JOISIE (Journal Of Information Systems And Informatics Engineering)
Vol. 9, No.1, Juni 2025, Hlm 109-122
Received: 20 April 2025 Revised: 6 Mei 2025, Accepted: 24 Juni 2025

p- ISSN: 2503-5304
e- ISSN: 2527-3116

ANALISIS *CLUSTERING* REKOMENDASI MATA KULIAH PEMINATAN BERDASARKAN KARIR ALUMI MENGGUNAKAN *MACHINE LEARNING*

Rahmat Maulana[1)], Fathoni[2*)]
[1,2]Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang email:
[1]09031282227088@student.unsri.ac.id, [2]fathoni@unsri.ac.id

Abstract

The Faculty of Computer Science at Sriwijaya University offers various elective courses designed to help students prepare for their future careers. However, many students struggle to choose courses that align with their career goals. This study aims to analyze the relationship between elective courses and alumni career paths to provide relevant course recommendations using the K-Means Clustering method. Based on tracer study data of undergraduate alumni (S1) prior to 2020, the clustering process was carried out using Google Colab and RapidMiner, with the number of clusters adjusted to match 18 career categories. Evaluation using the Silhouette Score indicated the best result at K = 3, but K = 18 was retained to preserve the granularity of career categories. A heatmap visualization was used to identify the dominance of certain courses within each cluster. The results show a relatively low clustering accuracy, with a Cluster Purity score of 21.34%, indicating that the majority of data did not align with actual career labels. Although the performance of the K-Means algorithm was suboptimal, this approach introduces a new alternative in academic recommendation systems based on alumni career data and may serve as an initial reference for students in selecting elective courses relevant to their desired careers.

Keywords: *Recommendation, Elective Courses, Student Career, K-Means Clustering*

Abstrak

Fakultas Ilmu Komputer Universitas Sriwijaya menyediakan berbagai mata kuliah peminatan yang dirancang untuk membantu mahasiswa mempersiapkan karir masa depan. Namun, banyak mahasiswa kesulitan memilih mata kuliah yang selaras dengan tujuan karir mereka. Upaya ini dilakukan guna menganalisis keterkaitan antara mata kuliah peminatan dan karir alumni guna memberikan rekomendasi pemilihan mata kuliah yang relevan, dengan menerapkan metode *K-Means Clustering*. Berdasarkan data *tracer study* alumni program sarjana (S1) sebelum tahun 2020, proses klasterisasi dilakukan menggunakan *Google Colab* dan *RapidMiner*, dengan jumlah klaster disesuaikan sebanyak 18 kategori karir. Evaluasi menggunakan *Silhouette Score* menunjukkan hasil terbaik pada K = 3, namun tetap digunakan K = 18 untuk mempertahankan kedetailan kategori karir. Visualisasi *heatmap* digunakan untuk mengidentifikasi dominasi mata kuliah di tiap klaster. Hasil penelitian menunjukkan bahwa akurasi klastering tergolong rendah, dengan nilai *Cluster Purity* sebesar 21,34%, yang mencerminkan ketidaksesuaian mayoritas data terhadap label karir aktual. Meskipun performa algoritma K-Means belum optimal, pendekatan ini memperkenalkan alternatif baru dalam sistem rekomendasi akademik berbasis data karir alumni, dan dapat digunakan sebagai referensi awal bagi mahasiswa dalam menentukan mata kuliah peminatan yang relevan dengan karir yang diinginkan.

Kata Kunci: Rekomendasi, Peminatan, Karir Mahasiswa, *K-Means Clustering*

1. PENDAHULUAN

Kemajuan teknologi telah mendorong pemanfaatan kecerdasan buatan yang semakin berkembang di berbagai bidang termasuk pendidikan tinggi. Salah satunya machine learning digunakan untuk menganalisis data dan mengubahnya menjadi informasi yang berguna dalam pengambilan keputusan (Eka, 2025). Data berperan dalam mendukung proses akademik serta membantu mahasiswa memilih mata kuliah peminatan yang selaras dengan kebutuhan karir. Mahasiswa perlu mempertimbangkan kesesuaian mata kuliah peminatan yang dipilih sebaiknya selaras dengan minat, kemampuan, serta kebutuhan pasar kerja (Nurchalia et al., 2023). Mata kuliah peminatan dalam kurikulum tidak hanya menyajikan pokok bahasan tiap mata kuliah, tetapi juga membimbing mahasiswa dalam memilih waktu yang tepat untuk memperdalam bidang yang diminati (Fernando et al., 2022). Pada Fakultas Ilmu Komputer Universitas

https://doi.org/10.35145/joisie.v9i1.4954

Sriwijaya (Fasilkom UNSRI), mata kuliah peminatan umumnya ditawarkan mulai semester enam, di mana mahasiswa program studi terkait dapat memilih sesuai dengan bidang yang diminati.

Mahasiswa memiliki kebebasan dalam memilih mata kuliah peminatan, namun sering kali mengalami kesulitan dalam menentukan pilihan yang sesuai dengan kurikulum karena keterbatasan informasi dan kompleksitas kriteria (Darmawan et al., 2024). Selain itu, tidak semua mahasiswa memiliki pemahaman yang jelas mengenai mata kuliah yang paling sesuai dengan karir yang diinginkan. Meskipun dosen pembimbing akademik dapat memberikan saran, keterbatasan waktu serta jumlah mahasiswa yang banyak membatasi efektivitas pendampingan (Fernando et al., 2022). Konsekuensinya, banyak mahasiswa memilih peminatan secara impulsif tanpa mempertimbangkan kemampuan akademik, sehingga menghadapi kesulitan perkuliahan dan kurang siap berkarir (Lestari & Perdana, 2020). Karena itu, pendekatan berbasis data diperlukan untuk membantu mahasiswa dalam pengambilan keputusan yang lebih terarah.

Maka dari itu, pendekatan berbasis data seperti *machine learning* dapat menawarkan solusi dengan mempelajari data menggunakan algoritma dan membuat keputusan berdasarkan pola-pola yang ditemukan dalam data (Ananto et al., 2023). Salah satu algoritma yang dapat digunakan adalah *K-Means Clustering*, yang mampu mengelompokkan data ke dalam beberapa cluster berdasarkan kesamaan tertentu. Metode K-Means merupakan algoritma *clustering* berbasis jarak yang membagi data ke dalam beberapa klaster secara *partitioning*, di mana setiap data akan dikelompokkan ke klaster tertentu berdasarkan jarak terdekat dengan pusat klaster yang diperbarui secara iteratif hingga mencapai konvergensi (Deti Karmanita & Billy Hendrik, 2023). Algoritma ini efektif dalam mengidentifikasi pola atau kelompok dalam data yang tidak berlabel (Darsono & Andrianti, 2022). Mengacu pada pemilihan mata kuliah peminatan, Algoritma *K-Means Clustering* dapat diterapkan untuk mengelompokkan alumni berdasarkan pola mata kuliah yang mereka ambil dan karir yang mereka jalani, sehingga dapat menentukan mata kuliah yang relevan dengan pilihan karir alumni.

