BAB II
TINJAUAN PUSTAKA

2.1 Penelitian relevan

Tabel 2. 1 Penelitian Relevan 1

Nama Penulis	Bernadus Gunawan Sudarsono & Sri Poedji Lestari
Tahun Terbit	2021
Judul Penelitian	Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means
Metode Penelitian	Pengumpulan Data: Data diperoleh dari mahasiswa calon penerima beasiswa dengan kriteria IPK, prestasi non akademik, disiplin, dan penghasilan orang tua. Data dikumpulkan melalui wawancara, dokumentasi, dan studi pustaka. Pra-pemrosesan Data: Data kualitatif diubah menjadi bentuk numerik agar dapat diproses oleh algoritma K-Means. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan mahasiswa ke dalam beberapa klaster berdasarkan kemiripan karakteristik menggunakan perhitungan jarak Euclidean dan proses iterasi hingga centroid stabil.
Hasil Penelitian	Hasil penelitian menunjukkan bahwa data mahasiswa dapat dikelompokkan ke dalam empat klaster berdasarkan kriteria yang digunakan. Klasterisasi membantu pihak yayasan dalam menentukan kelompok mahasiswa yang layak menerima beasiswa sesuai tingkat kebutuhan dan prestasi.
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan metode K-Means untuk mengelompokkan data penerima beasiswa. Selain itu, penelitian ini menunjukkan bahwa algoritma K-Means efektif dalam membantu proses pengambilan keputusan

	berbasis data untuk menentukan kelompok penerima bantuan pendidikan.

Penelitian pertama yang dilakukan oleh Sudarsono dan Sudarsono & Lestari (2021), berjudul *"Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means"*, mengusulkan metode pengelompokan mahasiswa penerima beasiswa menggunakan algoritma K-Means. Penelitian ini menekankan bahwa proses seleksi beasiswa dapat dilakukan secara lebih objektif dengan mengelompokkan mahasiswa berdasarkan karakteristik akademik dan ekonomi.Melalui tahapan pra-pemrosesan data, konversi nilai ke bentuk numerik, serta perhitungan jarak menggunakan Euclidean Distance, penelitian ini berhasil membentuk empat klaster mahasiswa dengan karakteristik yang berbeda. Hasil klasterisasi membantu pihak yayasan dalam menentukan prioritas pemberian beasiswa sesuai kebutuhan dan prestasi mahasiswa.Secara metodologis, penelitian ini memberikan dasar yang kuat bahwa algoritma K-Means mampu mengelompokkan data penerima beasiswa secara efektif dan dapat digunakan sebagai alat bantu pengambilan keputusan. Oleh karena itu, penelitian ini sangat relevan dengan penelitian yang dilakukan, yang juga bertujuan mengelompokkan penerima beasiswa menggunakan metode K-Means berdasarkan variabel tertentu.

Tabel 2. 2 Penelitian Relevan 2

Nama Penulis	Sindrawati, Dodi Syaripudin, & Agung Rachmat Raharja
Tahun Terbit	2024
Judul Penelitian	Penerapan Algoritma K-Means Clustering pada Data Nilai Siswa untuk Menentukan Kelompok Penerima Beasiswa
Metode Penelitian	Pengumpulan Data: Data nilai siswa kelas XII dikumpulkan berdasarkan beberapa mata pelajaran utama serta jurusan siswa.

	Pra-pemrosesan Data: Dilakukan proses data cleaning, seleksi atribut, serta transformasi data nominal menjadi numerik agar dapat diproses oleh algoritma K-Means. Metode Analisis: Penelitian menggunakan pendekatan CRISP-DM yang meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan implementasi. Algoritma K-Means digunakan untuk mengelompokkan siswa berdasarkan kemiripan nilai akademik menggunakan perhitungan jarak Euclidean
Hasil Penelitian	Penelitian menghasilkan tiga klaster dari 109 data siswa, yaitu klaster pertama berjumlah 26 siswa, klaster kedua 46 siswa, dan klaster ketiga 37 siswa. Klaster dengan rata-rata nilai tertinggi dianggap sebagai kelompok yang paling berpotensi menerima beasiswa, sedangkan klaster dengan nilai terendah tidak direkomendasikan sebagai penerima. Hasil klasterisasi dapat digunakan sebagai dasar pengambilan keputusan yang lebih objektif dalam penentuan penerima beasiswa
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan data calon penerima beasiswa berdasarkan atribut numerik. Metode pengelompokan berbasis kemiripan data dapat dijadikan referensi dalam penelitian ini untuk menentukan kelompok mahasiswa penerima beasiswa berdasarkan variabel Semester dan UKT.

Penelitian yang dilakukan oleh Sindrawati et al. (2024), berjudul *"Penerapan Algoritma K-Means Clustering pada Data Nilai Siswa untuk Menentukan Kelompok Penerima Beasiswa"* mengusulkan penggunaan teknik data mining untuk membantu sekolah dalam menentukan siswa yang layak menerima beasiswa. Penelitian ini menekankan bahwa pengambilan keputusan berbasis nilai akademik dapat dilakukan

secara lebih objektif melalui proses pengelompokan data menggunakan algoritma K-Means.Melalui pendekatan CRISP-DM, data siswa terlebih dahulu dibersihkan dan ditransformasikan ke bentuk numerik sebelum dilakukan proses klasterisasi menggunakan perhitungan jarak Euclidean. Hasil penelitian menunjukkan terbentuknya tiga klaster dengan karakteristik nilai yang berbeda, di mana klaster dengan nilai rata-rata tertinggi menjadi kelompok yang paling direkomendasikan sebagai penerima beasiswa. Temuan ini menunjukkan bahwa metode K-Means mampu mengidentifikasi kelompok siswa berdasarkan kemiripan prestasi akademik secara sistematis.Secara metodologis, penelitian ini sangat relevan dengan penelitian yang dilakukan karena sama-sama bertujuan mengelompokkan calon penerima beasiswa menggunakan algoritma K-Means. Perbedaannya terletak pada variabel yang digunakan, yaitu penelitian sebelumnya menggunakan nilai akademik siswa, sedangkan penelitian ini menggunakan variabel Semester (SMT) dan Uang Kuliah Tunggal (UKT) untuk menentukan kelompok mahasiswa penerima beasiswa.

Tabel 2. 3 Penelitian Relevan 3

Nama Penulis	Abu Salam, Diyan Adiatma, dan Junta Zeniarja
Tahun Terbit	2020
Judul Penelitian	Implementasi Algoritma K-Means dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS
Metode Penelitian	Pengumpulan Data: Data pendaftar beasiswa PPA tahun 2016 dari Biro Kemahasiswaan UDINUS sebanyak 441 mahasiswa. Pra-pemrosesan Data: Dilakukan seleksi atribut dengan menghapus atribut yang tidak relevan seperti NIM, nama, dan nomor rekening. Metode Analisis: Menggunakan algoritma K-Means untuk mengelompokkan data berdasarkan kemiripan karakteristik mahasiswa dengan perhitungan jarak Euclidean dan pembaruan centroid secara iteratif hingga konvergen

Hasil Penelitian	Penelitian menghasilkan dua klaster, yaitu klaster mahasiswa yang direkomendasikan menerima beasiswa sebanyak 154 orang dan klaster mahasiswa yang tidak direkomendasikan sebanyak 287 orang. Hasil rekomendasi menunjukkan kesesuaian dengan data riil penerima beasiswa sebanyak 113 mahasiswa, dengan tingkat akurasi sebesar 82%
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan data penerima beasiswa. Hasil penelitian menunjukkan bahwa metode clustering efektif dalam membantu proses seleksi beasiswa secara objektif dan berbasis data, sehingga dapat dijadikan referensi metodologis dalam penelitian pengelompokan penerima beasiswa berdasarkan variabel Semester dan UKT.

