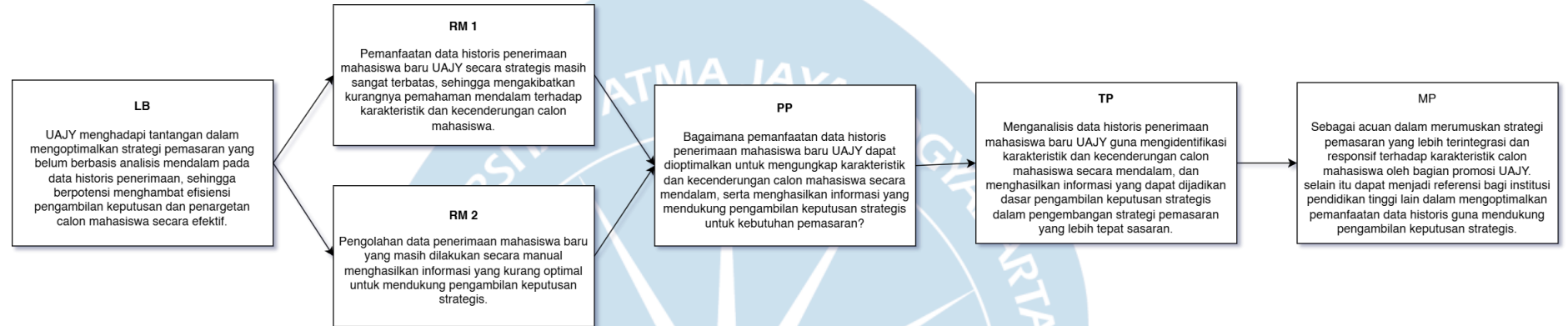


1. Penelitian ini difokuskan pada analisis pola karakteristik calon mahasiswa berdasarkan data pendaftar, dan tidak mencakup prediksi jumlah pendaftar untuk periode berikutnya.
2. Ruang lingkup data terbatas pada penerimaan mahasiswa baru program Sarjana (S1) UAJY dari lima tahun terakhir (2020/2021 s.d 2024/2025), tanpa melibatkan data dari jenjang atau periode lainnya.
3. Analisis hanya dilakukan terhadap atribut-atribut yang terdapat dalam data penerimaan, tanpa mencakup variabel eksternal atau faktor non-akademik lain yang mungkin memengaruhi strategi pemasaran universitas.
4. Penelitian ini tidak menyertakan pemodelan proses bisnis atau implementasi langsung strategi promosi, karena ruang lingkupnya terbatas pada analisis segmentasi berbasis data historis.
5. Perancangan strategi pemasaran dibatasi pada pendekatan yang sesuai dengan karakteristik dominan masing-masing segmen, sehingga tidak mencakup strategi di luar temuan hasil analisis.

1.6. Manfaat Penelitian

Informasi yang dihasilkan dan rekomendasi yang diberikan diharapkan dapat digunakan oleh bagian promosi UAJY sebagai acuan dalam merumuskan strategi pemasaran yang lebih terintegrasi dan responsif terhadap karakteristik calon mahasiswa. Hasil penelitian ini juga diharapkan dapat menjadi referensi bagi institusi pendidikan tinggi lain dalam meningkatkan pemanfaatan data historis guna mendukung pengambilan keputusan strategis.

1.7. Bagan Keterkaitan



Gambar 1.1. Bagan Keterkaitan

BAB II

TINJAUAN PUSTAKA

2.1. Studi Sebelumnya

Garonga dkk. [7] melakukan penelitian yang bertujuan untuk merancang strategi promosi penerimaan mahasiswa baru yang lebih tepat sasaran dengan mengelompokkan calon mahasiswa berdasarkan data kategorikal menggunakan algoritma *K-Modes*. Penelitian ini dilakukan di Universitas Kristen Indonesia Toraja dengan menggunakan data mahasiswa baru tahun akademik 2023/2024. Atribut yang digunakan dalam proses *clustering* mencakup asal daerah, asal sekolah, dan media promosi. Data dianalisis menggunakan pendekatan *Knowledge Discovery in Database* (KDD), dengan tahapan mulai dari pengumpulan data, *preprocessing*, hingga proses pemodelan dan evaluasi hasil. Metode *Elbow* digunakan untuk menentukan jumlah *cluster* optimal, yang pada akhirnya menghasilkan tiga *cluster*: *cluster 1* berisi 134 mahasiswa (mayoritas dari Toraja Utara, lulusan SMU, dan dipengaruhi brosur), *cluster 2* berisi 64 mahasiswa (dominan dari Tana Toraja, lulusan SMK, dan dipengaruhi oleh brosur serta kerabat), dan *cluster 3* berisi 23 mahasiswa (mayoritas dari Luwu Timur, lulusan SMU, dan terpengaruh oleh media sosial). Hasil analisis ini memberikan wawasan strategis yang dapat digunakan untuk menyesuaikan metode promosi berdasarkan preferensi informasi dan latar belakang calon mahasiswa di tiap segmen. Penelitian ini membuktikan bahwa algoritma *K-Modes* efektif dalam menangkap pola tersembunyi dari data kategorikal dan mendukung pengambilan keputusan pemasaran pendidikan tinggi secara *data-driven*.

