

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terkait

Berberapa penelitian telah dilakukan untuk mengidentifikasi mahasiswa yang berpotensi mengalami *drop out* dengan pendekatan data mining dan teknik pengelompokan (*clustering*). Salah satu penelitian menggunakan metode *decision tree* dan *deep learning* berbasis algoritma GRU dalam meramalkan potensi *drop out* berdasarkan data akademik, seperti IPK, SKS, dan status keaktifan. Hasilnya menunjukkan bahwa *decision tree* memberikan akurasi lebih tinggi dibanding *deep learning*, yang membuktikan bahwa pemodelan berbasis struktur keputusan dapat secara efektif mengklasifikasikan potensi *drop out* mahasiswa. Penelitian ini menyoroti pentingnya prediksi dini agar institusi dapat mengambil tindakan preventif terhadap mahasiswa yang berisiko [3].

Penelitian lain yang mengelompokkan mahasiswa guna mempermudah pembagian dosen pembimbing tugas akhir secara lebih tepat, terutama dalam mengatasi masalah ketidaksesuaian antara mahasiswa dan dosen pembimbing yang dapat menghambat kelulusan. Penelitian menggunakan metode data mining dengan algoritma *K-Means Clustering* karena dianggap sederhana dan efisien untuk pengelompokan data. Sampel yang digunakan adalah 22 mahasiswa semester 4 dari Prodi Manajemen Informatika UMPP. Proses analisis dilakukan menggunakan perangkat lunak RapidMiner, dan perbandingan jumlah *cluster* menunjukkan bahwa pembagian menjadi 5 *cluster* memberikan hasil terbaik dengan nilai *Davies Bouldin Index* terkecil (-1.126). Hasil menunjukkan bahwa dua *cluster*, yaitu *cluster_2* dan *cluster_4*, memiliki nilai *centroid* rendah pada beberapa mata kuliah dan IPK, yang mengindikasikan potensi risiko dalam penyelesaian tugas akhir. Kesimpulan dari penelitian ini adalah bahwa pembagian mahasiswa menjadi 5 kelompok dengan algoritma *K-Means* mampu membantu identifikasi mahasiswa yang membutuhkan perhatian khusus, sehingga dapat mendukung strategi

pembimbingan tugas akhir yang lebih efektif [4].

Selanjutnya pendekatan *clustering* dengan metode *K-Means* juga digunakan dalam sebuah studi yang mengelompokkan mahasiswa berdasarkan IPK, jenis kelamin, dan nilai mata kuliah. Dengan bantuan perangkat lunak RapidMiner, pengelompokan menghasilkan dua *cluster* utama, di mana *cluster* pertama mencerminkan mahasiswa dengan IPK rendah dan nilai kosong yang menunjukkan risiko *drop out* tinggi. Sedangkan *cluster* kedua berisi mahasiswa dengan performa akademik baik. Hasil ini digunakan sebagai dasar pengambilan kebijakan akademik seperti pendampingan belajar atau konseling [5].

Selanjutnya, penelitian lain menerapkan algoritma *K-Means* untuk menentukan kelompok mahasiswa yang berpotensi *drop out* berdasarkan IPK dan SKS. Hasil *clustering* menghasilkan tiga kategori, yaitu risiko rendah, sedang, dan tinggi. Temuan ini digunakan sebagai sistem pendukung keputusan bagi pihak prodi untuk memberikan intervensi lebih dini. Namun, disebutkan bahwa pemilihan jumlah *cluster* (nilai K) dan *centroid* awal sangat mempengaruhi kualitas pengelompokan, menjadi kelemahan dari metode ini yang perlu diperhatikan dalam implementasinya [6].

Sementara itu, salah satu studi menggunakan metode K-Medoids, yang dikenal lebih stabil dalam menangani *outlier* dibanding *K-Means*. Dengan atribut seperti IP semester awal, IPK, dan status beasiswa, penelitian ini menghasilkan dua *cluster* yang membedakan mahasiswa berdasarkan tren akademik dan kondisi finansial. Hasil evaluasi menunjukkan bahwa metode ini dapat memberikan hasil *clustering* yang lebih representatif terhadap kondisi mahasiswa [7].