Penelitian yang dilakukan oleh Salam et al. (2020), berjudul *"Implementasi Algoritma K-Means dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS"* mengkaji penggunaan metode K-Means untuk membantu proses seleksi penerima beasiswa secara lebih objektif. Penelitian ini menggunakan data pendaftar beasiswa PPA sebanyak 441 mahasiswa dan melakukan proses clustering berdasarkan atribut yang relevan setelah tahap pra-pemrosesan data.Hasil penelitian menunjukkan bahwa data dapat dikelompokkan menjadi dua klaster utama, yaitu kelompok mahasiswa yang direkomendasikan menerima beasiswa dan kelompok yang tidak direkomendasikan. Dari hasil clustering, sebanyak 154 mahasiswa masuk dalam klaster penerima beasiswa dan 287 mahasiswa tidak direkomendasikan. Setelah dibandingkan dengan data riil, diperoleh kesesuaian sebanyak 113 mahasiswa dengan tingkat akurasi sebesar 82%. Penelitian ini menunjukkan bahwa algoritma K-Means mampu membantu proses seleksi beasiswa menjadi lebih terstruktur dan berbasis data. Oleh karena itu, penelitian ini sangat relevan dengan penelitian yang dilakukan, karena sama-sama menggunakan metode K-Means untuk mengelompokkan calon penerima beasiswa berdasarkan karakteristik tertentu guna mendukung pengambilan keputusan yang lebih objektif.

Tabel 2. 4 Penelitian Relevan 4

Nama Penulis	Fadlil, A., Riadi, I., & Mulyana, Y.
Tahun Terbit	2022
Judul Penelitian	Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan
Metode Penelitian	Pengumpulan Data: Data diperoleh dari pendaftar program bantuan biaya pendidikan (KIP-Kuliah) di Universitas Muhammadiyah Tasikmalaya melalui wawancara dan dokumentasi. Variabel yang digunakan meliputi status DTKS, kepemilikan KIP/KKS, penghasilan orang tua, status orang tua, jumlah tanggungan, dan prestasi siswa. Pra-pemrosesan Data: Data nominal diubah menjadi numerik melalui proses inisialisasi agar dapat diolah menggunakan algoritma K-Means. Data juga dibersihkan dari nilai kosong atau anomali. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan data menjadi tiga klaster (K=3) dengan perhitungan jarak Euclidean Distance. Proses dilakukan secara iteratif hingga centroid stabil. Hasil manual kemudian divalidasi menggunakan tools RapidMiner untuk memastikan konsistensi hasil.
Hasil Penelitian	Penelitian Hasil klasterisasi menunjukkan terbentuk tiga kelompok data, yaitu klaster C0 sebanyak 109 data (57,1%), klaster C1 sebanyak 52 data (27,2%), dan klaster C2 sebanyak 30 data (15,7%). Klaster C2 direkomendasikan sebagai kelompok yang layak menerima bantuan, klaster C0 sebagai

	kelompok yang dipertimbangkan, dan klaster C1 sebagai kelompok yang tidak layak.
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan metode K-Means untuk mengelompokkan data penerima bantuan pendidikan berdasarkan beberapa variabel numerik. Hasil penelitian menunjukkan bahwa K-Means efektif dalam membantu pengambilan keputusan penentuan penerima bantuan secara objektif. Pendekatan ini dapat dijadikan referensi metodologis dalam penelitian pengelompokan penerima beasiswa berdasarkan variabel tertentu

Penelitian yang dilakukan oleh (Elektro et al., 2022) berjudul *"Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan"* membahas penggunaan metode K-Means untuk membantu penentuan penerima bantuan pendidikan secara lebih objektif. Penelitian ini dilatarbelakangi oleh banyaknya pendaftar bantuan KIP-Kuliah yang tidak seluruhnya berasal dari keluarga kurang mampu, sehingga diperlukan metode analisis data untuk mengelompokkan pendaftar berdasarkan kondisi ekonomi dan faktor pendukung lainnya.Dalam penelitian tersebut, data pendaftar yang terdiri dari berbagai atribut seperti penghasilan orang tua, status DTKS, kepemilikan kartu bantuan, jumlah tanggungan, dan prestasi siswa diolah menggunakan algoritma K-Means dengan jumlah klaster tiga. Proses klasterisasi dilakukan secara iteratif menggunakan jarak Euclidean hingga diperoleh centroid yang stabil. Hasil pengelompokan menunjukkan adanya tiga kategori utama, yaitu kelompok layak, dipertimbangkan, dan tidak layak menerima bantuan.Secara metodologis, penelitian ini sangat relevan karena menunjukkan bahwa K-Means mampu mengelompokkan data bantuan pendidikan secara efektif berdasarkan karakteristik tertentu. Temuan ini dapat menjadi acuan dalam penelitian pengelompokan penerima beasiswa, termasuk dalam menentukan kelompok prioritas penerima bantuan berdasarkan variabel yang digunakan.

Tabel 2. 5 Penelitian Relevan 5

Nama Penulis	Surya Darma, Sarjon Defit, Dedy Hartama, Wendi Robiansyah, Fahmi Firzada
Tahun Terbit	2020
Judul Penelitian	Penerapan Metode K-Means dalam Pengelompokan Jumlah Wisatawan Asing di Indonesia
Metode Penelitian	Pengumpulan Data: Data jumlah wisatawan asing per bulan tahun 2017–2019 diperoleh dari Badan Pusat Statistik (BPS). Pra-pemrosesan Data: Data diseleksi dan disusun berdasarkan periode waktu untuk dianalisis menggunakan metode data mining. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan data menjadi dua klaster, yaitu klaster tinggi dan rendah. Jarak antar data dihitung menggunakan Euclidean Distance, kemudian dilakukan iterasi hingga pusat klaster stabil. Implementasi juga dilakukan menggunakan perangkat lunak RapidMiner.
Hasil Penelitian	Hasil penelitian menunjukkan bahwa bulan Juni merupakan periode dengan jumlah kunjungan wisatawan asing tertinggi, sedangkan bulan lainnya tergolong rendah. Metode K-Means berhasil mengelompokkan data kunjungan wisatawan ke dalam dua kategori utama, yaitu tinggi dan rendah, sehingga dapat digunakan sebagai dasar analisis tren kunjungan wisata.
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan data numerik berdasarkan kemiripan karakteristik. Meskipun objek penelitian berbeda, metode yang digunakan identik dengan penelitian ini, yaitu clustering tanpa supervisi untuk menemukan pola dalam data. Oleh karena itu, penelitian ini dapat dijadikan referensi

	metodologis dalam penerapan K-Means untuk pengelompokan data penerima beasiswa berdasarkan variabel Semester dan UKT.

Penelitian yang dilakukan oleh Darma et al. (2020) berjudul *"Penerapan Metode K-Means dalam Pengelompokan Jumlah Wisatawan Asing di Indonesia"* menggunakan algoritma K-Means untuk mengelompokkan data kunjungan wisatawan asing berdasarkan periode bulanan. Penelitian ini menunjukkan bahwa metode K-Means mampu mengidentifikasi pola kunjungan wisata dengan membagi data menjadi klaster tinggi dan rendah, di mana bulan Juni merupakan periode dengan jumlah kunjungan tertinggi.Proses pengelompokan dilakukan dengan menghitung jarak Euclidean antara data dan pusat klaster, kemudian memperbarui centroid hingga konvergen. Implementasi menggunakan RapidMiner memperkuat validitas hasil analisis. Secara metodologis, penelitian ini menunjukkan bahwa K-Means efektif untuk mengelompokkan data numerik dan menemukan pola tersembunyi dalam dataset.Penelitian ini relevan dengan penelitian yang dilakukan karena sama-sama menggunakan algoritma K-Means sebagai metode utama untuk pengelompokan data. Perbedaannya terletak pada objek penelitian, di mana penelitian ini berfokus pada pengelompokan mahasiswa penerima beasiswa berdasarkan Semester dan UKT. Namun demikian, konsep clustering dan prosedur analisis yang digunakan memiliki kesamaan, sehingga penelitian tersebut dapat dijadikan landasan dalam penerapan metode K-Means pada penelitian ini.

