

## ANALISIS *BIG DATA* BEASISWA KIP-K MENGGUNAKAN *K-MEANS CLUSTERING*

Defi Pebriyanti<sup>1</sup>, Sarjon Defit<sup>2</sup>, Gunadi Widi Nurcahyo<sup>3</sup>

<sup>1,2,3</sup>Magister Teknik Informatika, Universitas Putra Indonesia YPTK Padang  
Jl. Raya Lubuk Begalung Padang, Sumatera Barat, Indonesia

e-mail: \*<sup>1</sup>[tanyadefify@gmail.com](mailto:tanyadefify@gmail.com), <sup>2</sup>[sarjond@yahoo.co.uk](mailto:sarjond@yahoo.co.uk), <sup>3</sup>[gunadiwidi@yahoo.co.id](mailto:gunadiwidi@yahoo.co.id)

### Abstract

The Kartu Indonesia Pintar Kuliah (KIP-K) Scholarship Program is a government initiative to provide higher education access to underprivileged students. It aims to reduce educational disparities and improve access for eligible students. However, the selection process faces challenges, particularly in identifying applicants who truly need financial aid. With the increasing number of applicants each year, a Big Data-based approach is essential to enhance selection efficiency and accuracy. This study analyzes KIP-K scholarship recipients' profiles using the K-Means Clustering method. This technique groups data based on attribute similarities, allowing an objective and data-driven selection process. The dataset, obtained from Universitas Prima Nusantara Bukittinggi (2024), consists of 479 applicants. It includes attributes such as academic performance, parental income, number of dependents, KIP-K card ownership, and achievements. Results indicate that recipients can be categorized based on document completeness, academic scores above 85, and more than three family dependents. Implementing K-Means Clustering improves the selection process by making it more objective, transparent, and efficient.

**Keyword:** Big Data, Data Mining, K-Means Clustering, Scholarship, Student Selection

### PENDAHULUAN

Era digital menghadirkan peran penting bagi *Knowledge Discovery in Databases* (KDD) dan *Big Data Mining* dalam menemukan informasi bernilai dari data dalam jumlah besar dan kompleks. KDD menggunakan algoritma matematika, AI, dan teknik data mining untuk menganalisis serta memprediksi pola di masa depan (Andi Diah Kuswanto *et al.*, 2024). KDD merupakan suatu cara yang digunakan dalam mendapatkan pengetahuan baru dari data yang terkumpul (Irawan *et al.*, 2025), proses analisis yang bertujuan untuk mengekstraksi pola dan hubungan baru dalam data yang luas dan bervariasi, menggunakan teknik statistik serta pembelajaran mesin (Neni Lusianah *et al.*, 2023). Salah satu metode yang umum digunakan dalam *Big Data Mining* adalah *Clustering*, yang mengelompokkan data berdasarkan karakteristik tertentu untuk memperoleh wawasan yang lebih mendalam (Mustika *et al.*, 2022).

Bidang pendidikan menunjukkan bahwa pengelompokan data menjadi semakin relevan, terutama dalam seleksi penerima beasiswa Kartu Indonesia Pintar Kuliah (KIP-K). Program KIP-Kuliah didukung oleh Permendikbud Nomor 10 Tahun 2020 tentang Program Indonesia Pintar, yang mengatur pelaksanaannya secara spesifik, serta Permendikbud Nomor 25 Tahun 2020 tentang Standar Satuan Biaya Operasional Pendidikan Tinggi yang menetapkan besaran bantuan bagi mahasiswa. Kedua peraturan ini berlandaskan pada PP Nomor 48 Tahun 2008 tentang Pendanaan Pendidikan, yang memberikan kerangka hukum bagi pendanaan pendidikan di Indonesia (Pendidikan, 2023). Program ini bertujuan untuk membantu mahasiswa dari keluarga kurang mampu agar dapat mengakses pendidikan tinggi (Sabrina Fitri Jasmine, 2023). Proses seleksi penerima beasiswa sering mengalami tantangan dalam menentukan kelayakan calon penerima secara efisien dan objektif. Pendekatan berbasis *K-Means Clustering* dapat digunakan untuk mengidentifikasi pola penerima beasiswa berdasarkan data akademik, ekonomi, serta faktor lainnya (Indriati *et al.*, 2024). Menurut Hendra Di Kesumal (2023) *K-Means Clustering* adalah salah satu algoritma pengajaran mesin yang paling populer dan

seederhana yang digunakan untuk memecahkan pengelompokan data (Hendra Di Kesuma1, 2023).

Penelitian mengenai penerapan *K-Means Clustering* dalam analisis *Big Data* telah berkembang pesat dalam berbagai bidang, termasuk pendidikan, kesehatan, dan bisnis. Algoritma ini digunakan untuk mengelompokkan data berdasarkan pola tertentu, sehingga dapat membantu dalam pengambilan keputusan yang lebih akurat dan efisien.

