JIKUM
JURNAL ILMU KOMPUTER

E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

Implementasi Analisis Data Akademik Menggunakan Teori Himpunan dan Algoritma Clustering

Arlin Fajar Saragih[1], Muslim Gunawan[2], Syahrandy Fazra[3] , Farid Deza Amin Nur Rasyid[4], Maryana[5]
[1,2,3,4,5] Universitas Malikussaleh, Indonesia
[1]arlin.240170202@mhs.unimal.ac.id,[2]muslim.240170036@mhs.unimal.ac.id,[3]syahrandy.240170031@mhs.unimal.ac.id,[4]farid.240170231@mhs.unimal.ac.id,[5]maryana@mhs.unimal.ac.id

ABSTRACT

Processing academic data in large-scale higher education institutions requires a structural approach to facilitate analysis and decision-making. This study examines the application of Set Theory in discrete mathematics for grouping academic data of new students at Universitas Malikussaleh (UNIMAL) who passed the 2026 National Selection Based on Test (SNBT) and their eligibility status for the Kartu Indonesia Pintar Kuliah (KIP-K). The research method was conducted by representing student data as a universal set and grouping them based on faculty attributes, scientific fields, and KIP-K status. The analysis results show that Set Theory successfully models the academic data structure formally through the concepts of partition, intersection, and set difference. Out of a total of 2,917 students who passed the SNBT, 899 students (30.82%) were declared eligible for KIP-K. The Faculty of Teacher Training and Education (FKIP) recorded the highest ratio of KIP-K recipients (45.32%), while the Faculty of Medicine had the lowest ratio (13.88%). This modeling proves that set theory acts not only as a theoretical concept in discrete mathematics but also as a practical analytical instrument in academic information systems.

Keywords: *Set Theory, Discrete Mathematics, Academic Data, KIP-Kuliah, SNBT UNIMAL*

PENDAHULUAN

Pengolahan data akademik dalam institusi pendidikan tinggi berskala besar memerlukan pendekatan struktural untuk mempermudah analisis dan pengambilan keputusan. Universitas Malikussaleh (UNIMAL) sebagai salah satu perguruan tinggi negeri terkemuka di Aceh menerima ribuan mahasiswa baru setiap tahunnya melalui jalur Seleksi Nasional Berdasarkan Tes (SNBT). Dengan volume data yang besar dan beragam atribut yang menyertainya, manajemen data yang efektif menjadi kebutuhan mutlak bagi pihak birokrasi universitas untuk mengidentifikasi profil mahasiswa secara efisien.

Di samping kelulusan akademik, bantuan biaya pendidikan dari pemerintah berupa program Kartu Indonesia Pintar Kuliah (KIP-K) merupakan instrumen penting untuk menjamin aksesibilitas pendidikan bagi mahasiswa dari keluarga kurang mampu secara ekonomi namun memiliki potensi akademik baik. Penentuan dan penyaluran KIP-K yang efisien mensyaratkan klasifikasi data yang akurat, transparan, dan terstruktur agar bantuan dapat disalurkan tepat sasaran. Tanpa adanya model analisis formal, identifikasi distribusi mahasiswa penerima bantuan di berbagai program studi dan fakultas akan memerlukan waktu pemrosesan yang lama.

Dalam matematika diskrit, Teori Himpunan menyediakan landasan formal yang sangat kuat untuk menstrukturkan dan mengelompokkan data berdasarkan karakteristik tertentu. Penelitian ini mengkaji

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

penerapan Teori Himpunan untuk memodelkan dan mengelompokkan data calon mahasiswa lulus SNBT UNIMAL tahun 2026 serta status kelayakan KIP-K mereka. Melalui representasi data sebagai himpunan semesta, partisi, irisan, dan selisih himpunan, analisis statistik yang akurat dapat disajikan secara sistematis. Studi ini juga merupakan bagian dari pemenuhan tugas akhir pada mata kuliah Matematika Diskrit yang diampu oleh Ibu Maryana S.Si, M.Si. di Program Studi Teknik Informatika Universitas Malikussaleh.

