026

Yang Menyatakan,



Dwiky Chandra Mulyo Utomo

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Dwiky Chandra Mulyo Utomo
NIM : 32602200058
Program Studi : Teknik Informatika
Fakultas : Teknologi Industri
Alamat Asal : Pangkalan Bun, Kalimantan Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul: Penerapan Algoritma Hybrid K-Means Dan Fuzzy C-Means Untuk Menganalisis Pola Minat Baca di Perpustakaan Provinsi Jawa Tengah. Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dalam pangkalan data dan dipublikasikan di internet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila di kemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung.

Semarang, 26 Februari 2026

Yang Menyatakan,



Dwiky Chandra Mulyo Utomo

KATA PENGANTAR

Dengan mengucap Syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan Rahmat dan karunia-Nya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “Penerapan Algoritma Hybrid K-Means dan Fuzzy C-Means untuk Menganalisis Pola Minat Baca di Perpustakaan Provinsi Jawa Tengah”. Laporan Tugas Akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang. Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor Universitas Islam Sultan Agung, Bapak Prof. Dr. H. Gunarto, S.H., M.H. yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri, Bapak Muhammad Qomaruddin, S.T., M. Sc., Ph.D.
3. Dosen Pembimbing, Bapak Andi Riansyah, ST., M.Kom yang telah meluangkan waktu, memberikan bimbingan, serta ilmu yang sangat berharga dalam penyusunan penelitian ini.
4. Orang tua penulis yang telah memberikan doa, semangat, serta dukungan yang tidak terhingga sehingga laporan ini dapat terselesaikan dengan lancar.
5. Serta kepada semua pihak yang tidak dapat saya sebutkan satu persatu yang telah membantu kelancaran penelitian ini.

Dengan segala kerendahan hati, penulis menyadari masih terdapat banyak kekurangan dari segi kualitas maupun dari ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini di masa mendatang.

Semarang, 26 Februari 2026



Dwiky Chandra Mulyo Utomo

DAFTAR ISI

| | |
|--|----------|
| HALAMAN JUDUL | i |
| LEMBAR PENGESAHAN LAPORAN TUGAS AKHIR..... | iii |
| SURAT PERNYATAAN KEASLIAN TUGAS AKHIR..... | iv |
| PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH | v |
| KATA PENGANTAR..... | vi |
| DAFTAR ISI..... | vii |
| DAFTAR GAMBAR..... | x |
| DAFTAR TABEL | xii |
| ABSTRAK | xiii |
| <i>ABSTRACT</i> | xiii |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Perumusan Masalah | 4 |
| 1.3 Pembatasan Masalah..... | 4 |
| 1.4 Tujuan Penelitian..... | 5 |
| 1.5 Manfaat Penelitian | 5 |
| 1.6 Sistematika Penulisan | 6 |
| BAB II TINJAUAN PUSTAKA DAN DASAR TEORI | 8 |
| 2.1 Tinjauan Pustaka | 8 |
| 2.2 Dasar Teori | 9 |
| 2.2.1 Minat Baca | 9 |
| 2.2.2 Data Mining | 9 |
| 2.2.3 <i>Clustering</i> | 10 |
| 2.2.4 <i>K-Means</i> | 10 |
| 2.2.5 <i>Fuzzy C-Means</i> (FCM)..... | 11 |
| 2.2.6 <i>Hybrid Clustering (K-Means + Fuzzy C-Means)</i> | 11 |

| | | |
|--|---|-----------|
| 2.2.7 | Klasifikasi (<i>Supervised Learning</i>)..... | 13 |
| 2.2.8 | <i>Light Gradient boosting Machine (LightGBM)</i> | 13 |
| 2.2.9 | Visualisasi Data Berbasis <i>Web</i> | 14 |
| BAB III METODOLOGI PENELITIAN | | 15 |
| 3.1 | Metode Penelitian..... | 15 |
| 3.1.1 | Studi Literatur | 16 |
| 3.1.2 | Pengumpulan Data | 17 |
| 3.1.3 | Pra-pemrosesan Data..... | 18 |
| 3.1.4 | Penerapan Model <i>Hybrid Clustering (K-Means + Fuzzy C-Means)</i> | 19 |
| 3.1.5 | <i>Profiling</i> dan Analisis Karakteristik..... | 21 |
| 3.1.6 | Transformasi dan Pelabelan Data (<i>Data Labeling</i>)..... | 22 |
| 3.1.7 | Alur Integrasi Clustering, Profiling, Labeling, dan Klasifikasi | 23 |
| 3.1.8 | Pembangunan Model Klasifikasi Prediktif Menggunakan <i>LightGBM</i> | 25 |
| 3.1.9 | Evaluasi Model..... | 26 |
| 3.1.10 | Implementasi Visualisasi..... | 27 |
| 3.2 | Rancangann Sistem | 28 |
| BAB IV HASIL DAN ANALISIS PENELITIAN..... | | 31 |
| 4.1 | <i>Library</i> yang Digunakan | 31 |
| 4.2 | <i>Dataset</i> yang Digunakan | 32 |
| 4.3 | Hasil <i>Preprocessing Dataset</i> | 33 |
| 4.4 | Hasil <i>Clustering</i> Menggunakan <i>Hybrid K-Means</i> dan <i>Fuzzy C-Means</i> | 35 |
| 4.4.1 | Penentuan Nilai <i>k</i> | 36 |
| 4.4.2 | Penentuan Nilai <i>m</i> | 37 |
| 4.4.3 | Hasil <i>Clustering Hybrid</i> | 37 |

| | | |
|---|---|-----------|
| 4.4.4 | Perbandingan Algoritma <i>Hybrid</i> dan Algoritma Tunggal | 41 |
| 4.5 | <i>Profiling</i> Pola Minat Baca Setiap <i>Cluster</i> | 42 |
| 4.6 | Analisis Mendalam Pola Minat Baca Masyarakat | 45 |
| 4.6.1 | Sintesis Karakteristik dan Dinamika Segmen Pembaca..... | 45 |
| 4.6.2 | Analisis Hubungan Antar Variabel (Dinamika Pola) | 46 |
| 4.7 | Penamaan Identitas (<i>Labeling</i>) Kluster | 47 |
| 4.8 | Klasifikasi dan Validasi Model Menggunakan <i>LightGBM</i> | 49 |
| 4.9 | Implikasi Manajerial | 54 |
| 4.10 | Implementasi <i>Dashboard Web</i> | 54 |
| 4.10.1 | Arsitektur Antarmuka dan Desain Pengalaman Pengguna (UI/UX) | 54 |
| 4.10.2 | Integrasi Arsitektur <i>AI Engine</i> dan <i>Pipeline Preprocessing</i> | 55 |
| 4.10.3 | Visualisasi <i>Dashboard</i> Utama dan Analisis Segmentasi..... | 56 |
| 4.10.4 | Sistem Prediksi Otomatis Anggota Baru..... | 58 |
| 4.10.5 | Analisis Strategis dan Rekomendasi Otomatis | 60 |
| 4.10.6 | Manajemen Data dan Fitur Ekspor Laporan | 62 |
| BAB V KESIMPULAN DAN SARAN | | 64 |
| 5.1 | Kesimpulan | 64 |
| 5.2 | Saran..... | 65 |
| DAFTAR PUSTAKA..... | | 66 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 3. 1 <i>Flowchart</i> Metode Penelitian..... | 16 |
| Gambar 3. 2 <i>Flowchart Hybrid Clustering</i> | 20 |
| Gambar 3. 3 <i>Flowchart</i> Proses <i>Clustering</i> ke Klasifikasi..... | 23 |
| Gambar 3. 4 <i>Flowchart</i> Rancangan Sistem | 29 |
| Gambar 4. 1 <i>Library Python</i> | 31 |
| Gambar 4. 2 <i>Dataset</i> Peminjaman Buku..... | 33 |
| Gambar 4. 3 Agregasi 2 Tahap | 34 |
| Gambar 4. 4 Kode <i>Transformasi</i> dan <i>Distribusi</i> Data..... | 35 |
| Gambar 4. 5 Hasil Tuning Nilai k | 36 |
| Gambar 4. 6 <i>Elbow Method</i> Nilai k | 36 |
| Gambar 4. 7 Ringkasan Performa Nilai m | 37 |
| Gambar 4. 8 Visualisasi 3D <i>Scatter Plot</i> | 38 |
| Gambar 4. 9 <i>Radar Chart</i> Fitur <i>Cluster</i> | 39 |
| Gambar 4. 10 Sebaran Preferensi Genre Buku | 40 |
| Gambar 4. 11 Data Hasil <i>Cluster</i> | 40 |
| Gambar 4. 12 Kecepatan Konvergensi..... | 41 |
| Gambar 4. 13 Visualisasi Profil Akhir Segmentasi Pembaca | 47 |
| Gambar 4. 14 Hasil Integrasi Data dan Arsitektur Fitur | 49 |
| Gambar 4. 15 Visualisasi Rasio Pembagian <i>Dataset</i> | 50 |
| Gambar 4. 16 Evaluasi Metrik Performa Model pada <i>Data Training</i> | 50 |
| Gambar 4. 17 Evaluasi Metrik Performa Model pada <i>Data Testing</i> | 51 |
| Gambar 4. 18 Visualisasi Matriks Kebingungan (<i>Confusion Matrix</i>) Data Uji.... | 52 |
| Gambar 4. 19 Variabel Penentu Segmentasi Pembaca..... | 52 |
| Gambar 4. 20 Visualisasi Stabilitas Pembelajaran Model (<i>Learning Curve</i>) | 53 |
| Gambar 4. 21 Kode Program Konfigurasi Halaman <i>Dashboard</i> | 55 |
| Gambar 4. 22 Kode Program Sinkronisasi <i>Engine</i> dan <i>Preprocessing</i> | 56 |
| Gambar 4. 23 Tampilan Halaman <i>Dashboard</i> Utama..... | 57 |
| Gambar 4. 24 Tampilan Menu Prediksi Anggota Baru | 59 |
| Gambar 4. 25 Tampilan Analisis Mendalam dan Rekomendasi | 60 |

Gambar 4. 26 Tampilan Fitur Ekspor Laporan 62



DAFTAR TABEL

| | |
|--|----|
| Tabel 3. 1 Atribut <i>Dataset</i> | 17 |
| Tabel 3. 2 Tiga Atribut Utama Fokus <i>Clustering</i> | 18 |
| Tabel 4. 1 Evaluasi Algoritma <i>K-Means</i> , <i>Fuzzy C-Means</i> dan <i>Hybrid Clustering</i> | 41 |
| Tabel 4. 2 <i>Profiling Cluster</i> Berdasarkan Usia dan Frekuensi | 43 |
| Tabel 4. 3 Identifikasi Genre Favorit Berdasarkan <i>Cluster</i> | 43 |
| Tabel 4. 4 <i>Profiling Cluster</i> Berdasarkan Keberagaman Minat Baca | 44 |



ABSTRAK

Pengelompokan pola minat baca masyarakat merupakan instrumen krusial bagi perpustakaan dalam merancang strategi layanan yang tepat sasaran. Penelitian ini bertujuan untuk mengembangkan sistem analisis pola minat baca di Perpustakaan Provinsi Jawa Tengah dengan menerapkan algoritma *Hybrid K-Means* dan *Fuzzy C-Means* (FCM). Hasil analisis terhadap 3.112 data peminjam aktif berhasil mengidentifikasi tiga segmen utama, yaitu Pembaca Muda Selektif (66,6%), Pembaca Mapan Antusias (18,4%), dan Pembaca Belia Eksploratif (15,0%). Temuan ini mengungkapkan adanya karakteristik yang kontras antar kelompok pemustaka berdasarkan dimensi usia, tingkat loyalitas kunjungan, serta keragaman preferensi genre buku. Untuk menghasilkan segmentasi tersebut secara akurat, metode yang digunakan mengintegrasikan keunggulan *K-Means* dalam menentukan pusat kluster awal untuk mengoptimalkan proses iterasi pada FCM, divalidasi menggunakan model klasifikasi *LightGBM*, serta diimplementasikan ke dalam *Executive Dashboard* berbasis *Streamlit*. Hasil evaluasi teknis menunjukkan bahwa metode hybrid secara signifikan lebih efisien dengan mencapai konvergensi hanya dalam 7 iterasi, jauh lebih cepat dibandingkan FCM konvensional dengan inialisasi acak yang memerlukan 19 iterasi. Validasi melalui *LightGBM* menghasilkan akurasi sebesar 99,68%, yang mengonfirmasi bahwa variabel usia merupakan faktor paling dominan dalam menentukan segmentasi minat baca masyarakat. Penelitian ini menyimpulkan bahwa pendekatan hybrid memberikan hasil pengelompokan yang stabil dan presisi tinggi, sehingga sangat relevan digunakan sebagai basis pengambilan keputusan strategis bagi pihak pengelola perpustakaan.

Kata Kunci: *Hybrid Clustering, K-Means, Fuzzy C-Means, LightGBM, Minat Baca.*

ABSTRACT

The grouping of reading interest patterns among the public is a crucial instrument for libraries in designing targeted service strategies. This study aims to develop a system for analyzing reading interest patterns at the Central Java Provincial Library by applying the Hybrid K-Means and Fuzzy C-Means (FCM) algorithms. The analysis of 3,112 active borrower data successfully identified three main segments, namely Selective Young Readers (66.6%), Enthusiastic Established Readers (18.4%), and Explorative Young Readers (15.0%). These findings reveal contrasting characteristics between reader

groups based on age, visit loyalty, and diversity of book genre preferences. To produce accurate segmentation, the method used integrates the advantages of K-Means in determining the initial cluster center to optimize the iteration process in FCM, validated using the LightGBM classification model, and implemented into a Streamlit-based Executive Dashboard. Technical evaluation results show that the hybrid method is significantly more efficient, achieving convergence in only 7 iterations, much faster than conventional FCM with random initialization, which requires 19 iterations. Validation through LightGBM resulted in an accuracy of 99.68%, confirming that age is the most dominant factor in determining the reading interest segmentation of the community. This study concludes that the hybrid approach is effective in determining the most dominant factors in public reading interest segmentation.

Keywords: *Hybrid Clustering, K-Means, Fuzzy C-Means, LightGBM, Reading Interest.*



BAB I

PENDAHULUAN

1.1 Latar Belakang

Minat baca masyarakat merupakan salah satu indikator fundamental dalam mengukur tingkat literasi sekaligus kualitas sumber daya manusia suatu bangsa. Literasi adalah fondasi utama pembangunan masyarakat yang berkualitas, di mana tingginya minat baca berpengaruh langsung pada kemampuan masyarakat dalam mengakses, memahami, serta memanfaatkan informasi sebagai modal pembangunan berkelanjutan di era digital (Pamungkas dkk., 2024). Secara nasional, tantangan literasi masih menjadi agenda besar pemerintah, yang kemudian direspon secara regional melalui penyediaan fasilitas publik yang masif, salah satunya oleh Perpustakaan Provinsi Jawa Tengah (Syarif, 2024).

Namun, keberadaan fasilitas tersebut belum memberikan dampak optimal bagi pengambilan keputusan strategis karena potensi datanya belum dimanfaatkan dengan baik. Berdasarkan pengamatan, Perpustakaan Provinsi Jawa Tengah memiliki akumulasi data transaksi peminjaman buku yang sangat besar. Masalah utamanya adalah data tersebut saat ini hanya berakhir sebagai data pasif tanpa adanya proses analisis pola perilaku pemustaka yang mendalam (Sinaga, 2024). Padahal, di balik ribuan baris data tersebut tersimpan pola kecenderungan masyarakat yang sangat berharga untuk menentukan kebijakan layanan perpustakaan (Pattee, 2024).

Kesenjangan antara melimpahnya data dengan minimnya informasi strategis mengakibatkan pihak pengelola perpustakaan kesulitan dalam memahami karakteristik riil pengunjungnya. Tanpa adanya pemetaan pola yang jelas, strategi pengadaan koleksi buku dan program promosi literasi dilakukan tanpa dasar data yang kuat (*data-driven insights*), sehingga berisiko tidak tepat sasaran. Masalah analisis ini menjadi semakin rumit karena perilaku masyarakat yang dinamis, di mana pemustaka seringkali memiliki

minat yang beririsan atau *multi-genre* atau menyukai lebih dari satu jenis buku (Mahmudi & Fatah, 2024).

Untuk membedah kompleksitas tersebut, teknik *data mining* dengan metode *Clustering* menjadi instrumen yang tepat. *Clustering* bekerja tanpa label kelas (*unsupervised*) dan sangat efektif diterapkan dalam analisis perilaku serta segmentasi preferensi masyarakat (Riansyah dkk., 2023). Dengan *Clustering*, ribuan data peminjaman dapat disederhanakan menjadi beberapa kelompok (segmen) yang memiliki kesamaan karakteristik, sehingga pola minat baca yang tadinya tersembunyi dapat teridentifikasi secara jelas.

Dalam pemilihan algoritma, *K-Means* sering menjadi pilihan utama karena kesederhanaan dan efisiensi komputasinya dalam menangani data berjumlah besar (Ikotun dkk., 2023). Namun, *K-Means* memiliki sifat *Hard Clustering* yang kaku, di mana setiap data dipaksa masuk ke dalam satu kelompok secara mutlak (0 atau 1) (Mussabayev dkk., 2023). Hal ini seringkali kurang presisi jika diterapkan pada data perilaku manusia yang bersifat ambigu, selain itu *K-Means* sangat sensitif terhadap penentuan inisialisasi pusat klaster awal yang acak (Zheng dkk., 2023).

Sebagai solusi atas kekakuan tersebut, algoritma *Fuzzy C-Means* (FCM) menawarkan fleksibilitas melalui konsep *Soft Clustering*. Dalam FCM, setiap data dihitung derajat keanggotaannya terhadap semua pusat klaster, sehingga sangat cocok untuk merepresentasikan pola minat baca yang saling beririsan atau *overlapping* (Jaelani dkk., 2025). Kendati demikian, FCM memiliki kelemahan utama yaitu membutuhkan waktu komputasi yang lama dan risiko terjebak pada solusi lokal (*local optima*) jika dimulai dengan inisialisasi pusat klaster yang tidak tepat (Zhang & Liu, 2025).

Guna menjamin stabilitas dan presisi analisis, penelitian ini menerapkan pendekatan *Hybrid Clustering* (*K-Means* + FCM). Pendekatan ini merupakan sinergi di mana *K-Means* digunakan pada tahap awal untuk menentukan pusat klaster (*centroid*) secara cepat dan global, yang kemudian digunakan sebagai titik inisialisasi awal bagi algoritma FCM untuk

memperhalus derajat keanggotaan data (Fan, 2023). Sinergi ini bertujuan menghasilkan pengelompokan yang lebih stabil, akurat, dan efisien dibandingkan hanya menggunakan metode tunggal.

Fokus utama penelitian ini tidak berhenti pada pembentukan kluster secara teknis, melainkan pada tahap *Profiling* untuk memberikan makna manajerial pada setiap kelompok yang terbentuk. Melalui analisis statistik deskriptif pada setiap kluster, penelitian ini mengubah *label numerik* yang abstrak menjadi identitas psikografis yang mudah dipahami, seperti 'Pembaca Muda Selektif' atau 'Pembaca Mapan Antusias'. Pengetahuan inilah yang menjadi jawaban atas kebutuhan manajemen perpustakaan untuk merancang strategi literasi yang lebih personal.

Setelah profil pola minat terbentuk, tantangan selanjutnya adalah memastikan sistem dapat menangani anggota baru secara instan. Oleh karena itu, penelitian ini mengintegrasikan model klasifikasi *LightGBM*. Algoritma ini dipilih karena mampu memproses data tabular secara efisien dengan tingkat akurasi yang tinggi dan waktu iterasi yang singkat (Kurniawan dkk., 2025). Model ini berfungsi sebagai mesin prediksi otomatis yang mampu mengklasifikasikan kategori minat baca anggota baru hanya berdasarkan data atributnya tanpa perlu melakukan proses *clustering* ulang dari awal.

Terakhir, seluruh hasil analisis cerdas ini diimplementasikan ke dalam sebuah *Executive Dashboard* berbasis *web* yang dirancang sebagai *Decision Support System* (DSS). Dengan adanya sistem visualisasi yang interaktif ini, pihak Perpustakaan Provinsi Jawa Tengah diharapkan memiliki alat bantu yang praktis untuk memantau profil pemustaka secara *real-time* dan mengambil kebijakan strategis berbasis data (*data-driven*) guna meningkatkan budaya literasi masyarakat Jawa Tengah secara berkelanjutan (Gopal & Panchal, 2023).