Tabel 2. 6 Penelitian Relevan 6

Nama Penulis	Fadlil, A., Riadi, I., & Mulyana, Y
Tahun Terbit	2022
Judul Penelitian	Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan

Metode Penelitian	Pengumpulan Data dilakukan melalui wawancara dan pengambilan data pendaftar KIP-Kuliah Universitas Muhammadiyah Tasikmalaya sebanyak 191 data. Pra-pemrosesan data dilakukan dengan mengubah data nominal menjadi numerik pada beberapa atribut seperti status DTKS, kepemilikan KIP/KKS, penghasilan orang tua, status orang tua, jumlah tanggungan, dan prestasi. Metode analisis menggunakan algoritma K-Means dengan jumlah klaster K = 3 serta perhitungan jarak Euclidean untuk menentukan kedekatan data terhadap centroid. Perhitungan dilakukan secara manual dan menggunakan perangkat lunak RapidMiner untuk validasi hasil.
Hasil Penelitian	Penelitian menghasilkan tiga klaster yaitu C0 sebanyak 109 data (57,1%), C1 sebanyak 52 data (27,2%), dan C2 sebanyak 30 data (15,7%). Berdasarkan karakteristiknya, klaster C2 direkomendasikan sebagai kelompok yang layak menerima bantuan biaya pendidikan, C0 sebagai kelompok yang dipertimbangkan, dan C1 sebagai kelompok yang tidak layak menerima bantuan.
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan metode K-Means untuk mengelompokkan data penerima bantuan pendidikan guna mendukung pengambilan keputusan. Pendekatan clustering berbasis atribut sosial ekonomi dapat menjadi referensi dalam penelitian ini yang mengelompokkan penerima beasiswa berdasarkan variabel Semester dan UKT.

Penelitian yang dilakukan oleh (Elektro et al., 2022), berjudul *"Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan"* bertujuan untuk mengoptimalkan penentuan penerima bantuan pendidikan melalui teknik data mining. Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan

191 data pendaftar berdasarkan beberapa atribut sosial ekonomi seperti penghasilan orang tua, status keluarga, kepemilikan kartu bantuan, jumlah tanggungan, dan prestasi.Melalui proses klasterisasi dengan K = 3 dan perhitungan jarak Euclidean, diperoleh tiga kelompok utama dengan karakteristik yang berbeda. Hasil penelitian menunjukkan bahwa satu klaster memiliki karakteristik paling layak menerima bantuan, sementara klaster lainnya berada pada kategori dipertimbangkan dan tidak layak. Validasi menggunakan RapidMiner menunjukkan hasil yang konsisten dengan perhitungan manual.Secara metodologis, penelitian ini memberikan landasan yang kuat dalam penerapan K-Means untuk pengelompokan data bantuan pendidikan. Oleh karena itu, penelitian ini sangat relevan dengan penelitian yang dilakukan, karena sama-sama bertujuan mengelompokkan penerima beasiswa menggunakan metode K-Means guna menghasilkan keputusan yang lebih objektif dan tepat sasaran.

Tabel 2. 7 Penelitian Relevan 7

Nama Penulis	Rahmah, S. A., & Antares, J
Tahun Terbit	2021
Judul Penelitian	Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering
Metode Penelitian	Pengumpulan Data: Data diperoleh dari calon mahasiswa penerima beasiswa yang meliputi nilai UAS, nilai rapor, status rumah, kondisi rumah, dan penghasilan orang tua. Pra-pemrosesan Data: Data dikonversi ke bentuk numerik agar dapat diproses menggunakan algoritma K-Means. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kemiripan karakteristik. Proses dilakukan melalui penentuan jumlah klaster (k), inisialisasi centroid, perhitungan jarak Euclidean, pembaruan centroid, dan iterasi hingga konvergen.

Hasil Penelitian	Hasil klasterisasi membagi calon penerima beasiswa menjadi tiga kategori, yaitu diterima, dipertimbangkan, dan ditolak. Dari 80 data pendaftar, diperoleh sekitar 16% diterima, 61% dipertimbangkan, dan 23% ditolak sebagai penerima beasiswa.
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan penerima beasiswa berdasarkan karakteristik tertentu. Hasil penelitian menunjukkan bahwa metode clustering dapat membantu proses seleksi secara objektif dan berbasis data, sehingga dapat dijadikan referensi dalam penelitian ini yang mengelompokkan penerima beasiswa berdasarkan variabel Semester (SMT) dan UKT.

Penelitian yang dilakukan oleh (Informatika et al., 2021) berjudul *"Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering"* mengusulkan metode pengelompokan calon penerima beasiswa menggunakan algoritma K-Means. Penelitian ini menekankan bahwa proses seleksi beasiswa dapat dilakukan secara lebih objektif dengan memanfaatkan teknik data mining untuk mengelompokkan data berdasarkan kesamaan karakteristik.Melalui penerapan algoritma K-Means, data pendaftar dianalisis menggunakan beberapa atribut yang berkaitan dengan kondisi akademik dan ekonomi, seperti nilai akademik dan penghasilan orang tua. Hasil penelitian menunjukkan bahwa data dapat dikelompokkan ke dalam tiga kategori utama, yaitu diterima, dipertimbangkan, dan ditolak sebagai penerima beasiswa. Mayoritas pendaftar berada pada kategori dipertimbangkan, yang menunjukkan perlunya analisis lebih lanjut dalam proses penentuan penerima.Secara metodologis, penelitian ini memberikan landasan yang kuat mengenai penggunaan algoritma K-Means dalam proses pengambilan keputusan berbasis data. Oleh karena itu, penelitian ini sangat relevan dengan penelitian yang dilakukan, karena sama-sama bertujuan mengelompokkan penerima beasiswa menggunakan metode K-Means, meskipun variabel yang digunakan berbeda, yaitu Semester dan UKT pada penelitian ini.

Tabel 2. 8 Penelitian Relevan 8

Nama Penulis	Alfandi, M. S., & Fatah, Z.
Tahun Terbit	2024
Judul Penelitian	Penerapan Data Mining Menggunakan Metode K-Means Clustering untuk Analisa Penjualan Toko Umama Hijab Kaliwates Jember
Metode Penelitian	Pengumpulan Data: Data yang digunakan berupa data penjualan produk hijab yang mencakup stok awal, stok terjual, dan stok akhir. Data diperoleh dari catatan transaksi toko. Pra-pemrosesan Data: Data dilakukan proses cleaning dan transformasi agar menjadi numerik dan siap diolah. Tahapan penelitian mengikuti proses Knowledge Discovery in Database (KDD) yang meliputi seleksi data, preprocessing, transformasi, data mining, dan evaluasi. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan data penjualan menjadi beberapa klaster berdasarkan kemiripan karakteristik penjualan produk. Proses clustering dilakukan menggunakan software RapidMiner dengan menentukan jumlah klaster K = 3.
Hasil Penelitian	Hasil penelitian menunjukkan bahwa data penjualan produk hijab berhasil dikelompokkan menjadi tiga kategori, yaitu produk sangat laris, laris, dan kurang laris. Dari total 100 data penjualan, diperoleh 24 produk sangat laris, 59 produk laris, dan 17 produk kurang laris. Metode K-Means terbukti mampu mengidentifikasi pola penjualan dan membantu pengambilan keputusan terkait strategi pemasaran dan pengelolaan stok.

Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan metode K-Means Clustering untuk mengelompokkan data berdasarkan karakteristik tertentu guna mendukung pengambilan keputusan. Perbedaannya terletak pada objek penelitian, di mana penelitian tersebut berfokus pada penjualan produk, sedangkan penelitian ini berfokus pada pengelompokan penerima beasiswa berdasarkan variabel Semester dan UKT. Namun, secara metodologis, penelitian tersebut memberikan referensi yang kuat mengenai penerapan K-Means dalam analisis data numerik

Penelitian yang dilakukan oleh Alfandi et al. (2024) berjudul *"Penerapan Data Mining Menggunakan Metode K-Means Clustering untuk Analisa Penjualan Toko Umama Hijab Kaliwates Jember"* membahas penerapan teknik data mining untuk mengelompokkan produk berdasarkan tingkat penjualan. Penelitian ini menggunakan algoritma K-Means dengan tahapan KDD yang meliputi seleksi data, preprocessing, transformasi, data mining, dan evaluasi.Hasil penelitian menunjukkan bahwa metode K-Means mampu mengelompokkan produk hijab menjadi tiga kategori, yaitu sangat laris, laris, dan kurang laris, berdasarkan data stok dan penjualan. Informasi ini dapat digunakan sebagai dasar dalam pengambilan keputusan terkait strategi pemasaran dan pengelolaan persediaan produk.Secara metodologis, penelitian ini relevan dengan penelitian yang dilakukan karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan data numerik guna memperoleh informasi yang bermanfaat. Perbedaannya terletak pada objek penelitian, di mana penelitian ini menerapkan K-Means pada data penerima beasiswa berdasarkan variabel Semester dan UKT. Oleh karena itu, penelitian tersebut dapat dijadikan referensi dalam penerapan metode clustering untuk analisis data pada bidang pendidikan.

