

BAB 2

TINJAUAN PUSTAKA

2.1 Landasan Kepustakaan

Dalam penelitian yang dilakukan sebelumnya (Budiman & Anto, 2019), dengan judul Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering), penelitian ini bertujuan mengelompokkan data calon mahasiswa melalui persebaran sesuai wilayah memakai K-Means clustering terbentuk 3 cluster yaitu cluster ke 1 merupakan kategori tinggi, cluster ke 2 termasuk kategori rendah dan cluster 3 termasuk kategori sedang. Selanjutnya melakukan strategi promosi pada calon mahasiswa baru agar tepat sasaran untuk setiap wilayah sesuai cluster yang sudah terbentuk, kemudian setelah terbentuk cluster maka dilakukan pengiriman tim yang sesuai dengan program studi yang paling banyak diminati oleh masyarakat serta melakukan promosi periklanan sesuai peminatan calon mahasiswa dengan melakukan penyelarasan, setelah memperoleh hasil dari pengelompokan sesuai cluster maka sistem akan menampilkan hasil dari pengelompokan daerah calon mahasiswa dalam bentuk dashboard.

Penelitian selanjutnya (Yunita, 2018), tentang Penerapan Data Mining dengan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri), pada tahap penelitiannya yg pertama Penentuan centroid di tahap awal algoritma K-Means sangat berpengaruh untuk hasil cluster seperti pada hasil pengujian yang dilakukan dengan memakai 516 dataset dengan centroid yang tidak sama memperoleh hasil yang akan terjadi cluster yang tidak sama juga, Selanjutnya melakukan pengelompokan data penerimaan mahasiswa memakai metode k-means Clustering, sehingga dihasilkan 3 cluster yaitu cluster ke 1 dengan jumlah 195 items, cluster ke 2 dengan jumlah 271 items, serta cluster ke 3 dengan jumlah 50 items. setelah itu melakukan taktik promosi periklanan untuk calon mahasiswa baru, yang nantinya akan mengikuti

cluster yang terbentuk berdasarkan program studi yang paling banyak diminati di masing-masing sekolah.

Penelitian selanjutnya (Asroni et al., 2018), tentang Penerapan Metode Clustering menggunakan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di UMY (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik) penelitian ini memiliki tujuan untuk melakukan Pendidikan Dokter serta Ilmu hubungan Internasional sebagai jurusan yang menjadi pilihan calon mahasiswa baru UMY. hasil penelitian ini sebagai tolak ukur pihak Universitas Muhammadiyah Yogyakarta buat melakukan strategi promosi ke calon mahasiswa baru.

Tabel 2. 1 Landasan Kepustakaan

No	Nama Penulis	Judul	Hasil
1.	Ramdani Budiman, Rudianto	Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering)	Hasil dari penelitian iterasi terjadi sebanyak 3 kali. Pada cluster pertama termasuk kelompok dengan peminatan yang paling tinggi, kemudian pada cluster kedua merupakan kelompok yang memiliki peminatan terendah, sedangkan pada cluster ketiga termasuk kelompok peminat sedang.
2.	Fitri Yunita	Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan	Hasil penelitian yang dilakukan dengan memakai 516 dataset mendapatkan tiga cluster seperti cluster pertama

		Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)	memperoleh 195 items, kemudian cluster kedua memperoleh 271 items, dan cluster ketiga mendapatkan 50 items.
3.	Asroni, Hidayatul Fitri, Eko Prasetyo	Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)	Hasil dari penelitian menggunakan data calon mahasiswa baru fakultas kedokteran dan ilmu keperawatan memperoleh 4 cluster, seperti pada Cluster 0 pada jurusan Ilmu Keperawatan, memperoleh 3.795 pendaftar dari total calon mahasiswa baru (maba) sebanyak 25.000. Selanjutnya pada cluster satu terdapat 8.296 pendaftar jurusan Pendidikan Dokter, dari total keseluruhan 25.000 calon maba. Berikutnya pada Cluster dua terdapat 7.427 pendaftar pada jurusan Kedokteran Gigi dari total keseluruhan 25.000 calon maba. Cluster yang terakhir

			yaitu cluster ketiga terdapat 5.482 pendaftar pada jurusan Pendidikan Dokter, dari total 25000 calon maba.
--	--	--	--