1.2 Perumusan Masalah

Perumusan masalah dalam penelitian ini berfokus pada pemanfaatan data transaksi peminjaman buku di Perpustakaan Provinsi Jawa Tengah yang selama ini hanya menjadi arsip pasif tanpa proses analisis pola perilaku pemustaka yang mendalam. Ketiadaan pengolahan data tersebut menyebabkan pihak manajemen kesulitan dalam merumuskan strategi layanan serta pengadaan koleksi buku yang benar-benar tepat sasaran berdasarkan kebutuhan riil masyarakat. Kendala teknis kemudian muncul pada penggunaan algoritma *Clustering* konvensional karena *K-Means* memiliki keterbatasan dalam menangani ambiguitas minat baca masyarakat yang bersifat *multi-genre* akibat sifat *hard clustering* yang dimilikinya. Selain itu, algoritma *Fuzzy C-Means* sangat bergantung pada inisialisasi pusat kluster awal yang tepat agar hasil analisisnya stabil dan tidak terjebak dalam *local optima*. Oleh karena itu, penelitian ini menerapkan metode *Hybrid K-Means* dan *Fuzzy C-Means* untuk menghasilkan pengelompokan pola minat baca yang lebih akurat serta presisi. Masalah selanjutnya berkaitan dengan aspek keberlanjutan fungsi sistem yaitu bagaimana memanfaatkan pola yang telah terbentuk untuk memprediksi kategori minat baca anggota baru secara otomatis menggunakan algoritma *LightGBM*. Tantangan implementasi terletak pada upaya menyajikan seluruh hasil analisis tersebut ke dalam platform visualisasi berbasis *Web* agar dapat digunakan secara praktis oleh pihak manajemen perpustakaan dalam pengambilan keputusan strategis.

1.3 Pembatasan Masalah

Agar penelitian lebih terarah, batasan masalah ditetapkan sebagai berikut:

1. Data yang digunakan adalah data peminjaman buku dari Perpustakaan Provinsi Jawa Tengah periode tahun 2024/2025.
2. Algoritma *Clustering* yang digunakan adalah *Hybrid K-Means + Fuzzy C-Means* dan Algoritma klasifikasi untuk prediksi anggota baru menggunakan algoritma *LightGBM*.

3. Evaluasi kualitas kluster menggunakan *Silhouette Coefficient*, *Partition Coefficient (PC)*, dan *Classification Entropy (CE)* sedangkan untuk kualitas Klasifikasi menggunakan *accuracy*, *precision*, *recall*, and *F1-score*.

1.4 Tujuan Penelitian

Mengembangkan dan mengimplementasikan algoritma *Hybrid K-Means* dan *Fuzzy C-Means* untuk menganalisis pola minat baca di Perpustakaan Provinsi Jawa Tengah guna menghasilkan segmentasi profil pembaca yang lebih presisi serta melakukan profiling karakteristik dominan pada setiap kelompok yang terbentuk. Penelitian ini juga bertujuan untuk membangun model prediksi menggunakan algoritma *LightGBM* yang diintegrasikan ke dalam sebuah *Dashboard* berbasis *Web* untuk mengklasifikasikan kategori minat baca anggota baru secara otomatis.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Memberikan kontribusi akademik mengenai penerapan metode *Hybrid Clustering* yang mengintegrasikan algoritma *K-Means* dan *Fuzzy C-Means* untuk menangani ambiguitas data, serta penggunaan *LightGBM* sebagai model klasifikasi untuk memetakan perilaku pembaca secara akurat.
2. Mempermudah pihak pengelola perpustakaan dalam menganalisis dan mengidentifikasi pola minat baca bagi anggota baru secara praktis melalui platform *Web*. Sistem ini memungkinkan klasifikasi kategori anggota dilakukan secara otomatis sehingga pihak perpustakaan tidak perlu melakukan proses analisis manual atau pengolahan data ulang dari awal.
3. Menyediakan wawasan berbasis data (*data-driven insights*) mengenai peta minat baca masyarakat, yang dapat digunakan sebagai acuan dalam pengadaan buku dan strategi promosi literasi.

1.6 Sistematika Penulisan

Penulisan dalam membuat laporan tugas akhir mengikuti sistematika penulisan sebagai berikut:

BAB I : PENDAHULUAN

Bab ini berisi latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, serta manfaat penelitian.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini menguraikan hasil penelitian terdahulu yang relevan, konsep dasar tentang minat baca, data mining, *Clustering*, algoritma *K-Means*, algoritma *Fuzzy C-Means*, *Hybrid Clustering*, serta visualisasi data berbasis *Web*.

BAB III : METODOLOGI PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, meliputi studi literatur, pengumpulan data, pra-pemrosesan data, penerapan model *Hybrid Clustering (K-Means + Fuzzy C-Means)*, evaluasi model, implementasi visualisasi berbasis *Web*, serta rancangan sistem penelitian.

BAB IV : HASIL DAN ANALISIS PENELITIAN

Bab ini memaparkan hasil dari penelitian yang dilakukan, termasuk pengolahan *Dataset*, langkah-langkah dalam mengoperasikan sistem, implementasi antar muka pengguna dan analisis yang dilakukan.

BAB V : KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian lebih lanjut maupun penerapan praktis pada bidang literasi masyarakat.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Penelitian dengan judul “Analisis Pengaruh Jenis Buku Terhadap Minat Baca Mahasiswa di Perpustakaan Ibrahimy dengan Algoritma *K-Means Clustering*” yang dilakukan oleh Mahmudi & Fatah (2024) menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan data minat baca berdasarkan jenis buku. Namun, hasil penelitian ini juga menegaskan kelemahan *K-Means* terhadap *Outlier* dan inisialisasi *centroid* yang tidak stabil, yang dapat menimbulkan hasil klasterisasi kurang representatif.

Selanjutnya, penelitian oleh Rilvani & Bangsa (2025) yang berjudul “Studi Literatur: Perbandingan Algoritma *K-Means* dan *Fuzzy C-Means* dalam Analisis *Clustering*” menyoroti bahwa *Fuzzy C-Means* lebih unggul dalam memberikan informasi yang lebih rinci mengenai struktur data yang kompleks, namun memerlukan iterasi yang lebih banyak.

Muhaimin (2024) dalam penelitiannya “Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan” menggunakan *K-Means* dengan berbagai variabel akademik dan demografis. Menerapkan *K-Means* untuk klasifikasi prestasi akademik siswa dan menemukan bahwa hasil *Clustering* sangat bergantung pada jumlah klaster awal (k). Hal ini menjadi salah satu tantangan utama dalam metode partisi, karena pemilihan k yang tidak optimal dapat menurunkan validitas hasil analisis.

Penelitian oleh Telaumbanua dkk. (2025) memperkenalkan pendekatan *Hybrid Clustering* yang menggabungkan beberapa algoritma untuk mengatasi keterbatasan metode tunggal. Hasilnya menunjukkan pendekatan *Enhanced Fuzzy C-Means*, yang menekankan pentingnya inisialisasi *centroid* yang tepat untuk menghindari hasil yang bias. Hal ini mendukung premis penelitian ini bahwa penggabungan *K-Means* sebagai inisiator bagi *Fuzzy C-Means* adalah solusi yang layak..

Sementara itu, Sinaga (2024) mengembangkan sistem berbasis *Web* untuk memvisualisasikan pola minat baca masyarakat. Sistem ini memudahkan pengguna memahami data secara interaktif, tetapi algoritma yang digunakan masih terbatas pada metode tunggal sehingga fleksibilitas analisisnya belum optimal.

Dari berbagai penelitian tersebut dapat disimpulkan bahwa algoritma *K-Means* memiliki keunggulan dalam hal efisiensi, namun sensitif terhadap penentuan inisialisasi awal. Di sisi lain, *Fuzzy C-Means* unggul dalam memberikan informasi keanggotaan yang lebih rinci (*Soft Clustering*), namun memiliki komputasi yang berat jika inisialisasinya acak. Selain itu, kebanyakan penelitian sebelumnya hanya berhenti pada tahap pengelompokan (*Clustering*) tanpa melanjutkannya ke tahap prediksi anggota baru.

Oleh karena itu, penelitian ini menawarkan pendekatan *Hybrid K-Means + Fuzzy C-Means* untuk mendapatkan klaster yang presisi dan stabil. Tidak berhenti di situ, penelitian ini juga mengintegrasikan algoritma *Light Gradient boosting Machine (LightGBM)* sebagai metode klasifikasi lanjutan untuk memprediksi kategori minat baca anggota baru secara otomatis, serta mengimplementasikannya dalam visualisasi berbasis *Web*.

2.2 Dasar Teori

2.2.1 Minat Baca

Minat baca merupakan kecenderungan psikologis seseorang untuk terlibat dalam aktivitas membaca secara konsisten. Pamungkas dkk. (2024) menjelaskan bahwa minat baca tidak hanya ditunjukkan melalui kemampuan teknis membaca, tetapi juga ketertarikan, motivasi, dan kebiasaan dalam mengakses bahan bacaan.

2.2.2 Data Mining

Data mining adalah proses menemukan pola atau pengetahuan dari kumpulan data berukuran besar dengan memanfaatkan teknik statistik, pembelajaran mesin, dan kecerdasan buatan (Hendrastuty, 2024). Proses ini

biasanya melibatkan tahapan mulai dari seleksi data, pembersihan data, transformasi, penerapan algoritma, hingga evaluasi hasil. Dengan demikian, data mining berperan penting dalam menggali informasi tersembunyi yang dapat digunakan untuk pengambilan keputusan.

2.2.3 Clustering

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Menurut Fadilah (2023), *Clustering* bekerja tanpa label kelas (*unsupervised*) dan sering diterapkan dalam analisis perilaku, segmentasi pasar, hingga pola minat baca. Pendekatan ini membantu peneliti memahami pola tersembunyi dalam data besar serta mengidentifikasi kelompok yang memiliki kesamaan tertentu, baik secara perilaku, geografis, maupun preferensi.

Riansyah dkk., (2023) menjelaskan bahwa penerapan metode Fuzzy *K-Means* dalam proses transformasi digital Usaha Mikro, Kecil, dan Menengah (UMKM) memungkinkan pengelompokan yang lebih fleksibel karena setiap entitas dapat memiliki tingkat keanggotaan berbeda pada tiap kelompok. Metode ini tidak hanya meningkatkan akurasi hasil analisis, tetapi juga memberikan dasar yang lebih kuat bagi pemerintah dan lembaga pendukung UMKM dalam merancang strategi pengembangan, peningkatan akses pasar, serta optimalisasi produktivitas sesuai karakteristik kelompok usaha.

2.2.4 K-Means

K-Means merupakan salah satu algoritma *Clustering* berbasis partisi yang bertujuan untuk membagi data ke dalam sejumlah kluster tertentu dengan meminimalkan jarak antara data dengan pusat kluster (*centroid*). Algoritma ini bekerja secara iteratif dengan cara menentukan jumlah kluster k , kemudian mencari pusat kluster terbaik berdasarkan distribusi data.

Menurut Rilvani & Bangsa (2025), *K-Means* sering digunakan karena kesederhanaan dan efisiensinya. Namun, kelemahan dari metode ini adalah sensitif terhadap keberadaan *Outlier* serta pemilihan jumlah kluster awal.

Persamaan jarak yang digunakan dalam *K-Means* umumnya adalah *Euclidean Distance* sebagai berikut:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (1)$$

Dengan keterangan:

1. $d(x_i, c_j)$: Jarak antara data ke-i dengan *centroid* ke-j
2. x_{ik} : Nilai atribut ke-k dari data ke-i
3. c_{jk} : Nilai atribut ke-k dari *centroid* ke j
4. n : Jumlah atribut

Fungsi objektif yang diminimalkan pada *K-Means* adalah *Sum of Squared Error* (SSE):

$$SSE = \sum_{j=1}^c \sum_{x_i \in S_j} \|x_i - c_j\|^2 \quad (2)$$

2.2.5 Fuzzy C-Means (FCM)

Fuzzy C-Means (FCM) adalah algoritma pengelompokan data di mana keberadaan setiap titik data dalam suatu kluster ditentukan oleh derajat keanggotaan. Berbeda dengan *K-Means* yang bernilai 0 atau 1, keanggotaan dalam FCM bernilai antara 0 hingga 1. Fungsi objektif FCM dirumuskan sebagai berikut (Syawali dkk., 2024):

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2 \quad (3)$$

Dengan keterangan:

1. N : Jumlah data
2. C : Jumlah kluster
3. m : Pembobot *fuzziness* ($m > 1$)
4. u_{ij} : Derajat keanggotaan data ke-i pada kluster ke-j
5. c_j : Pusat kluster ke-j

2.2.6 Hybrid Clustering (K-Means + Fuzzy C-Means)

Hybrid Clustering dalam penelitian ini mengintegrasikan efisiensi algoritma *K-Means* dalam menentukan partisi awal dengan fleksibilitas *Fuzzy C-Means* (FCM) dalam menangani ambiguitas data (Mchara dkk., 2025). Alur logika *hybrid* ini dilakukan melalui tiga tahapan matematis utama:

1. Tahap Inisialisasi K-Means (Pencarian Partisi Kaku)

Algoritma dimulai dengan menjalankan proses *K-Means* secara iteratif hingga konvergen untuk meminimalkan fungsi objektif *Sum of Squared Error* (SSE) guna mendapatkan pusat kluster (*centroid*) yang stabil:

$$SSE = \sum_{j=1}^c \sum_{x_i \in S_j} \|x_i - c_j\|^2 \quad (4)$$

Hasil dari tahap ini adalah Label Kluster (L_i) yang bersifat kaku (*hard partitioning*), di mana setiap data x_i hanya memiliki satu identitas kelompok.

2. Tahap Transisi (Pemetaan Matriks Keanggotaan)

Poin krusial dalam metode *hybrid* ini adalah transformasi label kaku dari K-Means menjadi Matriks Keanggotaan Awal ($u_{ij}^{(0)}$) untuk algoritma FCM. Berdasarkan implementasi program, label dikonversi menggunakan fungsi pemetaan sebagai berikut:

$$u_{ij}^{(0)} = \begin{cases} 1 + \epsilon, & \text{jika } L_i = j \\ \epsilon & \text{jika } L_i \neq j \end{cases} \quad (5)$$

Dengan keterangan ϵ merupakan nilai *smoothing* (0.01) untuk memastikan setiap data tetap memiliki derajat keanggotaan *fuzzy* awal yang valid dan tidak bernilai nol mutlak sebelum dinormalisasi.

3. Tahap Optimasi Fuzzy C-Means

Matriks keanggotaan awal ($u_{ij}^{(0)}$) hasil dari tahap transisi kemudian digunakan sebagai input utama untuk memulai iterasi FCM. Algoritma FCM kemudian melakukan optimasi pada fungsi objektif J_m untuk menghitung derajat keanggotaan u_{ij} yang lebih presisi:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2 \quad (6)$$

Pendekatan ini bertujuan untuk meminimalisir jumlah iterasi pada FCM dan menghindari jebakan local optima akibat inisialisasi pusat kluster yang dilakukan secara acak. Dengan menggunakan *centroid* atau label dari *K-Means* sebagai pemandu awal, algoritma dapat mencapai konvergensi dengan lebih cepat dan stabil.

2.2.7 Klasifikasi (*Supervised Learning*)

Klasifikasi adalah teknik data mining untuk memprediksi label kelas dari data baru berdasarkan model yang dilatih dari data sebelumnya. Berbeda dengan *Clustering* yang tidak memiliki label, klasifikasi termasuk dalam pembelajaran terawasi (*supervised learning*). Tujuan utamanya adalah menemukan fungsi target yang memetakan atribut input (fitur) ke label *output* (Afif Muhaimin, 2024).

2.2.8 *Light Gradient boosting Machine (LightGBM)*

LightGBM adalah kerangka kerja *gradient boosting* berbasis pohon keputusan yang menggunakan strategi pertumbuhan pohon secara *leaf-wise* (Gopal & Panchal, 2023). Berbeda dengan strategi *level-wise* yang menumbuhkan pohon secara horizontal, strategi *leaf-wise* memilih daun dengan tingkat pengurangan kehilangan (*loss reduction*) terbesar untuk diekspansi lebih lanjut, sehingga menghasilkan akurasi yang lebih tinggi dengan jumlah iterasi yang lebih sedikit. Menurut Kurniawan dkk. (2025), *LightGBM* menggunakan teknik *Gradient-based One-Side Sampling (GOSS)* dan *Exclusive Feature Bundling (EFB)* yang membuatnya jauh lebih cepat dibandingkan algoritma tradisional seperti *XGBoost*, namun tetap mempertahankan akurasi yang tinggi pada *Dataset* yang kompleks.

Model ini sangat efektif untuk menangani data berdimensi tinggi dan memberikan konvergensi yang lebih cepat dibandingkan algoritma boosting lainnya (Alsulamy, 2025). Keandalan model klasifikasi ini dievaluasi menggunakan metrik performa yang meliputi *accuracy*, *precision*, *recall*, dan *f1-score* yang dihitung melalui persamaan berikut (Riansyah et al., 2025):

1. *Accuracy*

Metrik untuk mengukur tingkat kedekatan hasil prediksi dengan nilai aktual secara keseluruhan.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Metrik ini memberikan gambaran umum mengenai seberapa sering model berhasil memprediksi kategori minat baca dengan benar terhadap keseluruhan data uji.

2. *Precision*

Metrik untuk mengukur keakuratan model dalam memprediksi kelas positif.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Precision mengukur ketepatan model dalam memberikan label tertentu pada data. *precision* menjawab: "Dari semua anggota yang diprediksi sebagai 'Pembaca Muda Selektif', berapa banyak yang benar-benar termasuk dalam kelompok tersebut?".

3. *Recall*

Metrik untuk mengukur keberhasilan model dalam memprediksi kembali kelas positif yang sebenarnya.

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

Recall mengukur kemampuan model dalam menemukan kembali seluruh anggota kelompok yang sebenarnya. Ini memastikan tidak ada pemustaka yang "terlewat" atau salah masuk kategori saat sistem melakukan prediksi otomatis pada anggota baru.

4. *F1-score*

Rata-rata harmonik antara *precision* dan *recall* untuk mengevaluasi model secara seimbang.

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

F1-score adalah rata-rata harmonik yang menyeimbangkan antara *Precision* dan *Recall*. Nilai ini memberikan evaluasi yang lebih adil jika terdapat ketidakseimbangan jumlah anggota antar kelompok, memastikan model tetap objektif dalam memprediksi kelompok mayoritas maupun minoritas.

2.2.9 Visualisasi Data Berbasis *Web*

Visualisasi data berbasis *Web* memungkinkan penyajian hasil analisis dalam bentuk grafik interaktif yang dapat diakses melalui *browser*. Sinaga (2024) menekankan bahwa visualisasi interaktif dapat membantu pengguna memahami pola data dengan lebih mudah dan intuitif.

BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

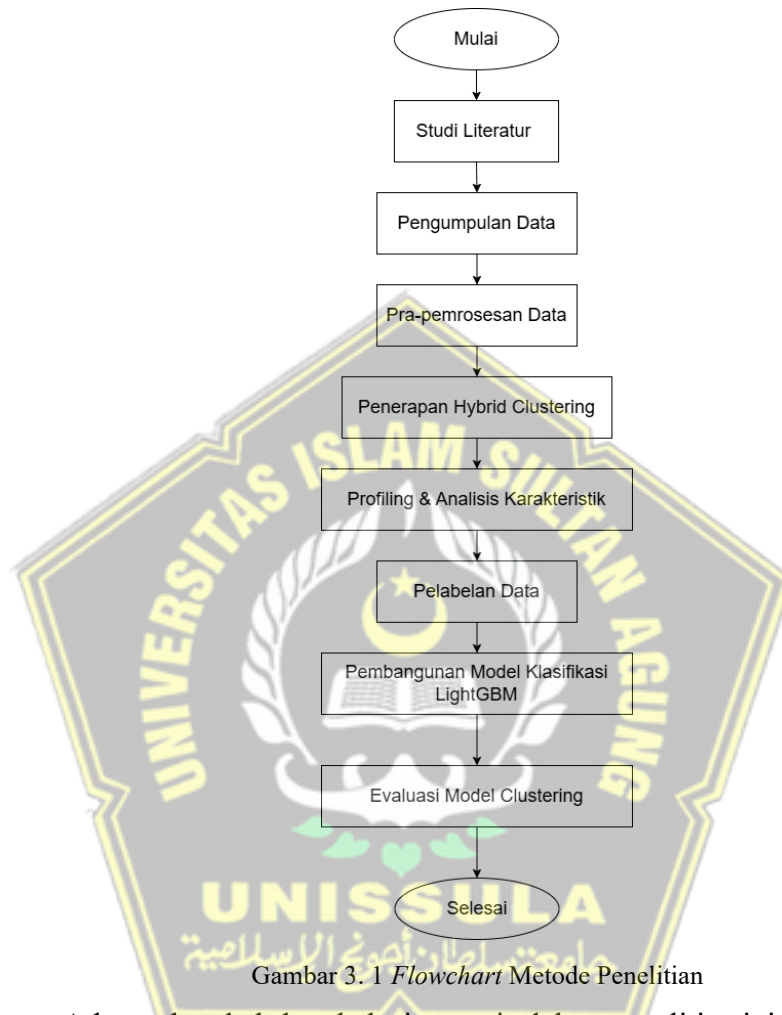
Penelitian ini menerapkan metode *Hybrid Clustering* yang mengintegrasikan algoritma *K-Means* dan *Fuzzy C-Means*(FCM) untuk menganalisis pola minat baca di Perpustakaan Provinsi Jawa Tengah berdasarkan data peminjaman buku di Perpustakaan Provinsi Jawa Tengah. Pemilihan pendekatan *Hybrid* ini dilakukan dengan mempertimbangkan karakteristik komplementer dari kedua algoritma tersebut guna mengatasi kelemahan masing-masing metode tunggal.