Berbagai penelitian telah dilakukan untuk meningkatkan akurasi dan relevansi rekomendasi pemilihan mata kuliah peminatan bagi mahasiswa. Penggunaan *algoritme Decision Tree* untuk merekomendasikan mata kuliah berdasarkan nilai prasyarat mahasiswa. Hasil penelitian menunjukkan bahwa algoritme ID3 memiliki akurasi tertinggi sebesar 86,90%, menjadikannya metode paling efektif dalam klasifikasi peminatan (Putu et al., 2019). Sementara itu, penerapan algoritme Apriori untuk mengidentifikasi pola pemilihan mata kuliah berdasarkan data historis mahasiswa, dengan tingkat confidence mencapai 80,16% (Khoerullah et al., 2019; Syahrul & Solichin, 2022). Pendekatan ini terbukti membantu mahasiswa dalam menentukan mata kuliah yang lebih sesuai dengan kompetensi dan aspirasi karier mereka. Penelitian terbaru menerapkan *algoritme K-Nearest Neighbor* (KNN) untuk membantu mahasiswa dalam menentukan konsentrasi peminatan yang sesuai dengan bakat dan tujuan karir mereka di Program Studi Manajemen Universitas Muhammadiyah Makassar (Nirmada, 2024). Sistem ini mengklasifikasikan peminatan mahasiswa berdasarkan nilai akademik dan preferensi mata kuliah, menghasilkan tingkat keberhasilan sebesar 76,71% dalam rekomendasi pemilihan mata kuliah.

Metode *machine learning* seperti klasifikasi dan *clustering* juga telah banyak digunakan dalam penelitian terkait rekomendasi peminatan. Penggunakan *algoritme* C5.0 dan C4.5 untuk menghasilkan rekomendasi dengan tingkat akurasi yang beragam, mulai dari 57% hingga 98% (Lusiah & Purwanto, 2023; Wahyudinnur et al., 2024) Selain itu, penelitian lain juga menerapkan *algoritme K-Means* untuk pengelompokan mahasiswa berdasarkan keahlian mereka (Afifuddin et al. 2019). Afifuddin & Nurjanah (2019) menguji algoritme K-Means pada dataset yang berisi 660 mahasiswa dan berhasil mengelompokkan mereka ke dalam tiga kategori keahlian yang sesuai. Pendekatan lain yang mengkombinasikan *K-Means* dengan *Collaborative Filtering* juga diterapkan oleh Fernando et al. (2022), yang meningkatkan akurasi rekomendasi hingga 9,92% dibandingkan metode sebelumnya. Selain itu, Budiman et al. (2024) menerapkan metode *Multi Attribute Utility Theory* (MAUT) dan menemukan bahwa peminatan terbaik berada di bidang Jaringan Komputer & Sistem Informasi dengan skor 0,545. Karmanita & Hendrik (2023) menggunakan *K-Means Clustering* dan berhasil mengelompokkan mahasiswa ke dalam tiga kelompok peminatan dengan validasi centroid sebesar 35,89%.

Berbeda dari penelitian sebelumnya yang umumnya hanya memanfaatkan nilai akademik, minat mahasiswa, atau data internal kampus dalam merekomendasikan peminatan, penelitian ini secara spesifik mengintegrasikan data tracer study alumni untuk membangun pemetaan antara pola pengambilan mata kuliah dan karir aktual yang dijalani. Namun, sebagian besar penelitian terdahulu belum secara langsung memanfaatkan data alumni sebagai acuan utama dalam pengambilan keputusan akademik, sehingga rekomendasi yang dihasilkan cenderung bersifat prediktif tanpa validasi nyata terhadap realitas dunia kerja.

https://doi.org/10.35145/joisie.v9i1.4954

Untuk itu, pendekatan berbasis machine learning digunakan, khususnya algoritma *K-Means Clustering*. Pemilihan *K-Means* didasarkan pada kemampuannya yang efektif dalam menangani data tanpa label (*unsupervised*) seperti data tracer study alumni, di mana tidak ada kelas target yang pasti, melainkan pola yang tersembunyi dalam kumpulan data. *K-Means* dapat mengelompokkan alumni berdasarkan kesamaan pola mata kuliah yang ditempuh dan karir yang dicapai, sehingga hasil klaster dapat digunakan sebagai dasar rekomendasi peminatan yang lebih realistis dan relevan dengan kebutuhan pasar kerja.

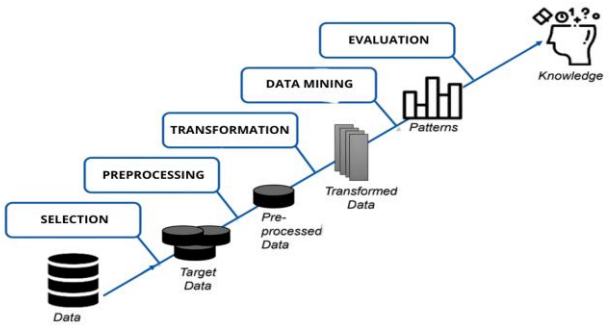
Dengan demikian, penelitian ini menghadirkan sudut pandang baru dalam pengambilan keputusan akademik, yaitu berbasis data historis alumni yang telah terbukti di dunia kerja, bukan sekadar prediksi internal. Hal ini menjadi kontribusi penting dalam merancang sistem rekomendasi peminatan yang kontekstual, berbasis bukti, dan berorientasi pada kesiapan karir mahasiswa.

2. METODE PENELITIAN

2.1 Teknik Pengumpulan Data dan Sumber Data

Teknik pengumpulan data dengan memanfaatkan data primer dari laporan atau dokumen resmi, yaitu data *tracer study* alumni yang dikumpulkan dan diolah langsung oleh kemahasiswaan Fasilkom Unsri. Data yang digunakan adalah data mengenai karir alumni dalam tahun 2020 kebelakang. Sementara itu, data mata kuliah berasal dari sistem akademik yang mencatat mata kuliah berdasarkan kurikulum yang ditempuh mahasiswa semasa studi. Pengumpulan data diperlukan untuk bahan penelitian dalam memberikan rekomendasi pemilihan mata kuliah peminatan yang sesuai dengan kecenderungan karir dan dapat menjadi acuan bagi mahasiswa dalam menentukan peminatan yang lebih tepat dan relevan terhadap tujuan karir masa depan.

2.2 Tahapan Penelitian



Gambar 1. Tahapan *Knowledge Discovery in Database*

Gambar di atas merupakan diagram alur tahapan pada proses penelitian ini. Proses data yang digunakan menerapkan pendekatan *Knowledge Discovery in Database* (KDD). Pendekatan ini melibatkan serangkaian langkah yang menggunakan hasil analisis pola dari data yang dikumpulkan, sehingga informasi yang dihasilkan dapat dipahami dengan lebih jelas dan mudah (Utomo et al., 2023). Nantinya, data yang diperoleh melalui proses KDD akan menjadi sebuah basis data yang dapat digunakan sebagai dasar dalam pengambilan Keputusan (Atalya et al., 2024).

2.2.1 *Data Selection*

Tahapan *data selection* merupakan proses awal dalam kerangka Knowledge Discovery in Databases (KDD) yang bertujuan untuk menyaring dan memilih data operasional yang relevan. Data yang dipilih pada tahap ini akan menjadi dasar dalam proses *data mining*, sehingga hanya informasi yang sesuai dengan kebutuhan analisis yang digunakan dalam penelitian (Alejandrino, 2021).