Kelima penelitian tersebut memberikan kontribusi signifikan dalam upaya deteksi mahasiswa berisiko *drop out*, semuanya cenderung berfokus pada data akademik seperti IPK dan SKS. Padahal, data absensi memiliki potensi sebagai indikator awal ketidakterlibatan mahasiswa dalam proses perkuliahan. Minimnya studi yang secara khusus memanfaatkan data absensi sebagai dasar dalam pengelompokan risiko *drop out* menjadi celah penelitian yang perlu

dijawab. Oleh karena itu, penelitian ini penting untuk dilakukan, yaitu dengan mengimplementasikan metode *K-Means* pada data absensi mahasiswa guna mendeteksi risiko *drop out* secara lebih dini, objektif, dan berbasis perilaku kehadiran yang konkret.

2.2 Landasan Teori

2.1.1 Metode *Elbow*

Metode *Elbow* merupakan salah satu teknik evaluasi yang digunakan untuk menentukan jumlah kluster optimal dalam algoritma *clustering*, khususnya pada algoritma *K-Means*[8]. Tujuan utama dari metode ini adalah untuk memperoleh jumlah kluster yang menghasilkan pemisahan data terbaik dengan mempertimbangkan nilai *Within-Cluster Sum of Squares* (WCSS)[9]. WCSS merupakan jumlah total dari kuadrat jarak antara data dan pusat klusternya (*centroid*), yang dirumuskan sebagai berikut:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

di mana K adalah jumlah kluster, C_i adalah himpunan data dalam kluster ke- i , x adalah data dalam kluster tersebut, dan μ_i adalah *centroid* kluster ke- i . Dalam pelaksanaannya, metode ini dilakukan dengan menjalankan algoritma *K-Means* untuk berbagai nilai k , kemudian memplot grafik nilai WCSS terhadap jumlah kluster. Pada grafik tersebut, titik tekukan (*elbow*) menunjukkan jumlah kluster optimal, karena setelah titik tersebut penurunan nilai WCSS mulai melambat dan tidak signifikan. Pemilihan jumlah kluster yang terlalu sedikit akan menyebabkan pengelompokan data yang kurang akurat, sedangkan jumlah kluster yang terlalu banyak dapat meningkatkan kompleksitas tanpa memberikan manfaat yang sebanding. Oleh karena itu, metode *Elbow* menjadi alat bantu penting dalam proses optimasi pengelompokan data untuk memperoleh hasil yang efisien dan representatif

2.1.2 *Min-Max Scaling*

Min-Max Scaling adalah metode normalisasi data yang digunakan untuk mengubah nilai suatu fitur ke dalam rentang tertentu, biasanya antara 0 dan 1, dengan tujuan menghilangkan perbedaan skala antar variabel sehingga setiap fitur memiliki kontribusi yang seimbang dalam analisis data[10]. Metode ini sangat penting pada algoritma yang berbasis jarak seperti *K-Means* dan *K-Nearest Neighbors*, karena perbedaan skala antar fitur dapat memengaruhi hasil pengelompokan atau prediksi[11]. Proses normalisasi dilakukan dengan menggunakan rumus:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

di mana X adalah nilai asli, X_{\min} adalah nilai minimum, dan X_{\max} adalah nilai maksimum dari fitur tersebut. Dengan menerapkan *Min-Max Scaling*, data menjadi lebih seragam dan algoritma dapat bekerja secara optimal tanpa bias terhadap fitur tertentu yang memiliki nilai besar.