Tabel 2. 9 Penelitian Relevan 9

Nama Penulis	Silvana Nazuah, Shofa Shofia Hilabi, Agustia Hananto, Baenil Huda, & Tukino

Tahun Terbit	2023
Judul Penelitian	Seleksi Penerimaan Beasiswa Dengan Metode K-Means Clustering Menggunakan Orange
Metode Penelitian	Pengumpulan Data: Data diperoleh melalui wawancara dan survei kepada pihak sekolah, serta menggunakan data nilai siswa yang telah direkap dari berbagai program keahlian. Dataset awal berjumlah 855 data dan diambil sampel sebanyak 120 data. Pra-pemrosesan Data: Dilakukan proses data cleaning untuk menghapus data tidak valid atau tidak relevan agar meningkatkan kualitas analisis. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan data menjadi beberapa klaster menggunakan tools Orange Data Mining. Penentuan jumlah klaster dilakukan secara acak sebanyak tiga klaster dengan kategori berhak menerima, tidak berhak menerima, dan dipertimbangkan.
Hasil Penelitian	Hasil penelitian menunjukkan bahwa data terbagi menjadi tiga klaster, yaitu klaster 1 sebanyak 58 siswa (48%) yang berhak menerima beasiswa, klaster 2 sebanyak 39 siswa (33%) yang tidak berhak menerima, dan klaster 3 sebanyak 23 siswa (19%) yang dipertimbangkan.
Keterkaitan Penelitian	Penelitian ini relevan karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan data penerima beasiswa berdasarkan beberapa atribut. Metode ini dapat menjadi referensi dalam menentukan kelompok mahasiswa berdasarkan karakteristik akademik dan ekonomi pada penelitian ini.

Penelitian yang dilakukan oleh Nazuah et al. (2023) berjudul *"Seleksi Penerimaan Beasiswa Dengan Metode K-Means Clustering Menggunakan Orange"*

membahas penerapan algoritma K-Means untuk membantu proses penentuan siswa yang layak menerima beasiswa. Penelitian ini menggunakan data nilai siswa, jarak rumah, penghasilan orang tua, dan kondisi orang tua sebagai atribut utama dalam proses klasterisasi.Melalui tahapan data cleaning dan analisis menggunakan tools Orange Data Mining, data dikelompokkan menjadi tiga klaster, yaitu berhak menerima, tidak berhak menerima, dan dipertimbangkan. Hasil penelitian menunjukkan bahwa sebagian besar siswa berada pada klaster yang berhak menerima beasiswa, sehingga metode K-Means terbukti efektif dalam membantu proses seleksi yang sebelumnya dilakukan secara manual.Secara metodologis, penelitian ini memiliki kesamaan dengan penelitian yang dilakukan penulis karena sama-sama menggunakan algoritma K-Means untuk mengelompokkan penerima beasiswa berdasarkan karakteristik tertentu. Oleh karena itu, penelitian ini dapat dijadikan referensi dalam proses pengelompokan data mahasiswa penerima beasiswa di Kabupaten Konawe Selatan.

Tabel 2. 10 Penelitian Relevan 10

Nama Penulis	Syah, F. A., Hasugian, S. P., Adafmi, M. V., Sibarani, A. D., & Lisnawita
Tahun Terbit	2024
Judul Penelitian	The Implementation of the K-Means Clustering Algorithm for Awarding Scholarships to Outstanding Students
Metode Penelitian	Pengumpulan Data: Data diperoleh dari catatan akademik mahasiswa, pengalaman kepemimpinan, nilai seleksi, serta prestasi akademik dan non-akademik. Pra-pemrosesan Data: Data dibersihkan dari nilai tidak lengkap, dilakukan normalisasi, penanganan missing value, dan transformasi data agar siap dianalisis. Metode Analisis: Algoritma K-Means digunakan untuk mengelompokkan mahasiswa berdasarkan tingkat kemiripan karakteristik akademik dan non-akademik menggunakan jarak Euclidean.

Hasil Penelitian	Penelitian menghasilkan dua klaster utama dari 200 data mahasiswa. Klaster 0 berisi 43 mahasiswa yang memenuhi kriteria penerima beasiswa, sedangkan Klaster 1 berisi 157 mahasiswa yang tidak memenuhi kriteria. Hasil ini menunjukkan bahwa metode K-Means mampu membantu proses seleksi beasiswa secara objektif dan berbasis data.
Keterkaitan Penelitian	Penelitian ini sangat relevan karena sama-sama menggunakan algoritma K-Means untuk pengelompokan penerima beasiswa berdasarkan variabel tertentu. Metodologi pengolahan data, tahap pra-pemrosesan, serta interpretasi hasil klaster dapat dijadikan referensi dalam penelitian ini yang mengelompokkan penerima beasiswa berdasarkan variabel Semester dan UKT.

Penelitian yang dilakukan oleh Syah et al. (2024) berjudul *"The Implementation of the K-Means Clustering Algorithm for Awarding Scholarships to Outstanding Students"* mengusulkan penggunaan metode K-Means untuk membantu proses seleksi penerima beasiswa secara objektif dan efisien. Penelitian ini memanfaatkan data akademik dan non-akademik mahasiswa, seperti nilai GPA, hasil seleksi, serta pengalaman kepemimpinan, untuk membentuk kelompok mahasiswa berdasarkan tingkat kemiripan karakteristik.Melalui proses pra-pemrosesan data dan penerapan algoritma K-Means, penelitian ini berhasil mengelompokkan 200 data mahasiswa menjadi dua klaster, yaitu klaster mahasiswa yang memenuhi syarat dan yang tidak memenuhi syarat sebagai penerima beasiswa. Hasilnya menunjukkan bahwa metode K-Means mampu mengidentifikasi pola dan karakteristik mahasiswa yang layak menerima bantuan pendidikan secara lebih objektif dibandingkan metode konvensional.Secara metodologis, penelitian ini memberikan dasar yang kuat dalam penggunaan K-Means untuk pengelompokan data beasiswa. Oleh karena itu, penelitian ini sangat relevan dengan penelitian yang dilakukan, karena sama-sama bertujuan mengelompokkan penerima beasiswa menggunakan pendekatan data mining, meskipun variabel yang digunakan dalam penelitian ini adalah Semester dan UKT.

2.1.1 Matrix Penelitian

Tabel 2. 11 Matrix Penelitian

No	Nama Penulis (Tahun)	Judul Penelitian	Sumber Data	Klasifikasi Masalah	Arsitektur / Metode	Kesimpulan
1	Sudarsono & Lestari (2021)	Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means	Data mahasiswa penerima beasiswa yayasan (IPK, prestasi non akademik, disiplin, penghasilan orang tua)	Pengelompokan penerima beasiswa berdasarkan kriteria akademik dan ekonomi	K-Means Clustering	Metode K-Means mampu mengelompokkan mahasiswa penerima beasiswa ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Hasil clustering membantu pihak yayasan dalam menentukan pembagian beasiswa secara lebih

						tepat dan merata.
2	Sindrawati, Syaripudin, & Raharja (2024)	Penerapan Algoritma K-Means Clustering pada Data Nilai Siswa untuk Menentukan Kelompok Penerima Beasiswa	Data nilai siswa kelas XII (nilai matematika, bahasa Indonesia, bahasa Inggris, jurusan)	Penentuan kelompok penerima beasiswa berdasarkan prestasi akademik	CRISP-DM + K-Means Clustering	Metode K-Means mampu mengelompokkan siswa menjadi beberapa kelompok berdasarkan kemiripan nilai sehingga membantu sekolah menentukan penerima beasiswa secara objektif
3	Abu Salam, Diyan Adiatma, Junta Zeniarja (2020)	Implementasi Algoritma K-Means dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa	Data pendaftar beasiswa PPA tahun 2016 sebanyak 441	Rekomendasi penerima beasiswa	Data MininCluster ing K-Means	Metode K-Means mampu mengelompokkan data pendaftar menjadi kelompok direkomendasikan dan