Beberapa penelitian sebelumnya telah menunjukkan efektivitas *K-Means Clustering* dalam menangani data besar dan kompleks. Diantaranya sebagai berikut. Ketepatan dalam seleksi penerima beasiswa menjadi tantangan utama dalam Program Indonesia Pintar (PIP) di SDN 020 PPU. Algoritma *K-Means Clustering* dalam penelitian ini diuji menggunakan perangkat lunak Weka, yang berperan sebagai alat bantu untuk memverifikasi hasil pengelompokan data dari proses perhitungan manual yang telah dilakukan sebelumnya, untuk menganalisis data 236 siswa berdasarkan jumlah tanggungan orang tua, nilai rapor, pekerjaan, dan penghasilan. Hasil penelitian yang dilakukan oleh Darlinda & Utamajaya (2022) menunjukkan bahwa 70 siswa merupakan penerima yang tepat sasaran, sedangkan 118 siswa tidak tepat sasaran dan 48 siswa masuk kategori lain. Pendekatan ini terbukti mampu meningkatkan objektivitas dalam seleksi penerima beasiswa (Darlinda & Utamajaya, 2022).

Pemanfaatan algoritma data mining semakin banyak diterapkan dalam proses seleksi beasiswa guna memastikan distribusi yang lebih tepat sasaran. Salah satu penelitian yang dilakukan oleh Yuliana *et al.* (2022) menggunakan *K-Means Clustering* untuk menyeleksi penerima beasiswa KIP-K di Universitas Nahdlatul Ulama Sunan Giri, Bojonegoro. Analisis dilakukan terhadap 346 pendaftar tahun akademik 2020/2021 dengan empat kriteria utama. Hasil *clustering* menunjukkan bahwa 219 mahasiswa (63,5%) layak menerima beasiswa, sedangkan 127 mahasiswa (36,5%) tidak memenuhi kriteria. Studi ini membuktikan bahwa metode *clustering* dapat memberikan hasil yang akurat dan mendekati data lapangan (Yuliana *et al.*, 2022).

Proses seleksi beasiswa yang masih dilakukan secara manual sering kali mengakibatkan ketidakefisienan dan ketidakakuratan. Untuk mengatasi permasalahan ini, penelitian menerapkan algoritma *K-Means Clustering* menggunakan Python pada Jupyter Notebook dalam seleksi penerima KIP-K di Universitas Papua. Dari 1.234 mahasiswa yang dianalisis, sebanyak 687 mahasiswa teridentifikasi sebagai penerima beasiswa, sedangkan 547 mahasiswa tidak memenuhi kriteria. Indriati *et al.* (2024) membuktikan bahwa metode ini mampu meningkatkan efisiensi dan akurasi dalam seleksi penerima beasiswa berdasarkan faktor akademik, ekonomi, dan prestasi (Indriati *et al.*, 2024).

Perkembangan teknologi analisis data terus mendorong peningkatan efisiensi dalam pengelompokan informasi skala besar. Salah satu penelitian yang berfokus pada optimasi algoritma *K-Means Clustering* dalam analisis *Big Data* dilakukan oleh Chen *et al.* (2023). Uji coba dengan dataset standar menunjukkan bahwa metode ini mampu meningkatkan kecepatan dan akurasi analisis hingga 95% dibandingkan teknik konvensional. Dengan hasil tersebut, *K-Means Clustering* terbukti sebagai solusi yang efektif dalam pemrosesan data besar secara lebih efisien dan akurat (Chen *et al.*, 2023).

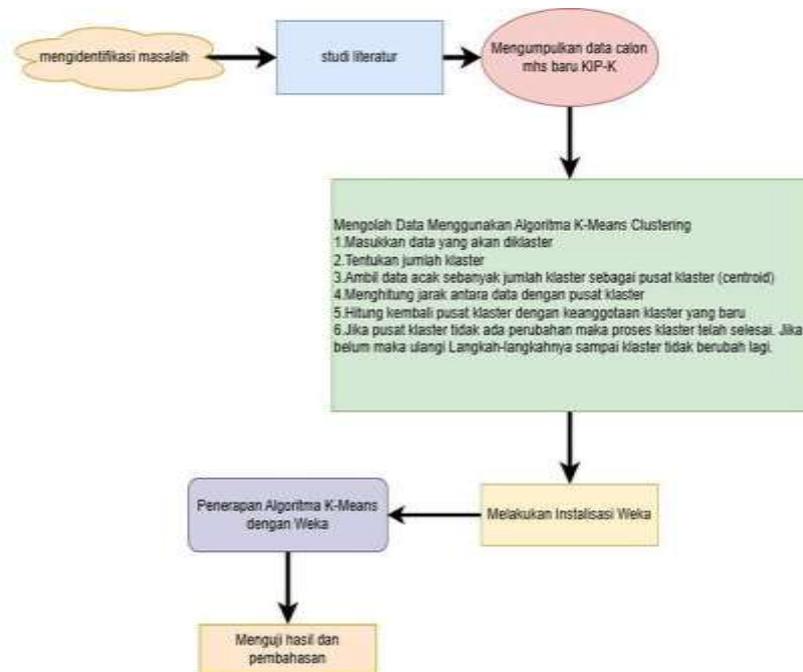
Penerapan teknologi kecerdasan buatan dalam dunia kesehatan semakin berkembang untuk meningkatkan akurasi layanan. Dalam penelitian ini, algoritma *K-Means Clustering* digunakan untuk mengoptimalkan pengelompokan pasien berdasarkan data sekunder dari sistem layanan kesehatan berbasis data. Rai *et al.* (2023) menunjukkan bahwa metode ini efektif dalam personalisasi perawatan serta distribusi sumber daya medis. Studi ini merekomendasikan pengembangan model hibrida dan integrasi algoritma ini ke dalam sistem layanan kesehatan untuk hasil yang lebih optimal (Rai *et al.*, 2023)