2.1.3 *Metode K-Means*

Algoritma *K-Means* adalah metode unsupervised learning yang digunakan untuk mengelompokkan data ke dalam K *cluster* berdasarkan kemiripan karakteristik[12]. Prosesnya diawali dengan menentukan jumlah *cluster* (K), lalu memilih K titik awal secara acak sebagai pusat *cluster* (*centroid*). Setiap data dihitung jaraknya ke *centroid* menggunakan rumus jarak Euclidean, yaitu:

$$d(x_i - c_j) = \sqrt{(x_{i1} - c_{j1})^2 + (x_{i2} - c_{j2})^2 + \dots + (x_{in} - c_{jn})^2}$$

di mana $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ adalah data ke- i , $c_j = (c_{j1}, c_{j2}, \dots, c_{jn})$ adalah data ke- j , dan $d = (x_i, c_j)$ adalah jarak Euclidean antara data dan *centroid*. Setelah pengelompokan ulang data, posisi *centroid* diperbarui menggunakan rata-rata dari semua data yang ada dalam *cluster* tersebut:

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in c} x_i$$

Dengan C_j sebagai *centroid* baru *cluster* ke- j , C_j adalah himpunan data yang termasuk dalam kluster ke- j . Dan $|c_j|$ adalah jumlah data dalam kluster ke- j . Proses pengelompokan dan pembaruan centroid ini diulang hingga hasil cluster stabil. Kelebihan algoritma ini adalah kesederhanaan dan kecepatannya dalam mengolah data besar, namun kelemahannya termasuk keharusan menentukan nilai K sejak awal, sensitif terhadap titik awal centroid, dan tidak optimal jika bentuk cluster tidak bulat atau mengandung outlier. K-Means banyak diterapkan dalam segmentasi data seperti kluster mahasiswa berdasarkan IPK dan kehadiran untuk mendeteksi risiko drop out[13].

2.1.4 Clustering

Clustering merupakan salah satu metode dalam unsupervised learning yang digunakan untuk mengelompokkan objek-objek data ke dalam sejumlah kelompok (*cluster*) berdasarkan kemiripan karakteristik atau atribut yang dimiliki[14]. Dalam proses *clustering*, data yang memiliki karakteristik serupa akan dikelompokkan ke dalam satu *cluster*, sedangkan data yang memiliki perbedaan signifikan akan dimasukkan ke dalam *cluster* yang berbeda. Tidak seperti metode supervised learning, *clustering* tidak memerlukan label kelas pada data, sehingga sangat sesuai digunakan dalam proses eksplorasi data atau ketika struktur data belum diketahui secara pasti.

Tujuan utama dari *clustering* adalah untuk memaksimalkan intra-*cluster* similarity (kemiripan antar data dalam satu kelompok) dan meminimalkan inter-*cluster* similarity (kemiripan antar kelompok yang berbeda). Oleh karena itu, metode ini banyak dimanfaatkan dalam berbagai bidang penelitian, seperti pengelompokan data pelanggan dalam dunia bisnis, deteksi pola dalam data medis, hingga identifikasi kelompok mahasiswa berdasarkan atribut akademik untuk keperluan evaluasi pendidikan.

Beberapa algoritma *clustering* yang umum digunakan dalam penelitian antara lain *K-Means*, *Hierarchical Clustering*, dan *DBSCAN*. Di antara algoritma tersebut, *K-Means* sering menjadi pilihan karena kesederhanaannya dalam implementasi serta efisiensinya dalam menangani data berukuran besar. Namun demikian, kelemahan utama dari metode *clustering* adalah keharusan untuk menentukan jumlah *cluster* di awal dan sensitivitas terhadap data *outlier*. Meskipun demikian, *clustering* tetap menjadi teknik yang andal dan sering digunakan dalam kajian ilmiah yang melibatkan analisis data tidak terlabel.

2.1.5 *Drop Out*

Drop out merupakan suatu kondisi di mana mahasiswa menghentikan proses pendidikan secara permanen sebelum menyelesaikan program studi yang ditempuh pada jenjang pendidikan tinggi [15]. Fenomena ini sering kali dijadikan sebagai indikator kinerja lembaga pendidikan, karena tingginya angka *drop out* dapat mencerminkan permasalahan sistemik dalam proses pembelajaran, pelayanan akademik, maupun kesejahteraan mahasiswa. Dalam kajian akademik, istilah *drop out* tidak hanya dipahami sebagai kegagalan individu, tetapi juga sebagai hasil interaksi kompleks antara faktor internal dan eksternal mahasiswa.