Khusnuliawati & Putri [8] melakukan penelitian untuk mengidentifikasi segmentasi pasar mahasiswa universitas dengan menggunakan analisis *cluster* berbasis variabel psikografis. Penelitian ini dilatarbelakangi oleh semakin ketatnya persaingan antar perguruan tinggi swasta di Indonesia, yang menuntut strategi promosi yang lebih akurat dan tersegmentasi. Data dikumpulkan dari 230 mahasiswa Universitas Sahid Surakarta menggunakan kuesioner dengan skala *Likert* 5 poin, yang mencakup 36 item terkait psikografis, seperti orientasi kerja, nilai pribadi, pengaruh keluarga, kondisi ekonomi, dan pengaruh promosi. Setelah dilakukan proses *preprocessing*, analisis

cluster diterapkan dengan dua algoritma: *K-Means* dan *K-Modes*. Hasil evaluasi menunjukkan bahwa algoritma *K-Means* memberikan performa yang lebih baik (DBI: 2.0604) dibanding *K-Modes* (DBI: 3.4786), dan berhasil mengelompokkan mahasiswa ke dalam enam segmen utama. Masing-masing segmen merepresentasikan motivasi dominan, seperti motivasi ekonomi, dorongan keluarga, kualitas kampus, nilai pribadi, manfaat maksimal saat kuliah, hingga mahasiswa yang netral. Segmentasi ini memberikan wawasan mendalam mengenai profil psikologis mahasiswa dan dapat dimanfaatkan untuk menyusun strategi promosi yang lebih tepat sasaran berdasarkan karakteristik psikografis tiap segmen.

Kurniawan dkk. [9] menerapkan metode *K-Modes clustering* untuk mengelompokkan mahasiswa berdasarkan gaya belajar mereka. Penelitian ini dilakukan terhadap 83 mahasiswa pada mata kuliah Kecerdasan Buatan yang mengisi kuesioner berisi 36 pertanyaan kategorikal mengenai preferensi belajar. Tujuan dari penelitian ini adalah untuk menyediakan segmentasi gaya belajar guna mendukung efektivitas proses pembelajaran dan pengembangan metode pengajaran yang sesuai dengan karakteristik mahasiswa. Proses clustering dilakukan setelah validasi data menggunakan uji validitas dan reliabilitas. Hasil pengelompokan menunjukkan bahwa mahasiswa terbagi ke dalam tiga kelompok utama, yakni kelompok dengan gaya belajar visual, auditori, dan kinestetik, dengan distribusi masing-masing sebesar 37, 31, dan 15 mahasiswa. Selain segmentasi, penelitian ini juga menghasilkan sebuah aplikasi web yang memfasilitasi input kuisisioner dan implementasi clustering. Pengujian sistem dilakukan melalui metode *black-box* untuk memastikan kelengkapan fungsional. Studi ini menegaskan bahwa metode *K-Modes* dapat menjadi pendekatan yang efektif dalam memahami keragaman karakteristik belajar mahasiswa dan menyarankan bahwa pendekatan pengajaran dapat dipersonalisasi berdasarkan hasil segmentasi tersebut.