Penelitian ini bertujuan untuk (1) Menganalisis data penerima beasiswa KIP-K menggunakan metode *K-Means Clustering* guna mengidentifikasi pola penerima beasiswa, (2) Menentukan kategori penerima beasiswa KIP-K berdasarkan faktor akademik, ekonomi, dan prestasi menggunakan *K-Means Clustering*, (3) Mengimplementasikan *Software Weka* untuk memproses data penerima beasiswa dan meningkatkan akurasi klasifikasi penerima KIP-K.

Penelitian ini memilih algoritma K-Means Clustering karena kesesuaian karakteristik datanya yang memiliki struktur tidak berlabel, serta efektivitas metode ini dalam mengelompokkan data berdasarkan kemiripan atribut, sehingga dapat mengelompokkan penerima beasiswa berdasarkan karakteristik tertentu dengan lebih sistematis dan akurat. Penerapan algoritma ini memungkinkan analisis pola distribusi beasiswa secara lebih objektif, serta mendukung proses seleksi yang lebih transparan, efisien, dan tepat sasaran.

## METODE PENELITIAN

Kerangka penelitian merupakan suatu struktur atau rangkaian konsep yang digunakan untuk merancang dan menyusun sebuah penelitian. Kerangka ini membantu peneliti mengidentifikasi variabel-variabel yang akan diteliti, hubungan antar variabel, serta memberikan arahan tentang bagaimana penelitian akan dilakukan. Rincian kerangka penelitian yang digunakan disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1. Tahapan penelitian ini menunjukkan tahapan penelitian yang dilakukan dalam analisis penerima beasiswa KIP-K menggunakan algoritma *K-Means Clustering*. Proses dimulai dengan mengidentifikasi masalah dan melakukan studi literatur untuk memahami konsep serta metode yang relevan. Selanjutnya, data calon mahasiswa penerima KIP-K dikumpulkan dan diolah menggunakan algoritma *K-Means*, yang mencakup proses klusterisasi berdasarkan karakteristik tertentu. Implementasi dilakukan menggunakan perangkat lunak *Weka*, diikuti dengan pengujian dan analisis hasil untuk mendapatkan wawasan yang lebih mendalam terkait pola penerima beasiswa.

### a. Mengidentifikasi Masalah

Identifikasi masalah merupakan langkah awal dalam penelitian yang bertujuan untuk menentukan serta merumuskan permasalahan pada objek penelitian. Proses ini penting agar penelitian tetap fokus dan tidak keluar dari batasan yang telah ditentukan. Dalam tahap ini, peneliti harus memahami masalah dalam ruang lingkup yang telah ditetapkan sehingga hasil penelitian dapat sesuai dengan harapan. Selain itu, penentuan tujuan dilakukan untuk memberikan arah yang jelas dalam penelitian, sehingga hasil yang diperoleh dapat dimanfaatkan oleh pihak yang berkepentingan.

## b. Studi Literatur

Tahap ini mencakup eksplorasi dan penggalian informasi terkait implementasi algoritma *K-Means Clustering*. Sumber informasi yang digunakan meliputi teori-teori, jurnal-jurnal ilmiah, buku-buku, dan referensi-referensi lain yang relevan dan mendukung. Studi literatur ini bertujuan untuk mendapatkan pemahaman yang komprehensif tentang topik penelitian dan memastikan kelancaran penelitian di masa yang akan datang.

## c. Mengumpulkan Data calon mahasiswa KIP-K

Dalam tahap pengumpulan data, peneliti mengumpulkan data dari Universitas Prima Nusantara Bukittinggi. Selain itu, peneliti juga melakukan observasi dan wawancara dengan tujuan untuk memperoleh informasi tambahan mengenai data tersebut.

## d. Mengolah Data Menggunakan Algoritma *K-Means Clustering*

Pada tahap ini, data yang telah dikumpulkan akan diolah menggunakan algoritma *K-Means Clustering* untuk mengelompokkan penerima beasiswa berdasarkan karakteristik tertentu. Proses ini mencakup serangkaian langkah sistematis yang bertujuan untuk menentukan jumlah kluster optimal, menghitung jarak antar data, serta mengelompokkan data berdasarkan kesamaan atribut. Adapun tahapan-tahapan yang dilakukan dalam proses ini.