		PPA di UDINUS	mahasiswa			tidak direkomendasikan menerima beasiswa secara efektif.
4	Fadlil, A., Riadi, I., & Mulyana, Y. (2022)	Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan	Data pendaftar program KIP-Kuliah sebanyak 191 data mahasiswa	Penentuan kelompok calon penerima bantuan biaya pendidikan agar tepat sasaran	Data Mining K-Means Clustering dengan jarak Euclidean	Metode K-Means menghasilkan 3 klaster: C0 = 10909 data (57,1%), C1 = 52 data (27,2%), C2 = 30 data (15,7%). Klaster C2 direkomendasikan sebagai kelompok paling layak menerima bantuan.
5	Surya Darma, Sarjon Defit, Dedy Hartama, Wendi	Penerapan Metode K-Means Dalam Pengelompokan Jumlah	Data jumlah wisatawan asing Indonesia tahun 2017–	Pengelompokan jumlah kunjungan wisatawan asing	Data Mining K-Means Clustering	Metode K-Means berhasil mengelompokkan data wisatawan menjadi

	Robiansyah, Fahmi Firzada (2020)	Wisatawan Asing Di Indonesia	2019 dari BPS	berdasarkan bulan		cluster tinggi dan rendah, di mana bulan Juni merupakan periode kunjungan tertinggi, sedangkan bulan lainnya termasuk kategori rendah
6	Fadlil, Riadi, & Mulyana (2022)	Penerapan Algoritma K-Means pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan	Data pendaftar KIP-Kuliah sebanyak 191 mahasiswa	Penentuan kelompok penerima bantuan pendidikan agar tepat sasaran	K-Means Clustering (K = 3), Euclidean Distance, RapidMiner	Metode K-Means mampu mengelompokkan data menjadi tiga klaster dengan distribusi C0 = 109 data, C1 = 52 data, dan C2 = 30 data. Klaster C2 direkomendasikan sebagai kelompok layak

						menerima bantuan, C0 sebagai dipertimbangkan, dan C1 sebagai tidak layak.
7	Rahmah & Antares (2021)	Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering	Data calon mahasiswa pendaftar beasiswa yayasan (nilai UAS, nilai rapor, kondisi ekonomi, dll.)	Seleksi dan rekomendasi penerima beasiswa	K-Means Clustering	Metode K-Means mampu mengelompokkan calon penerima beasiswa menjadi kategori diterima, dipertimbangkan, dan ditolak sehingga membantu proses seleksi secara objektif dan terstruktur
8	Alfandi & Fatah (2024)	Penerapan Data Mining Menggunak	Data penjualan produk hijab	Analisis penjualan dan pengelomp	Data Mining (KDD) + K-Means	Metode K-Means mampu mengelompo

		an Metode K-Means Clustering untuk Analisa Penjualan Toko Umama Hijab Kaliwates Jember	(stok awal, stok terjual, stok akhir) sebanyak 100 data	okan produk berdasarkan tingkat permintaan	Clustering (RapidMiner)	kkan produk hijab menjadi tiga kategori yaitu sangat laris, laris, dan kurang laris sehingga dapat mendukung strategi pemasaran toko
9	Nazuah et al. (2023)	Seleksi Penerimaan Beasiswa dengan Metode K-Means Clustering Menggunakan Orange	Data siswa SMKN 1 Karawang (120 sampel dari 855 data)	Penentuan kelayakan penerima beasiswa	K-Means Clustering menggunakan tools Orange Data Mining	Metode K-Means mampu mengelompokkan siswa menjadi 3 kategori: layak menerima, tidak layak, dan dipertimbangkan. Cluster 1 berisi 58 siswa (48%), Cluster 2 berisi 39 siswa (33%), dan Cluster 3

						berisi 23 siswa (19%).
10	Syah, F. A., Hasugian, S. P., Adafmi, M. V., Sibarani, A. D., & Lisnawita (2024)	The Implementation of the K-Means Clustering Algorithm for Awarding Scholarships to Outstanding Students	Data akademik mahasiswa, nilai IPK, prestasi akademik dan non-akademik, pengalaman organisasi	Seleksi penerima beasiswa berprestasi	K-Means Clustering + RapidMiner	Algoritma K-Means mampu mengelompokkan mahasiswa menjadi dua klaster, yaitu layak menerima beasiswa dan tidak layak, sehingga membantu pengambilan keputusan secara objektif dan berbasis data.

2.2 Landasan Teori

Pada bab ini berisikan mengenai semua teori yang akan digunakan sebagai acuan dalam melakukan penelitian. Berikut merupakan beberapa teori yang digunakan :

2.2.1 Beasiswa

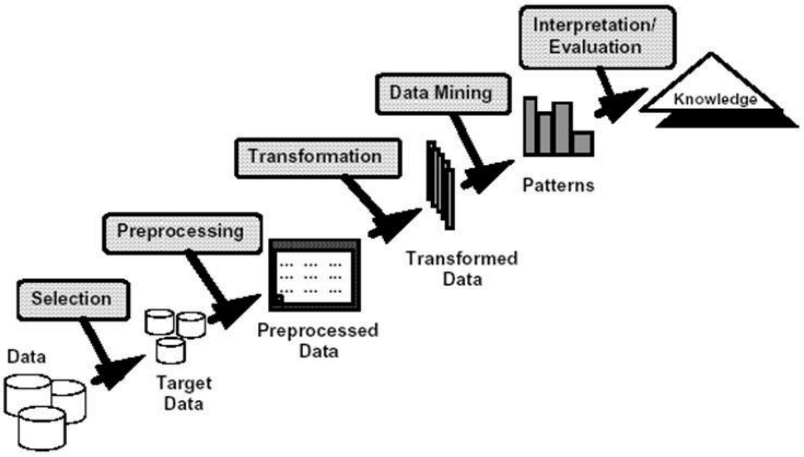
Beasiswa merupakan bantuan biaya pendidikan yang diberikan oleh pemerintah maupun pihak swasta kepada peserta didik yang sedang atau akan menempuh pendidikan, dengan tujuan membantu keberlangsungan studi sekaligus mendorong peningkatan prestasi akademik maupun nonakademik. Pemberian beasiswa biasanya ditujukan kepada siswa atau mahasiswa yang memiliki keterbatasan ekonomi atau memiliki prestasi tertentu sehingga layak memperoleh dukungan finansial untuk melanjutkan pendidikan. Selain sebagai bentuk bantuan, beasiswa juga berfungsi sebagai penghargaan atas capaian akademik serta motivasi agar penerima dapat mempertahankan atau meningkatkan kinerjanya selama masa studi. Dalam praktiknya, proses penentuan penerima beasiswa harus dilakukan secara selektif dan berdasarkan kriteria yang jelas agar bantuan dapat tepat sasaran, terutama ketika jumlah pendaftar melebihi kuota yang tersedia (Sulistiani & Utami, 2018)

2.2.2 Data Mining

Data mining merupakan suatu proses pengolahan data yang bertujuan untuk menemukan informasi penting dari sekumpulan data berukuran besar.Data mining diartikan sebagai proses pengumpulan dan pengolahan data untuk mengekstrak informasi yang dapat mendukung pengambilan keputusan.Dengan kata lain, data mining tidak hanya berfungsi sebagai proses penyimpanan data, tetapi lebih menekankan pada upaya menggali pola atau hubungan yang tersembunyi di dalam data tersebut.Teknik data mining digunakan untuk menelusuri basis data berukuran besar guna menemukan pola baru yang bermanfaat.Namun demikian, tidak semua aktivitas pencarian data dapat dikategorikan sebagai data mining, karena proses ini memerlukan metode analisis tertentu yang sistematis. Oleh sebab itu, data mining sering dikaitkan dengan teknik statistik, komputasi, dan algoritma tertentu yang mampu mengidentifikasi pola secara otomatis. (Alfandi et al., 2024).