Yulianton dkk. [10] melakukan penelitian untuk mengelompokkan mahasiswa Universitas Stikubank berdasarkan data kategorikal menggunakan algoritma *K-Modes clustering*. Penelitian ini menggunakan data mahasiswa jenjang D3 dan S1 dari tahun ajaran 2018 hingga 2019 yang diambil dari sistem informasi *Smart Campus*. Atribut yang digunakan dalam proses *clustering* mencakup program studi, jenis kelamin, golongan darah, asal kota, pendidikan dan pekerjaan orang tua, serta jurusan sekolah. Setelah

dilakukan proses pembersihan data dan konversi nilai kategorikal, sebanyak 1.684 data mahasiswa diproses lebih lanjut. Proses *clustering* dilakukan menggunakan bahasa pemrograman Python melalui platform Google Colab dengan bantuan *library* seperti Pandas, Scikit-learn, Matplotlib, dan Seaborn. Untuk meningkatkan akurasi hasil *cluster*, peneliti menggunakan metode inisialisasi pusat *cluster* dari algoritma Cao dan menentukan jumlah *cluster* ideal berdasarkan nilai cost dari $k = 1$ hingga $k = 5$. Hasil terbaik diperoleh saat jumlah *cluster* adalah dua. *Cluster* pertama terdiri dari 1.367 mahasiswa, sedangkan *cluster* kedua sebanyak 317. *Cluster* pertama didominasi oleh mahasiswa dengan orang tua bekerja di sektor swasta dan berpendidikan SMA, serta ibu yang tidak bekerja. Hasil ini memberikan gambaran komprehensif mengenai profil mayoritas mahasiswa dan dapat dijadikan dasar untuk strategi promosi, pemetaan calon mahasiswa potensial, serta kebijakan manajemen universitas yang berbasis data.

Bellanov & Nurhayati [11] melakukan penelitian untuk menentukan strategi promosi kampus Universitas Katolik Darma Cendika (UKDC) dengan menerapkan metode *K-Means clustering*. Penelitian ini menggunakan data mahasiswa baru mulai tahun ajaran 2015 hingga 2022, dengan atribut utama berupa asal provinsi siswa dan pilihan program studi. Analisis dilakukan menggunakan perangkat lunak RStudio, dengan proses penentuan jumlah *cluster* optimal menggunakan metode *Elbow* dan *Silhouette Coefficient*. Hasil dari proses clustering membagi data menjadi dua *cluster*, di mana *cluster* pertama didominasi oleh calon mahasiswa yang berasal dari Provinsi Nusa Tenggara Timur, sedangkan *cluster* kedua berasal dari Provinsi Jawa Timur. Segmentasi ini digunakan untuk menyusun strategi promosi yang lebih efektif, dengan merekomendasikan fokus promosi di daerah yang memiliki kontribusi pendaftar tertinggi. Penelitian ini menunjukkan bahwa penerapan *K-Means clustering* mampu memberikan gambaran segmentasi geografis mahasiswa baru yang dapat dimanfaatkan dalam perencanaan promosi yang lebih tepat sasaran di masa depan.

Tabel 2.1. Penelitian Sebelumnya

Penulis & Tahun	Objek/Data Penelitian	Tujuan	Metode	Hasil
Garonga dkk. (2024)	Data mahasiswa baru UKI Toraja 2023/2024; atribut: asal daerah, asal sekolah, media promosi	Menyusun strategi promosi berdasarkan segmentasi pendaftar dengan pendekatan data kategorikal	<i>K-Modes Clustering</i> dengan <i>Elbow Method</i>	Terbentuk 3 <i>cluster</i> : (1) mayoritas dari Toraja Utara & SMU, dipengaruhi brosur; (2) dari Tana Toraja & SMK, dipengaruhi kerabat; (3) dari Luwu Timur & media sosial
Khusnuliawati & Putri (2023)	230 mahasiswa Universitas Sahid Surakarta, data psikografis (orientasi kerja, nilai, ekonomi, promosi, keluarga)	Menyusun segmentasi mahasiswa berbasis psikografis untuk promosi yang lebih terarah	<i>K-Means</i> vs. <i>K-Modes</i>	6 <i>cluster</i> dengan motivasi berbeda; <i>K-Means</i> menghasilkan DBI lebih baik; segmentasi mendalam: ekonomi, kualitas kampus, keluarga, dll.
Kurniawan dkk. (2024)	83 mahasiswa yang mengisi kuisioner 36 item terkait preferensi belajar (kategori: visual, auditori, kinestetik)	Mengelompokkan gaya belajar mahasiswa untuk mendukung efektivitas metode pengajaran	<i>K-Modes Clustering</i>	3 <i>cluster</i> : visual (37), auditori (31), kinestetik (15); digunakan untuk membangun aplikasi segmentasi dan validasi metode belajar
Yulianton dkk. (2021)	1.684 data mahasiswa Universitas Stikubank (D3 dan S1), atribut: prodi, kota asal, pendidikan	Mengelompokkan mahasiswa berdasarkan atribut kategorikal untuk mendukung kebijakan promosi dan akademik	<i>K-Modes Clustering</i> dengan <i>Elbow Method</i>	2 <i>cluster</i> : <i>cluster</i> 1 (1.367 mahasiswa) didominasi orang tua SMA, sektor swasta; <i>cluster</i> 2 (317) dengan karakteristik berbeda