2.2.2 *Preprocessing*

Setelah penugumpulan dataset, langkah berikutnya adalah *preprocessing* yang di mana dataset dianalisis untuk memilih atribut yang relevan dan membagi data menjadi data *training* dan data *testing*. Dalam hal ini, peneliti akan memilih fitur-fitur yang relevan untuk dianalisis lebih lanjut. Proses pengujian dilakukan dengan membagi dataset ke dalam dua bagian, yaitu data *training* dan data *testing*, dengan

https://doi.org/10.35145/joisie.v9i1.4954

berbagai rasio seperti 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, dan 10:(Mohammed & Alsunosi, 2022). Rasio 70:30 dan 80:20 sering digunakan sebagai standar dalam pelatihan model pembelajaran mesin karena memberikan keseimbangan antara ketepatan pelatihan dan kemampuan generalisasi model (Sivakumar et al., 2024). Dalam penelitian ini, pembagian yang diterapkan adalah 80% untuk data *training* dan 20% untuk data *testing*.

2.2.3 Transformation

Transformasi adalah proses mengubah data yang telah diseleksi agar sesuai dengan kebutuhan dalam *data mining*, bergantung pada jenis database, model informasi, dan tujuan analisis. Studi ini fokus pada transformasi data mentah menjadi format yang lebih relevan untuk analisis, termasuk penggabungan data dari berbagai sumber, normalisasi untuk konsistensi, dan *encoding* untuk kompatibilitas dengan model analisis. Transformasi fitur juga mencakup konstruksi ulang fitur dari ruang berdimensi tinggi menjadi lebih rendah untuk mempermudah analisis lanjutan (Bahri et al., 2021).

2.2.4 Data mining

Algoritma *K-Means* merupakan salah satu algoritma pengelompokkan (*clustering*) berbasiskan metode *non-hierarchy* yang mempartisi data dan membentuk satu atau lebih kelompok yang memiliki kesamaan (Haviluddin et al., 2021). Kelebihan utama algoritma *K-Means* terletak pada sifatnya yang mudah diimplementasikan, memiliki struktur yang sederhana, fleksibel dalam penerapannya, cepat melakukan proses komputasi, dan sering digunakan dalam proses *data mining* (Muningsih et al., 2021).

Algoritma *K-Means* untuk *clustering* bisa dilakukan dengan (Zhu et al., 2021) :

1) Tentukan jumlah *cluster* (*k*) yang akan dibentuk.
2) Pilih secara acak k titik sampel sebagai pusat awal (*initial cluster centers*).
3) Hitung jarak dari setiap titik data ke semua pusat *cluster* menggunakan *Euclidean Distance* untuk menentukan kedekatan dan keanggotaan cluster. Persamaan *Euclidean Distance* ditunjukkan dalam bentuk dua dimensi dan tiga dimensi sebagai berikut:

   persamaan (1):

$$d(x,y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

   persamaan (2):

$$d(x,y) = \sqrt{(x_1 - x_{21})^2 + (y_2 - y_2)^2 + (z_3 - z_3)^2}$$

   Keterangan:
   d(x,y) = Jarak antara objek data dan *centroid*.
   xk = Data ke-i pada atribut ke-k
   yk = Nilai *centroid* pada atribut ke-k.
   n = jumlah atribut data
4) Kelompokkan titik data ke *cluster* terdekat sesuai dengan jarak minimum.
5) Hitung ulang posisi pusat *cluster* sebagai rata-rata dari semua titik data yang masuk ke dalam *cluster* tersebut.
6) Ulangi proses perhitungan jarak dan pemutakhiran pusat *cluster* hingga hasil tidak lagi berubah, atau hingga mencapai batas maksimum iterasi.

2.2.5 Evaluation

Hasil *clustering* yang diperoleh dari algoritma *K-Means* perlu melalui tahap evaluasi untuk menilai seberapa baik klaster yang terbentuk merepresentasikan struktur alami dari data. Pengukuran tingkat kesalahan pada penelitian ini dengan menggunakan metrik *Silhouette Coefficient*, yaitu ukuran yang mengevaluasi konsistensi internal antar anggota klaster dibandingkan dengan klaster lain. Selain itu, versi lanjutan dari metrik ini, yaitu *cluster purity*, karena memberikan gambaran yang lebih jelas mengenai kualitas klaster dalam hal representasi kelas data yang ada. *Purity* mengukur seberapa baik data dalam setiap klaster dapat dikategorikan ke dalam kelas yang benar, dengan menghitung jumlah data yang sesuai

https://doi.org/10.35145/joisie.v9i1.4954

dengan kelas mayoritas dalam klaster tersebut (Suresh Sikhakolli & Kiran Sikhakolli DrDY, 2023). Sementara itu, Ahmadinejad et al., (2023) menunjukkan bahwa *purity* juga dapat dimanfaatkan dalam integrasi pengetahuan latar belakang selama proses *clustering*, guna meningkatkan interpretabilitas serta relevansi hasil terhadap struktur data yang sebenarnya.

3. METODE PENELITIAN

Pembahasan berikut menjabarkan analisis serta hasil dari tahapan penelitian, yaitu *data selection*, *data pre-processing, data transformation, data mining,* dan *data evaluation*, dengan memanfaatkan *tools RapidMiner* dan *Python* melalui *Google Collaboratory* untuk memperoleh kesimpulan terkait rekomendasi pemilihan mata kuliah peminatan yang sesuai dengan karir yang diinginkan oleh mahasiswa.

3.1 *Data Selection*

Proses pengumpulan data dilakukan dengan memanfaatkan data primer yang bersumber dari laporan atau dokumen resmi, yaitu data tracer study alumni yang dikumpulkan dan diolah langsung oleh pihak Kemahasiswaan Fakultas Ilmu Komputer (Fasilkom) Universitas Sriwijaya. Data yang digunakan merupakan informasi mengenai karir alumni hingga tahun 2020 ke belakang. Sementara itu, data mata kuliah berasal dari sistem akademik yang mencatat mata kuliah berdasarkan kurikulum yang ditempuh mahasiswa semasa studi.

3.1.1 Pemilihan Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari 1.725 entri data mahasiswa, yang di mana setiap entri merepresentasikan satu alumni. Hasil pemilihan atribut dataset ditampilkan pada Tabel 1.

Tabel 1. Hasil Pemilihan Dataset

No	Nim	Angkatan	Program Studi	Kurikulum	Kategori Instansi	Instansi	Kategori Karir	Mata Kuliah
1	21094	2016	SI	2015	Perusahaan Swasta	PT Jamkrindo	Teknologi Informasi dan Sistem	Manajemen Sistem Informasi, Mobile Technology, Knowledge Management, Sistem Pendukung Keputusan, E-Government.
2	21010	2016	SI	2015	Pemerintahan	Dinas PU…	Administrasi Umum	Manajemen Sistem Informasi, E-Government, Aplikasi Komputer Akuntansi, Sistem Pendukung Keputusan, Manajemen Rantai Suplai.
3	21049	2016	TI	2015	Pemerintahan	bppkad	Administrasi Umum	Sistem Pendukung Keputusan, Sistem Pakar, Basis Data Lanjut, Pemrograman Internet, Manajemen Jaringan.
4	21116	2016	TI	2015	BUMN	Pos Indonesia	Pengelolaan Data dan Informasi	Sistem Pakar, Basis Data Lanjut, Sistem Pendukung Keputusan, Jaringan Syaraf Tiruan, Pemrograman Internet.
5	21075	2016	SI	2015	Pemerintahan	Dinas Perindust…	Administrasi Umum	E-Government, Manajemen Sistem Informasi, Sistem Pendukung Keputusan, CRM, Manajemen Rantai Suplai.
6	21021	2016	SI	2015	Pemerintahan	BPS	Pengelolaan Data dan Informasi	Manajemen Sistem Informasi, Sistem Pendukung Keputusan, Knowledge Management, E-Government, CRM.
7	21127	2016	TI	2015	Perusahaan Swasta	PT. Rifan Fina...	Konsultasi dan Penyuluhan	Sistem Pendukung Keputusan, Sistem Pakar, Multimedia, Grafika Komputer,