2.2.3 Knowledge discovery in database (KDD)

Knowledge discovery in database (KDD) adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (pattem) dalam data, pola yang sudah ditemukan bersifat sah dan baru sehingga dapat bermanfaat dan dapat dimengerti. Proses dalam KDD terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap pre- processing, transformasi, data mining dan tahap evaluasi. Alfandi et al. (2024), Tahap seleksi dilakukan untuk menargetkan data yang digunakan untuk penelitian, tahap pre-procesing dapat dilakukan integrasi data atau penggabungan data serta dilakukan cleaning data, yaitu dengan menhilangkan noise, data redundan, inkonsistensi data, serta data yang tidak relavan, tahap transformasi adalah penggabungan data dan penyesuaian format data agar dapat diproses pada tahap data mining, tahap data mining dilakukan menggunakan algoritma yang cocok untuk permasalahan dalam data, serta tahap evaluasi yang digunakan untuk pengujian dalam data



Gambar 2. 1 *Knowledge discovery in database*

1. Data Selection

Menciptakan himpunan data target, pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing / Cleaning

Pre-processing dan cleaning data merupakan operasi dasar yang dilakukan seperti penghapusan noise. Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak. Data bisa diperkaya dengan data atau informasi ekternal yang relevan.

Normalisasi data merupakan salah satu tahapan preprocessing yang bertujuan untuk menyamakan perbedaan skala antar atribut agar data tidak timpang dan dapat dianalisis secara optimal. Proses ini mentransformasikan nilai data ke dalam rentang tertentu sehingga perbandingan antar fitur menjadi lebih seimbang dan tidak didominasi oleh atribut dengan nilai besar. Salah satu metode normalisasi yang umum digunakan adalah Min-Max Normalization, yaitu teknik yang memanfaatkan nilai minimum dan maksimum suatu atribut untuk mengubah data secara linier ke rentang baru, biasanya antara 0 sampai 1, tanpa mengubah hubungan antar nilai pada data. Dengan normalisasi, data yang kompleks dapat diproses dengan lebih baik dan menghasilkan performa analisis yang lebih akurat. Rumus Min-Max Normalization adalah:

$$x' = \frac{\text{minR} + (x - x_{min})(\text{maxR} - \text{minR})}{x_{max} - x_{min}}$$

di mana x'adalah nilai hasil normalisasi, $x$adalah data asli, x_{max}dan x_{min}masing-masing adalah nilai maksimum dan minimum atribut, sedangkan minR dan maxR merupakan batas rentang baru yang diinginkan (Dwididanti & Anggoro, 2022).

3. Tranformation

Merupakan proses integrasi pada data yang telah dipilih, sehingga data sesuai untuk proses data mining. Merupakan proses yang sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data Mining

Pemilihan tugas data mining merupakan pemilihan goal dari proses KDD misalnya karakterisasi, klasifikasi, regresi, clustering, asosiasi, dan lain-lain. Pemilihan tugas data mining merupakan pemilihan goal dari proses KDD misalnya karakterisasi, klasifikasi, regresi, clustering, asosiasi, dan lain-lain. Pemilihan teknik, metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation/ Evaluation

Yaitu penerjemahan pola-pola yang dihasilkan dari data mining. Pola informasi yang dihasilkan perlu ditampilkan dalam bentuk yang mudah dimengerti. Tahap ini melakukan pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya

2.2.4 *Clustering*

Clustering merupakan salah satu metode data mining yang digunakan dalam menganalisa data, dengan proses pengelompokkan data berdasarkan karakteristik atau kesamana sifat yang mirip ke dalam sebuah cluster atau data dengan tingkat kesamaan yang rendah terhadap cluster lain dengan memaksimalkan kemiripan antara objek dalam sebuah cluster dan meminimalkan kemiripan antar cluster (Irhamni & Damayanti, 2014). *Clustering* atau klasifikasi yang digunakan untuk membagi rangkaian data untuk menjadi beberapa kelompok berdasarkan dari kesamaan-kesamaan yang telah ditentukan sebelumnya. Kelompok-kelompok cluster yang dihasilkan dari pengelompokkan unsurunsur yang lebh kecil berdasakan adanya kemiripan satu sama lain.

Tujuan utama dari clustering yaitu mengidentifikasi struktur dalm data, berupa kelompok-kelompok yang tidak diketahui sebelumnya, atau objek yang ada pada data tersebut. Clustering menggunakan metode pengelompokkan data yang tidak terstruktur, seperti k-means, hierarchical clustering, dan density-based. Hasil dari clustering adalah kelompok-kelompok data yang memiliki kemiripan tinggi dan perbedaan yang rendah antar kelompok. Dan Clustering juga dikenal sebagai data segmentasi karena clustering

mempartisi banyak dataset ke dalam banyak kelompok berdasarkan kesamaannya (Kamti et al., 2024) .

2.2.5 K-Means

K-Means adalah salah satu algoritma yang sederhana dan sering digunakan dalam teknik pengelompokan karena proses perhitungannya efisien dan tidak membutuhkan banyak parameter. K-Means menggunakan k kelompok yang sudah ditentukan (k kelompok pertama sebagai pusat kelompok) dan terus-menerus melakukan proses perhitungan titik tengah (min) hingga suatu fungsi kriteria tercapai (kelompok tidak berubah lagi). Dalam teknik pengelompokan, proses untuk membedakan kelompok satu sama lain menggunakan satu algoritma yang disebut fungsi jarak, yaitu tahap persamaan atau perbedaan (Kusuma & Agani, 2015).

Langkah-langkah algoritma K-Means adalah sebagai berikut:

1. Menentukan jumlah cluster (k).
2. Menginisialisasi centroid awal secara acak.
3. Menghitung jarak setiap data terhadap masing-masing centroid.
4. Mengelompokkan data ke centroid dengan jarak terdekat.
5. Menghitung ulang centroid sebagai rata-rata data dalam cluster.

$$C_K = \left(\frac{1}{n_k}\right)\sum d_i$$

Dimana *k n* adalah jumlah dokumen dalam *cluster* k dan *i d* adalah dokumen dalam *cluster* k.

6. Mengulangi proses hingga centroid tidak berubah lagi.

Adapun karakteristik dari algoritma K-Means salah salah satu adalah sangat sensitif dalam menentukan titik pusat.Awal klaster karena K-Means membangkitkan titik pusat klaster awal secara random. Pada saat pembangkitan awal Titik pusat yang acak tersebut mendekati solusi akhir untuk pusat.Klaster, K-Means mempunyai posibilitas yang tinggi untuk menemukan titik pusat klaster yang tepat. Sebaliknya,

jika Titik pusat tersebut masih jauh dari solusi akhir pusat klaster.Maka besar kemungkinan ini menyebabkan hasil pengklasteran yang tidak tepat.Akibatnya K-Means tidak menjamin hasil pengklasteran yang unik. Inilah yang menyebabkan metode K-Means sulit untuk mencapai optimum global, akan tetapi hanya minimum lokal. Selain itu, algoritma K-Means hanya bisa digunakan untuk data yang bersifat numerik.Atributnya bernilai numeric.Pengukuran persamaan atau jarak merupakan tugas yang penting dalam proses analisis kelompok di mana hampir semua teknik pengelompokan menggunakan pengiraan matriks jarak (atau perbedaan). Algoritma K-Means juga menggunakan kaedah pengiraan ini bagi menjelaskan lagi persamaan bagi setiap corak kelompok. Matriks Jarak Euclidean merupakan salah satu jenis matriks jarak yang sering digunakan.berikut adalah rumus digunakan di dalam algoritma K-Means.

$$d(x, y)\sum^{n}_{i=1} = (xi - yi)^2$$

$d(x, y)$= jarak antara data x dan y

x_i= nilai atribut ke-i dari data x

y_i= nilai atribut ke-i dari data y

n = jumlah atribut

2.2.6 Elbow

Elbow dikenal sebagai salah satu teknik evaluasi yang banyak digunakan dalam analisis *clustering*, khususnya pada algoritma K-Means, untuk memperkirakan jumlah cluster (*k*) yang paling sesuai. Pendekatan ini memanfaatkan nilai kesalahan pengelompokan yang dihitung melalui Within Cluster Sum of Squares (WSS) atau sering disebut Sum of Squared Error (SSE). Ide dasarnya adalah membandingkan perubahan nilai SSE terhadap beberapa variasi jumlah cluster, kemudian mengidentifikasi titik perubahan terbesar pada grafik yang membentuk pola menyerupai siku (*elbow*).