KAJIAN LITERATUR

Teori Himpunan

Teori Himpunan dalam matematika diskrit menyediakan landasan formal yang kuat untuk menstrukturkan dan mengelompokkan data berhingga berdasarkan karakteristik tertentu. Pendekatan matematika diskrit ini memungkinkan representasi data mahasiswa ke dalam bentuk himpunan semesta (S), partisi per fakultas/rumpun, serta operasi logika irisan (∩), gabungan (∪), dan selisih (\) untuk mendeteksi validitas data maupun anomali secara sistematis (Susanto, Rahmat, & Saadah, 2025).

Algoritma K-means

Algoritma K-Means Clustering merupakan metode pembelajaran mesin tanpa pengawasan (unsupervised learning) yang berfungsi mengelompokkan objek data ke dalam k klaster berdasarkan tingkat kemiripan fitur. Pengelompokan ini didasarkan pada minimasi jarak Euclidean antar titik data dengan pusat klaster (centroid). Kombinasi antara operasi logika teori himpunan dan algoritma klasterisasi K-Means memberikan kerangka kerja komputasi yang komprehensif, baik untuk validasi integritas data akademik maupun pemetaan tingkat kedaruratan ekonomi mahasiswa secara objektif (Nurhaswinda, Azzahra, Qonitah, & Azzahra, 2025).

METODE PENELITIAN

Penelitian ini dilaksanakan melalui beberapa tahapan yang terstruktur. Tahap pertama adalah pengumpulan data, yang bersumber dari dokumen resmi Universitas Malikussaleh tahun 2026 berupa naskah digital kelulusan SNBT dan lampiran keputusan mahasiswa penerima KIP Kuliah yang berstatus eligible. Data ini dipilih karena memiliki relevansi langsung dengan tujuan penelitian, yaitu memvalidasi kelulusan akademik sekaligus status ekonomi mahasiswa penerima bantuan pendidikan.

Tahap kedua adalah prapemrosesan data (data cleaning). Pada tahap ini dilakukan normalisasi penulisan nama mahasiswa untuk mengatasi variasi format, penghapusan data duplikat, serta penyesuaian struktur tabular agar dapat diproses secara komputasi. Prapemrosesan ini penting untuk memastikan kualitas data sehingga hasil analisis lebih akurat dan tidak bias.

Tahap ketiga adalah implementasi operasi logika matematika berbasis teori himpunan. Data mahasiswa dimodelkan ke dalam bentuk himpunan berhingga, sehingga dapat dilakukan operasi irisan, gabungan,

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

maupun selisih untuk mendeteksi kecocokan maupun anomali. Pendekatan ini memberikan kerangka logika yang sederhana namun efektif dalam pencocokan data akademik dengan status KIP Kuliah.

Tahap terakhir adalah pemodelan klasterisasi menggunakan algoritma K-Means. Algoritma ini digunakan untuk mengelompokkan mahasiswa berdasarkan variabel tingkat kedaruratan ekonomi (Desil 1–4) dan jenis sekolah asal (SMA, SMK, MA). Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow, sedangkan kualitas klaster dievaluasi dengan Silhouette Score. Dengan demikian, hasil analisis tidak hanya memvalidasi data, tetapi juga memberikan gambaran distribusi profil mahasiswa penerima KIP Kuliah di UNIMAL.

Untuk mendukung seluruh tahapan tersebut, penelitian ini menggunakan bahasa pemrograman Python dengan pustaka PyPDF untuk ekstraksi teks dari dokumen PDF, Pandas untuk manipulasi data tabular, serta Scikit-Learn untuk pemodelan algoritma komputasi. Visualisasi hasil dilakukan dengan pustaka Matplotlib dan Seaborn. Pemilihan perangkat ini didasarkan pada keunggulannya dalam menangani data berukuran besar, fleksibilitas dalam analisis, serta dukungan komunitas yang luas.