https://doi.org/10.35145/joisie.v9i1.4954

								Basis Data Lanjut.
…	…	…	…	…	…	…	…	…
1725	26113	2021	SI	2021	Perusahaan Swasta	PT Mardel	Administrasi Umum	Manajemen Rantai Suplai (SCM), Manajemen Hubungan Pelanggan (CRM), Audit Sistem Informasi, Sistem Pendukung Keputusan, Psikologi Pengguna

3.1.2 Pemilihan Fitur atau Variabel

Pada tahap pemilihan fitur atau variabel, dilakukan proses penyaringan berdasarkan kelengkapan data dan relevansi terhadap tujuan penelitian. Setelah dilakukan pembersihan data, jumlah entri yang digunakan untuk analisis lebih lanjut disaring menjadi sebanyak 287 entri data mahasiswa yang memiliki informasi lengkap dan sesuai kebutuhan. Proses dan hasil pemilihan fitur tersebut disajikan pada Tabel 2.

Tabel 2. Hasil Pemilihan FItur atau Variabel

No	Program Studi	Kategori Karir	Mata Kuliah
1	SI	2016	Manajemen Sistem Informasi, Mobile Technology, Knowledge
2	SI	2016	Management, Sistem Pendukung Keputusan, E-Government.
3	TI	2016	Sistem Pendukung Keputusan, Sistem Pakar, Basis Data Lanjut, Pemrograman Internet, Manajemen Jaringan.
4	SI	2016	Sistem Pakar, Basis Data Lanjut, Sistem Pendukung Keputusan, Jaringan Syaraf Tiruan, Pemrograman Internet.
5	SI	2016	Manajemen Sistem Informasi, Sistem Pendukung Keputusan, Knowledge
6	TI	2016	Sistem Pendukung Keputusan, Sistem Pakar, Multimedia, Grafika Komputer, Basis Data Lanjut.
	…	…	…
287	SI	2021	Manajemen Rantai Suplai (SCM), Manajemen Hubungan Pelanggan (CRM), Audit Sistem Informasi, Sistem Pendukung Keputusan, Psikologi Pengguna

Berdasarkan Tabel 2, pemilihan fitur ini bertujuan untuk melihat korelasi antara mata kuliah yang ditempuh selama studi dengan jalur karir yang dijalani oleh alumni, sehingga dapat dianalisis lebih lanjut korelasi antara mata kuliah yang diambil oleh alumni dengan karir yang dijalani.

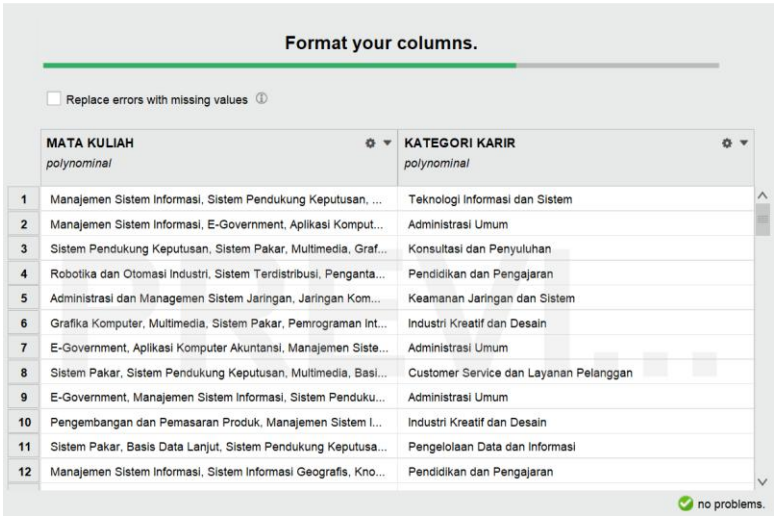
3.2 *Data Processing*

Pada tahap ini, dilakukan pemilihan dan persiapan data sebelum penerapan algoritma K-Means. Langkah-langkah pre-processing yang diimplementasikan adalah sebagai berikut.

3.2.1 Pemilihan Data Relevan

Hanya data alumni yang memuat informasi tentang mata kuliah peminatan dan kategori karir yang digunakan dalam analisis *clustering*. Gambar 2 menunjukkan hasil dari proses pemilihan data relevan tersebut, yang menjadi dasar dalam membentuk dataset untuk tahap analisis selanjutnya.

https://doi.org/10.35145/joisie.v9i1.4954



Gambar 2. Hasil Pemilihan Data Relevan

3.2.2 Format Data Konsisten

Dipastikan semua data numerik memiliki format yang sesuai untuk diolah dalam model *machine learning*. Pada kolom kategori karir dikonversi ke format numerik menggunakan *Mapping Final* yakni mengelompokkan atau meratakan *labeling* sesuai kategori yang relevan. Tabel 3 menyajikan hasil pengelompokan kategori karir tersebut secara terurut.

Tabel 3. Hasil pengelompokan kategori karir

Kategori Karir	Label Kategori
Administrasi Umum	1
Pengembangan Perangkat Lunak	2
Pendidikan dan Pengajaran	3
Teknologi Informasi dan Sistem	4
Pekerjaan Lapangan dan survey	5
Industri Kreatif dan Desain	6
Pengelolaan Data dan Informasi	7
Manajemen Proyek	8
Pengembangan Bisnis dan Penjualan	9
Pengelolaan Operasional Bisnis	10
Keamanan Jaringan dan Sistem	11
Sumber Daya Manusia dan Rekrutmen	12
Customer Service dan Layanan Pelanggan	13
Konsultasi dan Penyuluhan	14
Kesehatan dan Pelayanan Medis	15
Pemasaran Digital	16
Manajemen dan Kepemimpinan	17
Keuangan dan Akuntansi	18

Setelah melakukan atau meratakan *labeling,* Bentuk kolom dari atribut-atribut yang akan digunakan harus diubah menjadi bentuk yang jelas agar dapat terbaca oleh sistem *RapidMiner*. Perubahan format data tersebut dapat dilihat pada Gambar 3, yang menunjukkan kondisi tabel setelah dilakukan proses perapian.

https://doi.org/10.35145/joisie.v9i1.4954

LABEL KARIR	ın Managemen	atomii dan Fisiol	si Komputer	Akudit Sistem Informa	d Reality / Virtu	asis Data Lanjut	iologi Komputas	CRM	E-Government	rafika Kompute	Gudang Data	Infrastruktur TI
4	0	0	0	0	0	0	0	1	0	0	0	0
1	0	0	1	0	0	0	0	0	1	0	0	0
14	0	0	0	0	0	1	0	0	0	1	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	1	0	0
1	0	0	1	0	0	0	0	1	1	0	0	0
13	0	0	0	0	0	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	1	0	0	0
6	0	0	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	1	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	1	1	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0

