

# HYBRID CLUSTERING-CLASSIFICATION UNTUK PERSONALISASI REKOMENDASI UNIT KEGIATAN MAHASISWA BARU

Khania Luiza Cahya Tuluswati<sup>1)</sup> Dedi Trisnawarman<sup>2)</sup>

<sup>1) 2)</sup> Sistem Informasi Universitas Tarumanagara  
Letjen S. Parman St No.1, Jakarta 11440, Indonesia  
email : [khania.825220078@stu.untar.ac.id](mailto:khania.825220078@stu.untar.ac.id)<sup>1)</sup>, [dedit@fti.untar.ac.id](mailto:dedit@fti.untar.ac.id)<sup>2)</sup>

## ABSTRAK

Penelitian ini bertujuan membantu mahasiswa baru memilih Unit Kegiatan Mahasiswa (UKM) yang sesuai minat secara lebih akurat melalui pendekatan hybrid clustering-classification. Data diperoleh dari kuesioner PKKMB 2024, distandarisasi, dan selanjutnya dikelompokkan menggunakan algoritma K-Means untuk mengidentifikasi pola minat mahasiswa. Label klaster hasil pengelompokan ini ditambahkan sebagai fitur tambahan pada model klasifikasi Multinomial Logistic Regression (MLR) dan Random Forest (RF). Evaluasi kinerja model dilakukan dengan stratified k-fold cross-validation, mengukur akurasi, presisi, recall, dan F1-score. Selain itu, uji Chi-Square dan Cramer's V digunakan untuk menganalisis kekuatan asosiasi antar variabel kategorikal. Hasil eksperimen menunjukkan model hybrid (MLR dan RF) meningkatkan akurasi rekomendasi dibandingkan model baseline tanpa informasi klaster. Klaster minat memberikan konteks tambahan yang memperbaiki interpretabilitas rekomendasi. Sistem prototipe telah siap diintegrasikan ke dashboard layanan kemahasiswaan.

## Key words

rekomendasi UKM, clustering, klasifikasi, personalisasi, machine learning.

## 1. Pendahuluan

Unit Kegiatan Mahasiswa (UKM) berperan penting dalam pembentukan minat, bakat, kepemimpinan, serta jejaring sosial bagi mahasiswa [1]. Bagi mahasiswa baru, pemilihan UKM sering bersifat *trial-and-error* karena minimnya informasi yang terstruktur mengenai kecocokan minat pribadi dengan karakteristik UKM yang tersedia. Kondisi ini berpotensi menurunkan keterlibatan (engagement) dan retensi keanggotaan [2]. Diperlukan pendekatan berbasis data untuk mempersonalisasi rekomendasi UKM sehingga keputusan mahasiswa lebih terarah dan berdampak. Penelitian ini memanfaatkan data

kuesioner PKKMB tahun 2024 di Universitas X dengan sekitar  $\pm 5.238$  responden berskala *Likert* (1-5). Butir-butir pertanyaan merekam preferensi dan kecenderungan aktivitas yang relevan dengan berbagai kategori UKM, misalnya olahraga, seni & budaya, kerohanian, kewirausahaan, kepemimpinan, dan pengabdian Masyarakat. Skala sampel dan keragaman fitur memungkinkan segmentasi minat yang bermakna dan pemodelan rekomendasi yang adaptif di tingkat individu.

Penelitian sebelumnya menunjukkan penggunaan *machine learning* untuk memberikan rekomendasi/keputusan di konteks kemahasiswaan: Naïve Bayes untuk klasifikasi minat UKM [3], serta K-Means untuk mengelompokkan minat ekstrakurikuler di tingkat sekolah [4]. Pada domain pemilihan program studi, kombinasi K-Means sebagai segmentasi dengan model klasifikasi seperti Random Forest terbukti meningkatkan akurasi prediksi kecocokan jurusan [5]. Di tingkat global, pendekatan hybrid pada sistem rekomendasi pendidikan yang menggabungkan beberapa teknik, seperti clustering, collaborative filtering, dan pemodelan graf diakui efektif sekaligus menekankan isu fairness dan interpretabilitas [6].