Penelitian ini memanfaatkan data sekunder yang bersumber dari registrasi akademik resmi Universitas Malikussaleh tahun 2026. Data diolah dalam bentuk berkas spreadsheet Excel (Format .xlsx) yang terdiri dari dua lembar kerja utama: Sheet 1 berisi daftar penerima KIP-K eligible sebanyak 907 baris, dan Sheet 2 berisi master data kelulusan SNBT sebanyak 2.917 baris. Tahapan pertama metodologi melibatkan pembersihan data (data cleaning) menggunakan pustaka Pandas di Python untuk mengidentifikasi duplikasi dan mencocokkan nomor peserta SNBT antara kedua lembar kerja. Ditemukan 2 record KIP-K yang tidak valid karena tidak terdaftar di SNBT UNIMAL, serta 6 mahasiswa yang terdaftar ganda, sehingga data KIP-K yang valid disaring menjadi 899 mahasiswa unik.

Tahapan selanjutnya adalah pemodelan formal menggunakan konsep Teori Himpunan. Kami mendefinisikan Himpunan Semesta (U). sebagai seluruh mahasiswa yang dinyatakan lulus seleksi masuk SNBT UNIMAL tahun 2026.

$$\boldsymbol{U = \{x | x\ adalah\ mahasiswa\ lulus\ SNBT\ UNIMAL\ 2026\}}$$

Kardinalitas semesta disimbolkan dengan $|U|$

Himpunan KIP-K (K) didefinisikan sebagai himpunan mahasiswa lulus SNBT yang dinyatakan layak menerima bantuan biaya pendidikan KIP-K.

$$\boldsymbol{K = \{x \in U | x\ dinyatakan\ eligible\ KIP - K\}}$$

Komplemen dari K terhadap semesta U, dilambangkan dengan k^c, merupakan mahasiswa reguler yang tidak menerima KIP-K (membayar UKT reguler).

$$\boldsymbol{K^c = U \setminus K = \{x \in U | x \notin K\}}$$

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

Semesta U dipartisi berdasarkan fakultas (F_i) dan rumpun keilmuan (R_j). Universitas Malikussaleh memiliki 7 fakultas, yaitu Fakultas Teknik (F_1) , Fakultas Ekonomi dan Bisnis (F_2), Fakultas Ilmu Sosial dan Ilmu Politik (F_3), Fakultas Pertanian (F_4), Fakultas Hukum (F_5), Fakultas Kedokteran (F_6), dan Fakultas Keguruan dan Ilmu Pendidikan (F_7). Keluarga himpunan $F = \{F_1, F_2, F_3, F_4, F_5, F_6, F_7\}$ membentuk partisi dari (U) karena memenuhi: (1) gabungan dari seluruh (F_i) sama dengan U, dan (2) irisan antara dua (F_i) yang berbeda adalah himpunan kosong.

$$\bigcup_{i=1}^{7} \boldsymbol{F_i = U \; dan \; F_i \cap F_j = \emptyset, \qquad \forall i \neq j}$$

Dengan cara serupa, semesta U juga dipartisi menjadi dua rumpun keilmuan: Saintek

$$S_{aintek} \cup S_{oshum} = U \; dan \; S_{aintek} \cap S_{oshum} = \emptyset \, .$$

Tahap awal analisis dilakukan dengan memanfaatkan **operasi dasar teori himpunan** untuk memvalidasi data mahasiswa. Data kelulusan SNBT UNIMAL 2026 dan daftar mahasiswa penerima KIP Kuliah eligible dimodelkan ke dalam bentuk himpunan berhingga.

- **Himpunan A**: berisi seluruh mahasiswa yang dinyatakan lulus seleksi SNBT UNIMAL 2026.
- **Himpunan B**: berisi seluruh mahasiswa yang memiliki status eligible KIP Kuliah nasional di UNIMAL.