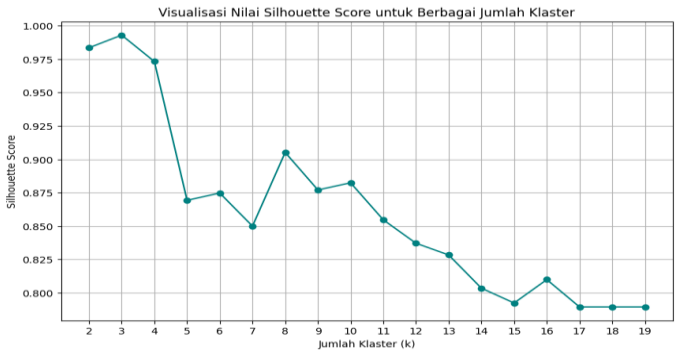
Gambar 3. Kondisi Tabel Setelah Dirapikan

3.3 *Data Transform*

Langka selanjutnya adalah melakukan beberapa proses transformasi data untuk mempersiapkan dataset sebelum proses klasterisasi di *RapidMiner*. Pada kolom label karir yang semula bersifat kategorikal, dilakukan proses *Label Encoding*, sehingga setiap jenis karir direpresentasikan dengan angka unik. Sementara itu, untuk kolom mata kuliah, digunakan pendekatan *Multi-Hot Encoding* dengan bantuan *MultiLabelBinarizer* dari pustaka *scikit-learn*. Dalam representasi ini, setiap baris menunjukkan satu label karir, dan kolom-kolom menunjukkan mata kuliah yang diambil selama kuliah. Nilai 1 menunjukkan bahwa mata kuliah tersebut diambil oleh alumni yang berkarir di bidang tersebut, sedangkan nilai 0 menunjukkan sebaliknya. Hasil dari proses ini adalah vektor biner yang menunjukkan kombinasi mata kuliah yang diambil oleh alumni untuk masing-masing label karir.

3.4 *Data Mining*

Proses *data mining* dalam penelitian ini dilakukan dengan algoritma *K-Means Clustering* untuk mengelompokkan karir alumni berdasarkan pola pengambilan mata kuliah selama masa studi. Pendekatan ini bertujuan untuk mengidentifikasi korelasi antara pola pembelajaran dan jalur karir pasca kelulusan. Penentuan jumlah klaster (K) disesuaikan dengan jumlah label karir, yaitu sebanyak 18 kategori. Evaluasi terhadap kualitas hasil klasterisasi dilakukan menggunakan metrik *Silhouette Score*, yang mengukur konsistensi internal dan pemisahan antar klaster. Nilai metrik ini berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan klaster yang kompak dan terpisah dengan baik. Visualisasi hasil evaluasi tersebut ditampilkan pada Gambar 4.



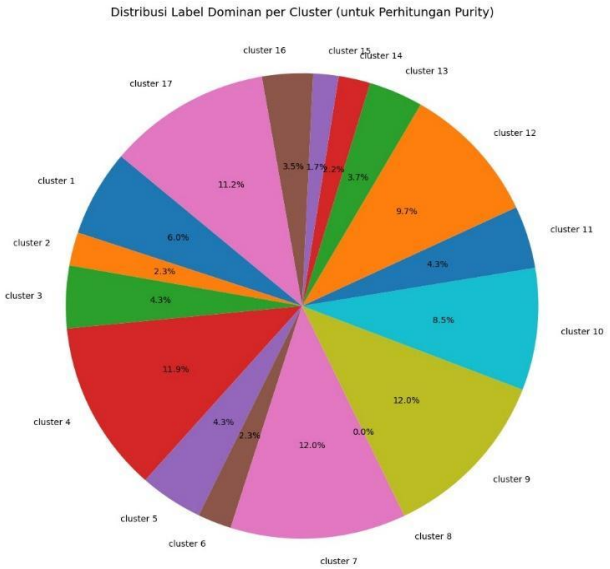
Gambar 4. Visualisasi Silhouette Score

Pada Gambar 4 disajikan visualisasi nilai *Silhouette Score* terhadap variasi jumlah klaster (*K*) yang digunakan dalam proses klasterisasi menggunakan algoritma K-Means. Grafik tersebut memperlihatkan kecenderungan nilai *Silhouette Score* mengalami penurunan seiring bertambahnya jumlah klaster. Nilai tertinggi dicapai pada $K = 3$ dengan skor mendekati 0,99, yang menunjukkan bahwa pembentukan tiga klaster memberikan pemisahan yang paling baik di antara observasi yang ada, berdasarkan kedekatan intra-klaster dan jarak antar-klaster. Meskipun terdapat fluktuasi nilai *Silhouette Score* pada rentang $K = 5$ hingga $K = 13$, nilai-nilai tersebut cenderung lebih rendah dibandingkan skor pada $K = 3$. Hal ini menunjukkan bahwa meskipun terdapat kemungkinan struktur klaster alternatif, konfigurasi tersebut tidak memberikan

https://doi.org/10.35145/joisie.v9i1.4954

pemisahan klaster yang lebih optimal. Penurunan skor yang terjadi pada $K > 3$ juga menunjukkan bahwa penambahan jumlah klaster tidak secara signifikan meningkatkan kualitas segmentasi data, bahkan cenderung memperburuk kohesi dalam masing-masing klaster.

Selanjutnya, untuk mengukur kualitas klasterisasi lebih lanjut, digunakan metrik *Cluster Purity*, yang menilai seberapa murni distribusi label karir dalam setiap klaster. Hasil dari visualisasi distribusi label karir alumni dalam klaster-klaster hasil klasterisasi *K-Means* dapat dilihat pada Gambar 5.