Studi-studi terdahulu menunjukkan bahwa algoritma seperti Naïve Bayes dan K-Means telah digunakan dalam sistem rekomendasi UKM maupun ekstrakurikuler. Kombinasi K-Means dengan Random Forest pada domain pemilihan jurusan juga terbukti meningkatkan akurasi klasifikasi. Studi global seperti Drushchak et al. (2025) menambahkan bahwa pendekatan hybrid mampu meningkatkan kualitas dan fairness sistem rekomendasi pendidikan [6]. Metode hybrid clustering-classification dipilih dalam penelitian ini karena mampu mengintegrasikan konteks minat mahasiswa (melalui segmentasi) ke dalam proses klasifikasi rekomendasi. Secara teknis, informasi hasil clustering (misalnya klaster minat olahraga, seni, atau pengabdian) ditambahkan sebagai fitur pada model supervised learning, yang memungkinkan model membuat prediksi lebih personal dan dapat dijelaskan. Keunggulan metode ini terletak pada peningkatan akurasi, transparansi rekomendasi, serta fleksibilitas dalam menangani keragaman preferensi

individu. Tantangannya adalah proses dua tahap yang lebih kompleks dan risiko *overfitting* jika jumlah klaster tidak optimal. Namun, hasil evaluasi dalam penelitian ini membuktikan bahwa manfaat pendekatan hybrid lebih dominan.

Namun, studi yang secara eksplisit menerapkan hybrid clustering-classification untuk personalisasi rekomendasi UKM masih terbatas. Padahal menggabungkan segmentasi (unsupervised) sebagai konteks/fitur dengan prediksi terarah (supervised) yang berpotensi meningkatkan akurasi sekaligus keterjelasan (interpretabilitas) rekomendasi. Celah ini menjadi fokus pada penelitian ini. Kontribusi penelitian:

1. Merumuskan kerangka hybrid clustering-classification untuk memberikan rekomendasi UKM;
2. Menunjukkan bahwa augmentasi label cluster sebagai fitur meningkatkan kinerja dibanding baseline tanpa informasi klaster;
3. Menyajikan prototipe sistem sebagai proof-of-concept yang siap diintegrasikan ke layanan kemahasiswaan.

Tujuannya adalah mengidentifikasi segmen minat dari data PKKMB 2024, memprediksi rekomendasi UKM per individu, membandingkan kinerja hybrid vs. baseline, dan mendemonstrasikan prototipe.

Ruang Lingkup dan Batasan, yaitu:

1. Data satu tahun (PKKMB 2024) pada Universitas X;
2. Variabel utama berupa butir *Likert* terkait minat/kecenderungan;
3. Kategori UKM disederhanakan ke beberapa tema;
4. Evaluasi berfokus pada metrik klasifikasi (akurasi, presisi, recall, F1) dan analisis asosiasi (Chi-Square, Cramer's V) tanpa menilai dampak jangka Panjang terhadap retensi UKM. Aspek etika dijaga melalui anonimisasi dan minimalisasi variable sensitif.

## 2. Metode Penelitian

### 2.1. Design Penelitian

Penelitian ini bersifat aplikatif-kuantitatif dengan rancangan hybrid clustering-classification untuk mempersonalisasikan rekomendasi UKM. Alur inti: prapemrosesan data → segmentasi minat (K-Means) → augmentasi fitur (penambahan label klaster) → klasifikasi rekomendasi → evaluasi & prototipe. Pilihan K-Means didasarkan pada penggunaannya yang luas dan kemampuannya membentuk segmen terukur dengan indeks validitas seperti silhouette/Calinski-Harabasz [7][8]. Komponen klasifikasi mengandalkan Multinomial Logistic Regression dan Random Forest

sebagai model kuat untuk klasifikasi multikelas pada data survei/edukasi [9] [10].

### 2.2. Data dan Variabel

Data berasal dari kuesioner PKKMB 2024 Universitas X ( $\pm 5.238$  responden). Variable utamanya berupa butir *Likert* 1-5 yang menyimpan minat/ketertarikan; demografi (program studi/fakultas, jenis kelamin, angkatan) bersifat opsional sebagai fitur tambahan. Label/target adalah kategori UKM. Jika tersedia kolom "UKM pilihan", itu dijadikan *ground truth*; apabila tidak, maka label direkayasa dari skor komposit per kategori UKM (olahraga, seni & budaya, kerohanian, kewirausahaan, komunitas/relawan, kepemimpinan). Praktik serupa menggunakan *proxy* label/segmentasi sebelum rekomendasi yang umumnya pada sistem rekomendasi berbasis clustering (misalnya K-Mean/FCM, PCA + K-Means) untuk mengatasi kelangkaan label atau *sparsity* [11] [12].

### 2.3. Tahapan Metode

Prapemrosesan adalah proses membersihkan data sehingga tidak ada duplikasi atau missing value, memvalidasi rentang (1-5), lalu distandarisasi dengan z-score sebelum K-Means agar metrik jarak adil antar fitur [13]. Respon tidak valid (*straight-lining*) ditandai/dieksekusi. Butir dipetakan ke kategori UKM untuk membantu interpretasi segmen dan perhitungan skor komposit.