Secara umum, faktor penyebab *drop out* dapat diklasifikasikan ke dalam dua kategori utama, yaitu faktor internal dan faktor eksternal. Faktor internal meliputi rendahnya motivasi belajar, kesulitan dalam memahami materi perkuliahan, kurangnya kemampuan manajemen waktu, serta masalah psikologis seperti stres atau depresi. Sementara itu, faktor eksternal mencakup kondisi sosial ekonomi yang tidak mendukung, keterbatasan fasilitas belajar, tanggung jawab pekerjaan atau keluarga, serta minimnya dukungan dari lingkungan sekitar. Faktor-faktor tersebut dapat saling berinteraksi dan memperburuk risiko mahasiswa untuk keluar dari sistem pendidikan sebelum waktunya.

Dalam pendidikan tinggi, studi-studi sebelumnya menunjukkan bahwa aspek akademik seperti Indeks Prestasi Kumulatif (IPK) dan tingkat kehadiran dalam perkuliahan memiliki keterkaitan erat dengan kemungkinan terjadinya

drop out. Mahasiswa yang menunjukkan pola kehadiran yang tidak konsisten serta pencapaian akademik yang rendah cenderung memiliki risiko lebih tinggi untuk mengalami putus studi. Oleh karena itu, diperlukan pendekatan analitik berbasis data untuk mengidentifikasi mahasiswa yang berada dalam kategori risiko, sehingga intervensi preventif dapat dilakukan secara lebih efektif. Salah satu pendekatan yang relevan untuk mengelompokkan mahasiswa berdasarkan karakteristik tersebut adalah metode *clustering*, seperti algoritma *K-Means*, yang mampu mengidentifikasi pola tersembunyi dalam data akademik dan kehadiran mahasiswa.

2.1.6 Mahasiswa

Mahasiswa merupakan individu yang sedang menempuh pendidikan formal pada jenjang perguruan tinggi, baik di tingkat diploma, sarjana, magister, maupun doktoral[16]. Dalam pendidikan tinggi, mahasiswa tidak hanya berperan sebagai peserta didik yang menerima ilmu pengetahuan, tetapi juga sebagai subjek yang aktif dalam proses pembelajaran, penelitian, dan pengabdian kepada masyarakat. Mahasiswa memiliki tanggung jawab untuk mengembangkan potensi diri secara akademik maupun non-akademik sebagai bagian dari persiapan menghadapi dunia kerja dan kehidupan sosial yang lebih luas.

Secara umum, mahasiswa berada dalam fase transisi dari remaja menuju dewasa, yang ditandai dengan peningkatan tanggung jawab, kemandirian, serta kemampuan mengambil keputusan. Masa ini merupakan periode penting dalam pembentukan karakter, identitas diri, dan orientasi masa depan. Oleh karena itu, perguruan tinggi tidak hanya berfungsi sebagai institusi akademik, tetapi juga sebagai lingkungan yang mendukung pengembangan kepribadian, keterampilan sosial, dan profesionalisme mahasiswa.

Dalam konteks manajemen pendidikan tinggi, pemantauan terhadap kondisi mahasiswa menjadi aspek krusial untuk menjamin kualitas dan efektivitas proses pembelajaran. Hal ini mencakup aspek akademik seperti capaian prestasi, tingkat kehadiran, partisipasi dalam kegiatan kampus, hingga aspek kesejahteraan mahasiswa. Upaya untuk mengidentifikasi permasalahan

yang dihadapi mahasiswa, termasuk risiko putus studi (*drop out*), menjadi penting untuk mendukung keberlangsungan studi mereka serta meningkatkan angka kelulusan sebagai indikator kualitas institusi pendidikan.