Penulis & Tahun	Objek/Data Penelitian	Tujuan	Metode	Hasil
	orang tua, dll.			
Bellanov & Nurhayati (2023)	Data pendaftaran mahasiswa baru Universitas Katolik Darma Cendika tahun 2015–2022	Menentukan fokus wilayah promosi kampus berdasarkan asal provinsi pendaftar	<i>K-Means Clustering</i> , dengan <i>Elbow Method</i> dan <i>Silhouette Coefficient</i>	Terbentuk 2 <i>cluster</i> : dominasi Provinsi NTT dan Jawa Timur untuk strategi promosi daerah

2.2. Dasar Teori

2.2.1. Data Mining

Data mining adalah proses menganalisis data dalam jumlah besar untuk menemukan pola, hubungan, atau wawasan yang berguna. Dalam perkembangannya, *data mining* telah menjadi bagian penting dari analitik data modern, terutama dengan meningkatnya volume data yang dihasilkan oleh teknologi digital. Proses ini sering digunakan untuk mendukung pengambilan keputusan berbasis data di berbagai bidang seperti bisnis, kesehatan, pendidikan, dan lainnya [12]. *Data Mining* merupakan komponen penting dalam *Business Intelligence* yang berperan dalam mendukung pengambilan keputusan strategis, terutama dalam menentukan strategi pemasaran. Melalui serangkaian proses *Knowledge Discovery in Database* (KDD) seperti *data preparation* dan *modeling*, data mentah diolah untuk menemukan *knowledge* atau pola informasi tersembunyi yang bermanfaat bagi kepentingan bisnis. *Knowledge* yang ditemukan dari proses *Data Mining* ini menjadi dasar bagi perusahaan untuk dapat menjalankan strategi *Knowledge-Based Marketing*. Pemilihan teknik atau tugas *Data Mining* sangat bergantung pada tujuan analisis dan jenis data yang tersedia. Beberapa teknik yang umum digunakan meliputi klasifikasi, *clustering*, dan penemuan aturan asosiasi [13]. Terdapat beberapa metodologi dalam *data mining* yang sering digunakan diantaranya:

- a. CRISP-DM (*Cross-Industry Standard Process for Data Mining*): Sebuah pendekatan standar untuk proyek data mining yang mencakup enam fase, mulai dari pemahaman bisnis hingga penerapan.
- b. KDD (*Knowledge Discovery in Databases*): Proses yang lebih luas mencakup pengumpulan data, *preprocessing*, eksplorasi pola, dan interpretasi hasil.
- c. SEMMA (*Sample, Explore, Modify, Model, Assess*): Metodologi yang berfokus pada eksplorasi dan pemodelan data untuk menghasilkan wawasan [14].

2.2.2. K-Modes Clustering

K-Modes Clustering adalah algoritma *clustering* berbasis partisi yang dikembangkan khusus untuk menangani data kategorikal. Algoritma ini pertama kali diperkenalkan oleh Zhexue Huang pada tahun 1998 sebagai perpanjangan dari algoritma *K-Means* yang secara orisinal hanya cocok untuk data numerik. Dalam konteks data penerimaan mahasiswa baru seperti jalur masuk, asal sekolah, atau program studi pilihan yang semuanya merupakan atribut kategorikal *K-Modes* menjadi pilihan algoritma yang lebih tepat dibandingkan *K-Means* [15].

Berbeda dengan *K-Means* yang menggunakan *mean* sebagai pusat *cluster*, *K-Modes* menggunakan mode/modus (nilai yang paling sering muncul) untuk setiap atribut dalam *cluster*. Hal ini penting karena nilai *mean* tidak relevan bagi data kategorikal. Selain itu, ukuran jarak yang digunakan bukan *Euclidean*, melainkan *simple matching dissimilarity*. Algoritma *K-Modes* bersifat iteratif dan terdiri dari tahapan-tahapan sebagai berikut [16]. Berikut adalah tahapan inti dalam menjalankan algoritma *K-Modes Clustering*.