### 1. Masukkan data yang akan dikluster

Pada tahap ini, data yang telah dikumpulkan dan diproses sebelumnya dimasukkan ke dalam sistem atau perangkat lunak yang akan digunakan untuk analisis *clustering*. Data yang digunakan harus bersih dari duplikasi, missing values, dan sudah dalam format yang sesuai untuk pengolahan lebih lanjut.

### 2. Tentukan Jumlah kluster

Jumlah kluster yang akan digunakan dalam proses *clustering* harus ditentukan terlebih dahulu. Jumlah kluster dipilih sebanyak dua karena disesuaikan dengan tujuan penelitian, yaitu untuk mengelompokkan calon penerima menjadi dua kategori utama: 'layak' dan 'tidak layak'. Meskipun jumlah kluster tergolong sederhana, penggunaan algoritma *K-Means Clustering* tetap relevan karena volume data yang dianalisis cukup besar, sehingga pendekatan *clustering* diperlukan untuk memproses dan mengelompokkan data secara efisien.

### 3. Menentukan *Centroid* Awal

Pusat kluster pada penelitian ini ditentukan secara acak. *Centroid* dapat diambil dari data yang telah ditetapkan. Berikut data *Centroid* Awal

### 4. Normalisasi *Centroid* Awal

Normalisasi data awal ditunjukkan agar data pada atribut yang memiliki rentang nilai yang jauh berbeda dapat sama-sama mendominasi perhitungan jarak *Euclidean*, Normalisasi memastikan bahwa semua atribut memiliki skala yang sama. Rumus normalisasi yang digunakan adalah sebagai berikut:

$$= \frac{x - x \min}{x \max - x \min} \quad (1)$$

### 5. Menghitung jarak setiap data ke pusat *Cluster*

Setelah menentukan nilai K dan pusat *Cluster* awal, langkah berikutnya adalah menghitung jarak setiap data terhadap pusat *Cluster*. Perhitungan ini dilakukan dengan menerapkan rumus *Euclidean Distance* untuk menentukan seberapa dekat atau jauh suatu data dari pusat *Cluster* yang telah ditetapkan, berikut rumus yang digunakan :

$$D_{ij} = \sqrt{(x_i - s_i)^2 + (x_j - s_j)^2} \quad (2)$$

Berdasarkan rumus diatas dilakukan perhitungan iterasi, iterasi merupakan perhitungan yang berulang-ulang. Langkah ini akan terus dilakukan hingga diperoleh pusat *cluster* yang stabil dan jarak data terhadap pusat *cluster* mencapai nilai optimal.

6. Melakukan pengelompokan ke dalam *Cluster*  
Data dikelompokkan berdasarkan Jarak terdekat menuju pusat *Cluster*.
7. Menentukan pusat *Cluster* baru  
Menetapkan lokasi *centroid* terbaru dengan menghitung rata-rata nilai dari data yang terdapat dalam *cluster*, berikut perhitungan yang akan dilakukan:

$$C = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Jika hasil iterasi terakhir tidak mengalami perubahan dibandingkan dengan iterasi sebelumnya, maka proses *clustering* telah mencapai kondisi stabil dan *konverge*.

#### e. Penerapan Algoritma *K-Means* dengan *Weka*

Tahap ini dataset yang telah diproses akan dimasukkan ke dalam *Software Weka* untuk menerapkan algoritma *K-Means* secara otomatis. *Weka* menyediakan berbagai fitur untuk melakukan *clustering*, termasuk visualisasi hasil dan evaluasi performa model. Proses ini mencakup pemilihan parameter *K-Means*, pemrosesan data, serta analisis hasil *clustering* untuk memastikan bahwa kelompok yang terbentuk sesuai dengan karakteristik yang diharapkan.

#### f. Hasil dan Pembahasan

Tahap akhir dalam penelitian ini melakukan pengujian Hasil yang diperoleh, baik melalui perhitungan manual maupun dengan bantuan *Software* untuk mengetahui apakah perhitungan dengan *Software* sesuai dengan hasil yang diharapkan. Analisis dilakukan dengan membandingkan hasil *clustering* yang diperoleh dari perhitungan manual dan implementasi menggunakan *Weka*. Evaluasi dilakukan untuk menilai akurasi, efektivitas, serta pola yang terbentuk dalam setiap klaster. Hasil *clustering* dianalisis untuk mengidentifikasi karakteristik masing-masing kelompok penerima beasiswa KIP-K. Kesimpulan dari penelitian ini akan menjadi dasar dalam memberikan rekomendasi yang lebih tepat dalam proses seleksi penerima beasiswa di masa mendatang.

### HASIL DAN PEMBAHASAN

Penelitian ini menerapkan algoritma *K-Means Clustering* dalam proses pengelompokan calon Pendaftar beasiswa KIP-K di Universitas Prima Nusantara Bukittinggi. Data yang digunakan mencakup variabel seperti kepemilikan KIP-K, KIS/PKH, SKTM, nilai rapor, penghasilan orang tua, jumlah tanggungan, serta prestasi akademik dan non-akademik. Data dari penelitian ini dapat dilihat pada tabel 1.