Algoritma *K-Means* dipilih sebagai inisiator karena keunggulannya dalam efisiensi komputasi dan kemampuannya menentukan pusat kluster (*centroid*) global secara cepat. Namun, *K-Means* memiliki keterbatasan karena sifatnya yang *Hard Clustering*, di mana setiap data dipaksa masuk ke dalam satu kelompok secara mutlak (0 atau 1), sehingga kurang mampu menangkap nuansa data perilaku manusia yang seringkali bersifat ambigu atau memiliki minat ganda.

Di sisi lain, algoritma *Fuzzy C-Means*(FCM) memiliki keunggulan dalam menerapkan konsep *Soft Clustering*, yang memungkinkan setiap data memiliki derajat keanggotaan pada beberapa kluster sekaligus, sehingga sangat cocok untuk merepresentasikan pola minat baca yang saling beririsan (*overlapping*). Kendati demikian, FCM memiliki kelemahan utama yaitu sensitif terhadap inisialisasi pusat kluster awal yang jika dilakukan secara acak dapat menyebabkan hasil terjebak pada solusi lokal (*local optimum*) dan membutuhkan waktu komputasi yang lebih lama.

Oleh karena itu, kombinasi keduanya diterapkan dengan mekanisme di mana hasil *centroid* dari *K-Means* digunakan sebagai titik inisialisasi awal bagi algoritma FCM. Pendekatan ini diharapkan dapat menghasilkan kualitas *Clustering* yang lebih stabil, akurat, serta efisien dalam memetakan kompleksitas pola minat baca masyarakat. Pendekatan ini diharapkan dapat

menghasilkan kualitas Clustering yang lebih stabil, akurat, serta efisien dalam memetakan kompleksitas pola minat baca masyarakat. Adapun langkah-langkah sistematis dalam penelitian ini ditunjukkan pada Gambar 3.1 berikut:



Gambar 3.1 Flowchart Metode Penelitian

Adapun langkah-langkah sistematis dalam penelitian ini adalah sebagai berikut:

3.1.1 Studi Literatur

Peneliti mencari berbagai sumber informasi terkait teori dan metode yang relevan dengan penelitian ini. Literatur yang dipelajari meliputi konsep minat baca, data mining, *Clustering*, algoritma *K-Means*, algoritma *Fuzzy C-Means*, serta *Hybrid Clustering*. Sumber literatur diperoleh dari artikel ilmiah, jurnal, tugas akhir, buku, dan situs resmi yang tersedia secara daring.

3.1.2 Pengumpulan Data

Pengumpulan data dilakukan dengan memperoleh *Dataset* dari Perpustakaan Provinsi Jawa Tengah dengan pengambilan data dari bulan November 2024 sampai dengan November 2025. *Dataset* yang digunakan dalam penelitian, mencakup sumber, volume, atribut, serta penentuan variabel utama yang akan membentuk pola minat baca masyarakat.

A. Sumber, Volume, dan Atribut *Dataset*

Dataset yang digunakan dalam penelitian ini adalah data sekunder berupa riwayat peminjaman buku oleh anggota dari Perpustakaan Provinsi Jawa Tengah di Kota Semarang.

1. Rentang Waktu Data: Periode tahun 2024/2025.
2. Volume Data: *Dataset* awal memiliki total 14.602 *entri* data peminjaman.
3. Atribut *Dataset*: Data mentah mencakup 4 atribut yang akan diproses:

Tabel 3. 1 Atribut *Dataset*

| No | Atribut Data | Keterangan |
|----|----------------------|--|
| 1 | Nama Lengkap | Nama orang yang meminjam buku. |
| 2 | Genre/Kategori buku | Jenis/kategori buku yang diminati (Kesusastraan, Bahasa, dll). |
| 3 | Frekuensi Peminjaman | Jumlah total peminjaman buku per anggota (tingkat aktivitas). |
| 4 | Usia Peminjam | Usia anggota yang melakukan peminjaman. |

B. Tiga Atribut Utama Fokus *Clustering*

Meskipun terdapat empat atribut dalam *Dataset*, tiga atribut berikut dipilih sebagai fokus utama dalam proses *Hybrid Clustering* (*K-Means* + *Fuzzy C-means*) karena ketiganya secara kolektif mendefinisikan dimensi paling esensial dari Pola Minat Baca seseorang:

Tabel 3. 2 Tiga Atribut Utama Fokus *Clustering*

| No | Fokus Utama | Dimensi Pola Minat Baca | Justifikasi Penggunaan dalam <i>Clustering</i> |
|----|----------------------|---|--|
| 1 | Genre/Kategori Buku | Dimensi 'Apa' yang dibaca (Preferensi Konten). | Merupakan indikator langsung terhadap minat subjek spesifik, membedakan kelompok berdasarkan tema bacaan (Minat tinggi pada Fiksi vs. Sains). |
| 2 | Frekuensi Peminjaman | Dimensi 'Intensitas' membaca (Tingkat Aktivitas). | Mengukur volume dan konsistensi peminjaman, yang membedakan kelompok pembaca aktif, sedang, dan pasif. |
| 3 | Usia Peminjam | Dimensi 'Siapa' yang membaca (Segmentasi Demografis). | Variabel demografis kunci yang paling kuat memengaruhi jenis dan intensitas minat baca, memungkinkan klasterisasi yang relevan untuk intervensi kebijakan. |

3.1.3 Pra-pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk memastikan data bersih dan siap diproses melalui serangkaian langkah sistematis sebagai berikut:

1. Pembersihan Data (*Data Cleaning*): Meliputi penyesuaian nama kolom (*renaming*), konversi tipe data usia dan frekuensi menjadi numerik, penyeragaman format penulisan genre buku menggunakan *Title Case*, serta penghapusan entri yang memiliki nilai kosong (*missing values*) pada atribut utama.

2. Agregasi dan *User Profiling*: Melakukan pengelompokan data secara bertahap. Tahap pertama berdasarkan nama dan genre untuk akumulasi frekuensi, dan tahap kedua untuk menentukan profil akhir setiap anggota yang mencakup rata-rata usia, total frekuensi, Genre Dominan, serta klasifikasi Tipe Pembaca (*Single* atau *Multi Genre*).
3. Rekayasa Fitur (*Feature Engineering*): Menghasilkan fitur baru berupa "Popularitas Genre" yang dihitung berdasarkan frekuensi kemunculan setiap genre dominan di seluruh populasi dataset untuk memperkuat dimensi preferensi dalam pemodelan.
4. Transformasi *Yeo-Johnson*: Menerapkan *PowerTransformer* dengan metode *Yeo-Johnson* untuk meminimalkan kemiringan distribusi data (*skewness*) dan menstabilkan varians, mengingat dataset diolah secara utuh tanpa melakukan pembuangan nilai ekstrem (*outliers*).
5. Normalisasi dan *Scaling*: Melakukan penskalaan fitur menggunakan *RobustScaler* yang berbasis pada *median* dan rentang *interkuartil* (IQR), sehingga setiap atribut memiliki skala yang seragam dan tangguh terhadap pengaruh nilai ekstrem dalam perhitungan jarak kluster.

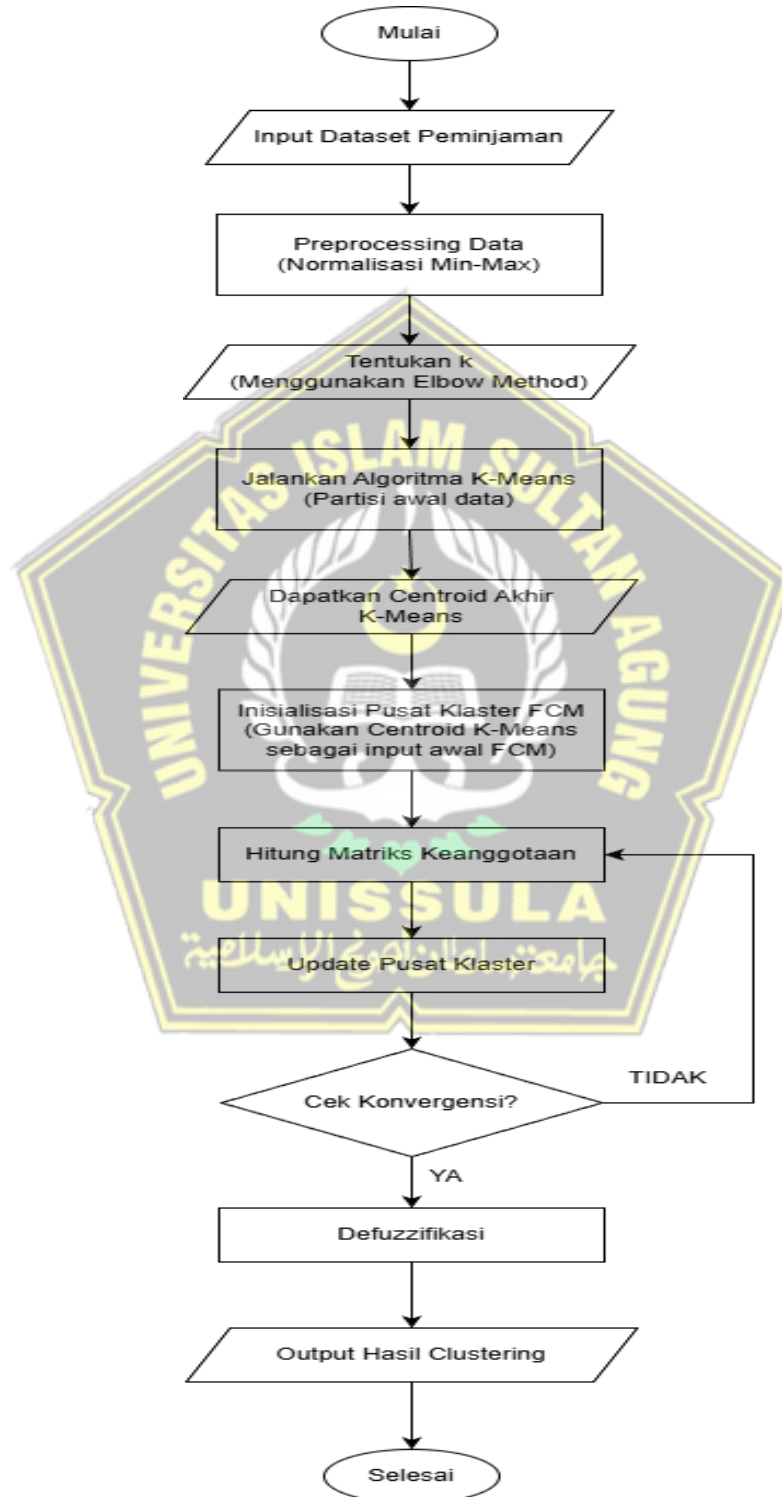
3.1.4 Penerapan Model *Hybrid Clustering (K-Means + Fuzzy C-Means)*

Model *Hybrid Clustering* digunakan untuk mengatasi keterbatasan masing-masing algoritma. Tahapannya adalah:

1. Penentuan K: Menentukan jumlah kluster optimal (k) menggunakan *Elbow Method*.
2. Fase 1 (*K-Means*): Menjalankan algoritma *K-Means* pada *Dataset* yang telah dinormalisasi. *Output* dari tahap ini adalah koordinat *Centroid final*.
3. Inisialisasi *Fuzzy C-Means*: Memasukkan koordinat *Centroid* hasil *K-Means* sebagai pusat kluster awal untuk algoritma *Fuzzy C-Means*.
4. Fase 2 (*Fuzzy C-Means*): Menjalankan algoritma *Fuzzy C-Means* untuk menghitung matriks keanggotaan (*Membership Matrix*) setiap data terhadap kluster.
5. Defuzzifikasi: Menentukan kluster dominan bagi setiap anggota perpustakaan berdasarkan nilai keanggotaan tertinggi, namun tetap

menyimpan informasi derajat keanggotaan lainnya untuk analisis lanjutan.

Flowchart model *Hybrid Clustering* ditunjukkan pada Gambar 3.2.



Gambar 3.2 Flowchart *Hybrid Clustering*

3.1.5 *Profiling dan Analisis Karakteristik*

Setelah data berhasil diklaster dan divalidasi kualitasnya, tahap selanjutnya adalah melakukan analisis profil (*profiling*) untuk mengidentifikasi serta menafsirkan karakteristik dominan dari setiap kelompok yang terbentuk. Proses ini bertujuan untuk mentransformasi label klaster numerik yang abstrak menjadi informasi perilaku pemustaka yang bermakna secara manajerial. Pencarian karakteristik ini dilakukan melalui pendekatan statistik deskriptif yang mencakup empat tahapan analisis utama sebagai berikut:

1. Analisis Statistik Numerik Dasar (*Central Tendency*): Analisis ini dilakukan dengan menerapkan fungsi agregasi mean, min, dan max pada variabel Usia dan Frekuensi Peminjaman untuk setiap klaster. Melalui perhitungan nilai rata-rata, peneliti dapat memetakan posisi demografi umur (misalnya kelompok anak-anak, remaja, atau dewasa) serta tingkat loyalitas anggota dalam meminjam buku secara kuantitatif. Nilai minimum dan maksimum digunakan untuk melihat rentang sebaran data agar profil yang dibentuk tetap berada dalam batas observasi yang valid.
2. Analisis Preferensi Kategori (*Rank-based Genre Analysis*): Untuk mengetahui identitas minat utama pada setiap klaster, dilakukan analisis terhadap variabel Genre. Metodenya adalah dengan menghitung akumulasi skor kepopuleran berdasarkan total frekuensi peminjaman pada setiap genre di dalam masing-masing klaster. Genre dengan frekuensi kumulatif tertinggi ditetapkan sebagai Genre Favorit, yang merepresentasikan identitas topik bacaan utama dari kelompok tersebut.
3. Analisis Keragaman Minat Baca (*Cross-Tabulation Analysis*): Analisis ini bertujuan untuk menentukan tipe perilaku membaca anggota klaster, apakah cenderung spesifik pada satu bidang atau bersifat eksploratif. Secara teknis, metode yang digunakan adalah tabulasi silang (*cross-tabulation*) antara label klaster dengan variabel Tipe Pembaca (*Single Genre* atau *Multi Genre*). Hasil tabulasi kemudian dinormalisasi untuk

mendapatkan persentase kontribusi setiap tipe pembaca terhadap total populasi dalam klaster tersebut.

4. Penentuan Karakteristik Karakter (*Logic-based Thresholding*): Setelah mendapatkan persentase keragaman, dilakukan klasifikasi karakter secara kualitatif. Klaster ditetapkan memiliki karakteristik "Eksploratif" jika proporsi pembaca Multi Genre mencapai ambang batas (threshold) $\geq 50\%$. Sebaliknya, jika proporsi tersebut berada di bawah 50%, maka klaster diklasifikasikan memiliki karakteristik "Spesifik". Tahap terakhir adalah mengompilasi seluruh temuan ke dalam profil ringkas yang mencakup jumlah anggota, persentase terhadap total populasi, serta pelabelan identitas semantik untuk kebutuhan pengambilan keputusan strategis.

3.1.6 Transformasi dan Pelabelan Data (*Data Labeling*)

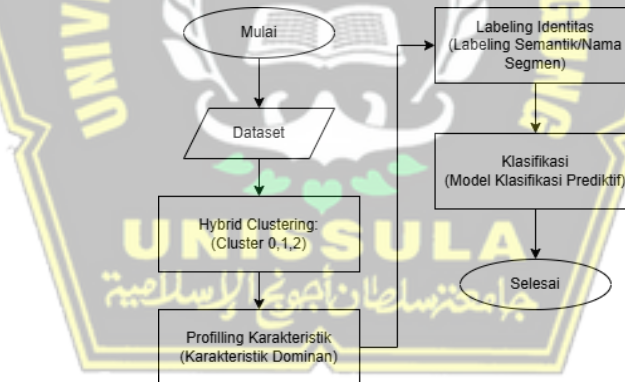
Tahap transformasi dan pelabelan data merupakan fase krusial dalam penelitian ini untuk menjembatani hasil komputasi numerik dengan interpretasi manajerial yang aplikatif. Proses ini bertujuan untuk mengonversi identitas klaster yang awalnya bersifat abstrak, seperti Cluster 0, 1, dan 2, menjadi label semantik deskriptif berdasarkan temuan signifikan pada tahap profiling yang telah dilakukan sebelumnya. Penamaan setiap segmen dilakukan dengan mempertimbangkan karakteristik dominan yang muncul pada dimensi demografi, tingkat aktivitas, dan variasi preferensi genre guna merepresentasikan perilaku nyata pemustaka secara akurat. Melalui mekanisme ini, dataset utama yang telah terlabeli akan memiliki variabel target (y) yang jelas, yang kemudian digunakan sebagai basis data untuk melatih model klasifikasi *LightGBM* agar mampu mengenali aturan keputusan (*decision rules*) secara otomatis.

Adapun label semantik yang direncanakan dalam penelitian ini mencakup tiga identitas utama, yaitu 'Pembaca Muda Selektif', 'Pembaca Mapan Antusias', dan 'Pembaca Belia Eksploratif'. Label 'Pembaca Muda Selektif' dirancang untuk merepresentasikan kelompok usia produktif dengan minat yang terfokus pada genre tertentu. Sementara itu, 'Pembaca Mapan

Antusias’ ditujukan untuk menggambarkan kelompok usia dewasa yang memiliki loyalitas serta frekuensi peminjaman tinggi, dan ‘Pembaca Belia Eksploratif’ digunakan untuk menandai kelompok pembaca usia anak-anak atau remaja yang menunjukkan rasa ingin tahu tinggi terhadap berbagai jenis bacaan. Dengan melakukan transformasi ini, keluaran sistem tidak lagi sekadar angka, melainkan informasi strategis yang mudah dipahami oleh pihak manajemen perpustakaan dalam merancang program peningkatan literasi yang tepat sasaran.

3.1.7 Alur Integrasi Clustering, Profiling, Labeling, dan Klasifikasi

Tahapan ini merupakan jembatan metodologis yang sangat krusial karena menghubungkan pendekatan unsupervised learning untuk penemuan pola tersembunyi dengan supervised learning untuk pembangunan model prediksi otomatis. Berdasarkan Gambar 3.3, alur ini dirancang agar data tidak hanya berhenti sebagai kelompok angka, melainkan bertransformasi menjadi informasi strategis yang dapat memprediksi profil pemustaka secara instan.



Gambar 3. 3 Flowchart Proses Clustering ke Klasifikasi

Adapun penjelasan detail mengenai keempat fase integrasi yang terdapat dalam alur penelitian ini adalah sebagai berikut:

1. Fase *Hybrid Clustering (K-Means + Fuzzy C-Means)*

Proses dimulai dengan menerapkan algoritma *Hybrid K-Means* dan *Fuzzy C-Means (FCM)* pada 3.112 data pemustaka aktif yang telah melalui tahap pra-pemrosesan. Pada tahap ini, *K-Means* berperan sebagai inisiator untuk menentukan koordinat pusat klaster (*centroid*) secara cepat guna menghindari jebakan *local optima* yang sering terjadi pada

inisialisasi acak. Hasil dari koordinat tersebut kemudian diteruskan ke algoritma FCM untuk menghitung derajat keanggotaan setiap data secara lebih halus, sehingga setiap individu mendapatkan identitas berupa nomor klaster (Cluster 0, 1, atau 2) berdasarkan nilai keanggotaan tertinggi. Keluaran utama dari fase ini adalah pemisahan data ke dalam tiga kelompok besar yang masih bersifat numerik dan abstrak.

2. Fase *Profiling* Karakteristik

Setelah setiap data memiliki identitas klaster, langkah berikutnya adalah melakukan bedah karakteristik melalui analisis profil atau *profiling*. Peneliti melakukan analisis statistik deskriptif terhadap variabel usia dan frekuensi peminjaman untuk melihat nilai rata-rata (*mean*), minimum, dan maksimum pada setiap kelompok yang terbentuk. Selain variabel numerik, dilakukan pula analisis preferensi kategori buku berdasarkan peringkat frekuensi peminjaman untuk mengidentifikasi genre favorit di setiap klaster. Fase ini bertujuan untuk mengubah label angka yang abstrak menjadi informasi perilaku pemustaka yang terukur secara kuantitatif, sehingga peneliti dapat memahami perbedaan riil antar kelompok, seperti perbedaan antara kelompok pembaca usia anak-anak dengan kelompok dewasa.