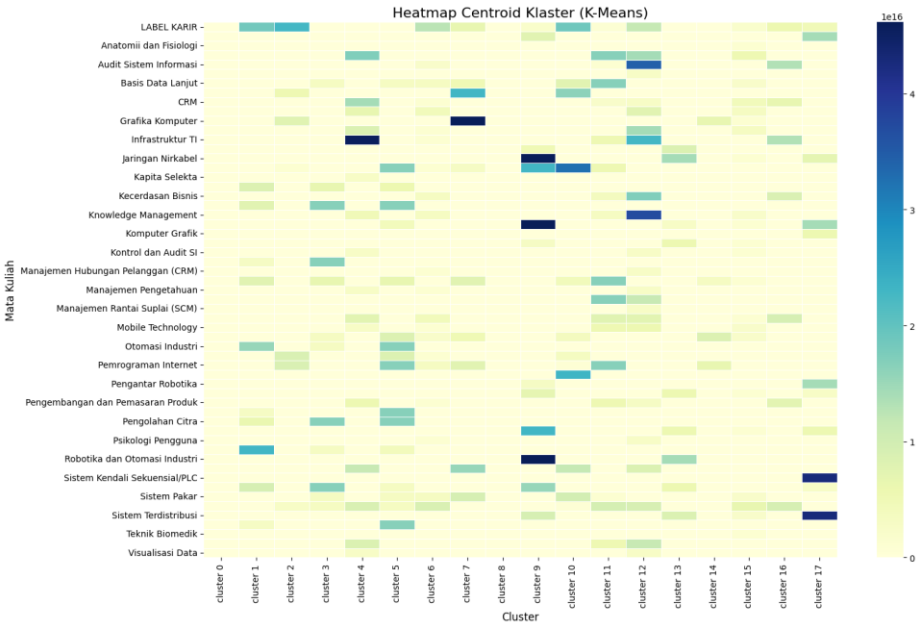


Gambar 5. Visualisasi *Cluster Purity*

Pada Gambar 5, ditampilkan distribusi label karir alumni dalam masing-masing klaster hasil proses klasterisasi menggunakan algoritma *K-Means*, dalam bentuk diagram pie. Setiap irisan pada diagram merepresentasikan proporsi satu jenis label karir dalam klaster tertentu. Terlihat bahwa beberapa klaster memiliki konsentrasi label dominan yang cukup besar, seperti *Cluster* 7, *Cluster* 9, dan *Cluster* 4 dengan proporsi masing-masing sebesar 12.0%, 12.0%, dan 11.9%. Hal ini menunjukkan bahwa label karir tertentu muncul lebih sering dibandingkan lainnya dalam klaster-klaster tersebut. Sebaliknya, terdapat pula klaster dengan proporsi dominan yang sangat kecil, seperti *Cluster* 8 dengan 0.0%, serta *Cluster* 2 dan 6 yang hanya sebesar 2.3%, yang menandakan heterogenitas label karir dalam klaster tersebut. Distribusi label karir alumni dalam masing-masing Nilai *Cluster Purity* sebesar 0.2134 (atau setara dengan 21.34%) mengindikasikan bahwa hanya sekitar seperlima dari data alumni yang berhasil dikelompokkan ke dalam klaster yang sesuai dengan label karir dominan.

Setelah memeriksa distribusi label karir alumni pada masing-masing klaster, langkah selanjutnya adalah menganalisis kontribusi setiap mata kuliah terhadap pembentukan klaster-karir tersebut. Gambar 6. *Heatmap Centroid Cluster* menyajikan visualisasi yang memberikan gambaran lebih mendalam mengenai pengaruh relatif setiap mata kuliah terhadap klaster-klaster yang terbentuk. Hal ini bertujuan untuk memahami bagaimana pola pengambilan mata kuliah berkorelasi dengan kecenderungan arah karir alumni.

https://doi.org/10.35145/joisie.v9i1.4954



Gambar 6. *Heatmap Centroid Cluster*

Pada Gambar 6 ditampilkan heatmap dari nilai centroid masing-masing klaster yang dihasilkan oleh algoritma *K-Means*. Visualisasi ini menunjukkan tingkat keterwakilan atau pengaruh relatif setiap mata kuliah terhadap masing-masing klaster karir yang terbentuk. Setiap baris mewakili satu mata kuliah, sedangkan setiap kolom menunjukkan satu klaster hasil *clustering*. Intensitas warna biru dalam setiap sel menunjukkan besarnya nilai *centroid*, yaitu seberapa kuat mata kuliah tersebut muncul dalam klaster tertentu. Semakin gelap warna pada sel, semakin tinggi keterkaitan mata kuliah tersebut dengan klaster yang bersangkutan. Pola intensitas ini mengindikasikan bahwa terdapat mata kuliah tertentu yang menonjol pada klaster tertentu, yang dapat ditafsirkan sebagai indikator hubungan antara pola pengambilan mata kuliah dengan kecenderungan arah karir alumni.

Setelah mendapatkan hasil klasterisasi, dilakukan pemetaan setiap klaster ke dalam label karir dominan. Pemetaan ini berdasarkan mata kuliah dominan (nilai tertinggi di *centroid*) pada masing-masing klaster. Pendekatan ini bertujuan untuk mengaitkan hasil klasterisasi dengan arah atau profil karir mahasiswa yang relevan secara akademik maupun praktis. Tabel 4. Hasil Pemetaan Klaster Mahasiswa terhadap Label Karir Dominan di bawah ini menunjukkan hasil pemetaan setiap klaster ke dalam label karir dominan berdasarkan mata kuliah yang paling dominan pada setiap *centroid* klaster.

Tabel 4. Hasil Pemetaan Klaster Mahasiswa terhadap Label Karir Dominan

No Cluster	Label Karir Dominan	Top 5 Mata Kuliah Dominan
0	Manajemen dan Kepemimpinan	E-Government, Sistem Pendukung Keputusan, Manajemen Sistem Informasi, Pengembangan dan Pemasaran Produk, Kecerdasan Bisnis, Sistem Multimedia.
1	Administrasi Umum	RFID, Otomasi Industri, Sistem Multimedia, Keamanan Jaringan Komputer, Manajemen Jaringan, Kecerdasan Buatan.
2	Pengembangan Perangkat Lunak	Pemrograman Internet, Pemrograman Berorientasi Objek Lanjut, Grafika Komputer, Biologi Komputasi, Sistem Pendukung Keputusan, Jaringan Syaraf Tiruan.
3	Pendidikan dan Pengajaran	Kecerdasan Buatan, Sistem Multimedia, Pengolahan Citra, Kriptografi, Manajemen Jaringan, Keamanan Jaringan Komputer.
4	Teknologi Informasi dan Sistem	Infrastruktur TI, Aplikasi Komputer Akuntansi, CRM, Sistem Informasi Geografis, Sistem Pendukung Keputusan, Teknologi Bergerak.
5	Pekerjaan Lapangan dan Survey	Kecerdasan Buatan, Jaringan Syaraf Tiruan, Otomasi Industri, Sistem Waktu Nyata, Pengenalan Pola, Pengolahan Citra.
6	Industri Kreatif dan Desain	Sistem Pendukung Keputusan, E-Government, Manajemen Sistem Informasi, Basis Data Lanjut, Pemrograman Internet, Sistem Pakar.
7	Pengelolaan Data dan Informasi	Grafika Komputer, Biologi Komputasi, Sistem Informasi Geografis, Sistem Pakar, Pemrograman Internet, Manajemen Jaringan.
8	Manajemen Proyek	Komputasi Berbasis Jaringan Komputer, Jaringan Nirkabel, Sistem Terdistribusi, Komputer Grafik, Pengolahan Citra Digital, Sistem Multimedia.
9	Pengembangan Bisnis dan Penjualan	Komputasi Berbasis Jaringan Komputer, Jaringan Nirkabel, Robotika dan Otomasi Industri, Jaringan Syaraf Tiruan, Pengolahan Citra Digital, Sistem Multimedia.

https://doi.org/10.35145/joisie.v9i1.4954

10	Pengelolaan Operasional Bisnis	Jaringan Syaraf Tiruan, Pemrograman Mikrokontroler, Biologi Komputasi, Sistem Informasi Geografis, Sistem Pakar, Basis Data Lanjut.
11	Keamanan Jaringan dan Sistem	Aplikasi Komputer Akuntansi, Basis Data Lanjut, Manajemen Rantai Suplai, Manajemen Jaringan, Pemrograman Internet, Sistem Pendukung Keputusan.
12	Sumber Daya Manusia dan Rekrutmen	Knowledge Management, Audit Sistem Informasi, Infrastruktur TI, Kecerdasan Bisnis, Gudang Data, Aplikasi Komputer Akuntansi.
13	Customer Service dan Layanan Pelanggan	Jaringan Nirkabel, Robotika dan Otomasi Industri, Sistem Terdistribusi, Jaringan Komputer Lanjut, Sistem Multimedia, Pengantar Sistem Embedded.
14	Konsultasi dan Penyuluhan	Multimedia, Grafika Komputer, Pemrograman Internet, Manajemen Jaringan, Sistem Pendukung Keputusan, Sistem Pakar.
15	Kesehatan dan Pelayanan Medis	Sistem Pendukung Keputusan, Aplikasi Komputer Akuntansi, CRM, Gudang Data, E-Government, Basis Data Lanjut.
16	Pemasaran Digital	Audit Sistem Informasi, Infrastruktur TI, Sistem Pendukung Keputusan, Manajemen Sistem Informasi, Kecerdasan Bisnis, Pengembangan dan Pemasaran Produk.
17	Keuangan dan Akuntansi	Sistem Kendali Sekuensial/PLC, Sistem Terdistribusi, Administrasi dan Managemen Sistem Jaringan, Komputasi Berbasis Jaringan Komputer, Pengantar Robotika, Jaringan Nirkabel.