2.2 Landasan Teori

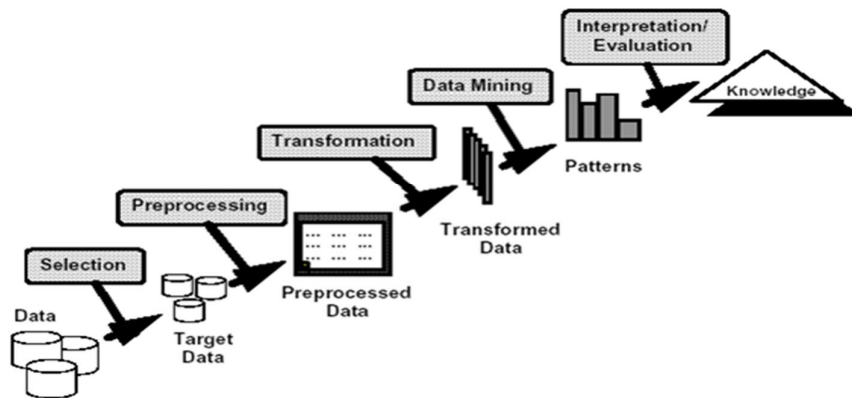
2.2.1 Strategi Promosi

Promosi adalah kegiatan pemasaran, melalui promosi calon mahasiswa baru akan mendapatkan informasi bahwa ada banyak program bagus di kampus. Banyak dari promosi memiliki pendanaan yang sama dengan agensi, semakin banyak dana yang dimiliki oleh suatu lembaga pendidikan, maka akan semakin sering pula terjadi kenaikan pangkat, yang juga sangat intensif. Akan tetapi uang tidak di atas segalanya. Keterbatasan pendanaan dapat diselesaikan melalui inovasi yang lebih cerdas dan tepat, salah satu pemecahan masalah yang dapat dilakukan adalah dengan menonjolkan prestasi kampus dan lain sebagainya. Kegiatan promosi yang dilakukan sangat erat keterkaitannya dengan penyebaran informasi yang akan dikomunikasikan kepada calon mahasiswa baru. Ada beberapa cara untuk menyampaikan strategi informasi ini, seperti menyebarkan brosur, menggunakan media sosial sebagai sarana periklanan, dll.

2.2.2 Data Mining

Data Mining adalah proses untuk menemukan sebuah informasi yang menarik dari data yang sudah terpilih dengan memakai metode tertentu. Algoritma pada data mining sangat bermacam-macam. Pemilihan algoritma yang tepat sangat tergantung pada tujuan dan penemuan pengetahuan yang lengkap dalam proses database (Anggreini, 2019).

Data mining merupakan bagian dari salah satu integral pada *Knowledge Discovery In Database*. Sebuah tahapan dalam proses menemukan pola yang terkandung dalam setiap pesan. Adapun tahapan yang dilakukan sebagai berikut.



Gambar 2. 1 Knowledge Discoveri in Database

- a. Data cleaning

Proses agar dapat menghilangkan data yang tidak lengkap, salah dan tidak konsisten dari pengumpulan data.
- b. Data integration

Proses integritas data yang menggabungkan data berulang.
- c. Data selection

Pada tahap ini dilakukan seleksi data, agar mendapatkan data yang relevan untuk dianalisis dari kumpulan data yang ada.
- d. Data transformation

Proses mengubah data yang terpilih kedalam bentuk program benamangan melalui manipulasi dan penyerangan data.
- e. Data mining

Proses menerapkan berbagai Teknik untuk mengekstrak berbagai pola yang mendasarinya untuk mendapatkan data yang berguna.
- f. Pattern evaluation

Pada tahapan ini dilakukan proses identifikasi terhadap pola menarik yang ditemukan sebelumnya bersumber dari dasar pengetahuan yang ada.
- g. Knowledge presentation

Tahapan terakhir dari proses ini, menggunakan teknik visualisasi yang

dirancang untuk membantu pengguna memahami dan menginterpretasikan hasil data mining.

2.2.3 K-means

K-means adalah sebuah metode pengelompokan data non-hierarkis pemisah data yang terdapat menjadi dua atau lebih kelompok. Partisi metode ini pengelompokan data sehingga data tersebut memasukkan fitur yang sama dan kelompok yang sama serta data yang sama fitur yang tidak selaras diklasifikasikan menjadi kelompok lain. Adapun tujuan paket data ini digunakan agar dapat meminimalkan fungsi tujuan yang ditetapkan pada kelompok serta maksimalkan perbedaan antar kelompok (Asroni et al., 2018).

K-means adalah sebuah metode klasterisasi yang terkenal dan telah banyak dipakai dalam berbagai bidang, karena praktis dan mudah dalam mengimplementasikan, kemudian *k-means* juga dapat mengkluster data yang memiliki ukuran besar, dapat menangani data outlier, dan kerumitan waktunya linear $O(nKT)$ dengan n ialah jumlah dokumen, k merupakan jumlah kluster, dan T ialah jumlah iterasi. *K-means* merupakan sebuah metode yang dapat melakukan pengklasteran secara partitioning yang mampu memisahkan data ke pada kelompok yang tidak sama. Dengan partitioning secara iteratif, *k-means* dapat meminimalkan akurasi jarak pada setiap kluster.