Clustering atau segmentasi minat diterapkan K-Means pada data yang telah distandarisasi. Nilai K diseleksi dari rentang 3-8 menggunakan Silhouette/Calinski-Harabasz dan dijalankan multi-inisialisasi untuk stabilitas [13]. Interpretasi segmen dilakukan dari rerata skor per cluster (misalnya olahraga, seni & budaya, komunitas/relawan) [14].

Hybrid atau augmentasi fitur dan klasifikasi adalah menambahkan label cluster (*one-hot*), opsional jarak ke pusat cluster disertakan sebagai fitur yang berkelanjutan. Lapisan prediksi memakai Multinomial Logistic Regression dengan regularisasi dan Random Forest, sebagai pembandingan dapat diuji *boosting* (misalnya CatBoost/XGBoost) [15][16][17].

Prototipe. *Pipeline* terlatih (scaler, pusat cluster, encoder, model) dikemas di *backend* (FastAPI/Flask) dengan *endpoint*/predict yang menerima jawaban *Likert* dan mengembalikan (kategori UKM, probabilitas, segmen, alasan singkat). Antarmuka sederhana menampilkan hasil dan siap integrasi ke *dashboard*.

### 2.4. Evaluasi dan Validasi

Skema pelatihan menggunakan stratified k-fold ( $k=5$ ) di data latih, dan uji akhir pada hold-out 20%. Metrik utama: akurasi, macro-precision, macro-recall, macro-f1, serta confusion matrix. Perbedaan hybrid dengan baseline adalah diuji signifikansinya (misal

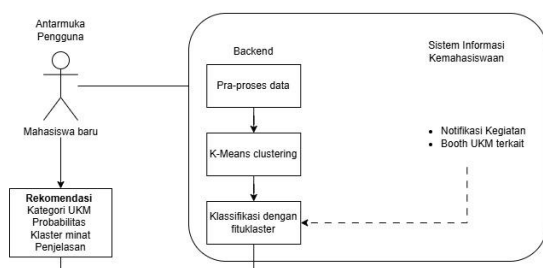
McNemar atau Wilcoxon pada skor per-fold). Analisis pendukung Chi-Square dan Cramer's V menilai keterkaitan fakultas/program studi dengan kategori UKM. Praktik evaluasi setara dengan pemilihan k, *feature importance* RF, dan pengaruh prapemrosesan (misal PCA untuk dimensi/*sparsity*) [11][12][18][19][20].

## 2.5. Evaluasi dan Validasi

Pendekatan hybrid clustering-classification ini direalisasikan dalam bentuk prototipe sistem yang siap digunakan oleh pengguna akhir (mahasiswa baru) dan diintegrasikan ke lingkungan layanan kampus. Arsitektur sistem prototipe digambarkan pada Gambar 1, yang terdiri dari beberapa komponen utama:

1. Antarmuka pengguna untuk input data kuesioner dan menampilkan rekomendasi;
2. Modul *backend* yang menjalankan *pipeline machine learning* (praproses data, K-Means clustering, dan model klasifikasi);
3. Modul integrasi dengan sistem informasi kemahasiswaan.

Alur kerja dimulai ketika mahasiswa baru mengisi jawaban kuesioner minat (*Likert* 1–5) melalui antarmuka. Selanjutnya, data dikirim ke server *backend* untuk diproses: dilakukan pembersihan dan transformasi (sesuai prosedur di Subbab 2.3), kemudian fitur-fitur minat mahasiswa di-cluster-kan menggunakan model K-Means yang telah dilatih sebelumnya. Berdasarkan nilai fitur dan pusat kluster terdekat, sistem menentukan segmen minat mahasiswa tersebut (misalnya termasuk kluster “olahraga” atau “seni budaya”). Label kluster itu lalu ditambahkan ke fitur mahasiswa (di-encode satu-hot). Selanjutnya, model klasifikasi (Multinomial Logit atau Random Forest yang telah dilatih) memprediksi kategori UKM yang paling sesuai untuk mahasiswa baru tersebut beserta probabilitas prediksi. Terakhir, hasil rekomendasi tersebut dikembalikan ke front-end untuk ditampilkan kepada pengguna.