Hasil ini menunjukkan bahwa kombinasi mata kuliah tertentu dapat menjadi indikator awal bagi pemetaan arah karir mahasiswa. Dengan demikian, proses klasterisasi tidak hanya menghasilkan pengelompokan data, tetapi juga memberikan insight strategis yang dapat digunakan sebagai dasar penyusunan kebijakan akademik, seperti pengembangan kurikulum peminatan, bimbingan karir, hingga rekomendasi mata kuliah yang relevan dengan tujuan karir mahasiswa.

3.5 *Evaluation*

Proses data mining dalam penelitian ini kualitas klasterisasi, tetapi juga mengkritisi tahapan awal yang mempengaruhi hasil, seperti pemilihan data, prapemrosesan, dan transformasi fitur. Data yang digunakan berupa riwayat pengambilan mata kuliah alumni, yang diasumsikan mencerminkan kecenderungan akademik yang memengaruhi arah karir. Namun, keputusan karir alumni tidak sepenuhnya dapat dijelaskan melalui data akademik, karena faktor eksternal seperti pengalaman organisasi, magang, preferensi pribadi, dan peluang kerja juga berperan, meskipun tidak tercakup dalam dataset.

Dalam proses *preprocessing*, data dikonversi menjadi bentuk vektor biner yang menyatakan apakah seorang alumni mengambil suatu mata kuliah atau tidak. Pendekatan ini menyederhanakan struktur data dan memungkinkan diterapkannya algoritma *K-Means*, namun menyisakan keterbatasan dalam hal kedalaman informasi. Informasi penting seperti nilai akhir, jumlah pengambilan, urutan semester, serta hubungan antarmata kuliah tidak dimasukkan dalam representasi data, sehingga kedalaman semantik dari setiap atribut menjadi terbatas. Hal ini berdampak pada kemampuan model dalam mengenali pola yang lebih kompleks yang mungkin berkorelasi dengan karir.

Selain itu, hanya mata kuliah peminatan yang dimasukkan ke dalam model tanpa seleksi atau reduksi fitur, mengakibatkan jumlah dimensi yang tinggi dan potensi *sparsity* data serta ketidakstabilan *centroid*. Tanpa penyaringan fitur relevan atau teknik reduksi seperti PCA, model kesulitan membedakan klaster secara optimal. Hal ini menjelaskan nilai *Cluster Purity* yang relatif rendah, yaitu 0.2134 (21.34%), yang menunjukkan hanya sebagian kecil alumni yang terkelompok sesuai dengan label karir dominan.

Meskipun nilai *Silhouette Score* yang tinggi pada K=3, hasil klasterisasi dengan 18 klaster tetap dipertahankan karena disesuaikan dengan jumlah kategori label karir. Namun, pendekatan ini membawa konsekuensi bahwa struktur klaster yang terbentuk tidak sepenuhnya mengikuti pola alami dari data, melainkan dipaksakan untuk sesuai dengan target tertentu. Hal ini berdampak pada kualitas segmentasi dan mengurangi koherensi internal antar data dalam klaster.

Berdasarkan analisis ini, pemilihan data dan tahapan prapemrosesan memiliki dampak signifikan terhadap kualitas akhir model klasterisasi. Untuk memperoleh hasil yang lebih representatif dan dapat diinterpretasikan dengan lebih baik, proses tersebut perlu disempurnakan melalui pemilihan fitur yang lebih selektif, penggunaan teknik reduksi dimensi, dan integrasi dengan atribut non-akademik. Dengan perbaikan tersebut, proses klasterisasi dapat menjadi alat yang lebih efektif dalam memetakan hubungan antara pola pembelajaran dan arah karir alumni secara lebih akurat dan aplikatif.

4. KESIMPULAN

Penelitian ini mengimplementasikan algoritma *K-Means Clustering* untuk memberikan rekomendasi mata kuliah peminatan berdasarkan kesesuaian dengan karir alumni. Berbeda dengan sistem rekomendasi konvensional yang umumnya hanya mempertimbangkan nilai akademik, pendekatan yang diusulkan dalam penelitian ini memanfaatkan data karir aktual alumni sebagai dasar pengelompokan, dengan harapan dapat

https://doi.org/10.35145/joisie.v9i1.4954

lebih mendukung proses pengambilan keputusan bagi mahasiswa dalam memilih mata kuliah yang sesuai dengan arah karir yang diinginkan.

Hasil klasterisasi menunjukkan nilai *Cluster Purity* sebesar 21,34%, yang mencerminkan bahwa hanya sebagian kecil data alumni yang berhasil terkelompokkan secara akurat sesuai dengan label karir yang aktual. Angka ini menunjukkan bahwa akurasi klastering menggunakan algoritma *K-Means* dalam penelitian ini tergolong rendah, jauh lebih rendah dibandingkan dengan akurasi rerata yang dicapai oleh metode lain dalam penelitian sebelumnya, yang berkisar sekitar 80%. Rendahnya nilai *Cluster Purity* ini mengindikasikan bahwa kualitas klaster yang dihasilkan sangat dipengaruhi oleh pemilihan fitur yang digunakan serta proses pra pemrosesan data yang belum optimal. Oleh karena itu, untuk meningkatkan performa sistem rekomendasi di masa yang akan datang, disarankan untuk melakukan seleksi fitur yang lebih cermat dan relevan, serta menerapkan teknik reduksi dimensi seperti *Principal Component Analysis* (PCA) atau *t-Distributed Stochastic Neighbor Embedding (t-SNE)*, yang diharapkan dapat meningkatkan struktur data dan memudahkan pengelompokan berdasarkan label karir yang lebih representatif.