Secara teoritis, ketika jumlah cluster bertambah, jarak antara data dan pusat cluster (*centroid*) akan semakin kecil. Kondisi tersebut menyebabkan nilai SSE terus menurun. Namun demikian, penurunan ini tidak selalu diikuti peningkatan kualitas

pengelompokan yang berarti. Setelah melewati batas tertentu, penambahan cluster hanya memberikan penurunan error yang relatif kecil. Oleh sebab itu, titik saat kurva mulai melandai dianggap sebagai jumlah cluster yang paling optimal karena telah mampu merepresentasikan struktur data tanpa menambah kompleksitas model secara berlebihan.

Menurut penelitian Maori (2023), metode Elbow dimanfaatkan sebagai teknik optimasi pada K-Means untuk menentukan jumlah cluster yang efisien dengan mempertimbangkan keseimbangan antara nilai SSE yang rendah dan kesederhanaan model. Visualisasi grafik memudahkan peneliti dalam melihat tren penurunan error pada setiap penambahan nilai *k*, sehingga keputusan pemilihan cluster dapat dilakukan secara lebih objektif dan terukur (Maori, 2023).

Secara matematis, SSE dihitung dengan menjumlahkan kuadrat selisih antara setiap data dan centroid cluster tempat data tersebut berada. Persamaan yang digunakan adalah sebagai berikut:

$$SSE = \Sigma_{k=1}^{K} \Sigma_{x_i \in C_k} (x_i - \varphi_k)^2$$

dengan:

C_k menyatakan himpunan data pada cluster ke-k,

x_i adalah data ke-i,

ϕ_k merupakan centroid atau nilai rata-rata cluster ke-k,

Ka dalah jumlah cluster.

Nilai SSE yang semakin kecil menunjukkan bahwa anggota dalam satu cluster memiliki tingkat kemiripan yang tinggi (homogen), sehingga hasil pengelompokan dinilai semakin baik. Dalam implementasinya, nilai *k* biasanya diuji secara bertahap, misalnya dari 2 hingga 6 atau lebih. Setiap nilai *k* dihitung SSE-nya, kemudian seluruh hasil diplot dalam bentuk grafik. Titik lengkungan paling tajam pada grafik tersebut menjadi indikasi jumlah cluster yang direkomendasikan.

Hasil penelitian pada jurnal yang sama memperlihatkan bahwa penurunan SSE paling signifikan terjadi pada *k = 3*. Temuan ini menunjukkan bahwa tiga cluster telah

cukup mewakili karakteristik data secara optimal tanpa meningkatkan kompleksitas model secara tidak perlu (Maori, 2023).

Berdasarkan uraian tersebut, metode Elbow memiliki beberapa keunggulan, antara lain prosedurnya sederhana, mudah diimplementasikan, serta memberikan visualisasi yang intuitif sehingga memudahkan interpretasi. Meskipun demikian, metode ini memiliki keterbatasan ketika kurva tidak menunjukkan titik siku yang jelas. Dalam kondisi demikian, diperlukan metode validasi tambahan, seperti Davies-Bouldin Index (DBI), guna memperkuat evaluasi hasil *clustering*.

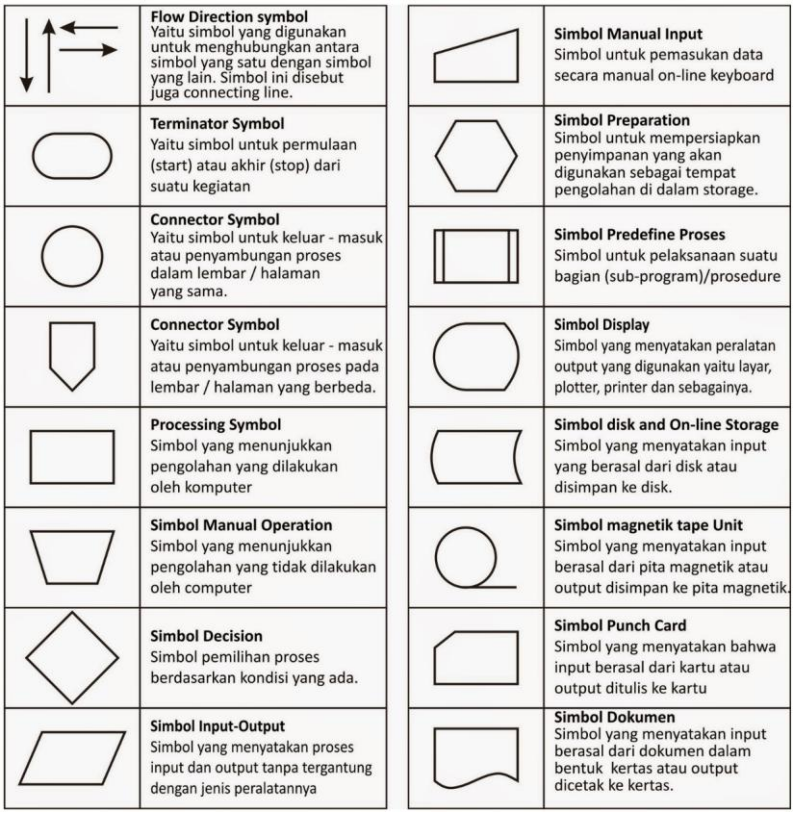
Dengan demikian, metode Elbow dapat dijadikan sebagai pendekatan awal yang efektif dalam menentukan jumlah cluster optimal pada algoritma K-Means, sehingga proses pengelompokan data menjadi lebih akurat, efisien, dan mudah dipahami.

2.2.7 Python

Python merupakan bahasa pemrograman tingkat tinggi yang dikembangkan oleh Guido van Rossum pada awal tahun 1990 di Amsterdam, Belanda, dan dikenal luas karena sintaksnya yang sederhana, mudah dipahami, serta mendukung pemrograman berorientasi objek sehingga cocok digunakan oleh pemula maupun pengembang profesional. Bahasa ini dirancang untuk meningkatkan efisiensi dan produktivitas dalam pengembangan perangkat lunak melalui struktur kode yang jelas dan ekspresif. Selain itu, Python memiliki pustaka standar yang sangat lengkap serta didukung oleh berbagai library tambahan seperti NumPy, Pandas, dan TensorFlow yang banyak dimanfaatkan dalam analisis data, kecerdasan buatan, dan pemrosesan informasi. Python juga digunakan secara luas dalam pengembangan aplikasi web, pemrosesan bahasa alami, otomatisasi sistem, hingga penelitian ilmiah karena fleksibilitas dan kemudahan implementasinya. Popularitas Python semakin meningkat karena kemampuannya untuk diterapkan pada berbagai bidang teknologi serta dukungan komunitas yang besar, sehingga menjadikannya salah satu bahasa pemrograman yang penting dalam pengembangan aplikasi modern (Alfarizi et al., 2023).

2.2.8 Flowchart

Flowchart atau diagram alir merupakan representasi visual yang digunakan untuk menggambarkan urutan langkah atau algoritma suatu proses secara sistematis dalam suatu sistem. Diagram ini menampilkan aliran kerja dari awal hingga akhir menggunakan simbol-simbol tertentu yang dihubungkan oleh garis, di mana setiap simbol memiliki arti khusus seperti proses, keputusan, input, dan output sehingga memudahkan pemahaman terhadap logika sistem sebelum diimplementasikan ke dalam program komputer. Dalam pengembangan sistem, flowchart berfungsi sebagai alat perancangan logis sekaligus dokumentasi yang membantu analis sistem menjelaskan rancangan kepada programmer atau pihak terkait agar alur kerja sistem dapat dipahami dengan jelas serta meminimalkan kesalahan sejak tahap perencanaan. Selain itu, flowchart juga digunakan untuk merancang proyek, mengelola alur kerja, memodelkan proses bisnis, dan menyajikan informasi secara visual sehingga lebih efisien dibandingkan penjelasan naratif. Dalam penyusunannya, flowchart menggunakan simbol-simbol standar yang umumnya dikelompokkan menjadi simbol arus, simbol proses, dan simbol input/output yang masing-masing menunjukkan arah aliran, kegiatan yang dilakukan, serta data yang masuk dan keluar dari sistem, sehingga diagram dapat dibaca secara konsisten oleh semua pengguna (Rosaly & Prasetyo, 2019).