Gambar 1 Arsitektur sistem prototipe rekomendasi UKM

*Backend* prototipe ini dikembangkan sebagai web service sederhana menggunakan teknologi Python (library scikit-learn untuk model MLR dan RF, serta framework FastAPI/Flask untuk API). *Pipeline* model yang telah dilatih (termasuk scaler, pusat kluster, encoder,

dan model klasifikasi) dikemas dalam *backend* tersebut sebagai sebuah *endpoint* API (misalnya /predict). *Endpoint* ini menerima masukan jawaban kuesioner dalam format tertentu dan mengembalikan hasil rekomendasi dalam format JSON yang memuat kategori UKM yang disarankan, probabilitas prediksi, label kluster minat, dan alasan atau fitur kunci sebagai interpretasi. Komponen *frontend* prototipe berupa antarmuka web sederhana (dapat berupa halaman web responsif) yang memungkinkan mahasiswa mengisi kuesioner dan menampilkan rekomendasi secara instan. Antarmuka menampilkan kategori UKM hasil rekomendasi beserta persentase keyakinan model, informasi segmen minat (kluster) mahasiswa, dan ringkasan alasan yang mendukung rekomendasi tersebut (misalnya “Anda memiliki minat tinggi di bidang olahraga, sehingga UKM X disarankan”). Gambar 2 menunjukkan contoh tampilan antarmuka prototipe yang dikembangkan, di mana mahasiswa baru dapat mengisi butir-butir pertanyaan minat dan memperoleh hasil rekomendasi secara *real-time*. Desain antarmuka mengutamakan kesederhanaan dan kejelasan: hasil rekomendasi ditampilkan dalam format yang mudah dibaca, disertai *highlight* pada aspek minat dominan yang memengaruhi rekomendasi untuk meningkatkan transparansi. Selain itu, prototipe ini telah disiapkan untuk integrasi lebih lanjut ke dalam *dashboard* Lembaga Kemahasiswaan kampus. Misalnya, hasil rekomendasi setiap mahasiswa dapat dikirim ke sistem *dashboard* kegiatan kemahasiswaan untuk ditindaklanjuti, seperti mengirim notifikasi tentang jadwal open house UKM atau info booth pendaftaran UKM yang sesuai. Dengan arsitektur terbuka ini, sistem rekomendasi UKM dapat menjadi modul layanan yang terpadu dalam ekosistem aplikasi kemahasiswaan yang lebih luas.

## 3. Hasil dan Pembahasan

### 3.1. Deskripsi Dataset & Pra-olah

Data diambil dari kuesioner PKKMB tahun 2024 di Universitas X dengan sekitar  $\pm 5.238$  responden. Setiap mahasiswa mengisi 20 pertanyaan dengan skala *Likert* (1-5) tentang minat/ketertarikan mahasiswa terhadap UKM berdasarkan kategorinya, yaitu olahraga, seni & budaya, kerohanian, kewirausahaan, komunitas/relawan, kepemimpinan, dll.

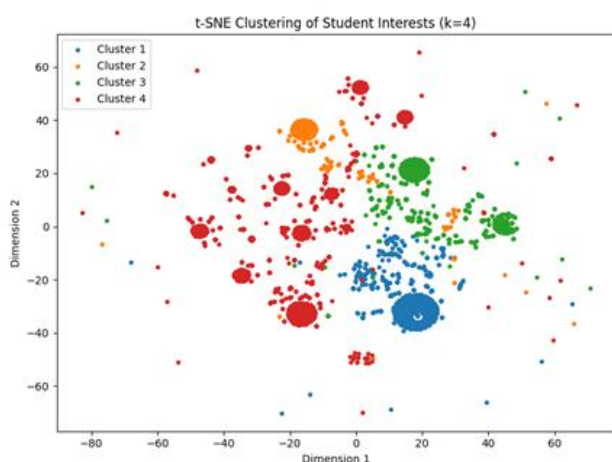
Jika ada kolom “UKM Pilihan”, itu dipakai sebagai label asli. Kalau tidak ada, maka akan dibuat label sementara dari skor gabungan per kategori UKM. Sebelum dianalisis, data akan diolah pada tahap pra-olah yang mencakup proses:

1. Penghapusan duplikasi;
2. Penanganan missing value;
3. Validasi rentang (1-5);
4. Standarisasi (z-score) sebelum clustering, agar metrik jarak setara antar fitur.

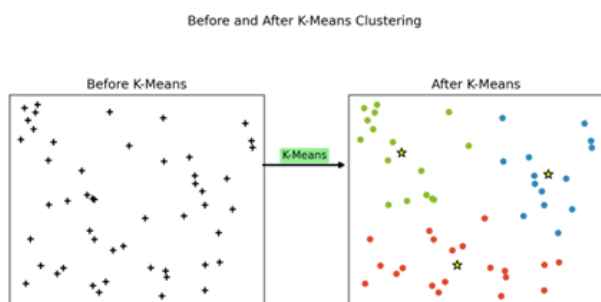
Pemetaan butir → kategori UKM dicantumkan pada lampiran untuk menjaga keterlacakan.