3. Fase *Labeling* Identitas Semantik

Temuan signifikan pada tahap *profiling* kemudian digunakan sebagai dasar utama untuk melakukan transformasi identitas melalui mesin pelabelan (*labeling engine*). Pada fase ini, nomor klaster angka (seperti Cluster 0, 1, dan 2) dikonversi menjadi label semantik yang deskriptif dan mencerminkan perilaku nyata di lapangan. Penamaan setiap segmen, seperti ‘Pembaca Muda Selektif’, ‘Pembaca Mapan Antusias’, atau ‘Pembaca Belia Eksploratif’, dilakukan dengan mempertimbangkan karakteristik dominan pada dimensi demografi dan tingkat aktivitas peminjaman. Dengan selesainya tahap ini, dataset kini memiliki variabel target (*y*) yang definitif, yang secara teknis mengubah

status data dari tidak terlabeli (*unlabeled*) menjadi data terlabeli (*labeled*) yang siap untuk tahap pembelajaran mesin selanjutnya.

4. Fase Klasifikasi Prediktif (*LightGBM*)

Alur integrasi ini diakhiri dengan pembangunan model klasifikasi otomatis menggunakan algoritma *Light Gradient Boosting Machine (LightGBM)*. Model ini dilatih menggunakan fitur independen berupa usia, frekuensi, dan popularitas genre untuk mempelajari aturan keputusan (*decision rules*) menuju *label* semantik yang telah dibentuk pada tahap sebelumnya. Tujuan utama dari fase ini adalah untuk memvalidasi stabilitas kluster sekaligus menciptakan mesin prediksi cerdas yang dapat diintegrasikan ke dalam *dashboard web*. Melalui fase klasifikasi ini, pihak perpustakaan tidak perlu lagi melakukan proses klustering ulang secara manual dikarenakan sistem secara otomatis akan mampu mengkategorikan profil anggota baru sesaat setelah data atributnya dimasukkan ke dalam sistem.

3.1.8 Pembangunan Model Klasifikasi Prediktif Menggunakan *LightGBM*

Setelah profil dari setiap kluster terbentuk, tahap selanjutnya adalah memasuki fase *Supervised Learning* (pembelajaran terawasi) untuk membangun sistem klasifikasi otomatis. Proses ini bertujuan untuk menciptakan model cerdas yang mampu memprediksi kategori minat baca anggota baru hanya berdasarkan data atributnya, tanpa perlu melakukan proses klustering ulang dari awal. Pembangunan model klasifikasi ini dilakukan menggunakan algoritma *Light Gradient boosting Machine (LightGBM)* yang mencakup tiga tahapan proses utama:

1. Persiapan dan Pembagian Data (*Data Splitting*)

Dilakukan dengan memisahkan variabel fitur (x) yang terdiri dari Usia, Frekuensi Peminjaman, dan Popularitas Genre, serta variabel target (y) yaitu Label Kluster. Data kemudian dibagi menjadi dua bagian menggunakan rasio 70:30 secara *stratified*. Tujuannya adalah menyediakan 70% data untuk melatih model (*Training*) agar mempelajari pola, dan menyisihkan 30% data untuk pengujian (*Testing*)

guna memvalidasi keakuratan prediksi model pada data yang belum pernah dilihat sebelumnya.

2. Pelatihan Model (*Model Training*)

Algoritma *LightGBM* digunakan untuk mempelajari hubungan pola non-linear antara atribut demografis (Usia) dan perilaku (Frekuensi & Genre) terhadap kategori klaster yang sudah terbentuk dari tahap sebelumnya. Tujuannya adalah agar sistem dapat membentuk aturan keputusan (*decision rules*) yang optimal dalam menentukan klasifikasi pemustaka secara otomatis, apakah termasuk ke dalam kategori ‘Pembaca Muda Selektif’, ‘Pembaca Mapan Antusias’, atau ‘Pembaca Belia Eksploratif’. Dengan proses ini, sistem mampu mengenali karakteristik unik seperti dominasi usia maupun tingginya loyalitas untuk melakukan prediksi pada data anggota baru.

3. Evaluasi Kinerja dan Analisis Fitur (*Evaluation & Feature Importance*)

Dilakukan dengan mengukur metrik akurasi serta menganalisis *Confusion Matrix* untuk melihat detail kesalahan prediksi antar kelas. Selain itu, dilakukan analisis *Feature Importance* untuk mengurutkan variabel mana yang paling berkontribusi dalam pengambilan keputusan model. Tujuannya adalah untuk memastikan model memiliki performa yang handal (akurat) sekaligus transparan dalam menjelaskan faktor dominan yang mempengaruhi klasifikasi minat baca.

3.1.9 Evaluasi Model

Evaluasi model dilakukan untuk mengukur kualitas hasil Clustering serta performa dari model klasifikasi yang telah dibangun. Metode evaluasi dalam penelitian ini mencakup dua aspek utama:

1. Evaluasi Kualitas Klaster (*Unsupervised*)

Tahap ini bertujuan untuk memastikan bahwa kelompok (klaster) yang terbentuk memiliki validitas yang tinggi sebelum dilanjutkan ke tahap klasifikasi:

- a. *Silhouette Coefficient*: Digunakan untuk mengukur kualitas klaster berdasarkan tingkat kepadatan internal dan pemisahan antar-klaster.

- b. *Partition Coefficient (PC) & Partition Entropy (PE)*: Metrik khusus ini digunakan untuk mengukur kualitas validitas kluster pada metode *Fuzzy*. Semakin besar nilai PC dan semakin kecil nilai PE, maka semakin baik kualitas derajat keanggotaan kluster yang dihasilkan.
 - c. Perbandingan Hasil: Dilakukan perbandingan antara hasil *Clustering* tunggal (*K-Means* dan *Fuzzy C-Means*) dengan *Hybrid Clustering* untuk mengetahui peningkatan akurasi dan kestabilan pemetaan pola minat baca.
2. Evaluasi Performa Klasifikasi (*Supervised*)
- Setelah data terlabeli maka akan dilakukan proses klasifikasi dengan algoritma *LightBGM* adapun kinerja algoritma *LightGBM* diukur menggunakan metrik berikut:
- a. *Accuracy*: Metrik untuk mengukur tingkat kedekatan hasil prediksi dengan nilai aktual secara keseluruhan.
 - b. *Precision*: Metrik untuk mengukur keakuratan model dalam memprediksi kelas positif secara tepat.
 - c. *Recall*: Metrik untuk mengukur keberhasilan model dalam memprediksi kembali seluruh data yang sebenarnya berada pada kelas positif.
 - d. *F1-score*: Nilai rata-rata harmonik antara *precision* dan *recall* yang digunakan untuk mengevaluasi performa model secara lebih seimbang, terutama jika terdapat ketidakseimbangan jumlah data antar kluster.

3.1.10 Implementasi Visualisasi

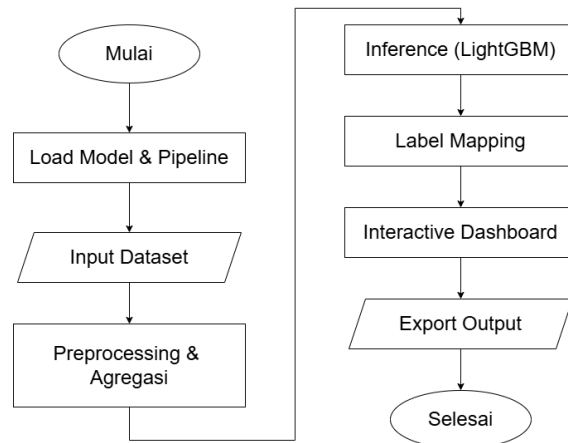
Tahap akhir dari penelitian ini adalah pembangunan sistem visualisasi dalam bentuk *Executive Dashboard* interaktif menggunakan framework Streamlit. Berbeda dengan visualisasi statis, sistem ini dirancang sebagai *Decision Support System (DSS)* yang mampu mengintegrasikan skrip sains data dengan antarmuka pengguna secara *real-time* guna mendukung pengambilan keputusan oleh pengelola perpustakaan. Proses dimulai dengan penyediaan fitur data uploader untuk *dataset* format *.csv* atau *.xlsx*, yang

kemudian diproses secara otomatis oleh *core engine* aplikasi melalui tahapan pembersihan dan agregasi data. Keunggulan utama sistem ini terletak pada kemampuannya melakukan *real-time inference* menggunakan model LightGBM yang dimuat melalui file .pkl, sehingga kategori profil anggota baru dapat diprediksi secara instan sesaat setelah data diunggah ke dalam sistem.

Selanjutnya, hasil analisis tersebut disajikan melalui berbagai grafik dinamis berbasis *library Plotly*, meliputi *Radar Chart* untuk pemetaan karakteristik simultan (usia, frekuensi, dan keragaman), serta *Pie Chart* dan *Bar Chart* untuk memvisualisasikan proporsi populasi tiap klaster dan dominasi genre buku secara interaktif. Selain aspek visual, sistem ini dilengkapi dengan fitur rekomendasi strategis otomatis yang memberikan saran kebijakan spesifik, seperti fokus pengadaan buku atau strategi kampanye promosi, berdasarkan karakteristik segmen yang terdeteksi paling dominan maupun yang paling tidak aktif. Seluruh hasil klasifikasi yang telah melalui proses pelabelan semantik seperti ‘Pembaca Muda Selektif’, ‘Pembaca Mapan Antusias’, dan ‘Pembaca Belia Eksploratif’ dapat diekspor oleh pengguna ke dalam format laporan Microsoft Excel (.xlsx) untuk kebutuhan dokumentasi resmi pihak perpustakaan.

3.2 Rancangann Sistem

Rancangan sistem ini menggambarkan alur kerja Smart Library Analytics System yang mengintegrasikan hasil analisis data (Hybrid Clustering) ke dalam model prediktif (LightGBM) dan disajikan melalui Executive Dashboard. *Flowchart* rancangan sistem ditunjukkan pada Gambar 3.4.



Gambar 3. 4 *Flowchart* Rancangan Sistem

Penjelasan detail tahapan sistem berdasarkan Gambar 3.4 adalah sebagai berikut:

1. Tahap *Input Data*

Sistem menerima *input* berupa *dataset* sekunder peminjaman buku (format .csv atau .xlsx). Atribut utama yang diproses meliputi:

- a. Identitas: Nama Lengkap (sebagai kunci agregasi).
- b. Demografi: Usia Peminjam.
- c. Aktivitas: Frekuensi Peminjaman.
- d. Preferensi: Genre Buku.

2. Tahap Pemrosesan (*Core Engine*)

Tahap ini merupakan inti dari aplikasi yang dibangun menggunakan *framework Streamlit* dengan langkah-langkah:

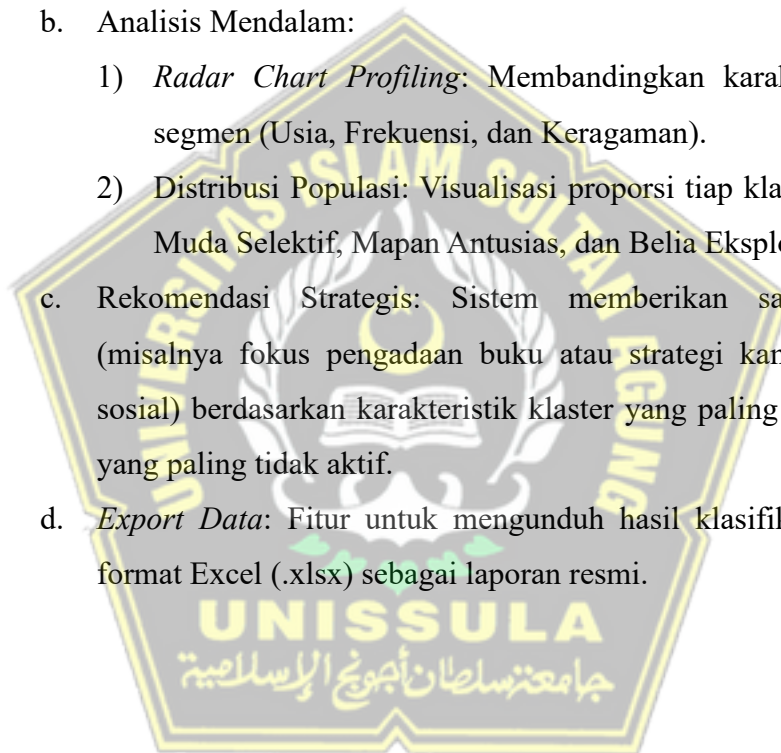
- a. *Preprocessing & Agregasi*: Melakukan pembersihan data, konversi tipe data numerik, dan agregasi data berdasarkan nama pengguna. Sistem secara otomatis menentukan Genre Dominan untuk setiap individu menggunakan fungsi mode.
- b. Sinkronisasi Mode: Memuat (*loading*) resource yang telah dilatih sebelumnya (*Pre-trained Model*), yaitu file `model_lgbm_ta.pkl` dan *metadata* `pipeline_resources.pkl`.
- c. Prediksi Klasifikasi (*LightGBM*): Data baru yang diunggah akan diklasifikasikan secara *real-time* ke dalam salah satu dari tiga

segmen psikografis menggunakan algoritma *LightGBM* berdasarkan fitur usia, frekuensi, dan genre.

3. Tahap *Output* (Visualisasi & *Decision Support*)

Output sistem disajikan dalam bentuk *Interactive Dashboard* yang terbagi menjadi tiga fungsi utama:

- a. Ringkasan Eksekutif: Menampilkan metrik utama seperti total sampel, rata-rata usia, rata-rata kunjungan, dan genre paling populer di perpustakaan.
- b. Analisis Mendalam:
 - 1) *Radar Chart Profiling*: Membandingkan karakteristik antar segmen (Usia, Frekuensi, dan Keragaman).
 - 2) Distribusi Populasi: Visualisasi proporsi tiap kluster (Pembaca Muda Selektif, Mapan Antusias, dan Belia Eksploratif).
- c. Rekomendasi Strategis: Sistem memberikan saran otomatis (misalnya fokus pengadaan buku atau strategi kampanye media sosial) berdasarkan karakteristik kluster yang paling dominan atau yang paling tidak aktif.
- d. *Export Data*: Fitur untuk mengunduh hasil klasifikasi ke dalam format Excel (.xlsx) sebagai laporan resmi.



BAB IV

HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil dan analisis mendalam mengenai penerapan algoritma *Hybrid K-Means* dan *Fuzzy C-Means* dalam memetakan pola minat baca di Perpustakaan Provinsi Jawa Tengah. Pembahasan diawali dengan tahap persiapan data dan identifikasi pustaka pemrograman, dilanjutkan dengan proses pemodelan klustering serta profilisasi karakteristik unik dari setiap kelompok yang terbentuk. Bab ini diakhiri dengan pengujian validitas model menggunakan algoritma *LightGBM* serta demonstrasi implementasi model ke dalam dashboard berbasis *web* sebagai alat pendukung keputusan strategis.

4.1 *Library* yang Digunakan

Dalam penelitian ini, beberapa *Library* Python digunakan untuk membantu proses analisis pola minat baca pada google collab:

```
import io
import os
import time
import warnings
from IPython.display import display
import numpy as np
import pandas as pd
from scipy.stats import mstats
from google.colab import files
from sklearn.preprocessing import StandardScaler, PowerTransformer, RobustScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, accuracy_score, classification_report, confusion_matrix
from sklearn.model_selection import train_test_split
import skfuzzy as fuzz
import lightgbm as lgb
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
```

Gambar 4. 1 *Library* Python

Berdasarkan Gambar 4.1, dapat dijelaskan bahwa pengembangan sistem ini mengintegrasikan berbagai pustaka (*Library*) Python yang dikelompokkan menurut peran fungsionalnya dalam alur kerja data. Tahap awal melibatkan penggunaan pustaka standar dan utilitas sistem seperti `io` dan `os` untuk manajemen aliran data serta `file`, `time` untuk pemantauan durasi eksekusi program, serta `warnings` guna memastikan tampilan laporan hasil eksekusi tetap bersih dari pesan peringatan sistem. Selain itu, digunakan pula pustaka `IPython.display` dan `google.colab` untuk mendukung fungsionalitas

interaktif di lingkungan kerja *Cloud*, yang memudahkan proses pemuatan data serta presentasi objek secara rapi. Fondasi pengolahan data dan analisis statistik dibangun menggunakan pustaka `numpy` untuk operasi numerik berbasis matriks, `pandas` untuk manipulasi *Dataset* dalam format `DataFrame`, serta pustaka `scipy.stats` (`mstats`) guna menangani analisis statistik tingkat lanjut, terutama pada data yang memiliki nilai ekstrem (*Outliers*).

Pada tahap persiapan data (*preprocessing*), pustaka `StandardScaler`, `PowerTransformer`, dan `RobustScaler` diterapkan untuk menormalisasi skala fitur agar seimbang, sementara `train_test_split` digunakan untuk membagi *Dataset* menjadi data latih dan data uji secara proporsional. Inti dari kecerdasan buatan yang dibangun dalam penelitian ini diimplementasikan melalui algoritma *K-Means* untuk pengelompokan data, `skfuzzy` guna mengakomodasi logika *fuzzy* dalam proses *Clustering*, serta *LightGBM* sebagai algoritma *gradient boosting* untuk efisiensi klasifikasi data. Performa model kemudian diukur menggunakan metrik `silhouette_score` untuk mengevaluasi kualitas pengelompokan, serta `accuracy_score`, `classification_report`, dan `confusion_matrix` untuk menguji akurasi prediksi. Seluruh hasil analisis tersebut dipresentasikan secara visual melalui pustaka `matplotlib.pyplot` dan `seaborn` untuk grafik statis, serta `plotly.express` dan `graph_objects` guna menghasilkan visualisasi interaktif yang memungkinkan eksplorasi data secara lebih mendalam.

4.2 *Dataset* yang Digunakan

Pada tahap pengumpulan data, ekstraksi metadata dilakukan dari sistem informasi Perpustakaan Provinsi Jawa Tengah dengan izin yang sudah diberikan oleh pihak perputakaan. *Dataset* yang diperoleh memiliki beberapa atribut utama yang merepresentasikan karakteristik setiap dokumen. Gambar 4.2 menunjukkan sebagian *Dataset* yang akan digunakan. Jumlah *Dataset* yang digunakan berjumlah sebanyak 14.602 *entri*.

| Nama_Lengkap | Usia_Peminjam | Genre_Buku | Frekuensi_Peminjam |
|----------------------------------|---------------|-----------------|--------------------|
| ALYA ASHALINA FIRDAUSI | 8 | Kesusastraan | 1 |
| ALYA ASHALINA FIRDAUSI | 8 | Kesusastraan | 1 |
| AQILA MISHA IZZATUNNISA PRASETYO | 8 | Kesenian | 1 |
| AQILA MISHA IZZATUNNISA PRASETYO | 8 | Kesenian | 1 |
| ALYA ASHALINA FIRDAUSI | 8 | Karya Umum | 1 |
| ALYA ASHALINA FIRDAUSI | 8 | Ilmu-ilmu Murni | 1 |
| ALYA ASHALINA FIRDAUSI | 8 | Ilmu-ilmu Murni | 1 |
| ALYA ASHALINA FIRDAUSI | 8 | Ilmu Sosial | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |
| RAHAYU | 67 | Kesusastraan | 1 |

Gambar 4. 2 *Dataset* Peminjaman Buku

Dari hasil pengumpulan data ini, dapat disimpulkan bahwa *Dataset* yang diperoleh berisi informasi lengkap mengenai dokumen-dokumen peminjaman buku di Perpustakaan Provinsi Jawa Tengah. Adapun 3 atribut yang digunakan antara lain:

1. Usia_Peminjam : Menjadi indikator untuk melihat distribusi minat baca berdasarkan kelompok generasi.
2. Genre_Buku: Menentukan preferensi topik (seperti Kesusastraan, Kesenian, atau Ilmu Sosial) yang paling sering diakses.
3. Frekuensi_Peminjam: Mengukur seberapa aktif masyarakat dalam memanfaatkan fasilitas perpustakaan.

4.3 Hasil *Preprocessing Dataset*

Tahap *preprocessing* merupakan langkah krusial untuk mengubah data transaksi perpustakaan yang masih mentah menjadi *Dataset* yang siap diolah oleh model *Hybrid K-Means* dan *Fuzzy C-Means*. Proses ini bertujuan untuk meningkatkan kualitas data, dan memastikan setiap fitur memiliki skala yang seragam sehingga proses klastering menghasilkan pengelompokan pola minat baca yang akurat di Kota Semarang. Proses ini dibagi menjadi beberapa tahapan utama sebagai berikut:

1. Normalisasi dan Pembersihan Data Dasar

Langkah awal dalam *preprocessing* adalah melakukan penyeragaman nama atribut untuk memudahkan pemanggilan variabel dalam program. Atribut Nama_Lengkap, Usia_Peminjam, Genre_Buku, dan Frekuensi_Peminjaman diubah menjadi format yang lebih ringkas. Selain itu, dilakukan konversi tipe data numerik pada atribut usia dan

frekuensi serta penghapusan data yang kosong (*missing values*) menggunakan fungsi `dropna`. Hal ini dilakukan untuk memastikan integritas data sebelum dilakukan perhitungan statistik.