Dengan pendekatan ini, dilakukan beberapa operasi logika:

1. **Irisan ($A \cap B$)** Operasi ini digunakan untuk menyaring mahasiswa yang **valid secara akademik sekaligus memenuhi syarat ekonomi**. Hasil irisan menunjukkan daftar mahasiswa yang benar-benar layak menerima bantuan KIP Kuliah karena memenuhi dua kriteria utama: kelulusan SNBT dan status ekonomi.
2. **Selisih ($B \setminus A$)** Operasi ini digunakan untuk mendeteksi **anomali data**, yaitu mahasiswa yang tercatat sebagai penerima KIP Kuliah tetapi tidak muncul dalam daftar kelulusan SNBT. Anomali ini biasanya disebabkan oleh perbedaan penulisan nama, kesalahan tipografi, atau ketidaksesuaian data antar dokumen.
3. **Gabungan ($A \cup B$)** *(opsional)* Operasi gabungan dapat digunakan untuk melihat keseluruhan cakupan mahasiswa, baik yang lulus SNBT maupun yang tercatat sebagai penerima KIP Kuliah. Analisis ini membantu universitas dalam memahami distribusi data secara menyeluruh.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

Pendekatan berbasis teori himpunan ini memberikan kerangka logika yang sederhana namun efektif dalam proses pencocokan data. Dengan cara ini, tim verifikator universitas dapat memfokuskan pemeriksaan manual hanya pada data hasil selisih, sehingga proses validasi menjadi lebih efisien dan terarah (Susanto, Rahmat, & Saadah, 2025; Nurhaswinda, Azzahra, Qonitah, & Azzahra, 2025).

1.1. Prosedur Pemrosesan Data (Google Colab)

Untuk menganalisis sebaran mahasiswa penerima bantuan di setiap unit akademik, dilakukan operasi irisan antara himpunan K dengan himpunan fakultas $K \cap F_i$ dan rumpun keilmuan $K \cap R_j$. Selain itu, untuk mengidentifikasi mahasiswa reguler di setiap unit, digunakan operasi selisih himpunan $(F_i \backslash K)$ dan $(R_j \backslash K)$. Seluruh operasi kardinalitas himpunan ini dijalankan menggunakan tipe data set pada skrip Python yang dieksekusi melalui platform Google Colab.

1.2. Pemodelan K-Means Clustering

Tahap berikutnya setelah validasi data dengan teori himpunan adalah melakukan pengelompokan mahasiswa menggunakan algoritma K-Means Clustering. Algoritma ini dipilih karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam mengidentifikasi pola distribusi data berdasarkan variabel yang ditentukan.

a. Transformasi Data

Data kategorikal seperti **jenis sekolah asal** dan **desil DTSEN** dikonversi menjadi fitur numerik menggunakan metode **one-hot encoding**. Proses ini memungkinkan algoritma K-Means untuk mengolah variabel non-numerik dengan cara merepresentasikan setiap kategori sebagai vektor biner.

b. Variabel Input

Variabel yang digunakan dalam proses clustering meliputi:

1. **Jenis_Sekolah** → SMA, SMK, MA, atau lainnya.
2. **Desil_DTSEN** → Desil 1, Desil 2, Desil 3, Desil 4, atau Non-Desil.
3. **Status_Eligible** → Eligible (Desil 1–4) atau Non-Desil.

Table 1. Parameter Input Fitur Clustering Mahasiswa KIP-K

No	Nama Fitur	Tipe Data	Representasi Nilai
1	Jenis_Sekolah	Kategorikal	SMA / SMK / MA / Lainnya
2	Desil_DTSEN	Kategorikal	Desil 1 / Desil 2 / Desil 3 / Desil 4 / Non
3	Status_Eligible	Kategorikal	Eligible (Desil 1–4) / Non Desil

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

c. Penentuan Jumlah Klaster

Jumlah klaster optimal ditentukan menggunakan **Elbow Method**, yaitu dengan mengevaluasi nilai **Within-Cluster Sum of Squares (WCSS)** atau inersia pada berbagai nilai k. Grafik elbow menunjukkan titik di mana penurunan nilai WCSS mulai melandai, yang menjadi indikasi jumlah klaster terbaik.