### 3.2. Deskripsi Dataset & Pra-olah

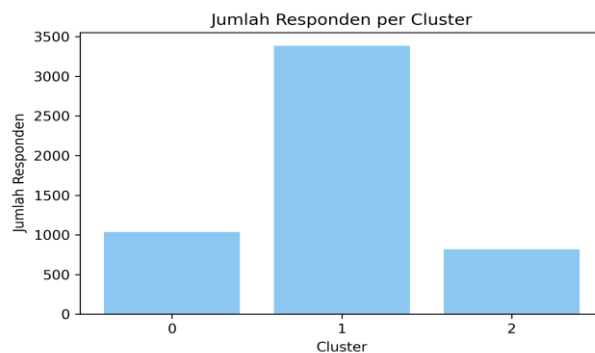
Segmentasi minat dilakukan menggunakan metode K-Means pada data yang sudah distandarisasi. Penentuan jumlah kluster ( $k$ ) dieksplor untuk  $k = 3-8$ , lalu dipilih berdasarkan Silhouette Score dan/atau Calinski-Harabasz (CH), semakin tinggi nilainya, maka kualitas kluster semakin baik [21][8][22]. Untuk visualisasi, data diproyeksikan ke dua dimensi menggunakan PCA agar pola antar kluster mudah dilihat sehingga keputusan kluster tetap didasarkan pada ruang fitur asli (*Likert*)[22][11].



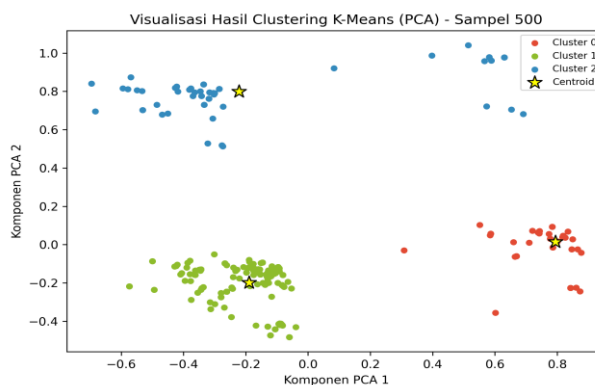
Gambar 2 Visualisasi kluster mahasiswa



Gambar 3 Sebaran data sebelum dan sesudah proses K-Means clustering



Gambar 4 Jumlah responden per kluster



Gambar 5 Visualisasi K-Means dari sample data menggunakan PCA

### 3.3. Deskripsi Dataset & Pra-olah

Tabel 1 merangkum rata-rata skor *Likert* P1–P5 per kluster (skala 1–5) untuk menggambarkan profil kecenderungan tiap segmen mahasiswa. Secara umum, pola yang muncul selaras dengan rancangan metodologi:

1. satu kluster cenderung kuat pada aspek operasional/praktis namun lebih rendah pada aspek afektif (P3);
2. satu kluster afektifnya tinggi namun membutuhkan penguatan operasionalisasi (P1, P2, P4);
3. satu kluster tinggi merata (*champions*). Ringkasan ini menjadi dasar segmentasi intervensi dan personalisasi rekomendasi UKM.

Tabel 1 Rata-rata skor *Likert*

Kluster	P1	P2	P3	P4	P5	Deskripsi
0	-	-	-	-	-	praktis namun belum menjiwai
1	-	-	-	-	-	Merasakan nilai, butuh operasionalisasi
2	-	-	-	-	-	Selaras menyeluruh

Penjelasan Tabel 1 adalah sebagai berikut:

1. Klaster 0 menunjukkan P1, P2, P4, P5 relatif lebih tinggi daripada P3 → kuat di aspek praktis/operasional, namun internalisasi nilai (P3) masih rendah.
2. Klaster 1 memperlihatkan P3 tertinggi dengan P1, P2, P4 lebih moderate/ragu → resonansi nilai sudah terbentuk, perlu konkretisasi langkah.
3. Klaster 2 menampilkan P1–P5 tinggi merata → keseimbangan nilai–operasional terbaik (champions).

### 3.4. Deskripsi Dataset & Pra-olah

Untuk prediksi kategori UKM per individu, label klaster hasil K-Means ditambahkan sebagai fitur tambahan (*one-hot*) yang dapat pula ditambahkan jarak ke centroid. Dua model pokok digunakan:

1. Multinomial Logistic Regression (MLR): sederhana dan mudah dijelaskan [23].
2. Random Forest (RF): kuat dan stabil, serta dapat menyediakan *feature importance* [24].

Skema evaluasi: 80% latih/20% uji. Metrik dilaporkan: akurasi, macro-precision, macro-recall, macro-F1, dan confusion matrix. Perbandingan Hybrid dan Baseline dinilai dengan uji McNemar pada prediksi berpasangan sehingga pada skema k-fold dapat memakai Wilcoxon signed-rank.