2. Agregasi Bertahap dan Profiling Pengguna

Setelah data dibersihkan, langkah berikutnya adalah melakukan transformasi dari data transaksi menjadi data profil individu, yang dapat dilihat pada Gambar 4.3.

 ✦ LANGKAH 2: AGREGASI & PROFILING

◆ Tahap 1: Grup berdasarkan nama + genre ...

| | nama | genre | usia | freq |
|---|--------------------------|------------------------|------|------|
| 0 | KEVIN MARCELLO PASHA | Filsafat Dan Psikologi | 18.0 | 6 |
| 1 | KEVIN MARCELLO PASHA | Ilmu Sosial | 18.0 | 1 |
| 2 | AAD HARITS SANDI SAPUTRA | Teknologi | 24.0 | 1 |
| 3 | AAN NAISABURY | Agama | 40.0 | 1 |
| 4 | AAQILATUNNAJAH | Filsafat Dan Psikologi | 18.0 | 3 |

◆ Tahap 2: Profil User & Penentuan Genre Dominan ...

✦ Hasil Akhir Agregasi (Full Data):

| | nama | usia | freq | Genre_List | Genre_Dominan | Tipe_Pembaca |
|---|--------------------------|------|------|---------------------------------------|------------------------|--------------|
| 0 | KEVIN MARCELLO PASHA | 18.0 | 7 | [Filsafat Dan Psikologi, Ilmu Sosial] | Filsafat Dan Psikologi | Multi Genre |
| 1 | AAD HARITS SANDI SAPUTRA | 24.0 | 1 | [Teknologi] | Teknologi | Single Genre |
| 2 | AAN NAISABURY | 40.0 | 1 | [Agama] | Agama | Single Genre |
| 3 | AAQILATUNNAJAH | 18.0 | 4 | [Filsafat Dan Psikologi, Geografi] | Filsafat Dan Psikologi | Multi Genre |
| 4 | AAS FIBRIYANTI | 42.0 | 3 | [Kesusastaan, Filsafat Dan Psikologi] | Kesusastaan | Multi Genre |

✓ Total User Terproses: 3112

Gambar 4. 3 Agregasi 2 Tahap

Berdasarkan Gambar 4.3, sistem melakukan agregasi dalam dua tingkatan. Pertama, data dikelompokkan berdasarkan nama dan genre untuk menghitung total frekuensi peminjaman. Kedua, dilakukan pembentukan profil pengguna yang mencakup penentuan "Genre Dominan" dan kategori "Tipe Pembaca". Hasil akhir dari proses ini menghasilkan *Dataset* terfilter sebanyak 3.112 user yang merepresentasikan populasi aktif peminjam di perpustakaan.

3. Transformasi Distribusi dan Penskalaan (*Scaling*)

Tahap akhir dari *preprocessing* adalah menyiapkan fitur dalam skala yang seragam melalui proses transformasi distribusi dan penskalaan, sebagaimana ditunjukkan pada Gambar 4.4.

```

genre_pop = df_grouped['Genre_Dominan'].value_counts().to_dict()
df_grouped['genre_popularity'] = df_grouped['Genre_Dominan'].map(genre_pop)

features = ['usia', 'freq', 'genre_popularity']
X_raw = df_grouped[features].values

pt = PowerTransformer(method='yeo-johnson')
X_transformed = pt.fit_transform(X_raw)

scaler = RobustScaler()
X_scaled = scaler.fit_transform(X_transformed)

X_final_df = pd.DataFrame(X_scaled, columns=['usia_s', 'freq_s', 'genre_s'])

```

Gambar 4. 4 Kode *Transformasi dan Distribusi Data*

Berdasarkan Gambar 4.4, data melalui tahap *Yeo-Johnson Transformation* untuk memperbaiki distribusi agar mendekati normal (*Gaussian*) dan *RobustScaler* untuk menyamakan rentang nilai antar fitur. Hasil transformasi ini menghasilkan tiga variabel utama yang siap digunakan sebagai input model, yaitu *usia_s*, *freq_s*, dan *genre_s*. Dengan skala yang telah seragam, setiap fitur memiliki bobot yang adil dalam menentukan jarak antar data pada saat proses klustering menggunakan *Hybrid K-Means* dan *Fuzzy C-Means* dilakukan.

4.4 Hasil *Clustering* Menggunakan *Hybrid K-Means* dan *Fuzzy C-Means*

Pendekatan *Hybrid Clustering* dilakukan dengan mengintegrasikan keunggulan algoritma *K-Means* dan *Fuzzy C-Means* (FCM). Dalam metodologi ini, algoritma *K-Means* digunakan pada tahap awal untuk menentukan pusat klaster (*centroid*) yang akan dijadikan nilai awal (*initialization*) bagi matriks keanggotaan (*u*) pada *Fuzzy C-Means*.

Tujuan utama dari penggabungan ini adalah untuk mengatasi sensitivitas *Fuzzy C-Means* terhadap inisialisasi acak, sehingga proses konvergensi fungsi objektif J_m dapat dicapai lebih cepat dan stabil. Hasil akhir dari pemodelan ini memberikan gambaran segmentasi masyarakat berdasarkan variabel usia dan frekuensi membaca dengan tingkat keanggotaan *fuzzy* (*membership degree*) yang akurat.

4.4.1 Penentuan Nilai k

Pada tahap awal pemodelan, dilakukan pencarian jumlah kluster optimal (k) menggunakan algoritma *K-Means* melalui dua pendekatan utama, yaitu *Elbow Method* dan *Silhouette Coefficient*.

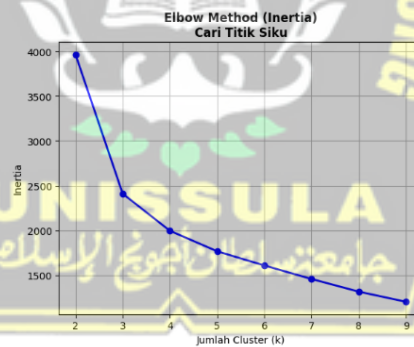
```

=====
=== TAHAP 1: TUNING K (ELBOW METHOD dengan K-MEANS) ===
=====
K      Inertia (SSE)      Silhouette Score
-----
2      3795.06             0.4407
3      2316.61             0.4426
4      1905.98             0.3262
5      1685.15             0.2873
6      1537.99             0.2755
7      1395.12             0.2790
8      1258.06             0.2718
9      1150.32             0.2746
-----
★ Saran Sistem: K = 3 (Silhouette Score tertinggi: 0.4426)

```

Gambar 4. 5 Hasil Tuning Nilai k

Berdasarkan Gambar 4.5, pengujian dilakukan pada rentang nilai k antara 2 sampai 9 untuk melihat stabilitas pengelompokan. Gambar 4.5 menunjukkan perbandingan nilai *Inertia* dan *Silhouette Score* yang dihasilkan oleh sistem.



Gambar 4. 6 *Elbow Method* Nilai k

Berdasarkan grafik *Elbow Method* pada Gambar 4.6, penurunan nilai *Inertia* yang paling signifikan terjadi hingga titik $k=3$, di mana setelah titik tersebut grafik mulai melandai (membentuk siku). Hal ini didukung oleh skor *Silhouette* yang mencapai nilai tertinggi pada $k=3$ dengan skor 0.4426. Dengan demikian, dapat disimpulkan bahwa pembagian *Dataset* ke dalam 3 kelompok memberikan hasil pemisahan yang paling optimal.

4.4.2 Penentuan Nilai m

Setelah nilai k ditetapkan, tahap selanjutnya adalah menentukan parameter keberatan (*fuzziness exponent*) atau nilai m. Parameter ini sangat krusial dalam algoritma *Fuzzy C-Means* karena menentukan tingkat kekaburan transisi antar kluster. Pengujian dilakukan dengan membandingkan beberapa kandidat nilai m, yaitu 1.2, 1.5, 1.8, 2.0, 2.2, dan 2.5. Analisis validitas *fuzzy* dilakukan menggunakan tiga metrik, yaitu *Silhouette Score*, *Partition Coefficient* (PC), dan *Classification Entropy* (CE). Gambar 4.8 menunjukkan ringkasan performa dari setiap nilai m.

| m | Silhouette | PC | CE | Iterasi |
|-----|------------|--------|--------|---------|
| 1.2 | 0.4478 | 0.9717 | 0.0485 | 7.0 |
| 1.5 | 0.4459 | 0.8824 | 0.2273 | 9.0 |
| 1.8 | 0.4403 | 0.7372 | 0.4853 | 13.0 |
| 2.0 | 0.4376 | 0.6463 | 0.6316 | 17.0 |
| 2.2 | 0.4278 | 0.5722 | 0.7456 | 20.0 |
| 2.5 | 0.4108 | 0.492 | 0.8653 | 26.0 |

Gambar 4. 7 Ringkasan Performa Nilai m

Pada Gambar 4.7, nilai m=1.2 dipilih sebagai parameter terbaik karena memiliki karakteristik berikut:

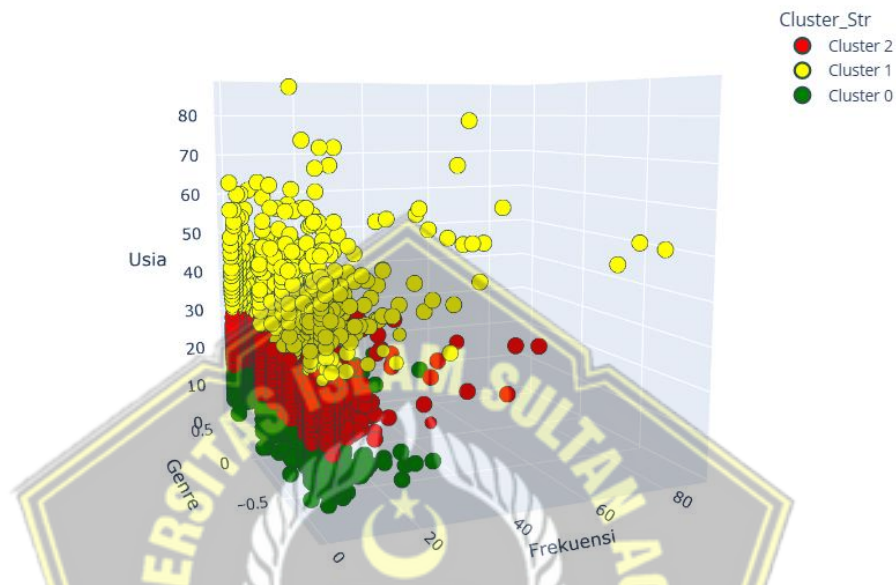
1. *Partition Coefficient* (PC): Mencapai nilai tertinggi sebesar 0.9717, yang menunjukkan tingkat kejelasan partisi yang sangat baik.
2. *Classification Entropy* (CE): Mencapai nilai terendah sebesar 0.0465, yang berarti tingkat ketidakpastian dalam pengelompokan sangat minim.
3. Kecepatan Konvergensi: Algoritma mencapai titik stabil hanya dalam 7 iterasi, jauh lebih cepat dibandingkan nilai m lainnya.

4.4.3 Hasil *Clustering Hybrid*

Tahap akhir adalah eksekusi *Hybrid Clustering* dengan mengintegrasikan pusat kluster dari *K-Means* sebagai inisialisasi awal bagi *Fuzzy C-Means* menggunakan parameter optimal k=3 dan m=1.2. Hasil pemodelan ini memberikan gambaran komprehensif mengenai pola minat baca berdasarkan variabel usia dan frekuensi kunjungan. Untuk memperjelas hasil pengelompokan, dilakukan tiga pendekatan visualisasi utama sebagai berikut:

1. Visualisasi Sebaran Data 3D

Visualisasi ini bertujuan untuk melihat distribusi seluruh data dalam ruang tiga dimensi berdasarkan fitur yang telah distandarisasi (*usia_s*, *freq_s*, dan *genre_s*), sebagaimana ditunjukkan pada Gambar 4.8.



Gambar 4. 8 Visualisasi 3D Scatter Plot

Berdasarkan Gambar 4.8, *Dataset* terbagi ke dalam tiga karakteristik utama dengan skema warna kustom:

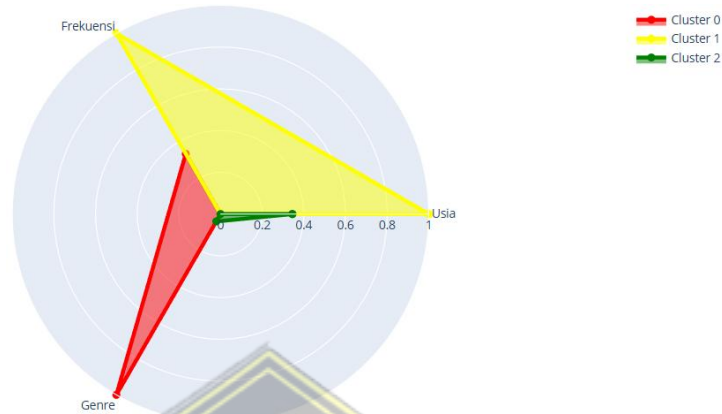
- Klaster 2 (Merah): Dominan diisi oleh kelompok usia remaja akhir hingga dewasa muda (rentang 15–28 tahun).
- Klaster 1 (Kuning): Mewakili kelompok usia dewasa hingga lanjut usia (rentang 30–60 tahun) dengan frekuensi peminjaman yang cenderung lebih tinggi.
- Klaster 0 (Hijau): Terdiri dari kelompok usia anak-anak dan remaja awal (di bawah 15 tahun).

Ukuran titik pada grafik ini merepresentasikan nilai Membership Degree (derajat keanggotaan). Semakin besar ukuran titik, semakin kuat posisi data tersebut berada pada pusat klaster yang bersangkutan.

2. Profil Karakteristik Klaster

Untuk memahami profil rata-rata dari setiap klaster secara lebih mendalam, dilakukan pemetaan fitur menggunakan *Radar Chart* yang

dinormalisasi pada rentang 0-1, sebagaimana ditunjukkan pada Gambar 4.10.



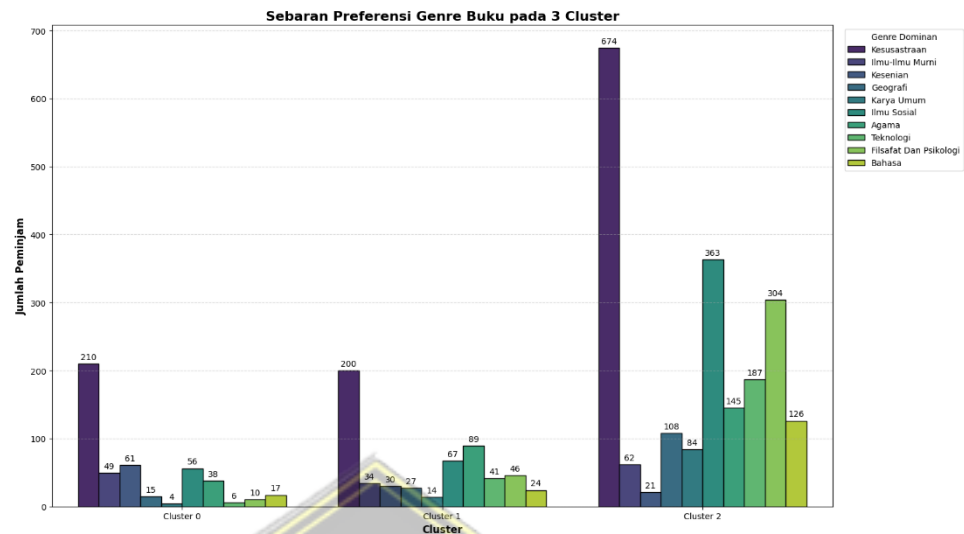
Gambar 4. 9 Radar Chart Fitur Cluster

Berdasarkan Gambar 4.9, terlihat perbedaan yang signifikan antar kelompok:

- a. Klaster 1 (Kuning) menunjukkan dominasi pada aspek Usia dan Frekuensi, yang menandakan bahwa kelompok dewasa/lansia adalah peminjam yang paling aktif.
- b. Klaster 2 (Hijau) berada pada posisi moderat yang menyeimbangkan antara frekuensi dan keragaman genre pada kelompok usia produktif.
- c. Klaster 0 (Merah) menonjol pada aspek Genre, menunjukkan adanya pola minat pada kategori buku tertentu yang sangat spesifik di kalangan anak-anak/remaja.

3. Distribusi Preferensi Genre per Klaster

Selain karakteristik numerik, distribusi genre buku juga dianalisis untuk melihat kategori apa yang paling diminati oleh setiap kelompok, sebagaimana ditunjukkan pada Gambar 4.10.



Gambar 4. 10 Sebaran Preferensi Genre Buku

Berdasarkan Gambar 4.10, dapat disimpulkan bahwa genre Kesusastaan merupakan kategori yang paling banyak diminati di semua klaster, terutama pada Klaster 2 (Merah) dengan jumlah peminjam mencapai lebih dari 661 orang. Namun, pada Klaster 1 (Kuning), terlihat sebaran yang lebih merata pada genre lain seperti Agama dan Filsafat, mencerminkan kematangan minat baca pada usia dewasa.

4. Validasi Keanggotaan dan Evaluasi Model

| | nama | usia | freq | Genre_Dominan | Cluster | Membership_Degree |
|---|--------------------------|------|------|------------------------|---------|-------------------|
| 0 | KEVIN MARCELLO PASHA | 18.0 | 7 | Filsafat Dan Psikologi | 2 | 0.999714 |
| 1 | AAD HARITS SANDI SAPUTRA | 24.0 | 1 | Teknologi | 2 | 0.999801 |
| 2 | AAN NAISABURY | 40.0 | 1 | Agama | 1 | 0.999947 |
| 3 | AAQILATUNNAJAH | 18.0 | 4 | Filsafat Dan Psikologi | 2 | 0.999978 |
| 4 | AAS FIBRIYANTI | 42.0 | 3 | Kesusastaan | 1 | 0.999998 |

Gambar 4. 11 Data Hasil Cluster

Keunggulan metode hybrid ini terlihat pada atribut *Membership Degree* yang dihasilkan. Sebagaimana ditunjukkan pada Gambar 4.11, setiap data memiliki nilai derajat keanggotaan yang menunjukkan seberapa kuat data tersebut termasuk ke dalam suatu klaster. Sebagai contoh, subjek "Kevin Marcello Pasha" diklasifikasikan ke Klaster 2 dengan derajat keyakinan 0.9995, yang menandakan akurasi pengelompokan yang sangat tinggi oleh sistem. Secara keseluruhan, model mencapai nilai Silhouette akhir sebesar 0.4427.

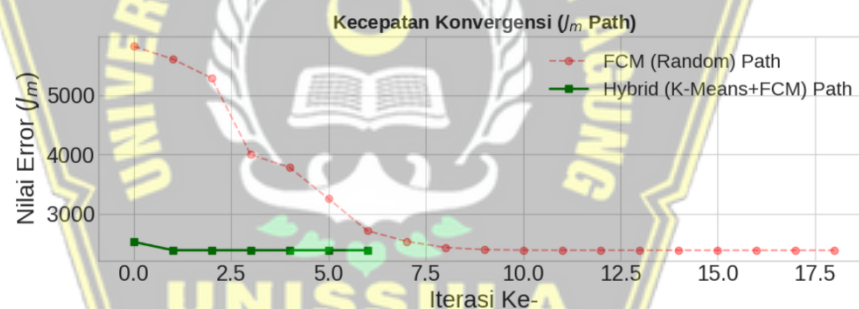
4.4.4 Perbandingan Algoritma *Hybrid* dan Algoritma Tunggal

Evaluasi performa dilakukan untuk membuktikan keunggulan metode *Hybrid Clustering* dibandingkan dengan algoritma tunggal dalam memetakan pola minat baca. Pengujian ini dilakukan dengan membandingkan tiga algoritma utama menggunakan parameter optimal yang telah ditemukan, yaitu $k=3$ dan $m=1.2$. Hasil perbandingan metrik evaluasi dapat dilihat pada Tabel 4.1 di bawah ini:

Tabel 4. 1 Evaluasi Algoritma *K-Means*, *Fuzzy C-Means* dan *Hybrid Clustering*

| Metode | <i>Silhouette</i> | PC(<i>Fuzzy</i>) | CE(<i>Fuzzy</i>) | Iterasi | Waktu(s) |
|----------------|-------------------|--------------------|--------------------|---------|----------|
| <i>K-Means</i> | 0.4479 | - | - | 8 | 0.0439 |
| FCM | 0.4478 | 0.9717 | 0.0485 | 19 | 0.0551 |
| <i>Hybrid</i> | 0.4478 | 0.9717 | 0.0485 | 7 | 0.0285 |

Berdasarkan data pada Tabel 4.1, serta ringkasan visual yang ditampilkan pada Gambar 4.12, terdapat beberapa poin analisis krusial:



Gambar 4. 12 Kecepatan Konvergensi

1. Analisis Kecepatan Konvergensi (J_m Path)

Berdasarkan grafik "Kecepatan Konvergensi" pada Gambar 4.12, terlihat perbedaan yang signifikan antara jalur *Fuzzy C-Means* dengan *Hybrid (K-Means+FCM)*:

- Fuzzy C-Means* memulai pencarian pusat kluster dengan nilai error (J_m) yang sangat tinggi (di atas 5000) dan membutuhkan 18 iterasi untuk mencapai titik stabil.
- Hybrid Clustering* memulai dengan nilai error yang jauh lebih rendah (di bawah 2500) karena pusat kluster telah diinisialisasi oleh *K-Means* terlebih dahulu. Hal ini menyebabkan algoritma hanya

membutuhkan 7 iterasi untuk konvergen. Hal ini membuktikan bahwa penggunaan *K-Means* sebagai inisiator pusat kluster mampu mempercepat proses komputasi secara signifikan.