Tabel 1. Tabel data Calon Pendaftar Beasiswa KIP-K

No	Kode Pendaftar	Kartu Kipk	Kartu Kis/Pkh	Sktm	Rata Nilai Rapor	Penghasilan Orang Tua	Jumlah Tanggungan	Prestasi Akademik	Prestasi Non Akademik
1	001	Ada	Ada	Ada	87.85	1200000	5	Ada	Ada
2	002	Ada	Ada	Ada	87.86	1000000	4	Ada	Ada
3	003	Ada	Ada	Ada	85.21	1500000	5	Tidak	Tidak
4	004	Ada	Tidak	Ada	87.45	1000000	3	Ada	Ada
5	005	Ada	Tidak	Ada	75.55	1000000	1	Tidak	Tidak
6	006	Ada	Tidak	Ada	87.86	1000000	4	Tidak	Tidak
7	007	Ada	Ada	Ada	92.41	700000	5	Tidak	Tidak
8	008	Ada	Tidak	Ada	81.45	1000000	2	Ada	Tidak
9	009	Ada	Tidak	Ada	84	700000	2	Ada	Ada

10	010	Tidak	Ada	Ada	81.45	1000000	4	Tidak	Tidak
...	...	...	...	...	...	...	...	...	...
479	479	Ada	Ada	Ada	89	800000	3	Tidak Ada	Tidak Ada

Sumber: Data UPN Bukittinggi, 2024

Selanjutnya melakukan perubahan data non numerik menjadi numerik sebagaimana ditampilkan pada tabel 2.

Tabel 2. Tabel dataset Calon Pendaftar Beasiswa KIP-K

No	Kode Pendaftar	Attr1	Attr2	Attr3	Attr4	Attr5	Attr6	Attr7	Attr8
1	001	1	1	1	2	2	3	1	1
2	002	1	1	1	2	2	2	1	1
3	003	1	1	1	2	2	3	0	0
4	004	1	0	1	2	2	2	1	1
5	005	1	0	1	1	2	1	0	0
6	006	1	0	1	2	2	2	0	0
7	007	1	1	1	3	3	3	0	0
8	008	1	0	1	2	2	2	1	0
9	009	1	0	1	2	3	2	1	1
10	010	0	1	1	2	2	2	0	0
...	...	...	...	...	...	...	...	...	...
479	479	1	1	1	2	2	2	0	0

Sumber: Dataset UPN Bukittinggi, 2024

Selanjutnya melakukan penentuan Jumlah kluster yang akan diterapkan dengan metode *K-Means* yang berjumlah 2 Kluster. Langkah selanjutnya menentukan *centroid* awal yang mana ditentukan secara acak. *Centroid* dapat diambil dari data yang telah ditetapkan pada tabel 3.

Tabel 3. *Centroid* Awal

Centroid	Data Ke-i	Attr 1	Attr 2	Attr 3	Attr 4	Attr 5	Attr 6	Attr 7	Attr 8
Cluster 1	463	1	1	0	2	2	3	0	0
Cluster 2	476	1	1	1	3	2	2	1	0

Sumber: Data Set Yang Diolah, 2024

Setelah menentukan *centroid* awal dilakukan normalisasi *centroid* Awal yang bertujuan agar data pada atribut yang memiliki rentang nilai yang jauh berbeda dapat sama-sama mendominasi perhitungan jarak *Euclidean*, sehingga di dapat hasil normalisasi terhadap *centroil* awal pada tabel 4.

Tabel 4. Normalisasi *centroid* awal

Centroid	Data Ke-i	Attr 1	Attr 2	Attr 3	Attr 4	Attr 5	Attr 6	Attr 7	Attr 8
Cluster 1	463	1	1	0	0.5	0	1	0	0
Cluster 2	476	1	1	1	1	0	0.5	1	0

Sumber: Data Set Yang Diolah, 2024

Setelah menentukan nilai K dan pusat *Cluster* awal, langkah berikutnya adalah menghitung jarak setiap data ke pusat *Cluster* menggunakan rumus *Euclidean Distance*, berikut rumus yang digunakan :

$$D_{ij} = \sqrt{(x_i - s_i)^2 + (x_j - s_j)^2}$$

Berdasarkan rumus diatas dilakukan perhitungan iterasi, iterasi merupakan perhitungan yang berulang-ulang, berikut perhitungan iterasi ke-1.