Hasil (n=5,238):

Tabel 2 Perbandingan kinerja hybrid dan baseline

Pendekatan	Model	Akurasi	Macro-Precision	Macro-Recall
Baseline	MLR	0,960	0,949	0,941
Baseline	RF	0,994	0,992	0,991
Hybrid	MLR	0,981	0,975	0,964
Hybrid	RF	0,992	0,992	0,990

### 2.4. Implikasi Praktis

1. Segmentasi sebagai konteks latar: Label klaster menyediakan sinyal konteks yang membantu model membedakan kategori UKM yang sesuai (misalnya kategori komunitas dengan kategori kepemimpinan), terutama ketika fitur terbatas [11].
2. Transparansi rekomendasi: Random Forest memberikan *feature importance*, sedangkan Multinomial Logistic Regression menyediakan koefisien. Keduanya memudahkan penyusunan alasan singkat (segmen & butir dominan) pada setiap rekomendasi, sehingga meningkatkan trust pengguna [25].
3. Penargetan layanan kemahasiswaan:
  - a. Klaster 0: intervensi afektif menggunakan role model, storytelling dampak visi-misi sehingga meningkatkan P3.

- b. Klaster 1: operasionalisasi & literasi menggunakan use case yang konkret, SOP yang menautkan aktivitas ke indikator visi-misi sehingga meningkatkan P1, P2, P4.
  - c. Klaster 2: champions menggunakan mentor lintas unit dan diseminasi praktik baik.
4. Integrasi sistem: Output (kategori UKM + probabilitas + segmen + alasan) dihubungkan ke *dashboard* UKM agar segera dapat ditindaklanjuti (notifikasi booth/agenda relevan) [26].

### 2.5. Keterbatasan dan Arah Lanjutan

Hasil penelitian ini menunjukkan efektivitas pendekatan hybrid untuk personalisasi rekomendasi UKM, namun ada beberapa keterbatasan yang perlu diperhatikan, sebagai berikut:

1. Proses pelabelan UKM pilihan sebagai *ground truth* masih terbatas. Apabila data *ground truth* pilihan UKM mahasiswa tidak tersedia, penelitian ini mengandalkan label komposit yang direkayasa dari skor minat kategori (proxy) sebagai target klasifikasi. Pendekatan ini berpotensi menimbulkan bias atau ketidakakuratan, misalnya apabila mahasiswa memiliki minat beragam yang tidak tercermin oleh satu kategori UKM dominan. Untuk itu, evaluasi tambahan perlu dilakukan dengan uji ablasi fitur, misalnya mengukur kinerja model tanpa fitur klaster atau tanpa variabel demografis yang gunanya untuk memastikan bahwa peningkatan kinerja memang berasal dari informasi klaster dan bukan artefak dari *proxy* label.
2. Cakupan data penelitian masih terbatas pada satu institusi (Universitas X) dalam satu periode (PKKMB 2024). Model yang dihasilkan mungkin mengandung bias spesifik terhadap kultur atau karakteristik kampus tersebut. Hal ini memengaruhi generalisasi: sistem belum teruji untuk konteks kampus lain atau angkatan tahun berbeda.
3. Evaluasi yang dilakukan berfokus pada metrik akurasi dan evaluasi offline (berbasis data historis), tanpa melibatkan umpan balik langsung dari pengguna atau pengukuran dampak jangka panjang. Dampak sesungguhnya dari penggunaan sistem rekomendasi ini, misalnya apakah rekomendasi yang lebih personal dapat meningkatkan keterlibatan dan retensi mahasiswa dalam UKM yang belum diteliti secara menyeluruh.

Berdasarkan keterbatasan di atas, beberapa arah pengembangan lanjutan diusulkan untuk meningkatkan kualitas dan daya guna sistem rekomendasi UKM ini:

Dengan melakukan serangkaian agenda di atas, sistem rekomendasi UKM berbasis hybrid clustering-

classification ini dapat terus ditingkatkan akurasi, interpretabilitas, dan utilitasnya. Penerapan luas di berbagai kampus serta evaluasi berkelanjutan akan mendorong terciptanya layanan rekomendasi yang andal untuk mendukung pengembangan minat dan bakat mahasiswa baru di Indonesia.