2. Analisis Perbandingan Metrik Evaluasi Algoritma

Berdasarkan tabel "Perbandingan Metrik Evaluasi Algoritma" menunjukkan bahwa ketiga metode menghasilkan nilai *Silhouette Score* yang hampir identik, yaitu di kisaran 0.4478.

- a. Nilai *Partition Coefficient* (PC) yang tinggi (0.9717) dan *Classification Entropy* (CE) yang rendah (0.0485) pada metode *Hybrid* menunjukkan bahwa tingkat ketidakpastian (*fuzziness*) antar kluster sangat minim.
- b. Meskipun *Fuzzy C-Means* memiliki skor validitas yang serupa, metode *Hybrid* lebih unggul karena memberikan hasil yang lebih stabil dan tidak bergantung pada inisialisasi acak yang berisiko terjebak pada *Local optima*.

Secara keseluruhan, algoritma *Hybrid (K-Means + Fuzzy C-Means)* dipilih sebagai model terbaik untuk penelitian ini. Metode ini menggabungkan kecepatan partisi kaku dari *K-Means* dengan kehalusan derajat keanggotaan dari *Fuzzy C-Means*. Dengan waktu eksekusi yang lebih cepat daripada *C-Means* tunggal dan hasil yang lebih detail daripada *K-Means* tunggal, model ini terbukti efektif dalam mengatasi tantangan pengelompokan pola minat baca di Perpustakaan Provinsi Jawa Tengah yang memiliki persebaran data yang kompleks.

4.5 Profiling Pola Minat Baca Setiap Cluster

Setelah proses klustering selesai dilakukan, tahap selanjutnya adalah melakukan profilisasi (*profiling*) untuk memberikan makna kualitatif pada setiap kelompok data yang terbentuk. Tahap ini bertujuan untuk mengidentifikasi karakteristik unik dari setiap kluster berdasarkan variabel usia, frekuensi, dominasi genre, dan tingkat keragaman minat baca di Perpustakaan Provinsi Jawa Tengah.

1. Analisis Statistik Numerik Dasar (*Central Tendency*)

Langkah awal dalam profilisasi adalah melihat sebaran nilai rata-rata (*mean*), nilai minimum, dan nilai maksimum pada setiap kluster untuk variabel usia dan frekuensi peminjaman, sebagaimana ditunjukkan pada Tabel 4.2.

Tabel 4. 2 *Profiling Cluster* Berdasarkan Usia dan Frekuensi

| <i>Cluster</i> | Usia | | | Frekuensi | | |
|----------------|-------------|------------|------------|-------------|------------|------------|
| | <i>mean</i> | <i>min</i> | <i>max</i> | <i>mean</i> | <i>min</i> | <i>max</i> |
| 0 | 10.3 | 2.0 | 15.0 | 5.0 | 1 | 56 |
| 1 | 40.1 | 29.0 | 86.0 | 7.4 | 1 | 89 |
| 2 | 20.9 | 15.0 | 29.0 | 3.9 | 1 | 72 |

Berdasarkan Tabel 4.2, terlihat perbedaan demografis yang jelas di antara ketiga kluster. *Cluster 1* merupakan kelompok dengan usia rata-rata tertua (40,1 tahun) sekaligus memiliki tingkat loyalitas tertinggi dengan frekuensi peminjaman rata-rata 6,4 kali. Sebaliknya, *Cluster 2* mewakili kelompok usia paling muda (rata-rata 10,5 tahun), sementara *Cluster 0* mendominasi secara jumlah dengan rata-rata usia dewasa muda (20,9 tahun).

2. Analisis Preferensi Kategori (*Rank-based Genre Analysis*)

Tahap kedua bertujuan mengidentifikasi identitas minat utama dengan menghitung akumulasi skor kepopuleran genre di setiap kluster. Berdasarkan analisis peringkat, genre dengan frekuensi kumulatif tertinggi ditetapkan sebagai "Genre Favorit" sebagaimana terlihat pada Tabel 4.3:

Tabel 4. 3 Identifikasi Genre Favorit Berdasarkan *Cluster*

| <i>Cluster</i> | Genre Favorit | Skor Kepopuleran |
|----------------|---------------|------------------|
| 0 | Kesusastaan | 1299 |
| 1 | Kesusastaan | 1758 |
| 2 | Kesusastaan | 3227 |

Berdasarkan Tabel 4.3, ditemukan bahwa genre Kesusastaan menjadi primadona di seluruh kluster. Hal ini menunjukkan bahwa buku bertema

sastra, novel, dan cerita memiliki daya tarik universal bagi seluruh kelompok usia masyarakat di Kota Semarang, baik dari kalangan anak-anak, remaja, hingga dewasa lanjut.

3. Analisis Keragaman Minat Baca (*Cross-Tabulation Analysis*)

Profiling juga dilakukan untuk mengukur apakah seorang pembaca cenderung hanya menyukai satu jenis buku (*Single Genre*) atau gemar mengeksplorasi berbagai kategori (*Multi Genre*). Data tingkat keragaman minat ini disajikan pada Tabel 4.4.

Tabel 4. 4 *Profiling Cluster* Berdasarkan Keberagaman Minat Baca

| <i>Cluster</i> | Tipe Pembaca | |
|----------------|--------------------|---------------------|
| | <i>Multi Genre</i> | <i>Single Genre</i> |
| 0 | 56.4% | 43.6% |
| 1 | 55.1% | 44.9% |
| 2 | 38.1% | 61.9% |

Berdasarkan data pada Tabel 4.4, dapat disimpulkan karakteristik keragaman setiap klaster:

- a. Cluster 2 (Spesifik): Memiliki kecenderungan minat baca yang sangat terfokus, di mana 61,9% anggotanya hanya meminjam satu jenis genre buku saja.
- b. Cluster 1 & 0 (Eksploratif): Menunjukkan karakteristik pembaca yang lebih terbuka terhadap berbagai jenis bacaan, dengan persentase *Multi Genre* masing-masing sebesar 56,4% dan 55,1%.

4. Penentuan Karakteristik Karakter (*Logic-based Thresholding*)

Berdasarkan seluruh hasil analisis di atas, maka dapat ditarik kesimpulan karakteristik tiap klaster sebagai berikut:

- a. Cluster 2 (Pembaca Muda Selektif): Merupakan kelompok terbesar (66,6% atau 2.074 orang) dengan rata-rata usia 20,9 tahun. Kelompok ini memiliki minat yang sangat kuat dan spesifik pada satu genre tertentu, khususnya Kesusastraan.
- b. Cluster 1 (Pembaca Senior Eksploratif): Terdiri dari 572 orang (18,4%) dengan rata-rata usia 40,0 tahun. Kelompok ini adalah

pembaca paling aktif (frekuensi 7,4x) dan gemar mengeksplorasi berbagai jenis genre buku.

- c. Cluster 0 (Pembaca Anak Eksploratif): Merupakan kelompok anak-anak dan remaja awal (rata-rata usia 10,3 tahun) berjumlah 466 orang (15,0%). Meskipun masih muda, mereka menunjukkan minat eksplorasi genre yang cukup tinggi (56,4% *Multi Genre*).

4.6 Analisis Mendalam Pola Minat Baca Masyarakat

Berdasarkan hasil klastering dan *profiling* yang telah dilakukan terhadap 3.112 data pemustaka aktif, penelitian ini berhasil mengidentifikasi struktur pola minat baca yang nyata di Perpustakaan Provinsi Jawa Tengah. Analisis ini tidak hanya melihat angka secara parsial, melainkan menyatukan dimensi demografi, aktivitas, dan preferensi menjadi sebuah narasi pola perilaku sebagai berikut:

4.6.1 Sintesis Karakteristik dan Dinamika Segmen Pembaca

Setiap segmen yang diidentifikasi melalui algoritma *Hybrid Clustering* membawa profil psikografis yang berbeda, yang dapat dijelaskan sebagai berikut:

1. Segmen Pembaca Muda Selektif (*Cluster 2* - Mayoritas 66,6%)
 - a. Karakteristik Dominan: Kelompok ini didominasi oleh individu dengan rata-rata usia 20,9 tahun (mahasiswa dan pekerja muda) yang memiliki fokus minat sangat spesifik pada satu jenis genre buku saja (*Single Genre* mencapai 61,9%).
 - b. Kelebihan: Merupakan basis massa terbesar bagi perpustakaan, yang memberikan volume sirkulasi buku tertinggi pada kategori populer seperti Kesusastaan.
 - c. Kekurangan: Memiliki tingkat loyalitas atau frekuensi kunjungan terendah dibandingkan segmen lain, yaitu hanya rata-rata 3,9 kali dalam setahun. Hal ini menunjukkan bahwa meskipun jumlahnya besar, keterikatan (*engagement*) mereka terhadap layanan perpustakaan masih bersifat transaksional dan belum berkelanjutan.

2. Segmen Pembaca Mapan Antusias (*Cluster 1 - Loyal 18,4%*)
 - a. Karakteristik Dominan: Mewakili kelompok usia dewasa matang (rata-rata 40,1 tahun) yang menunjukkan perilaku membaca sangat eksploratif lintas berbagai kategori buku (*Multi Genre* mencapai 55,1%).
 - b. Kelebihan: Merupakan kelompok yang paling aktif dan loyal dengan frekuensi peminjaman tertinggi mencapai rata-rata 7,4 kali. Kelompok ini adalah pilar stabilitas aktivitas literasi di perpustakaan.
 - c. Kekurangan: Jumlah anggotanya merupakan yang paling sedikit (572 orang), yang mengindikasikan adanya tantangan regenerasi pembaca loyal dari kelompok usia di bawahnya
3. Segmen Pembaca Belia Eksploratif (*Cluster 0 - Potensial 15,0%*)
 - a. Karakteristik Dominan: Kelompok anak-anak dan remaja awal (rata-rata usia 10,3 tahun) yang mulai menunjukkan kecenderungan eksplorasi genre yang tinggi (56,4% *Multi Genre*) meskipun frekuensi peminjamannya masih di tingkat moderat.
 - b. Kelebihan: Menunjukkan rasa ingin tahu yang besar terhadap berbagai jenis bacaan sejak usia dini, yang merupakan modal awal penting dalam pembentukan budaya literasi berkelanjutan.
 - c. Kekurangan: Pola minat mereka masih dalam tahap perkembangan dan sangat bergantung pada ketersediaan koleksi yang variatif, sehingga berisiko menurun jika tidak didukung dengan strategi pengadaan buku yang sesuai dengan psikologi perkembangan mereka.

4.6.2 Analisis Hubungan Antar Variabel (Dinamika Pola)

Ditemukan bahwa Usia merupakan variabel paling dominan dalam menentukan pola minat baca masyarakat. Terdapat pola dinamika yang menarik antara usia, frekuensi, dan keragaman minat:

1. Korelasi Usia dan Keragaman: Terlihat adanya transisi perilaku di mana kelompok pembaca yang lebih tua (*Cluster 1*) cenderung memiliki minat

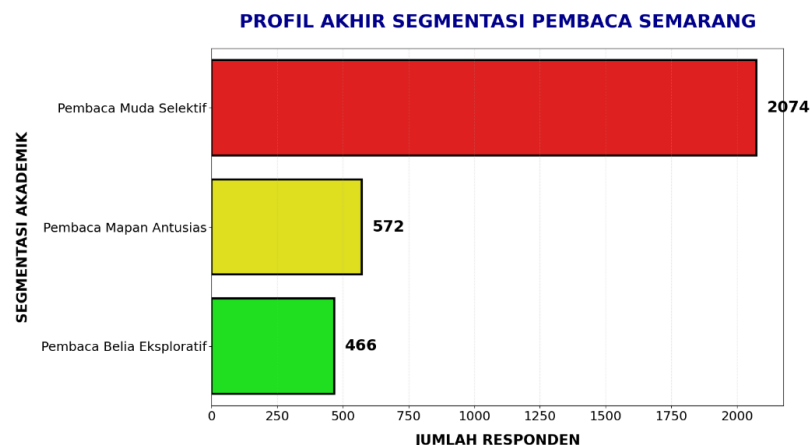
yang lebih terbuka dan luas (*Multi Genre*) dibandingkan kelompok usia produktif (Cluster 2) yang lebih terfokus. Hal ini menunjukkan bahwa pola minat baca berkembang dari sifatnya yang fungsional-akademis pada usia muda menjadi eksploratif-hobi pada usia matang.

2. Dominasi Genre Universal: Kesusastaan menjadi genre yang paling diminati di seluruh klaster. Analisis ini membuktikan bahwa karya sastra memiliki daya tarik universal sebagai "pintu masuk" utama masyarakat Semarang ke dalam dunia literasi, sebelum akhirnya mereka mengeksplorasi genre lain yang lebih spesifik seperti Agama atau Teknologi.

4.7 Penamaan Identitas (*Labeling*) Kluster

Setelah karakteristik setiap kelompok teridentifikasi melalui tahap profilisasi, langkah selanjutnya adalah memberikan identitas atau pelabelan (*labeling*) pada setiap klaster. Proses pelabelan ini bertujuan untuk memberikan nama yang representatif bagi setiap kelompok masyarakat berdasarkan karakteristik usia, loyalitas (frekuensi), dan perilaku eksplorasi genre buku yang telah dianalisis sebelumnya.

Penamaan identitas kluster dilakukan secara dinamis menggunakan mesin pelabelan (*labeling engine*) yang memetakan statistik rata-rata setiap klaster ke dalam tiga kategori segmen pembaca utama di Kota Semarang, sebagaimana ditunjukkan pada Gambar 4.13.



Gambar 4. 13 Visualisasi Profil Akhir Segmentasi Pembaca

Berdasarkan hasil pemrosesan pada Gambar 4.13, berikut adalah uraian mendalam mengenai identitas akhir dari ketiga segmen pembaca yang terbentuk:

1. Pembaca Muda Selektif (Cluster 2):

Merupakan segmen terbesar dengan jumlah anggota mencapai 2.074 orang (66,6%). Nama "Muda" dipilih karena rata-rata usia berada pada 20,9 tahun, sedangkan label "Selektif" merepresentasikan karakter pembaca yang memiliki minat sangat spesifik pada genre tertentu (terutama Kesusastraan) dan jarang mengeksplorasi genre lain.

2. Pembaca Mapan Antusias (Cluster 1):

Terdiri dari 572 orang (18,4%) dengan rata-rata usia 40,1 tahun. Identitas "Mapan" merujuk pada kelompok usia dewasa hingga lanjut usia, sementara label "Antusias" diberikan karena kelompok ini memiliki loyalitas tertinggi dengan frekuensi peminjaman rata-rata 7,4 kali serta memiliki perilaku membaca yang eksploratif terhadap berbagai genre.

3. Pembaca Belia Eksploratif (Cluster 2):

Berjumlah 466 orang (15,0%) dengan rata-rata usia paling muda, yaitu 10,5 tahun. Nama "Belia" mewakili kategori anak-anak dan remaja awal, sedangkan "Eksploratif" menunjukkan bahwa meskipun frekuensi peminjamannya moderat (5,0 kali), kelompok ini memiliki rasa ingin tahu yang tinggi dengan persentase *Multi Genre* mencapai 56,4%.

Pemberian label identitas ini mengonfirmasi bahwa model *Hybrid Clustering* berhasil mengidentifikasi struktur komunitas pembaca di Perpustakaan Provinsi Jawa Tengah secara detail. Dengan adanya segmentasi akademik ini, pihak pengelola perpustakaan dapat merancang strategi layanan yang lebih personal dan tepat sasaran bagi setiap kelompok masyarakat.

4.8 Klasifikasi dan Validasi Model Menggunakan *LightGBM*

Tahap klasifikasi merupakan bagian integral dari validasi hasil *Clustering*. Dengan membangun model klasifikasi, kita dapat menguji sejauh mana variabel independen mampu membedakan setiap kelompok yang telah dibentuk oleh algoritma *Hybrid K-Means* dan *Fuzzy C-Means*. Penggunaan *LightGBM* (*Light Gradient boosting Machine*) dipilih karena efisiensinya dalam menangani data tabular dan kemampuannya memproses fitur kategorikal secara asli (*Native support*).

1. Integrasi Data dan Arsitektur Fitur (*Preprocessing*)

Tahap integrasi data dan *preprocessing* merupakan fondasi utama dalam membangun model klasifikasi yang efisien.



| | usia | freq | Genre_Dominan |
|---|------|------|------------------------|
| 0 | 18.0 | 7 | Filsafat Dan Psikologi |
| 1 | 24.0 | 1 | Teknologi |
| 2 | 40.0 | 1 | Agama |
| 3 | 18.0 | 4 | Filsafat Dan Psikologi |
| 4 | 42.0 | 3 | Kesusastraan |

Gambar 4. 14 Hasil Integrasi Data dan Arsitektur Fitur

Pada Gambar 4.14, terlihat bahwa sistem mengintegrasikan tiga fitur utama yaitu usia, frekuensi peminjaman (*freq*), dan *Genre_Dominan* sebagai variabel independen untuk menentukan target kluster. Secara teknis, atribut *Genre_Dominan* dikonversi menjadi tipe data *category* untuk memanfaatkan fitur asli *LightGBM* dalam menangani data kategorikal secara langsung. Pendekatan ini memungkinkan algoritma untuk bekerja lebih optimal tanpa perlu menambah kompleksitas dimensi data melalui proses *one-hot encoding*, sehingga proses pengenalan pola minat baca menjadi lebih cepat dan akurat.

2. Strategi Pembagian *Dataset* (*Data Splitting*)

Proses pembagian *Dataset* dilakukan untuk menjamin objektivitas dan menguji kemampuan generalisasi model terhadap data baru.

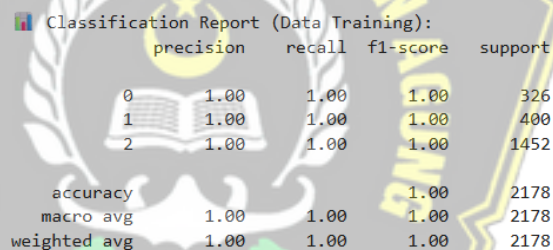
[STEP 2/7] PROGRESS: SPLITTING DATASET...
 ✓ Data Split Selesai (Train: 2178 | Test: 934)

Gambar 4. 15 Visualisasi Rasio Pembagian *Dataset*

Berdasarkan dokumentasi pada Gambar 4.15, *Dataset* dibagi menggunakan rasio 70:30 yang menghasilkan 2.178 data untuk tahap pelatihan (*Training*) dan 934 data untuk tahap pengujian (*Testing*). Penggunaan metode pembagian ini sangat penting dalam pengembangan model kecerdasan buatan untuk memastikan bahwa sistem benar-benar mempelajari pola perilaku pembaca secara mendalam dan tidak hanya menghafal *Dataset* yang sudah ada (*Overfitting*).

3. Analisis Performa Pelatihan (*Training Evaluation*)

Evaluasi kinerja pada tahap pelatihan menunjukkan bahwa model *LightGBM* mampu menangkap karakteristik setiap klaster dengan tingkat presisi yang sempurna.



| Classification Report (Data Training): | | | | |
|--|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 1.00 | 1.00 | 1.00 | 326 |
| 1 | 1.00 | 1.00 | 1.00 | 400 |
| 2 | 1.00 | 1.00 | 1.00 | 1452 |
| accuracy | | | 1.00 | 2178 |
| macro avg | 1.00 | 1.00 | 1.00 | 2178 |
| weighted avg | 1.00 | 1.00 | 1.00 | 2178 |

Gambar 4. 16 Evaluasi Metrik Performa Model pada *Data Training*

Sebagaimana ditunjukkan pada Gambar 4.16, hasil *Classification Report* untuk data *Training* mencapai nilai 1.00 pada semua metrik evaluasi, termasuk *precision*, *recall*, dan *f1-score*. Dominasi data pada klaster 2 dengan dukungan 1.452 sampel memberikan informasi mengenai profil kelompok terbesar, di mana model berhasil memvalidasi seluruh logika pemisahan klaster tanpa adanya kesalahan prediksi sedikit pun pada tahap pembelajaran ini.