**Iterasi 1**

$$C1 = \sqrt{((1 + 1)^2 + (1 + 1)^2 + (1 + 0)^2 + (0.5 + 0.5)^2 + (0 + 0)^2 + (1 + 1)^2 + \dots)}$$

$$= 1.732050808$$

$$C2 = \sqrt{((1 + 1)^2 + (1 + 1)^2 + (1 + 1)^2 + (0.5 + 1)^2 + (0 + 0)^2 + (1 + 0.5)^2 + \dots)}$$

$$= 1.224744871$$

$$C1 = \sqrt{((1 + 1)^2 + (1 + 1)^2 + (1 + 0)^2 + (0.5 + 0.5)^2 + (0 + 0)^2 + (0.5 + 1)^2 + \dots)}$$

$$= 1.802775638$$

$$C2 = \sqrt{((1 + 1)^2 + (1 + 1)^2 + (1 + 1)^2 + (0.5 + 1)^2 + (0 + 0)^2 + (0.5 + 0.5)^2 + \dots)}$$

$$= 1.118033989$$

$$C1 = \sqrt{((1 + 1)^2 + (1 + 1)^2 + (1 + 0)^2 + (0.5 + 0.5)^2 + (0 + 0)^2 + (1 + 1)^2 + \dots)}$$

$$= 1$$

$$C2 = \sqrt{((1 + 1)^2 + (1 + 1)^2 + (1 + 1)^2 + (0.5 + 1)^2 + (0 + 0)^2 + (1 + 0.5)^2 + \dots)}$$

$$= 1.224744871$$

Iterasi 1, jarak ke pusat *cluster* C1 dan C2 terus berubah, mencerminkan proses pengelompokan data. Langkah ini berlanjut hingga pusat *cluster* stabil dan jarak data mencapai nilai optimal.

Selanjutnya melakukan pengelompokan kedalam *Cluster* Data dikelompokkan berdasarkan jarak yang paling pendek ke pusat *Cluster*. Hasil pengelompokan data dapat ditemukan pada Tabel 5.

Tabel 5. Pengelompokan data Iterasi 1

No	Kode Pendaftar	C1	C2	Jarak Terdekat	Cluster
1	001	1.732050808	1.224744871	1.224744871	C2
2	002	1.802775638	1.118033989	1.118033989	C2
3	003	1	1.224744871	1	C1
4	004	2.061552813	1.5	1.5	C2
5	005	1.802775638	1.802775638	1.802775638	C1
6	006	1.5	1.5	1.5	C1
7	007	1.118033989	1.118033989	1.118033989	C1
8	008	1.802775638	1.118033989	1.118033989	C2
9	009	2.061552813	1.5	1.5	C2
10	010	1.118033989	1.118033989	1.118033989	C1
...	...	...	...	...	...
479	479	1.118033989	1.118033989	1.118033989	C1

Sumber: Data Set Yang Diolah, 2024

Selanjutnya menghitung Pusat *Cluster* yang baru ditentukan dengan menghitung rata-rata nilai dari data yang tersedia berikut perhitungan yang akan dilakukan

$$C = \frac{1}{n} \sum_{i=1}^n x_i$$

$$C1a = \frac{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + \dots}{292}$$

$$C1b = \frac{1 + 1 + 1 + 0 + 0 + 0 + 1 + 0 + 0 + 1 + \dots}{292}$$

$$C1c = \frac{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + \dots}{292}$$

...

$$C2a = \frac{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + \dots}{187}$$

$$C2b = \frac{1 + 1 + 1 + 0 + 0 + 0 + 1 + 0 + 0 + 1 + \dots}{187}$$

$$C2c = \frac{1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + \dots}{187}$$

...

Hasil perhitungan rata-rata pada setiap dimensi menentukan posisi *centroid* baru untuk *cluster* C1 dan C2 dan digunakan sebagai referensi iterasi berikutnya. Proses ini berulang hingga *centroid* stabil. Hasil perhitungan menunjukkan pusat *cluster* baru pada Tabel 6.

Tabel 6. *Centroid* Baru

Centroid	Total	Attr 1	Attr 2	Attr 3	Attr 4	Attr 5	Attr 6	Attr 7	Attr 8
C1	292	0.80822	0.60274	0.85274	0.5411	0.09932	0.7089	0.01027	0.00342
C2	187	0.81283	0.69519	0.9893	0.67112	0.07487	0.6123	0.76471	0.02139

Sumber: Data Set Yang Diolah, 2024

Iterasi kedua dilakukan dengan Mengukur jarak data terhadap pusat *Cluster* yang baru dengan menerapkan rumus *Euclidean Distance* Prosesnya sama seperti iterasi pertama, tetapi dengan pusat *cluster* terbaru. Hasilnya ditampilkan pada Tabel 7.

Tabel 7. Pengelompokan Data Iterasi 2

No	Kode Pendaftar	C1	C2	Jarak Terdekat	Cluster
1	001	1.511717696	1.151649316	1.151649316	C2
2	002	1.498063584	1.090227322	1.090227322	C2
3	003	0.559184811	0.947886418	0.559184811	C1
4	004	1.565143429	1.256570708	1.256570708	C2
5	005	1.107732465	1.390893568	1.107732465	C1
6	006	0.69070342	1.072922391	0.69070342	C1
7	007	0.7222131	0.988618059	0.7222131	C1
8	008	1.206865058	0.788511695	0.788511695	C2
9	009	1.565143429	1.256570708	1.256570708	C2
10	010	0.521144665	0.872231693	0.521144665	C1

...	...	...	...	...	...
479	479	0.521144665	0.872231693	0.521144665	C1

Sumber: Data Set Yang Diolah, 2024

Jika masih ada perpindahan data setelah iterasi sebelumnya, perhitungan diulang hingga stabil. Perbandingan iterasi pertama dan kedua menunjukkan perubahan anggota *cluster*. Iterasi ketiga dilakukan dengan menghitung kembali jarak data menggunakan *Euclidean Distance*, dengan pusat *cluster* terbaru yang ditampilkan pada Tabel 8.