#### 4. Kesimpulan

Pendekatan hybrid clustering-classification yang memanfaatkan K-Means sebagai segmen minat dan model klasifikasi seperti Multinomial Logistic Regression dan Random Forest terbukti efektif dalam merekomendasikan UKM secara personal. Dari sekitar 5.238 responden, segmentasi K-Means mengidentifikasi kluster dengan pola kecenderungan yang berbeda, misalnya kluster pecinta olahraga, seni & budaya, dan pengabdian/komunitas, sehingga membantu meningkatkan akurasi prediksi ketika label kluster ditambahkan sebagai fitur. Evaluasi menunjukkan kinerja model hybrid melampaui baseline tanpa informasi kluster, baik dalam akurasi maupun macro-precision/recall. Selain itu, analisis asosiasi seperti Chi-Square dan Cramer's V menegaskan adanya kaitan antara program studi/fakultas dan kategori UKM.

Secara praktis, sistem prototipe yang dikembangkan dapat memberikan rekomendasi UKM lengkap dengan probabilitas prediksi, segmen minat, dan alasan singkat, sehingga memudahkan mahasiswa baru dalam memilih UKM yang sesuai. Implementasi sistem ini juga memungkinkan integrasi dengan dasbor layanan kemahasiswaan untuk tindak lanjut berupa notifikasi kegiatan atau booth pendaftaran yang relevan. Meski demikian, penelitian ini hanya mencakup satu kampus dalam satu periode; studi lanjutan perlu menguji generalisasi antar tahun dan institusi serta menilai dampak jangka panjang terhadap retensi atau keaktifan mahasiswa di UKM.

#### REFERENSI

- [1] A. Nadhira Tsurayya, R. Febriana, and M. A. Wildan, "Peran Unit Kegiatan Mahasiswa (UKM) dan Konsep Kepemimpinan terhadap Pengembangan Diri Mahasiswa FEB UTM," *J. Adm. Kant.*, vol. 12, no. 2, pp. 158–173, 2024.
- [2] T. Pipit Mulyah, Dyah Aminatun, Sukma Septian Nasution, Tommy Hastomo, Setiana Sri Wahyuni Sitepu, "Panduan Pengenalan Kehidupan Kampus bagi Mahasiswa Baru (PKKMB) 2024," *J. GEEJ*, vol. 7, no. 2, 2024.
- [3] N. Ailmi, Z. Saharuna, E. Tungadi, T. Elektro, P. Negeri, and U. Pandang, "Metode Klasifikasi Pada Aplikasi Pendukung Keputusan Untuk Pemilihan Unit Kegiatan Mahasiswa," 2020.
- [4] A. Fanani, S. Aminah, and F. E. Purwiantono, "Sistem Informasi Klasifikasi Minat Siswa Terhadap Ekstrakurikuler Untuk Pembinaan Lomba Dengan Metode K-Means Clustering," *Smatika J.*, vol. 13, no. 01, pp. 166–175, 2023, doi: 10.32664/smatika.v13i01.920.
- [5] A. Mulyana, Y. Hermawan, and N. J. Saputri, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Rekomendasi Pilihan Program Studi Pada Mahasiswa Baru (Studi Kasus di Institut Bisnis dan Informatika Kesatuan)," *KERNEL J. Ris. Inov. Bid. Inform. dan Pendidik. Inform.*, vol. 5, no. 1, pp. 60–72, 2024, doi: 10.31284/j.kernel.2024.v5i1.7624.
- [6] N. Drushchak, V. Tyshchenko, and N. Polyakovska, "Towards Responsible AI in Education: Hybrid Recommendation System for K-12 Students Case Study," *Proc. - 2025 IEEE/ACM Int. Work. Responsible AI Eng. RAIE 2025*, pp. 61–68, 2025, doi: 10.1109/RAIE66699.2025.00015.
- [7] N. Dwitianti, S. A. Kumala, and S. D. Handayani, "Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms," *J. RESTI*, vol. 8, no. 6, pp. 768–778, 2024, doi: 10.29207/resti.v8i6.5514.
- [8] Samidi, R. Y. Suladi, and D. Kusumaningsih, "Comparison of the RFM Model's Actual Value and Score Value for Clustering," *J. RESTI*, vol. 7, no. 6, pp. 1430–1438, 2023, doi: 10.29207/resti.v7i6.5416.
- [9] D. Nastiti, "Peran Organisasi Mahasiswa Dalam Pembentukan Sikap Demokratis," *Prima Magistra J. Ilm. Kependidikan*, vol. 4, no. 1, pp. 64–76, 2023, doi: 10.37478/jpm.v4i1.2433.
- [10] Suparyati, E. Utami, and A. H. Muhammad, "Applying Different Resampling Strategies In Random Forest Algorithm To Predict Lumpy Skin Disease," *J. RESTI*, vol. 6, no. 4, pp. 555–562, 2022, doi: 10.29207/resti.v6i4.4147.
- [11] A. N. D. Andra and Z. K. A. Baizal, "E-commerce Recommender System Using PCA and K-Means Clustering," *J. RESTI*, vol. 6, no. 1, pp. 57–63, 2022, doi: 10.29207/resti.v6i1.3782.
- [12] M. H. M. Maruyama, L. W. Silveira, J. P. M. de Oliveira, I. Gasparini, and V. Maran, "Hybrid Recommender System for Educational Resources to the Smart University Campus Domain," *Int. Conf. Comput. Support. Educ. CSEDU - Proc.*, vol. 1, no. Csedu, pp. 47–56, 2023, doi: 10.5220/0011841900003470.
- [13] M. Gagolewski, M. Bartoszek, and A. Cena, "Are cluster validity measures (in) valid?," *Inf. Sci. (Ny)*, vol. 581, pp. 620–636, 2021, doi: 10.1016/j.ins.2021.10.004.
- [14] R. Ananda, Muhammad Zidny Naf'an, Amalia Beladinna Arifa, and Auliya Burhanuddin, "Recommendation System for Specialization Selection Using K-Means Density Canopy," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 172–179, 2020, doi: 10.29207/resti.v4i1.1531.
- [15] A. N. A. Aldania, A. M. Soleh, and K. A. Notodiputro, "A Comparative Study of CatBoost and Double Random Forest for Multi-class Classification," *J. RESTI*, vol. 7, no. 1, pp. 129–137, 2023, doi: 10.29207/resti.v7i1.4766.
- [16] V. R. S. Nastiti, Y. Azhar, and R. S. Putri, "Logistic Regression Using Hyperparameter Optimization on COVID-19 Patients' Vital Status," *J. RESTI*, vol. 7, no. 3, pp. 681–687, 2023, doi: 10.29207/resti.v7n3.4868.
- [17] M. Aucancela, A. G. Briones, and P. Chamoso, "Educational Recommender Systems: A Systematic Literature Review," *Barcelona Conf. Educ. 2023 Off. Conf. Proc.*, no. November 2023, pp. 933–951, 2023,