4. Validasi Model pada Data Independen (*Testing Evaluation*)

Keandalan model semakin terbukti melalui hasil pengujian pada data uji yang berfungsi sebagai representasi data riil di masa depan.

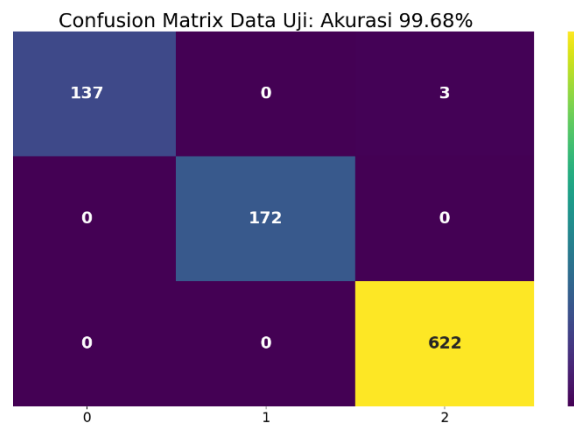
| Classification Report (Data Testing): | | | | |
|---------------------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| 0 | 1.00 | 0.98 | 0.99 | 140 |
| 1 | 1.00 | 1.00 | 1.00 | 172 |
| 2 | 1.00 | 1.00 | 1.00 | 622 |
| accuracy | | | 1.00 | 934 |
| macro avg | 1.00 | 0.99 | 1.00 | 934 |
| weighted avg | 1.00 | 1.00 | 1.00 | 934 |

Gambar 4. 17 Evaluasi Metrik Performa Model pada *Data Testing*

Gambar 4.17 memperlihatkan bahwa model tetap mempertahankan performa puncaknya dengan nilai akurasi mencapai 99,68 pada 934 sampel uji. Konsistensi nilai *f1-score* sebesar 0.99 pada klaster 0 (140 sampel), 1.00 pada klaster 1 (172 sampel), dan 1.00 pada klaster 2 (622 sampel) menegaskan bahwa model klasifikasi ini bersifat sangat tangguh (robust) dan memiliki tingkat kepercayaan yang sangat tinggi untuk diimplementasikan dalam sistem perpustakaan.

5. Ketepatan Prediksi Antar Klaster (*Confusion Matrix*)

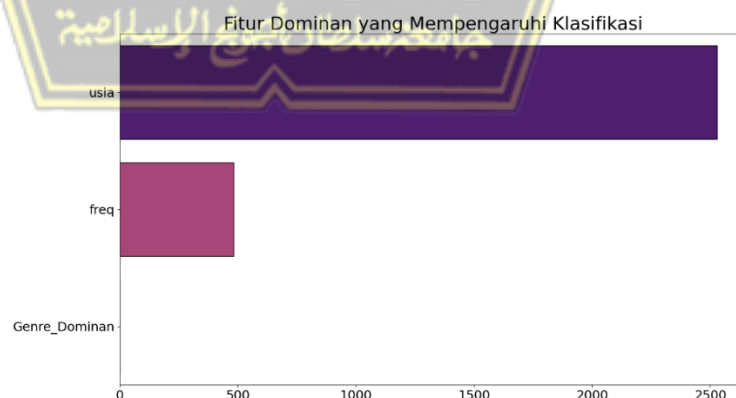
Validasi visual melalui *Confusion Matrix* memberikan bukti konkret mengenai ketepatan sistem dalam memetakan label klaster yang sebenarnya. Penggunaan *Confusion Matrix* berukuran 3x3 dalam penelitian ini didasarkan pada jumlah kelas target yang terdiri dari tiga klaster berbeda (multiklasifikasi), yaitu Cluster 0, Cluster 1, dan Cluster 2. Berbeda dengan matriks 2x2 yang hanya digunakan untuk klasifikasi biner, matriks 3x3 ini memungkinkan peneliti untuk memantau performa prediksi pada setiap kategori secara spesifik. Hal ini krusial untuk memastikan tidak adanya misclassification atau tumpang tindih antar kelompok, sehingga batas pemisah antara variabel usia, frekuensi, dan genre dominan dapat divalidasi secara presisi pada setiap dimensi kelas. Untuk hasil validasi visual dapat kita lihat pada Gambar 4.19.



Gambar 4. 18 Visualisasi Matriks Kebingungan (*Confusion Matrix*) Data Uji Berdasarkan Gambar 4.18, terlihat bahwa seluruh data terdistribusi secara sempurna pada garis diagonal walaupun adanya angka di luar jalur tersebut, hanya kecil saja kemungkinan terjadinya misklasifikasi antar kluster. Hasil ini menunjukkan bahwa variabel usia, frekuensi, dan genre dominan memiliki batas pemisah yang sangat kontras dan jelas (*linearly separable*), sehingga sistem klasifikasi dapat membedakan segmentasi pembaca secara presisi.

6. Analisis Kontribusi Variabel (*Feature Importance*)

Analisis kontribusi fitur memberikan wawasan strategis mengenai variabel apa yang paling memengaruhi segmentasi masyarakat dalam minat baca.



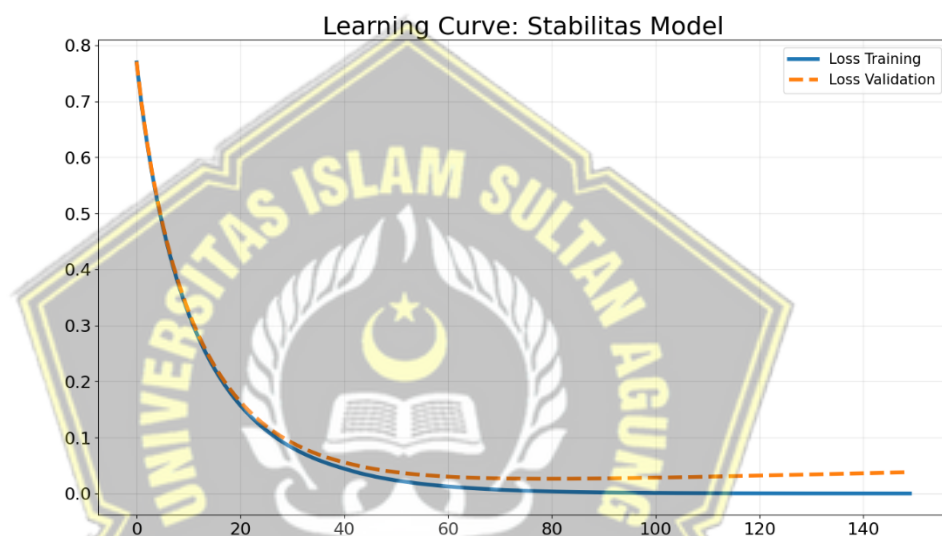
Gambar 4. 19 Variabel Penentu Segmentasi Pembaca

Gambar 4.19, menunjukkan bahwa variabel usia merupakan faktor penentu yang paling dominan dengan skor kepentingan tertinggi, yang kemudian diikuti oleh frekuensi peminjaman dan preferensi genre

dominan. Temuan ini secara ilmiah membuktikan bahwa di Perpustakaan Provinsi Jawa Tengah, demografi umur pengunjung adalah identitas yang paling kuat dalam membentuk pola minat baca masyarakat, melebihi variabel perilaku lainnya.

7. Stabilitas Pembelajaran Model (*Learning Curve*)

Stabilitas proses pembelajaran model dipantau melalui kurva kehilangan (log-loss) untuk memastikan model berkembang secara konsisten selama iterasi berlangsung.



Gambar 4. 20 Visualisasi Stabilitas Pembelajaran Model (*Learning Curve*)

Gambar 4.20 memperlihatkan bahwa kurva *Loss Training* dan *Loss Validation* menurun secara eksponensial dan saling berhimpit walaupun ada sedikit celah (*gap*) yang tidak terlalu lebar di antara keduanya. Fenomena ini merupakan indikator utama bahwa model berada dalam kondisi yang sangat stabil dan telah mencapai konvergensi yang optimal, sehingga model ini sangat layak digunakan sebagai mesin prediksi otomatis untuk mendukung pengambilan keputusan layanan perpustakaan.

4.9 Implikasi Manajerial

Hasil analisis kluster dan model klasifikasi LightGBM memberikan sejumlah implikasi strategis bagi pengelolaan perpustakaan di Kota Semarang untuk meningkatkan efektivitas layanan literasi secara keseluruhan. Berdasarkan akurasi model sebesar 99,68%, perpustakaan dapat mengintegrasikan sistem klasifikasi ini ke dalam aplikasi peminjaman untuk memberikan rekomendasi buku otomatis yang disesuaikan dengan profil kluster pengguna. Mengingat genre Sastra mendominasi semua kluster, manajemen perlu memprioritaskan pengadaan koleksi dalam genre ini dengan penyesuaian berdasarkan segmen usia, seperti menyediakan literatur anak-anak untuk Kluster 0 dan literatur akademik yang lebih spesifik untuk Kluster 2. Selain itu, optimasi anggaran dapat dicapai dengan fokus pada program promosi untuk Kluster 2 (Pembaca Muda Selektif), yang memiliki basis keanggotaan terbesar tetapi frekuensi peminjaman terendah sebesar 3,9 buku, untuk merangsang minat membaca siswa. Akhirnya, penggunaan pendekatan Explainable AI (XAI) memberikan transparansi bagi manajemen untuk memahami bahwa usia dan frekuensi merupakan indikator kunci dalam merumuskan kebijakan layanan pelanggan yang lebih akurat dan terarah.

4.10 Implementasi *Dashboard Web*

Tahap akhir dari penelitian ini adalah transformasi model eksperimental ke dalam bentuk aplikasi praktis berupa *Executive Dashboard* berbasis *Web*. *Dashboard* ini dirancang menggunakan paradigma *Decision Support System* (DSS), di mana hasil algoritma *Hybrid Clustering* dan *LightGBM* tidak hanya ditampilkan sebagai angka statis, tetapi diolah menjadi informasi strategis untuk mendukung pengambilan keputusan di Perpustakaan Provinsi Jawa Tengah.

4.10.1 Arsitektur Antarmuka dan Desain Pengalaman Pengguna (UI/UX)

Antarmuka *Dashboard* dikembangkan menggunakan kerangka kerja (*Framework*) *Streamlit*, sebuah pustaka berbasis Python yang dioptimalkan untuk pembuatan aplikasi *Web* data secara cepat dan responsif. Sistem ini

dikonfigurasi dengan tema profesional melalui penyisipan *Custom CSS* untuk menciptakan lingkungan kerja yang bersih dan intuitif bagi pengguna tingkat manajerial.

```
st.set_page_config(page_title="Library AI Executive Dashboard", page_icon="📖", layout="wide")

# CSS
st.markdown("""
<style>
.main { background-color: #F1F5F9; }
.stMetric { background: white; padding: 20px; border-radius: 12px; box-shadow: 0 4px 6px
-1px rgba(0,0,0,0.1); border-left: 5px solid #10B981; }
.card { background: white; padding: 25px; border-radius: 16px; border: 1px solid #E2E8F0;
margin-bottom: 20px; box-shadow: 0 2px 4px rgba(0,0,0,0.05); }
h1, h2, h3 { color: #0F172A; }
</style>
""", unsafe_allow_html=True)
```

Gambar 4. 21 Kode Program Konfigurasi Halaman *Dashboard*

Gambar 4.21 menunjukkan proses konfigurasi awal aplikasi, termasuk pengaturan tata letak (layout wide) dan penyisipan CSS kustom. Penggunaan gaya visual seperti latar belakang #F1F5F9 dan desain kartu metrik yang memiliki bayangan halus bertujuan untuk meningkatkan keterbacaan data (*data readability*). *Dashboard* ini juga dilengkapi dengan bilah sisi (*sidebar*) yang berfungsi sebagai navigasi menu utama, mencakup *Dashboard* Utama, Analisis Mendalam, hingga fitur Ekspor Data. Penyantuman identitas peneliti dan instansi Universitas Islam Sultan Agung (UNISSULA) berfungsi sebagai standar dokumentasi akademik yang formal.

4.10.2 Integrasi Arsitektur *AI Engine* dan *Pipeline Preprocessing*

Keunggulan teknis dari sistem ini terletak pada kemampuannya melakukan sinkronisasi otomatis dengan model kecerdasan buatan yang telah divalidasi. Sistem memuat aset cerdas berupa file model `lgbm_ta.pkl` (model klasifikasi *LightGBM*) dan `pipeline_resources.pkl` (kamus label dan fitur) menggunakan pustaka *Joblib*.

```

@st.cache_resource
def load_ai_assets():
    """Memuat Model & Resources Pipeline dari Colab"""
    try:
        model = joblib.load("model_lgbm_ta.pkl")
        pipeline = joblib.load("pipeline_resources.pkl")
        return model, pipeline
    except:
        return None, None

def preprocess_data(uploaded_file):
    """Preprocessing & Outlier Filter (Sinkronisasi 3088 data)"""
    df_raw = pd.read_csv(uploaded_file) if uploaded_file.name.endswith(".csv") else pd.read_excel(uploaded_file)

    # Mapping Kolom
    rename_map = {"Nama_Lengkap": "nama", "Usia_Peminjam": "usia", "Frekuensi_Peminjaman": "freq", "Genre_Buku": "genre"}
    df_raw = df_raw.rename(columns={k: v for k, v in rename_map.items() if k in df_raw.columns})

    # Cleaning & Agregasi
    df_raw["freq"] = pd.to_numeric(df_raw["freq"], errors="coerce")
    df_raw["usia"] = pd.to_numeric(df_raw["usia"], errors="coerce")
    df_raw["genre"] = df_raw["genre"].astype(str).str.strip().str.title()
    df_clean = df_raw.dropna(subset=["nama", "usia", "genre", "freq"])

    df_grouped = df_clean.groupby("nama").agg({
        "usia": "mean", "freq": "sum",
        "genre": lambda x: x.value_counts().index[0]
    }).reset_index().rename(columns={"genre": "Genre_Dominan"})

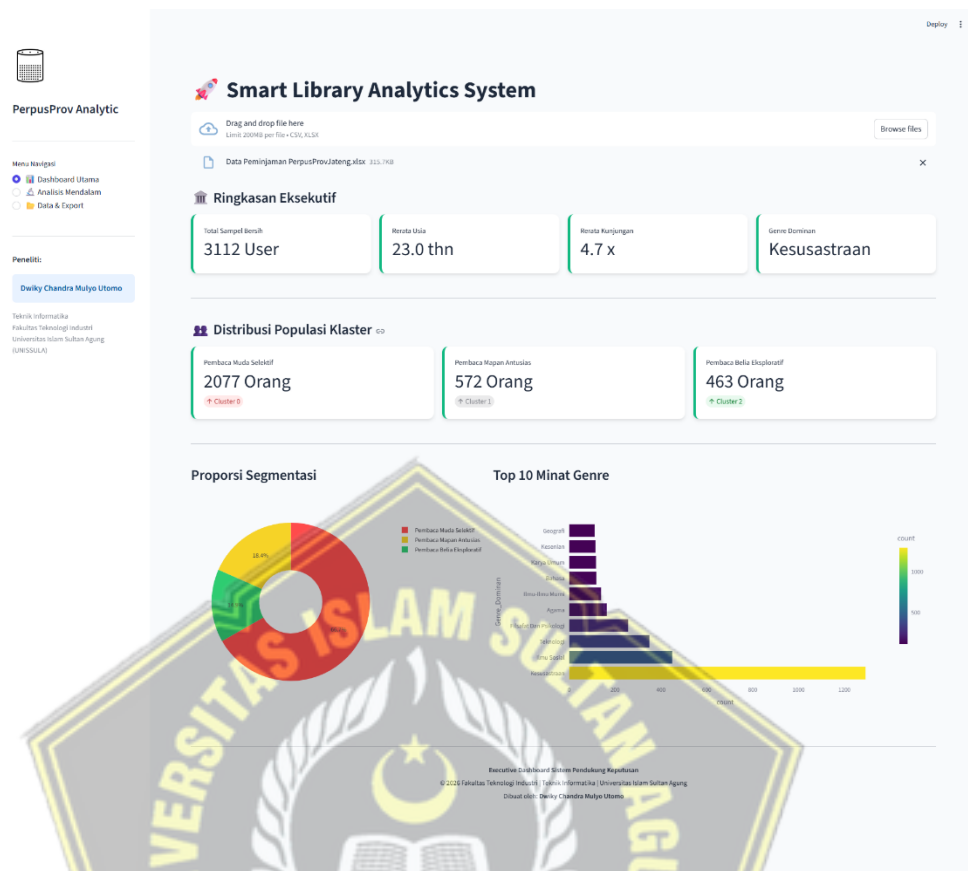
```

Gambar 4. 22 Kode Program Sinkronisasi *Engine* dan *Preprocessing*

Gambar 4.22 menampilkan fungsi pemuatan aset dan fungsi `preprocess_data` yang dirancang untuk memastikan data yang diunggah oleh pengguna melewati tahapan pembersihan yang identik dengan tahap eksperimen. Hal ini mencakup pemetaan kolom mentah menjadi variabel standar, penanganan nilai kosong (*missing values*). Mekanisme `@st.cache_resource` memastikan model hanya dimuat satu kali ke dalam memori, sehingga proses prediksi dapat berjalan secara instan. Sinkronisasi ini menjamin bahwa data yang diproses di *Dashboard* akan selalu konsisten dengan standar 3.112 sampel bersih yang digunakan pada tahap validasi model.

4.10.3 Visualisasi *Dashboard* Utama dan Analisis Segmentasi

Setelah *Dataset* berhasil diproses dan diklasifikasikan oleh model *LightGBM*, *Dashboard* akan menampilkan baris ringkasan eksekutif yang memuat metrik utama secara otomatis untuk memberikan wawasan cepat bagi manajemen.



Gambar 4. 23 Tampilan Halaman *Dashboard* Utama

Berdasarkan Gambar 4.23, antarmuka dashboard terdiri dari beberapa komponen fungsional utama sebagai berikut:

1. Sistem Pengunggahan Data dan Navigasi:

Pada bagian atas, terdapat fitur *File Uploader* yang memungkinkan pengguna mengunggah *dataset* dalam format *.xlsx* atau *.csv*. Pada sisi kiri (*sidebar*), tersedia menu navigasi yang mencakup *Dashboard* Utama, Analisis Mendalam, dan *Data & Export*. Selain itu, sidebar ini juga menampilkan identitas peneliti sebagai standar formal dokumentasi akademik.

2. Ringkasan Eksekutif (*Key Performance Indicators*):

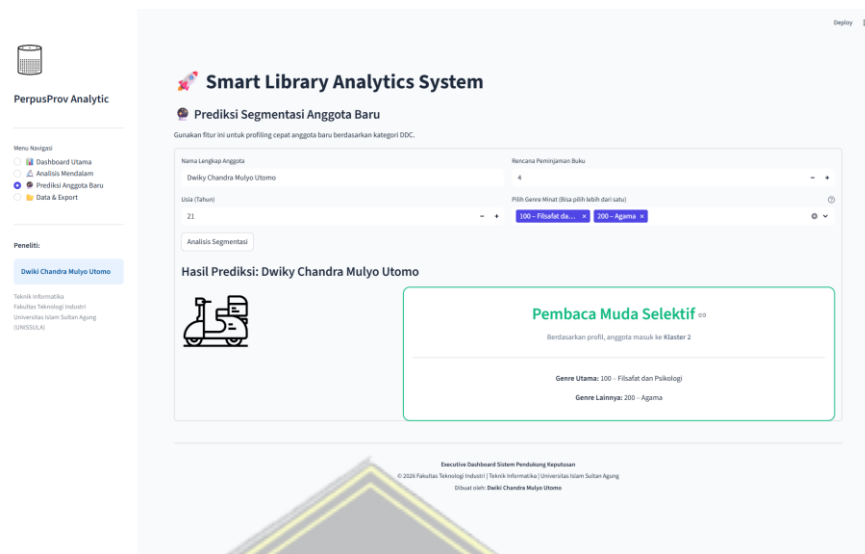
Baris pertama menampilkan empat metrik utama (KPI) secara otomatis:

- a. Total Sampel Bersih: Menampilkan jumlah data setelah melalui proses *filtering* (3.112 *User*).

- b. Rerata Usia: Menunjukkan rata-rata umur pengguna yang aktif (23 tahun).
 - c. Rerata Kunjungan: Memberikan gambaran frekuensi peminjaman rata-rata (4,7 kali).
 - d. Genre Terpopuler: Menampilkan kategori buku yang paling diminati secara global (Kesusastaan).
3. Distribusi Populasi Klaster:
- Fitur utama yang ditambahkan adalah panel distribusi populasi yang menyajikan jumlah anggota spesifik pada setiap kelompok. Klaster Pembaca Muda Selektif (Cluster 0) mendominasi dengan jumlah 2.077 orang, diikuti oleh Pembaca Mapan Antusias (Cluster 1) sebanyak 572 orang, dan Pembaca Belia Eksploratif (Cluster 2) sebanyak 463 orang. Informasi ini sangat krusial bagi pihak manajemen perpustakaan untuk mengetahui skala prioritas layanan pada setiap segmen masyarakat.
4. Analisis Visual Proporsi dan Preferensi:
- Bagian bawah dashboard menyajikan dua grafik interaktif:
- a. Proporsi Segmentasi (Donut Chart): Menunjukkan bahwa mayoritas pembaca berasal dari Klaster 2 dengan persentase mencapai 66,7%. Skema warna kustom (Merah, Kuning, Hijau) digunakan untuk menjaga konsistensi identitas tiap klaster.
 - b. Top 10 Minat Genre Buku (Bar Chart): Memvisualisasikan tingkatan minat genre secara urut, di mana genre Kesusastaan memiliki bar terpanjang, disusul oleh Ilmu Sosial dan Teknologi.