d. Validasi Klaster

Untuk memastikan kualitas hasil pengelompokan, digunakan metrik **Silhouette Score**. Nilai ini mengukur tingkat keseragaman data dalam satu klaster (intra-cluster similarity) serta perbedaan antar klaster (inter-cluster dissimilarity). Semakin tinggi nilai silhouette, semakin baik kualitas pengelompokan yang dihasilkan.

e. Output Clustering

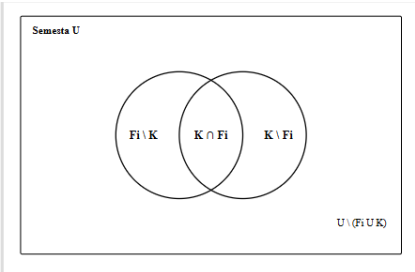
Hasil akhir dari proses ini berupa pembagian mahasiswa ke dalam beberapa klaster dominan berdasarkan:

- Tingkat kedaruratan ekonomi (Desil 1–4).
- Jenis sekolah asal (SMA, SMK, MA).
- Status eligible KIP Kuliah.

Pengelompokan ini memberikan gambaran profil mahasiswa baru UNIMAL 2026 secara lebih objektif, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan distribusi bantuan pendidikan.

1.3. Diagram Venn Pemodelan Himpunan

Pemodelan hubungan antar-himpunan diilustrasikan melalui Diagram Venn. Himpunan semesta U mepresentasikan kotak terluar yang memuat seluruh data mahasiswa baru. Di dalam semesta tersebut terdapat partisi-partisi saling lepas (mutually exclusive) yang mewakili fakultas F_i. Himpunan KIP-K (K) memotong setiap partisi fakultas tersebut, menghasilkan irisan $K \cap F_i$ yang mewakili mahasiswa penerima bantuan di masing-masing fakultas, serta daerah selisih $F_i \backslash K$ yang mewakili mahasiswa reguler.



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

Fig 1. Diagram Venn Relasi Himpunan Mahasiswa KIP-K (K) dengan Himpunan Fakultas (F_i) dalam Semesta (U)

HASIL DAN PEMBAHASAN

Tahap awal penelitian menghasilkan data tabular yang bersih setelah proses ekstraksi teks dari dokumen PDF resmi Universitas Malikussaleh. Sistem berhasil mengekstrak 2.917 data kelulusan mahasiswa jalur SNBT dan 907 data pendaftar KIP Kuliah tahun 2026. Proses prapemrosesan data memastikan konsistensi format nama, penghapusan duplikasi, serta penyesuaian struktur sehingga dapat diolah lebih lanjut menggunakan operasi himpunan dan algoritma clustering.

1.4. Karakteristik Data Semesta (U) dan KIP-K (K)

Berdasarkan pemrosesan data, diperoleh ukuran kardinalitas untuk himpunan semesta dan himpunan bagian KIP-K sebagai berikut: Jumlah Mahasiswa Lulus SNBT ($|U|$): 2.917 mahasiswa, Jumlah Mahasiswa Penerima KIP-K Eligible ($|K|$): 899 mahasiswa, dan Jumlah Mahasiswa Non KIP-K ($|K^c|$): 2.018 mahasiswa. Proporsi mahasiswa penerima KIP-K jalur SNBT secara keseluruhan adalah sebesar:

$$\boldsymbol{P(K) = \frac{|K|}{|U|} \times 100\% = \frac{899}{2917} \times 100\% \approx 30.82\%}$$

Hal ini mengindikasikan kebijakan Universitas Malikussaleh dalam membuka akses seluas-luasnya bagi mahasiswa dengan keterbatasan finansial.