- doi: 10.22492/issn.2435-9467.2023.74.
- [18] A. F. Mohamed Nafuri, N. S. Sani, N. F. A. Zainudin, A. H. A. Rahman, and M. Aliff, "Clustering Analysis for Classifying Student Academic Performance in Higher Education," *Appl. Sci.*, vol. 12, no. 19, 2022, doi: 10.3390/app12199467.
- [19] J. C. Sanguino, O. Mariño, N. Cardozo, R. Manrique, and M. Linares-Vásquez, "A course hybrid recommender system for limited user information scenarios," *J. Educ. Data Min.*, vol. 14, no. 3, pp. 162–188, 2022, doi: 10.5281/zenodo.7304829.
- [20] D. Ariatmanto and A. M. Rifai, "The Impact of Feature Extraction in Random Forest Classifier for Fake News Detection," *J. RESTI*, vol. 8, no. 6, pp. 730–736, 2024, doi: 10.29207/resti.v8i6.6017.
- [21] R. Ananda and Achmad Zaki Yamani, "Determination of Initial K-means Centroid in the Process of Clustering Data Evaluation of Teaching Lecturers," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 3, pp. 544–550, 2020, doi: 10.29207/resti.v4i3.1896.
- [22] A. A. C. Rindu, R. Astriratma, and A. Zaidiah, "K-Means Algorithm Implementation for Project Health Clustering," *J. RESTI*, vol. 7, no. 5, pp. 1064–1076, 2023, doi: 10.29207/resti.v7i5.5181.
- [23] س. ع. جلال and ل. ع. كيورك, "Using Multinomial Logistic Regression Model To Predict The Most Important Factors Affecting In The Students' Selection For Specialization In Secondary Level For Some Schools Erbil City," *مجلة جامعة كركوك للعلوم الإدارية والاقتصادية*, no. October 2023, p. 1, 2023, doi: 10.32894/1913-013-002-001.
- [24] I. Wardhana, M. Ariawijaya, V. A. Isnaini, and R. P. Wirman, "Gradient Boosting Machine, Random Forest and Light GBM for Classification of Dried Beans," *J. RESTI*, vol. 6, no. 1, pp. 92–99, 2022, doi: 10.29207/resti.v6i1.3682.
- [25] H. N. Pradani and F. Mahananto, "Studi Literatur Human Activity Recognition (HAR) Menggunakan Sensor Inersia," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1193–1206, 2021, doi: 10.29207/resti.v5i6.3665.
- [26] A. A. Mutezar and U. Salamah, "Pengembangan Sistem Manajemen Event Pameran Karya Mahasiswa Menggunakan Metode Extreme Programming," *J. RESTI*, vol. 5, no. 4, pp. 809–819, 2021, doi: 10.29207/resti.v5i4.3249.