4.10.4 Sistem Prediksi Otomatis Anggota Baru

Fitur "Prediksi Anggota Baru" dirancang sebagai modul operasional yang memungkinkan pihak perpustakaan melakukan profilisasi instan terhadap anggota yang baru mendaftar. Dengan memanfaatkan kemampuan real-time inference dari model *LightGBM*, sistem dapat menentukan kategori segmentasi pemustaka hanya berdasarkan data atribut demografi dan rencana peminjamannya.



Gambar 4. 24 Tampilan Menu Prediksi Anggota Baru

Berdasarkan Gambar 4.24, fitur prediksi ini memiliki beberapa komponen fungsional utama sebagai berikut:

1. Formulir Profiling Anggota Baru:

Sistem menyediakan antarmuka input yang komprehensif bagi petugas perpustakaan, yang mencakup:

- a. Nama Lengkap: Digunakan untuk identitas personal dalam laporan prediksi.
- b. Usia (Tahun): Atribut kunci yang menjadi faktor penentu utama dalam pembentukan pola minat baca.
- c. Rencana Peminjaman Buku: Estimasi frekuensi kunjungan anggota untuk mengukur potensi loyalitas.
- d. Pilihan Genre Minat: Pengguna dapat memilih lebih dari satu genre berdasarkan standar kategori *Dewey Decimal Classification* (DDC) melalui fitur *multi-select*.

2. Mekanisme Real-time Inference:

Setelah tombol "Analisis Segmentasi" ditekan, core engine aplikasi secara otomatis mengonversi data input menjadi format DataFrame yang sesuai dengan arsitektur model klasifikasi. Sistem akan mengambil genre pertama yang dipilih sebagai "Genre Utama" untuk diproses oleh model *LightGBM* yang telah dimuat sebelumnya melalui file .pkl.

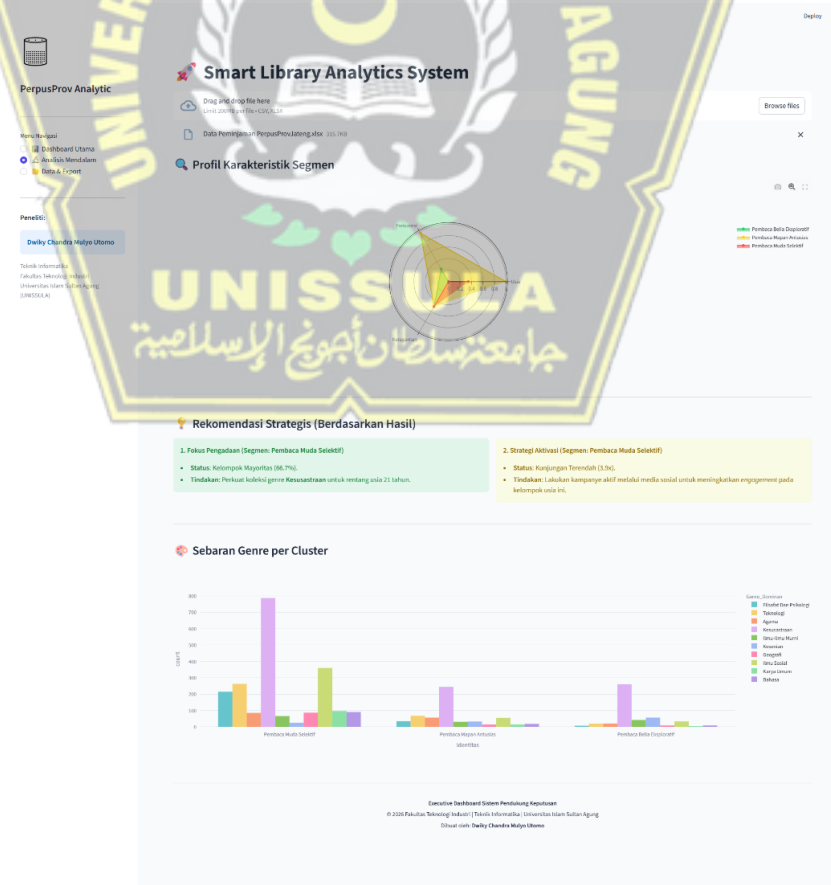
3. Visualisasi Hasil Prediksi Semantik:

Output prediksi disajikan secara interaktif dalam sebuah kotak hasil (prediction box) yang menonjolkan label identitas semantik:

- Identitas Kluster: Menampilkan nama segmen yang mudah dipahami, seperti "Pembaca Muda Selektif", beserta nomor klasternya.
- Ikona Representatif: Sistem menampilkan ilustrasi visual yang berbeda untuk setiap kluster guna mempermudah pengenalan profil secara visual.
- Balloons Effect*: Memberikan umpan balik visual positif saat proses analisis berhasil diselesaikan dengan sukses.

4.10.5 Analisis Strategis dan Rekomendasi Otomatis

Fitur paling inovatif dalam *Dashboard* ini adalah menu "Analisis Mendalam" yang menyediakan profil karakteristik setiap segmen menggunakan *Radar Chart* dan sistem rekomendasi otomatis.



Gambar 4. 25 Tampilan Analisis Mendalam dan Rekomendasi

Berdasarkan visualisasi pada Gambar 4.25, terdapat tiga komponen utama yang mendukung proses pengambilan keputusan strategis:

1. Profil Karakteristik Segmen (*Radar Chart*):

Sistem menggunakan *Radar Chart* untuk memetakan tiga variabel utama secara simultan: Usia, Frekuensi, dan Keragaman Minat Baca. Visualisasi ini memberikan gambaran "postur" perilaku setiap klaster:

- a. Pembaca Belia Eksploratif (Hijau): Menunjukkan dominasi pada aspek keragaman genre meskipun berada pada rentang usia paling rendah.
- b. Pembaca Mapan Antusias (Kuning): Memiliki jangkauan luas pada frekuensi kunjungan dan usia.
- c. Pembaca Muda Selektif (Merah): Terpusat pada titik usia produktif dengan profil keragaman yang cenderung spesifik.

2. Rekomendasi Strategis (Berdasarkan Hasil):

Puncak dari fungsionalitas cerdas sistem ini adalah panel rekomendasi yang dihasilkan secara otomatis oleh logika sistem:

- a. Fokus Pengadaan: Sistem mendeteksi Pembaca Muda Selektif sebagai kelompok mayoritas (66,7%). Berdasarkan temuan ini, menyarankan tindakan prioritas berupa penguatan koleksi genre Kesusastaan untuk rentang usia 21 tahun guna mempertahankan loyalitas kelompok terbesar.
- b. Strategi Aktivasi: Sistem secara cerdas mengidentifikasi kelompok dengan kunjungan terendah (3,9x) dan memberikan peringatan strategis. Tindakan yang disarankan adalah melakukan kampanye aktif melalui media sosial untuk meningkatkan keterikatan (*engagement*) pada segmen tersebut.

3. Sebaran Genre per Cluster (*Grouped Bar Chart*):

Bagian akhir halaman menyajikan distribusi genre buku yang lebih detail untuk setiap klaster. Grafik ini memvalidasi preferensi genre masing-masing kelompok, di mana genre Kesusastaan tampil sebagai genre dominan di seluruh segmen, namun dengan proporsi yang berbeda.

Visualisasi ini memungkinkan pengelola perpustakaan untuk melihat pola minat baca secara mikro, sehingga promosi buku dapat dilakukan lebih personal (misalnya, menonjolkan genre Agama atau Teknologi pada klaster tertentu yang menunjukkan minat lebih tinggi).

Melalui integrasi ketiga komponen ini, dashboard tidak hanya berfungsi sebagai alat visualisasi, tetapi bertransformasi menjadi penasihat strategis digital. Hal ini mempermudah pihak Perpustakaan Provinsi Jawa Tengah dalam menyusun program kerja berbasis data (*data-driven*) yang lebih akurat dan tepat sasaran.

4.10.6 Manajemen Data dan Fitur Ekspor Laporan

Sebagai bentuk dukungan terhadap kebutuhan administrasi, *Dashboard* menyediakan modul untuk melihat *Dataset* lengkap dan mengunduhnya ke dalam format laporan resmi.

| no | nama | usia | kluster |
|----|--------------------------------|------|-------------------------|
| 0 | KEVIN MARCELLO PASIA | 18 | 7. Pribat Dan Psikologi |
| 1 | AMIR HAKITS SAHRI LAPUSIA | 24 | 1. Teknologi |
| 2 | AMI WASABURY | 40 | 1. Agama |
| 3 | AADILTIWALAH | 18 | 4. Pribat Dan Psikologi |
| 4 | AAS FERRYKATI | 42 | 3. Kesehatan |
| 5 | FABRYN YUGAR KHALIDIN | 8 | 2. Akademik |
| 6 | ARIS NAY TIGAR PRASAKA | 20 | 1. Ilmu-Ilmu Bumi |
| 7 | ABDILAY KASUN SYAFIK | 13 | 2. Ilmu-Ilmu Bumi |
| 8 | ABDILLAH QOTRULHIDAYAT KURNIAH | 7 | 2. Akademi |
| 9 | ARSI LAY SA MANALFARDY | 19 | 1. Gangguan |

Gambar 4. 26 Tampilan Fitur Ekspor Laporan

Berdasarkan Gambar 4.26, modul *Data & Export* memiliki beberapa keunggulan teknis dan administratif sebagai berikut:

1. Integrasi Label Prediksi Otomatis: Tabel menyajikan data mentah yang telah diperkaya dengan kolom Identitas, yang merupakan output langsung dari model klasifikasi *LightGBM*. Setiap baris data kini memiliki label definitif seperti Pembaca Muda Selektif, Pembaca Mapan Antusias, atau Pembaca Belia Eksploratif. Hal ini memudahkan staf

perpustakaan dalam mengenali profil setiap anggota secara individu tanpa perlu melakukan perhitungan manual.

2. **Transparansi Analisis:** Dengan menampilkan variabel usia, frekuensi, dan genre dominan berdampingan dengan label identitas, sistem memberikan transparansi mengenai alasan di balik segmentasi yang dilakukan oleh sistem. Hal ini memungkinkan pengelola perpustakaan untuk melakukan validasi data secara instan sebelum mengambil kebijakan layanan tertentu.
3. **Fitur Ekspor Laporan Mandiri:** Terdapat tombol Unduh Laporan Excel (.xlsx) yang terletak di bagian bawah tabel. Fitur ini secara otomatis mengonversi data hasil pengolahan yang bersifat dinamis menjadi format dokumen statis atau spreadsheet yang siap digunakan sebagai lampiran dalam laporan bulanan instansi Perpustakaan Provinsi Jawa Tengah.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil memetakan tiga segmen profil pembaca utama di Perpustakaan Provinsi Jawa Tengah berdasarkan analisis terhadap 3.112 data pemustaka aktif. Segmen pertama diidentifikasi sebagai Pembaca Muda Selektif yang mendominasi populasi sebesar 66,6%, di mana kelompok usia produktif ini memiliki minat yang sangat spesifik pada genre Kesusastaan namun menunjukkan loyalitas kunjungan yang paling rendah. Segmen kedua adalah Pembaca Mapan Antusias sebesar 18,4% yang mewakili kelompok usia dewasa hingga lanjut usia dengan tingkat loyalitas tertinggi serta perilaku membaca yang sangat eksploratif lintas berbagai genre. Sementara itu, segmen ketiga yaitu Pembaca Belia Eksploratif mencakup 15,0% populasi yang terdiri dari anak-anak dan remaja awal dengan rasa ingin tahu tinggi terhadap keragaman buku meskipun frekuensi kunjungannya masih bersifat moderat.

Dari sisi teknis, penerapan algoritma Hybrid K-Means dan Fuzzy C-Means terbukti secara signifikan meningkatkan efisiensi proses pengelompokan pola minat baca tersebut. Penggunaan centroid dari K-Means sebagai inisialisasi awal berhasil mempercepat konvergensi sistem, di mana model hybrid mencapai titik stabil hanya dalam 7 iterasi dibandingkan metode Fuzzy C-Means konvensional yang memerlukan 19 iterasi. Keandalan hasil klastering ini diperkuat oleh validasi model klasifikasi LightGBM yang mencapai nilai akurasi sebesar 99,68% yang berarti data terdistribusi dengan baik sehingga error terhadap pemetaan klasifikasi sangat rendah pada data uji. Analisis kontribusi fitur pada model tersebut mengonfirmasi bahwa variabel usia merupakan faktor paling dominan yang memengaruhi pola minat baca di masyarakat, diikuti oleh frekuensi peminjaman dan preferensi genre buku.

Seluruh rangkaian analisis cerdas ini telah diimplementasikan ke dalam sebuah Executive Dashboard berbasis Web menggunakan framework Streamlit yang dirancang sebagai Decision Support System (DSS). Sistem ini memberikan solusi praktis bagi pihak Perpustakaan Provinsi Jawa Tengah dalam melakukan pemantauan profil pengunjung secara real-time. Dengan adanya sistem visualisasi yang interaktif ini, pihak manajemen perpustakaan dapat merumuskan kebijakan strategis berbasis data, baik dalam pengadaan koleksi buku maupun strategi promosi literasi yang lebih personal dan tepat sasaran bagi setiap segmen masyarakat.

5.2 Saran

Berikut adalah beberapa saran yang diajukan untuk pengembangan penelitian selanjutnya:

1. Integrasi Sistem Rekomendasi Personal: Sistem klasifikasi yang telah dibangun dapat dikembangkan lebih lanjut menjadi mesin rekomendasi buku otomatis (*Recommendation System*) yang terintegrasi dengan akun anggota perpustakaan, sehingga setiap pengunjung dapat menerima saran bacaan melalui email atau aplikasi sesuai dengan profil klaster mereka.
2. Ekspansi Variabel Data: Menambahkan variabel baru seperti durasi waktu kunjungan, format buku (fisik vs digital), dan tingkat kepuasan layanan ke dalam *Dataset* dapat memberikan dimensi baru yang lebih mendalam dalam memahami profil psikografis pembaca.
3. Otomasi *Pipeline Data*: Implementasi teknologi *Real-time Data Pipeline* disarankan agar *Dashboard* eksekutif dapat terhubung langsung dengan basis data transaksi perpustakaan secara otomatis, tanpa memerlukan proses pengunggahan file secara manual.

DAFTAR PUSTAKA

- Afif Muhaimin. (2024). Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor Dan Kedisiplinan. *Etheses*, 1–67. <http://etheses.uin-malang.ac.id/id/eprint/66134>
- Alsulamy, S. (2025). Predicting construction delay risks in Saudi Arabian projects: A comparative analysis of CatBoost, XGBoost, and LGBM. *Expert Systems with Applications*, 268. <https://doi.org/10.1016/j.eswa.2024.126268>
- Asyifa Tasya Fadilah. (2023). *Implementasi Algoritma K-Means Clustering Untuk Targeting*. *Ads*. <https://repository.uinjkt.ac.id/dspace/bitstream/123456789/72806/1/ASYIFA%20TASYA%20FADILAH-FST.pdf>
- Fan, C. L. (2023). AI-Enhanced Defect Identification in Construction Quality Prediction: Hybrid Model of Unsupervised and Supervised Machine Learning. *Procedia Computer Science*, 230, 112–119. <https://doi.org/10.1016/j.procs.2023.12.066>
- Gopal, N., & Panchal, D. (2023). Fuzzy decision support system for sustainable operational performance optimization for boiler unit in milk process industry. *Applied Soft Computing*, 135, 109983. <https://doi.org/10.1016/J.ASOC.2023.109983>
- 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. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>

- Imam Adlin Sinaga, P. M. S. R. A. M. P. (2024). *Rancang Bangun Sistem Informasi Perpustakaan Berbasis Website Desa Bakaran Batu Kecamatan Lubuk Pakam*. 4(1), 1678–1691. <https://doi.org/10.31004/innovative.v4i1.7936>
- Jaelani, Octaviana, & Rilvani, E. (2025). Studi Literatur: Perbandingan Algoritma K-Means dan Fuzzy C-Means dalam Analisis Clustering. *Journal of Computer Science and Technology JCS-TECH*, 5(2), 52–58. <https://mail.journalunwidha.com/index.php/jcstech/article/view/382/256>
- Kurniawan, H., Dwi Akbar, A., Svensons, N., Jaya Antonio, Y., Karnila, S., & Safitri, E. (2025). Evaluasi Performa Random Forest, XGBoost, dan LightGBM dalam Diagnosis Dini Diabetes Mellitus. *Jurnal JUPITER*, 17(93), 835–844.
- Lyra Bumantara Syarif, A. A. H. E. S. (2024). Pengembangan Layanan Perpustakaan Daerah Kabupaten Semarang Berdasarkan Preferensi Masyarakat. *SINOV*, 6(1), 116–124. <https://doi.org/10.55606/sinov.v6i1.806>
- Mahmudi Mahmudi, & Zaehol Fatah. (2024). Analisis Pengaruh Jenis Buku Terhadap Minat Baca Mahasiswa di Perpustakaan Ibrahimy dengan Algoritma K-Means Clustering. *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, 3(1), 53–65. <https://doi.org/10.59024/jiti.v3i1.1013>
- Mchara, W., Manai, L., Khalfa, M. A., Raissi, M., Dimassi, W., & Hannachi, S. (2025). A hybrid deep learning framework for global irradiance prediction using fuzzy C-Means, CNN-WNN, and Informer models. *Cleaner Engineering and Technology*, 28. <https://doi.org/10.1016/j.clet.2025.101061>
- Mussabayev, R., Mladenovic, N., Jarboui, B., & Mussabayev, R. (2023). How to Use K-means for Big Data Clustering? *Pattern Recognition*, 137. <https://doi.org/10.1016/j.patcog.2022.109269>
- Pattee, A. (2024). Identifying the hypothetical adolescent in library and information science literature: Describing the reader constructed in recommendations of adult books for young adults. *Library and Information Science Research*, 46(3). <https://doi.org/10.1016/j.lisr.2024.101315>
- Riansyah, A., Ghufro, Fitriyah, L., & Suyanto. (2025). Implementation of XGBoost for diabetes mellitus risk prediction based on health history.

International Journal of Advances in Applied Sciences, 14(4), 1028–1039.
<https://doi.org/10.11591/ijaas.v14.i4.pp1028-1039>

Riansyah, A., Qomaruddin, M., Indriastuti, M., & Sagaf, M. (2023). *Clustering Digital Transformation of Small and Medium Enterprises (SMEs) Using Fuzzy K-means Method*. (September), 540–544.
<https://doi.org/10.1109/eecsi59885.2023.10295664>

Syawali, Y., Arnita, Haikal Hafiz Rangkuti, M., Aprianda Mayadi, K., Silvi Pratiwi, K., & Fahrur Razi, M. (2024). Penerapan Algoritma DBSCAN Untuk Analisis Demografis dan Pengeluaran Pelanggan Mall. *JURNAL MATEMATIKA SAINS SAHOMBU*.
<https://jumas.ourhope.biz.id/ojs/index.php/JM/article/download/27/14>

Telaumbanua, S. A. B., Setiadi, F., & Nurjanah, S. (2025). Analisis Clustering Menggunakan Metode Enhanced Fuzzy C-Means Clustering Dengan Algoritma Rock Pada Student Performance Dataset. *Bit-Tech*, 7(3), 984–994.
<https://doi.org/10.32877/bt.v7i3.2287>

Tiara Dwi Pamungkas, Ayu Nur Intani, & Ichsan Fauzi Rachman. (2024). Upaya Meningkatkan Kesadaran Literasi Digital Pada Mahasiswa Untuk Mencapai Pembangunan Manusia Berkelanjutan. *Inspirasi Dunia: Jurnal Riset Pendidikan Dan Bahasa*, 3(2), 151–161.
<https://doi.org/10.58192/insdun.v3i2.2099>

Zhang, X., & Liu, F. (2025). Probabilistic linguistic three-way decisions: Integrating prospect theory with fuzzy possibilistic C-means clustering. *Fuzzy Sets and Systems*, 517. <https://doi.org/10.1016/j.fss.2025.109442>

Zheng, X., Tang, C., Liu, X., & Zhu, E. (2023). Multi-view clustering via matrix factorization assisted k-means. *Neurocomputing*, 534, 45–54.
<https://doi.org/10.1016/j.neucom.2023.03.004>