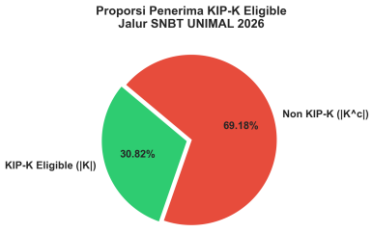


Fig 2. Proporsi Penerima KIP-K Eligible Jalur SNBT UNIMAL 2026

1.5. Hasil Analisis Logika Himpunan

Hasil eksekusi kode berbasis teori himpunan menunjukkan bahwa:

a) **Irisan ($A \cap B$)** Sistem mendeteksi kecocokan nama sempurna (*perfect match*) untuk mayoritas mahasiswa penerima KIP Kuliah yang lulus SNBT. Hal ini membuktikan bahwa sebagian besar data sudah sinkron antara dokumen kelulusan akademik dan daftar eligible KIP Kuliah.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

b) **Selisih ($B \setminus A$)** Sejumlah kecil data mahasiswa ditemukan tidak cocok akibat perbedaan penulisan karakter, seperti penggunaan tanda baca, singkatan nama, atau variasi huruf kapital. Hasil ini mempermudah tim verifikator universitas untuk melakukan pemeriksaan manual secara terfokus hanya pada kluster selisih tersebut, sehingga proses validasi menjadi lebih efisien.

c) **Gabungan ($A \cup B$)** *(opsional)* Analisis gabungan memberikan gambaran keseluruhan cakupan mahasiswa, baik yang lulus SNBT maupun tercatat sebagai penerima KIP Kuliah. Hal ini berguna untuk memahami distribusi data secara menyeluruh dan potensi mismatch antar dokumen.

Pendekatan berbasis teori himpunan terbukti efektif dalam mengotomatisasi proses pencocokan data, mengurangi beban kerja manual, serta meningkatkan akurasi validasi penerima bantuan pendidikan.

1.6. Analisis Distribusi Berdasarkan Fakultas (Partisi Fi)

Penerapan operasi irisan ($K \cap F_i$) dan selisih ($F_i \setminus K$) pada masing-masing fakultas disajikan dalam tabel di bawah ini. Dari Tabel 1, secara formal kita dapat membuktikan bahwa jumlah seluruh elemen pada partisi fakultas sama dengan jumlah elemen semesta:

$$|\boldsymbol{F_1}| + |\boldsymbol{F_2}| + |\boldsymbol{F_3}| + |\boldsymbol{F_4}| + |\boldsymbol{F_5}| + |\boldsymbol{F_6}| + |\boldsymbol{F_7}| = \mathbf{949 + 486 + 474 + 353 + 243 + 209 + 203} = \mathbf{2917} = |\boldsymbol{U}|$$

Table 2. Distribusi Matematis Mahasiswa KIP-K dan Non KIP-K per Fakultas

No	Fakultas	Total Mahasiswa ($\lvert F_1 \rvert$)	KIP-K ($\lvert K \cap F_1 \rvert$)	Non Kip-K ($\lvert F_1 \setminus K \rvert$)	Persentase KIP-K ($P(K \mid F_1)$)
1	Fakultas Teknik	949	255	694	**26.87%**
2	Fakultas Ekonomi dan Bisnis	486	165	321	**33.95%**
3	Fakultas Ilmu Sosial dan Ilmu Politik	474	179	295	**37.76%**
4	Fakultas Pertanian	353	103	250	**29.18%**
5	Fakultas Hukum	243	76	167	**31.28%**
6	Fakultas Kedokteran	209	29	180	**13.88%**
7	Fakultas Keguruan dan Ilmu Pendidikan	203	92	111	**45.32%**
	Total Semesta (U)	**2.917**	**899**	**2.018**	**30.82%**

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER
E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