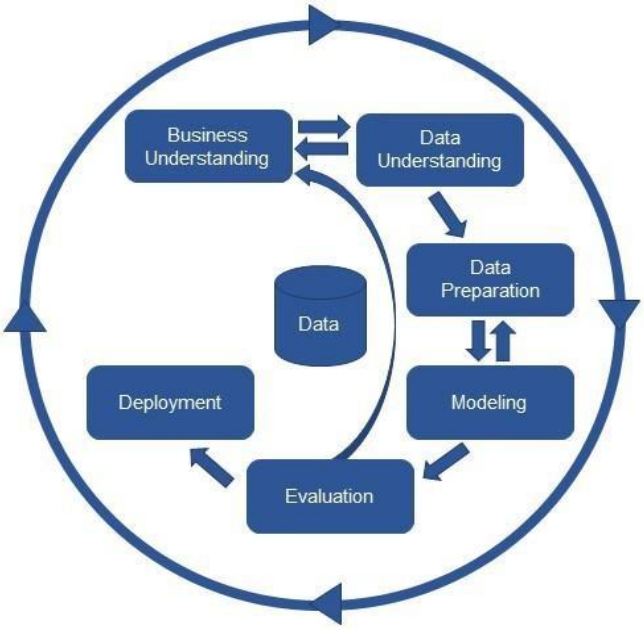
Gambar 2. 2 Flowchart

2.2.9 *Cross-Industry Standard Process for Data Mining* (CRISP-DM)

Cross-Industry Standard Process for Data Mining (CRISP-DM) merupakan metodologi standar yang digunakan dalam pengembangan proyek data mining. Metodologi ini pertama kali diperkenalkan pada tahun 1996 oleh konsorsium industri yang terdiri dari SPSS, DaimlerChrysler, dan NCR. CRISP-DM dirancang sebagai kerangka kerja yang sistematis dan terstruktur untuk membantu penyelesaian permasalahan data mining, baik dalam konteks bisnis maupun penelitian akademik.

CRISP-DM menggambarkan siklus hidup proyek data mining yang terdiri atas enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Keenam tahapan tersebut bersifat iteratif dan saling berkaitan, sehingga memungkinkan peneliti untuk kembali ke tahap

sebelumnya apabila hasil yang diperoleh belum sepenuhnya memenuhi tujuan penelitian yang telah ditetapkan.



Gambar 2. 3 CRISP-DM

Gambar diatas menunjukkan tahapan-tahapan dalam metodologi CRISP-DM yang diawali dengan tahap Business Understanding. Tahap ini berfokus pada pemahaman terhadap permasalahan yang dihadapi serta penetapan tujuan penelitian. Pada tahap ini juga ditentukan ruang lingkup permasalahan yang akan diselesaikan menggunakan pendekatan data mining (Rifai et al., 2019).

Tahap Data Understanding bertujuan untuk memperoleh pemahaman yang mendalam terhadap data yang digunakan dalam penelitian. Aktivitas yang dilakukan pada tahap ini meliputi proses pengumpulan data, deskripsi karakteristik data, eksplorasi data, serta identifikasi kualitas data sebelum dilakukan proses pengolahan lebih lanjut (Rifai et al., 2019).

Selanjutnya, tahap Data Preparation merupakan proses pengolahan data yang dilakukan sebelum pemodelan. Pada tahap ini dilakukan seleksi data, pembersihan data (data cleaning), transformasi data, serta pembentukan dataset akhir yang siap digunakan

dalam proses analisis. Tahap ini memiliki peran yang sangat penting karena kualitas data yang baik akan sangat memengaruhi hasil model yang dihasilkan (Rifai et al., 2019).

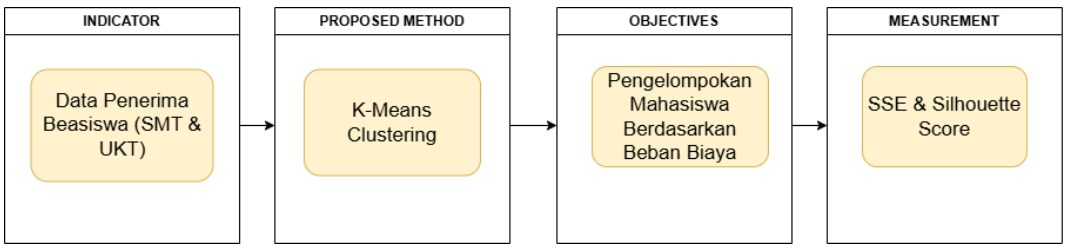
Tahap Modeling merupakan fase penerapan teknik atau algoritma data mining yang dipilih sesuai dengan tujuan penelitian. Pada tahap ini dilakukan pembangunan model serta pengaturan parameter agar model yang dihasilkan dapat bekerja secara optimal (Rifai et al., 2019).

Tahap Evaluation dilakukan untuk menilai kinerja model yang telah dibangun. Evaluasi bertujuan untuk memastikan bahwa model yang dihasilkan telah sesuai dengan tujuan penelitian serta memenuhi kriteria kinerja yang telah ditetapkan (Rifai et al., 2019).

Tahap terakhir dalam metodologi CRISP-DM adalah Deployment, yaitu proses penerapan hasil data mining ke dalam bentuk laporan, sistem, atau aplikasi yang dapat dimanfaatkan oleh pengguna. Pada tahap ini, pengetahuan yang dihasilkan dari proses data mining disajikan sebagai dasar dalam mendukung pengambilan keputusan (Rifai et al., 2019).

2.3 Kerangka Pemikiran

Pada penelitian ini dikembangkan sebuah kerangka pemikiran yang digambarkan secara terurut mulai dari proses analisis data hingga proses menarik kesimpulan dari hasil penelitian. Kerangka pemikiran ini menguraikan langkah-langkah dan pertimbangan yang diperlukan untuk memastikan hasil yang akurat sebagai berikut :



Gambar 2. 4 Kerangka Pemikiran

1. Indikator

Indikator dalam penelitian ini adalah data penerima beasiswa di Kabupaten Konawe Selatan yang terdiri dari variabel Semester (SMT) dan Uang Kuliah Tunggal (UKT). Kedua variabel tersebut dipilih karena mewakili dua aspek penting, yaitu perkembangan akademik mahasiswa dan besaran biaya pendidikan yang ditanggung. Data yang tersedia masih berupa data administratif tanpa pengelompokan tertentu, sehingga diperlukan proses analisis untuk menemukan pola atau segmentasi berdasarkan kemiripan karakteristik mahasiswa

2. Proposed Method

Metode yang digunakan dalam penelitian ini adalah K-Means Clustering. Metode ini dipilih karena mampu mengelompokkan data berdasarkan kedekatan jarak antar nilai variabel. Dalam penerapannya, data Semester dan UKT terlebih dahulu dinormalisasi agar memiliki skala yang seimbang sebelum dilakukan proses pengelompokan. K-Means bekerja dengan menentukan centroid awal, menghitung jarak antar data dan centroid menggunakan Euclidean Distance, serta memperbarui centroid hingga mencapai kondisi konvergen. Melalui tahapan tersebut diperoleh kelompok mahasiswa dengan karakteristik yang relatif homogen.

3. Objectives

Tujuan utama dari penelitian ini adalah menghasilkan pengelompokan mahasiswa penerima beasiswa berdasarkan tingkat beban biaya pendidikan. Dengan memanfaatkan nilai Semester dan UKT, penelitian ini bertujuan untuk mengidentifikasi pola distribusi mahasiswa serta membentuk segmentasi yang dapat memberikan gambaran kondisi akademik dan finansial secara lebih terstruktur. Hasil pengelompokan diharapkan dapat mendukung proses evaluasi dan pengambilan keputusan dalam pengelolaan program beasiswa daerah.

4. Measurement

Pengukuran kualitas hasil clustering dilakukan menggunakan nilai Sum of Squared Error (SSE) dan Silhouette Score. SSE digunakan untuk mengetahui tingkat

kedekatan data terhadap centroid dalam setiap cluster, sehingga dapat menunjukkan tingkat kekompakan kelompok. Sementara itu, Silhouette Score digunakan untuk menilai seberapa baik suatu data berada dalam cluster yang terbentuk dibandingkan dengan cluster lainnya. Kedua metrik tersebut menjadi dasar dalam mengevaluasi apakah hasil pengelompokan telah terbentuk secara optimal.