Tabel 8. Centroid Baru Iterasi 2

Centroid	Total	Attr 1	Attr 2	Attr 3	Attr 4	Attr 5	Attr 6	Attr 7	Attr 8
C1	333	0.79279	0.5946	0.87988	0.58859	0.09309	0.66517	0	0.003
C2	146	0.84932	0.73973	0.96575	0.59932	0.08219	0.68493	1	0.0274

Sumber: Data Set Yang Diolah, 2024

Setelah pusat *Cluster* baru ditentukan Selanjutnya, lakukan iterasi ketiga untuk menghitung jarak masing-masing data ke pusat *Cluster* yang baru. Hasil dari iterasi ketiga dapat dilihat pada tabel 9.

Tabel 9. Pengelompokan Data Iterasi 3

No	Kode Pendaftar	C1	C2	Jarak Terdekat	Cluster
1	001	1.531126661	1.073994662	1.073994662	C2
2	002	1.503167994	1.043262211	1.043262211	C2
3	003	0.591907813	1.099208376	0.591907813	C1
4	004	1.564833284	1.252137411	1.252137411	C2
5	005	1.099301127	1.551415222	1.099301127	C1
6	006	0.674321299	1.273829901	0.674321299	C1
7	007	0.715378411	1.165737529	0.715378411	C1
8	008	1.206113267	0.789077066	0.789077066	C2
9	009	1.564833284	1.252137411	1.252137411	C2
10	010	0.515286352	1.069200898	0.515286352	C1
...	...	...	...	...	...
479	479	0.521144665	0.872231693	0.521144665	C1

Sumber: Data Set Yang Diolah, 2024

Setelah dilakukan pengelompokan data iterasi 3 di dapatlah centroid baru iterasi 3 dapat dilihat pada tabel 10.

Tabel 10. Centroid baru

Centroid	Total	Attr 1	Attr 2	Attr 3	Attr 4	Attr 5	Attr 6	Attr 7	Attr 8
C1	333	0.7928	0.5946	0.8799	0.5886	0.0931	0.6652	0	0.003
C2	146	0.8493	0.7397	0.9658	0.5993	0.0822	0.6849	1	0.0274

Sumber: Data Set Yang Diolah, 2024

Hasil pengelompokan pada iterasi kedua dan ketiga tidak berubah, menandakan *clustering* telah stabil. Untuk penelitian ini terdapat tiga iterasi yang dilakukan dan telah mendapatkan hasil yang *konvergen*. *Cluster* 1 menunjukkan penerima KIP-K yang tidak tepat sasaran sebanyak 333 orang, sedangkan *cluster* 2 menunjukkan penerima yang tepat 146 orang.

Interpretasi ini diperoleh dari pengamatan terhadap nilai rata-rata atribut pada klaster yang direkomendasikan, di mana atribut seperti nilai rapor (rata-rata di atas 85), jumlah tanggungan keluarga >3, dan kelengkapan dokumen beasiswa menunjukkan dominasi dalam klaster yang dianggap ‘tepat sasaran’

## KESIMPULAN

Penelitian ini menerapkan algoritma *K-Means Clustering* dalam analisis penerima beasiswa Kartu Indonesia Pintar Kuliah (KIP-K) untuk mengidentifikasi pola dan karakteristik calon penerima. Dari hasil pengelompokan, penerima beasiswa dikategorikan ke dalam dua kelompok utama: penerima yang tepat sasaran dan tidak tepat sasaran. Proses iterasi menunjukkan bahwa stabilitas *clustering* tercapai pada iterasi ketiga, di mana *cluster 1* mengelompokkan 333 penerima yang tidak tepat sasaran, sementara *cluster 2* mengelompokkan 146 penerima yang memenuhi kriteria. Hasil penelitian menunjukkan bahwa calon penerima yang paling layak mendapatkan KIP-K adalah mereka yang memenuhi persyaratan lengkap, memiliki nilai akademik di atas 85, dan memiliki lebih dari tiga tanggungan keluarga.

Dengan demikian penerapan algoritma K-Means Clustering mampu mendukung proses seleksi penerima beasiswa KIP-K secara lebih objektif, transparan, dan terstruktur. Dari hasil pengelompokan, *cluster 2* teridentifikasi sebagai kelompok calon mahasiswa yang memenuhi kriteria kelayakan berdasarkan atribut akademik, ekonomi, dan administratif. Sehingga *cluster 2* direkomendasikan kepada Universitas Prima Nusantara Bukittinggi sebagai dasar pengambilan keputusan dalam penyaluran beasiswa, guna memastikan bantuan diterima oleh mahasiswa yang benar-benar membutuhkan secara adil dan tepat sasaran.