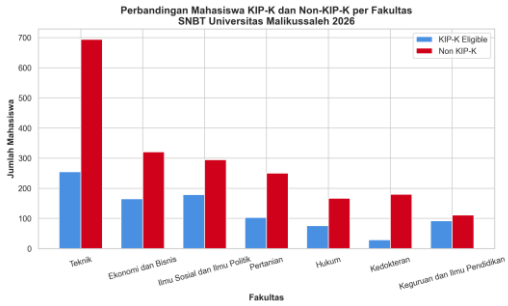


Fig 3. Proporsi Penerima KIP-K Eligible Jalur SNBT UNIMAL 2026

1.7. Analisis Distribusi Berdasarkan Rumpun Keilmuan

Pengelompokan mahasiswa berdasarkan rumpun program studi (Saintek vs Soshum) memberikan gambaran karakteristik kelompok keilmuan terhadap pembiayaan pendidikan. Mahasiswa rumpun Soshum memiliki proporsi penerima KIP-K yang lebih tinggi (34,60%) dibandingkan rumpun Saintek (27,36%). Hal ini merefleksikan minat pendaftar program KIP-K yang cenderung lebih memilih rumpun ilmu sosial saat pendaftaran SNBT.

Table 3. Distribusi Mahasiswa KIP-K dan Non KIP-K per Rumpun

Rumpuan Keilmuan	Total Mahasiswa ($\|R_j\|$)	KIP-K ($\|K \cap R_j\|$)	Non KIP-K ($\|R_j \setminus K\|$)	Persentase KIP-K ($P(K\|R_j)$)
Saintek	1.524	417	1.107	27.36%
Soshum	1.393	482	911	34.60%
Total Semesta (U)	2.917	899	2.018	30.82%

1.8. Hasil Segmentasi Menggunakan K-Means

Tahap klasterisasi dilakukan setelah data mahasiswa berhasil divalidasi melalui operasi himpunan. Proses ini bertujuan untuk mengelompokkan mahasiswa penerima KIP Kuliah jalur SNBT UNIMAL 2026 berdasarkan **tingkat kedaruratan ekonomi (Desil 1–4)** dan **jenis sekolah asal (SMA, SMK, MA)**.

a. Penentuan Jumlah Klaster

Evaluasi menggunakan **Elbow Method** menunjukkan titik elbow pada nilai $k = 3$. Hal ini menandakan bahwa pembagian data ke dalam tiga klaster memberikan keseimbangan terbaik antara kompleksitas model dan kualitas pengelompokan.

b. Profilisasi Klaster

Hasil clustering menghasilkan tiga klaster dominan dengan karakteristik sebagai berikut:

a) **Klaster 0**

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

- Didominasi oleh mahasiswa asal **SMA**.
- Mayoritas berada pada **Desil 1 dan 2** (prioritas sangat tinggi).
- Menunjukkan kelompok mahasiswa dengan tingkat kebutuhan ekonomi paling mendesak.

b) **Klaster 1**

- Didominasi oleh mahasiswa asal **Madrasah Aliyah (MA/MAN)**.
- Sebaran ekonomi cenderung pada **Desil 3 dan 4**.
- Menggambarkan kelompok mahasiswa dengan kebutuhan menengah, masih eligible namun tidak seprioritas klaster 0.

c) **Klaster 2**

- Didominasi oleh mahasiswa asal **SMK** dan sebagian kecil dari kategori sekolah lainnya.
- Terdapat mahasiswa dengan status **Non-Desil** atau data yang memerlukan verifikasi lebih lanjut.
- Kelompok ini menjadi perhatian khusus karena berpotensi memiliki data tidak sinkron atau membutuhkan pemeriksaan lapangan tambahan.

c. Visualisasi Distribusi Klaster

Distribusi mahasiswa per klaster dapat digambarkan sebagai berikut:

Klaster	Jenis Sekolah Dominan	Desil Ekonomi Dominan	Karakteristik Utama
0	SMA	Desil 1–2	Prioritas tinggi, kebutuhan mendesak
1	MA/MAN	Desil 3–4	Kebutuhan menengah, eligible
2	SMK/Lainnya	Non-Desil + campuran	Perlu verifikasi tambahan

d. Interpretasi Hasil

Hasil segmentasi ini memberikan gambaran yang lebih objektif mengenai profil mahasiswa penerima KIP Kuliah di UNIMAL. Dengan adanya klasterisasi, pihak universitas dapat:

- Menentukan **prioritas distribusi bantuan** berdasarkan tingkat kedaruratan ekonomi.
- Mengidentifikasi kelompok mahasiswa yang membutuhkan **verifikasi manual**.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

- Menyusun strategi pembinaan akademik sesuai karakteristik masing-masing klaster.