5. DAFTAR PUSTAKA

Afifuddin, R. N., & Nurjanah, D. (2019). Sistem Rekomendasi Pemilihan Mata kuliah Peminatan Menggunakan Algoritma K-means dan Apriori (studi kasus: Jurusan S1 Teknik Informatika Fakultas Informatika). *EProceedings of Engineering*, *6*(144), 2359. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/8682

Ahmadinejad, N., Chung Corresp, Y., Liu Corresp, L., Authors, C., Chung, Y., & Liu, L. (2023). J-score: A robust measure of clustering accuracy. *PeerJ Computer Science*, *9*, 1545. https://doi.org/https://doi.org/10.7717/peerj-cs.1545

Alejandrino, J. C. (2021). Application of Data Mining in Knowledge Management: A Review. *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*, *10*(4), 2690–2696. https://doi.org/https://doi.org/10.30534/ijatcse/2021/061042021

Ananto, D. T., Duta Mahardewantoro, D., Mustafa, F., Ardianto, M. G., Rafi, M. M., Zein, R. A., Saputra, O. E., Mujiastuti, R., Rosanti, N., & Adharani, Y. (2023). Edukasi dan Pelatihan Pengenalan Machine Learning dan Computer Vision Untuk Mengeksplorasi Potensi Visual. *Prosiding Seminar Nasional Pengabdian Masyarakat LPPM UMJ*, 049. https://jurnal.umj.ac.id/index.php/semnaskat/article/view/19491

Atalya, M., Leza, A., Utami, W., Anugrah, P., & Dewi, C. (2024). Prediksi Prestasi Siswa Smas Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery In Database Dan Algoritma Regresi Linier Berganda. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).

Bahri, M., Bifet, A., Maniu, S., & Gomes, H. M. (2021). Survey on Feature Transformation Techniques for Data Streams. *International Joint Conferences on Artifical Intelligence (IJCAI)*, *20*, 4796–4802. https://doi.org/https://doi.org/10.24963/ijcai.2020/668

Budiman, A., Dwi Lestari, Y., & Eka, M. (2024). Penerapan Metode MAUT dalam Pemilihan Peminatan pada Program Studi Teknik Informatika. *Jurnal UNITEK*, *17*(2), 2024. https://doi.org/https://doi.org/10.52072/unitek.v17i2.921

Darmawan, B., Pamungkas, D. P., & Mahdiyah, U. (2024). Rekomendasi Pendukung Keputusan Pemilihan Matakuliah Dengan Kombinasi Dari Metode MOORA dan TOPSIS. *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 520*, *8*, 2549–7952. https://proceeding.unpkediri.ac.id/index.php/inotek/

Darsono, V., & Andrianti, A. (2022). Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Penerapan Data Mining Algoritma K-Means Untuk Rekomendasi Pemilihan Bidang Studi Perguruan Tinggi Pada Siswa SMKN 1 Kota Jambi. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, *2*, 161. https://doi.org/https://doi.org/10.33998/jakakom.2022.2.2.80

Deti Karmanita, & Billy Hendrik. (2023). Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Peminatan Mata Kuliah. *Jurnal Ilmiah Dan Karya Mahasiswa*, *1*(6), 01–10. https://doi.org/10.54066/jikma.v1i6.1028

https://doi.org/10.35145/joisie.v9i1.4954

Fernando, E., Mudjiraharjo, P., & Aswin, M. (2022). Implementasi Pendekatan Collaborative Filtering Dan K-Means Clustering Pada Sistem Rekomendasi Mata Kuliah Implementation Of Collaborative Filtering And K-Means Clustering Approaches In The Course Recommendation System. *Jurnal Informatika Dan Komputer) Akreditasi KEMENRISTEKDIKTI*, *5*(2). https://doi.org/10.33387/jiko

Haviluddin, H., Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir. *Informatika Mulawarman : Jurnal Ilmiah Ilmu Komputer*, *16*(1), 13. https://doi.org/http://dx.doi.org/10.30872/jim.v16i1.5182

Khoerullah, A. R., Nurjanah, D., & Romadhony, A. (2019). Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Association Rule dan Ant Colony Optimization (Studi Kasus Mata Kuliah di Jurusan Teknik Informatika Universitas Telkom). *E-Proceeding of Engineering*, *6*, 9597–9617.

Lestari, Y. D., & Perdana, A. (2020). Penerapan Metode Waspas Dalam Menentukan Pemilihan Peminatan Pada Program Studi Teknik Informatika Application Of The Waspas Method In Determining The Selection Of Specialty In Informatics Engineering Programs. In *Jurnal Ilmu Komputer dan Sistem Komputer Terapan (JIKSTRA)* (Vol. 01, Issue 02).

Lusiah, R. N., & Purwanto, I. (2023). Sistem Informasi Geografis Fasilitas Kesehatan di Tuntungan Berbasis Android. *Bulletin of Computer Science Research*, *3*(3), 257–262. https://doi.org/10.47065/bulletincsr.v3i3.244

Mohammed, M., & Alsunosi, R. (2022). Effect of Selecting Validation Dataset on Building Random Forest and Decision Tree Models. *AlQalam Journal of Medical and Applied Sciences*, *5*(2), 470–478. https://doi.org/10.5281/zenodo.7113928

Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, *9*(1). https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428

Nirmada, P. (2024). *Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Konsentrasi Mahasiswa Program Studi Manajemen Universitas Muhammadiyah Makassar* [Universitas Muhammadiyah Makassar]. https://digilibadmin.unismuh.ac.id/upload/40843-Full_Text.pdf

Nurchalia, L. P., Ghifari, Y., Limbong, J. A., & Setiawati, L. (2023). Professional analysis of Educational Technology students with appropriate specializations. *Inovasi Kurikulum*, *20*(2), 193–204. https://doi.org/10.17509/jik.v20i2.53904

Putu, I., Iswara, P., Farhan, F., Kumara, W., Supianto, A. A., Informasi, S., Komputer, I., & Brawijaya, U. (2019). Rekomendasi Pengambilan Mata Kuliah Pilihan Untuk Mahasiswa Sistem Informasi Menggunakan Algoritme Decision Tree. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, *6*(3), 341–384. https://doi.org/https://doi.org/10.25126/jtiik.201963892

Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ Computer Science*, *10*, e2245. https://doi.org/10.7717/peerj-cs.2245

Suresh Sikhakolli, S., & Kiran Sikhakolli DrDY, A. (2023). Effective Purity Method for Measuring the Clustering Accuracy and its Illustration. *International Journal of Computer Applications*, *185*(9), 975–8887. https://doi.org/https://doi.org/10.5120/ijca2023922752

Syahrul, A., & Solichin, A. (2022). Rekomendasi Pemilihan Mata Kuliah Dalam Pengisian Rencana Studi Mahasiswa Dengan Penerapan Algoritma Apriori. *Teknologi Informasi Dan Komputer*, *6*(1), 79–88. https://doi.org/10.31961/eltikom.v6i1.549

Utomo, Y. B., Kurniasari, I., & Yanuartanti, I. (2023). Penerapan Knowledge Discovery In Database Untuk Analisa Tingkat Kecelakaan Lalu Lintas. *Jurnal Teknik Informatika Kaputama (JTIK)*, *7*(1), 171–180. https://doi.org/https://doi.org/10.59697/jtik.v7i1.61

Wahyudinnur, R. A., Arriyanti, E., & Wahyuni. (2024). Sistem Rekomendasi Peminatan Pada Prodi Teknik Informatika STMIK Widya Cipta Dharma Menggunakan Algoritma C5.0. *SEBASTIK Journal*, *28*, 1410–3737. https://doi.org/https://doi.org/10.46984/sebatik.v28i2.0000

Zhu, A., Hua, Z., Shi, Y., Tang, Y., & Miao, L. (2021). An improved k-means algorithm based on evidence distance. *Entropy*, *23*(11), 1550. https://doi.org/https://doi.org/10.3390/e23111550

https://doi.org/10.35145/joisie.v9i1.4954