## SARAN

Berdasarkan hasil penelitian ini, berikut beberapa saran yang dapat diberikan untuk penelitian selanjutnya:

1. Penelitian ini hanya menggunakan salah satu teknik data mining dalam penggalian pengetahuan baru (*knowledge*) yaitu menggunakan algoritma *K-Means Clustering* untuk menganalisa penerima beasiswa Kartu Indonesia Pintar Kuliah (KIP-K) untuk mengidentifikasi pola dan karakteristik calon penerima. Penelitian selanjutnya dapat dibandingkan metode ini dengan algoritma lainnya untuk meningkatkan akurasi dan efektivitas analisis.
2. Untuk meningkatkan keakuratan hasil analisis penelitian berikutnya dapat memasukkan variabel lain sebagai faktor tambahan dalam pengelompokan calon penerima beasiswa.

## DAFTAR PUSTAKA

- Andi Diah Kuswanto, Azumardi Nabil Fadhila, Paulus Tri Setiawan, Muhammad Kevin Setiawan, & Dody Renal Syahputra. (2024). Penerapan K-Means Clustering Untuk Menentukan Jumlah Pengangguran Berdasarkan Umur. *Repeater: Publikasi Teknik Informatika Dan Jaringan*, 2(3), 135–146. <https://doi.org/10.62951/repeater.v2i3.116>
- Chen, S., Tian, Y., Huang, R., & Dai, Z. (2023). Optimization Simulation of Big Data Analysis Model Based on K-means Algorithm. *Proceedings - 2023 International Conference on Networking, Informatics and Computing, ICNETIC 2023*, 764–768. <https://doi.org/10.1109/ICNETIC59568.2023.00162>
- Darlinda, D., & Utamajaya, J. N. (2022). Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means Clustering. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 167. <https://doi.org/10.30865/jurikom.v9i2.3971>
- Hendra Di Kesuma1, S. H. (2023). Penerapan Data Mining Menggunakan Algoritma K-Means

*Clustering dalam Pengelompokan Penerima Beasiswa KIP Kuliah.*

- Indriati, E., Suharyani Azisa, N., Ivo Sihombing, E., & Sukma Dewi Mokodompit, Z. (2024). Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Status Penerima KIP Kuliah Mahasiswa Universitas Papua. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3458–3463. <https://doi.org/10.36040/jati.v7i6.8222>
- Irawan, I., Rahman, R., Wibowo, A., Studi, P., Ilmu, M., Informasi, F. T., Luhur, U. B., & Utara, P. (2025). Pengelompokan Transaksi Kartu Debit Perbankan Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Dan Informatika (Simika)*, 8(1), 99–110.
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Habis, I., Guntoro, Agnes Manuhutu, M., Ridwan, M., Hozairi, Khrisna Wardhani, A., Alim, S., Romli, I., Religia, Y., Octafian, D. T., Utan Sufandi, U., & Ernawati, Ii. (2022). *Data Mining 2.Pdf*.
- Neni Lusianah, Ade Irma Purnamasari, & Bani Nurhakim. (2023). Implementasi Algoritma K-Means Dalam Pengelompokan Jumlah Wisatawan Akomodasi Di Jawa Barat. *Jurnal Ekonomi, Bisnis Dan Manajemen*, 2(1), 254–268. <https://doi.org/10.58192/ebismen.v2i1.682>
- Pendidikan, P. L. P. (2023). *Program Indonesia Pintar (PIP) Pendidikan Tinggi Pedoman Pendaftaran Kartu Indonesia Pintar Kuliah KIP Kuliah Merdeka*.
- Rai, A. K., Aswal, U. S., Saravanan, V., Shalini, N., Dwivedi, S. P., & Kumar, N. (2023). Patient Clustering Optimization With K-Means In Healthcare Data Analysis. *International Conference on Artificial Intelligence for Innovations in Healthcare Industries, ICAIHI 2023, 1*, 1–7. <https://doi.org/10.1109/ICAIIHI57871.2023.10489428>
- Sabrina Fitri Jasmine. (2023). Pengaruh Beasiswa KIP-K Terhadap Prestasi Belajar Mahasiswa Manajemen Pendidikan Angkatan 2021 Universitas Negeri Surabaya. *Jurnal Pendidikan, Bahasa Dan Budaya*, 2(2), 61–70. <https://doi.org/10.55606/jpbb.v2i2.1437>
- Usanto. (2023). Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP. *Building of Informatics, Technology and Science (BITS)*, 5(1). <https://doi.org/10.47065/bits.v5i1.3411>
- Yuliana, D. T., Fathoni, M. I. A., & Kurniawati, N. (2022). Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering. *Journal Focus Action of Research Mathematic (Factor M)*, 5(1), 127–141. [https://doi.org/10.30762/f\\_m.v5i1.570](https://doi.org/10.30762/f_m.v5i1.570)