1. Inisialisasi: Menentukan jumlah *cluster* (*K*) dan memilih *K* objek data secara acak untuk dijadikan pusat *cluster* awal (*modes*).
2. Alokasi Objek: Menghitung ketidaksamaan antara setiap objek data dengan pusat *cluster* menggunakan ukuran *simple matching*. Objek kemudian dialokasikan ke *cluster* yang memiliki ketidaksamaan paling kecil.
3. Pembaruan *Modes*: Setelah semua objek dialokasikan, pusat *cluster* diperbarui dengan nilai mode terbaru dari atribut-atribut anggotanya.
4. Iterasi: Proses dialokasikan dan diperbarui diulang sampai tidak terjadi lagi perpindahan anggota *cluster*, atau mode tidak berubah.

Dissimilarity pada *K-Modes* dihitung menggunakan *simple matching*, yang dirumuskan sebagai berikut:

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j)$$

Notasi:

$d(X, Y)$ = Jarak ketidaksamaan antara objek X dan Y.

m = Jumlah atribut pada setiap objek.

$\delta(x_j, y_j)$ = Fungsi perbandingan sederhana, dengan ketentuan:

- $\delta(x_j, y_j) = 0$, jika nilai atribut ke- j sama ($x_j = y_j$)
- $\delta(x_j, y_j) = 1$, jika nilai atribut ke- j berbeda ($x_j \neq y_j$)

Dengan kata lain, jarak antara dua objek adalah jumlah total atribut yang nilainya tidak cocok (berbeda). Semakin kecil nilai $\delta(x_j, y_j)$ semakin mirip kedua objek tersebut [17].

Dalam implementasinya, algoritma *K-Modes* memiliki sejumlah kelebihan yang membuatnya unggul dalam menangani data kategorikal. Pertama, *K-Modes* memiliki efisiensi komputasi yang tinggi karena hanya menggunakan operasi perbandingan sederhana (*simple matching*) dan perhitungan modus, sehingga sangat sesuai untuk dataset berukuran besar dengan banyak atribut kategorikal. Kedua, algoritma ini secara khusus dirancang untuk data kategorikal, sehingga memberikan hasil *clustering* yang lebih akurat dan relevan dibandingkan algoritma seperti *K-Means* yang kurang cocok untuk tipe data tersebut. Ketiga, proses iteratif pada *K-Modes* cenderung cepat mencapai konvergensi karena perubahan mode biasanya terjadi dalam beberapa langkah saja [18].

2.2.3. Segmentasi Calon Mahasiswa

Segmentasi pasar merupakan salah satu pendekatan yang digunakan untuk memahami karakteristik kelompok calon mahasiswa. Dengan menerapkan segmentasi, institusi pendidikan dapat menyesuaikan pendekatan komunikasi, promosi, atau layanan berdasarkan karakteristik yang dimiliki oleh masing-masing segmen. Dalam konteks ini, *data mining* berperan penting dalam

mengelompokkan calon mahasiswa berdasarkan pola-pola yang terekam dalam data historis penerimaan. Salah satu metode yang umum digunakan dalam proses segmentasi ini adalah analisis *clustering*. Tujuan utama dari segmentasi calon mahasiswa adalah untuk mengarahkan strategi promosi dan pengembangan layanan yang lebih sesuai dengan kebutuhan, preferensi, dan latar belakang masing-masing kelompok. Melalui segmentasi, calon mahasiswa tidak hanya dipahami sebagai individu, tetapi sebagai bagian dari kelompok dengan karakteristik serupa yang dapat ditargetkan secara lebih efektif [19].

2.2.4. Strategi Pemasaran Berbasis Data

Strategi pemasaran berbasis data (*data-driven marketing*) adalah pendekatan yang menggunakan analisis kuantitatif dan kualitatif untuk merancang, melaksanakan, dan meningkatkan efektivitas kampanye pemasaran. Dengan memanfaatkan data dari sumber seperti perilaku pelanggan, transaksi historis, media sosial, dan sensor IoT, strategi ini memungkinkan bisnis membuat keputusan yang presisi, mengurangi risiko, dan meningkatkan *return on investment* (ROI). Dalam era digital, pendekatan ini menjadi kunci kompetitif bagi organisasi untuk memahami *market*, mempersonalisasi interaksi, dan mengukur dampak kampanye secara *real-time*. Strategi pemasaran berbasis data didefinisikan sebagai kerangka kerja yang mengintegrasikan data dari berbagai sumber (internal, eksternal, *real-time*) untuk menginformasikan setiap tahap pemasaran mulai dari perencanaan hingga evaluasi. Data digunakan tidak hanya sebagai alat pendukung, tetapi sebagai fondasi utama dalam pengambilan keputusan [20][21]