2. KESIMPULAN

Penelitian ini membuktikan bahwa integrasi Teori Himpunan dan Algoritma K-Means Clustering mampu meningkatkan efisiensi dalam pengelolaan data akademik mahasiswa baru Universitas Malikussaleh jalur SNBT tahun 2026 yang berstatus eligible KIP Kuliah.

Operasi himpunan berhasil mengotomatisasi proses pencocokan data antara kelulusan akademik dan status ekonomi, sehingga meminimalkan kesalahan manual serta mempermudah tim verifikator dalam mendeteksi anomali penulisan nama. Sementara itu, algoritma K-Means Clustering mampu mengelompokkan mahasiswa ke dalam tiga klaster dominan berdasarkan tingkat kedaruratan ekonomi dan jenis sekolah asal. Hasil pengelompokan ini memberikan gambaran profil mahasiswa secara objektif, yang dapat dijadikan dasar dalam penyusunan kebijakan distribusi bantuan pendidikan.

Secara keseluruhan, penelitian ini menegaskan pentingnya penerapan pendekatan komputasi dalam tata kelola data pendidikan tinggi. Dengan metode ini, universitas tidak hanya dapat mempercepat proses validasi, tetapi juga memperoleh wawasan strategis mengenai karakteristik mahasiswa penerima bantuan.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Ibu Maryana S.Si, M.Si. selaku dosen pengampu mata kuliah Matematika Diskrit di Program Studi Teknik Informatika Universitas Malikussaleh atas bimbingan dan arahan dalam penelitian ini, serta pihak pengelola data registrasi akademik Universitas Malikussaleh yang telah memberikan akses data kelulusan SNBT tahun 2026.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

JIKUM
JURNAL ILMU KOMPUTER

E-ISSN: 3090-739X
Volume 2, **Number** 2, November 2026

REFERENSI

Aprudi, S., & Etriyanti, E. (2026). Penerapan Metode K-Means Clustering untuk Pengelompokan Peminatan Unit Kegiatan Mahasiswa. Jurnal Pustaka Data, 6(2). https://doi.org/10.55382/jurnalpustakadata.v6i2.1864 (doi.org in Bing)

Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques. Morgan Kaufmann.

Manuhutu, J. V., Makaruku, Y. N., & Halauwet, D. (2026). Analisis dan Visualisasi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma K-Means Clustering dan Silhouette Score Berdasarkan Asal Daerah. Progresif: Jurnal Ilmiah Komputer, 22(2). https://doi.org/10.35889/progresif.v22i2.3663 (doi.org in Bing)

Nurhaswinda, N., Azzahra, G., Qonitah, M., & Azzahra, N. (2025). Pemahaman Dasar Himpunan dalam Matematika: Konsep, Operasi dan Penerapannya di Dunia Nyata. Cendekia Pendidikan, 4(1). https://doi.org/10.36841/cendekiapendidikan.v4i1.6137 (doi.org in Bing)

Susanto, H., Rahmat, Y. S., & Saadah, R. (2025). Peranan Himpunan sebagai Fondasi Matematika Diskrit dan Implikasinya terhadap Pengembangan Sistem Komputasi. Journal of Computer Science and Information Technology, 3(1). https://doi.org/10.70248/jcsit.v3i1.3465 (doi.org in Bing)

Viola, A., & Syabila, Y. (2025). Analisis Efektivitas Program KIP Kuliah dalam Meningkatkan Prestasi Akademik Mahasiswa. Institut Syekh Abdul Halim Hasan Binjai.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Kaufman, L., & Rousseeuw, P. J. (2009). Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281–297.

Munir, R. (2012). Matematika Diskrit (Revisi Kelima). Bandung: Informatika Bandung.

Rosen, K. H. (2019). Discrete Mathematics and Its Applications (8th ed.). McGraw-Hill Education.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.