2.2.4 Algoritma K-Means

Algoritma K-Means adalah sebuah teknik untuk melakukan pengelompokan data hampir sama, jika berbeda maka akan menjadi kelompok dengan lain. *K-Means* menggunakan himpunan (k grup pertama adalah centroid) dan akan terus menerus melalui sebuah proses menghitung pada titik tengah (min) membuat sesuatu menjadi fungsi standar hasil yang dicapai oleh kelompok ialah konstan. Pada teknik mengelompokkan, menghitung kelompok divisi eksekusi menggunakan algoritma artinya, yang disebut fungsi jarak adalah langkah sama atau berbeda (Royal, 2021).

Dalam penelitian John Oscar Baik. Langkah-langkah yang dijelaskan pada tahun 2013 Saat melakukan clustering menggunakan algoritma K-Means, langkah-langkahnya sebagai berikut:

1. Memilih jumlah cluster k.
2. Inisialisasi k pusat cluster dengan berbagai cara. Ini sering digunakan secara *random*, dengan mengisi pusat cluster yang nilai awalnya berisi angka *random*.
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dari dua objek ini berdasarkan jarak dari objek tersebut. Dari dekatnya Data ke cluster tertentu ditentukan oleh jarak antara data dan pusat cluster. Pada tahap ini, setiap data perlu dihitung untuk pusat setiap cluster. jarak terdekat antara sekelompok data beberapa akan menentukan data cluster mana yang harus dimasuki.

Untuk menghitung jarak semua data di setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang menggunakan rumus dibawah ini:

$$d(i, j) = \sqrt{\sum(x_{ki} - \mu_{kj})^2}$$

Keterangan :

$d(i, j)$ = merupakan jarak data ke I pusat cluster j

x_{ki} = merupakan data ke I pada atribut ke K

μ_{kj} = merupakan titik pusat ke j pada atribut k

4. Menghitung ulang pusat cluster dengan anggota cluster yang sekarang. Pusat cluster merupakan rata-rata dari semua data/ objek pada cluster tertentu. Juga dapat menggunakan nilai tengah klaster. Jadi rata-rata bukan satu-satunya yang dapat dijadikan acuan agar bisa digunakan.
5. Menugaskan setiap objek memakai pusat cluster terbaru. Jika pusat cluster tidak akan berubah lagi maka proses clustering dianggap selesai atau kembali pada langkah 3 sampai pusat cluster tidak berubah-berubah lagi. (sistematis) pemahaman tentang subjek atau objek penelitian, seperti: mencoba mencari jawabannya bisa dihitung ilmu pengetahuan dan termasuk ke sahananya.

2.2.5 Clustering

Clustering merupakan suatu teknik pengelompokan data yang dapat menganalisis data untuk memecahkan masalah saat mengelompokkan data atau menetapkan dataset menjadi subset. Tentang teknologi pengelompokan target bersifat khusus kasus distribusi (objek, orang, peristiwa, dll). Didalam kelompok untuk mencapai derajat konektivitas antara anggota cluster yang sama ialah kekuatan di antara anggota cluster berbeda. Ada dua metode teknologi cluster dalam pengelompokan, yaitu hierarkis pengelompokan dan pengelompokan non-hierarkis. Pengelompokan hierarkis adalah sebuah metode yang memulai pengelompokan menggunakan 2 data atau lebih yang memiliki persamaan, maka proses berlanjut ke objek lain yang berdekatan keduanya, proses ini berlanjut hingga cluster membentuk semacam pohon yang di dalamnya terdapat hierarki atau level antar objek sangat mirip hingga paling tidak mirip. Akan tetapi logikanya, semua objek akhirnya akan membentuk suatu cluster (Nasution, 2019).

2.2.6 Rapidminer

Rapidminer merupakan alat yang dapat digunakan untuk menganalisis perangkat lunak pengolah data mining. *RapidMiner* juga merupakan bagian lingkungan (environment) yang digunakan dalam machine learning, data mining, text mining, serta predictive analytics. Machine learning adalah algoritma untuk evolusi perilaku komputer berdasarkan data empiris seperti sensor atau basis data. Data mining adalah sebuah proses meringkas pola dari kumpulan data besar dengan menggabungkan metode statistik, kecerdasan buatan dan database (Anggreini, 2019).