

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Algoritma *K-Means* adalah algoritma pengelompokan *iterative* yang melakukan partisi *set* data ke dalam sejumlah *cluster* yang diawal sudah ditetapkan. Algoritma *K-Means* mudah diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Parameter yang harus dimasukkan ketika menggunakan algoritma *K-Means* adalah nilai K. Nilai K umumnya digunakan berdasarkan informasi yang diketahui sebelumnya mengenai sebenarnya berapa banyak *cluster* yang muncul dalam X, berapa banyak yang digunakan untuk penerapannya, atau jenis *cluster* dicari dengan melakukan percobaan dengan beberapa nilai K. Untuk *set* data di dalam X dapat dikelompokkan berdasarkan konsep kedekatan atau kemiripan, namun kuantitas yang digunakan untuk mengukurnya adalah ketidakmiripan. Metrik yang umum digunakan untuk ketidakmiripan tersebut adalah *Euclidean* (Adeo & Sembodo, 2021).

Teknik *Clustering Data Mining* seperti *K-Means* adalah salah satu algoritma yang banyak diterapkan oleh peneliti di antaranya menggunakan algoritma *clustering* pengelompokan data, pemetaan data, klasifikasi data, dan sebagainya (Harahap, *et al.*, 2022). Algoritma *K-Means* adalah salah satu metode *clustering* melalui *attribute*

*numeric*. Tahapan dalam implementasi algoritma *K-Means* terdiri dari penentuan jumlah *cluster*, selanjutnya menentukan pusat awal *cluster* dengan mengambil data secara acak (Wibowo, *et al.*, 2021).

KDD (*Knowledge Discovery in Database*) atau penemuan pengetahuan dalam basis data, merupakan sebuah yang luas untuk menemukan pengetahuan dalam data dan menekankan aplikasi tingkat tinggi dari teknik penambangan data. KDD merupakan bidang yang menarik bagi para peneliti di berbagai bidang, seperti kecerdasan buatan, *machine learning*, pengenalan pola, basis data, statistik, akuisisi pengetahuan untuk sistem profesional, dan visualisasi data. KDD menggunakan beberapa algoritma yang bersifat belajar mandiri (*self-learning*) untuk menyimpulkan pola yang berguna dari data yang diproses (Harjono, *et al.*, 2023).

Saat ini perkembangan teknologi informasi telah berkembang sangat pesat mengakibatkan ketersediaan dan keberagaman data pun semakin meningkat. Pemanfaatan data yang ada tidak cukup hanya mengandalkan data operasional saja, tetapi diperlukan suatu analisis data untuk menggali informasi – informasi yang ada. Karena banyaknya data mahasiswa saat ini belum dimanfaatkan secara maksimal dan efisien, pihak prodi masih merasa kesulitan dalam menganalisis evaluasi tingkat kelulusan mahasiswa dikarenakan data yang belum terintegrasi dalam sebuah basis data. Padahal pihak prodi perlu untuk melakukan evaluasi tingkat kelulusan untuk meningkatkan dan mempertahankan kualitas kelulusan dan akreditasi program studi, sehingga untuk memaksimalkan data informasi kelulusan diperlukan data data tersebut dapat dimanfaatkan sebagai sumber informasi strategis bagi program studi untuk

melakukan klasifikasi tingkat kelulusan mahasiswa dengan menggunakan teknik *Data Mining* (Fernanda, *et al.*, 2021).

Algoritma pengelompokan *K-Means* mengklasifikasikan warna, pola, bentuk, dan konten gambar komputer, serta memanfaatkan keunggulan invarian pengambilan konten gambar untuk meningkatkan akurasi pencarian gambar. Hasil percobaan menunjukkan bahwa algoritma pengelompokan *K-Means* yang diusulkan dapat dengan cepat menyatukan hasil ke target kueri dan memiliki ketahanan yang lebih baik (Yu & Liu, 2022).

Algoritma *K-Means* digunakan untuk melakukan segmentasi gambar dengan tujuan mengidentifikasi pola korosi pada permukaan logam. Algoritma *K-Means* merupakan algoritma *clustering* yang digunakan untuk membagi data ke dalam  $k$  kelompok berdasarkan atribut yang dimiliki. Dalam konteks penelitian ini, algoritma *K-Means* digunakan untuk membagi gambar menjadi beberapa segmen yang memungkinkan untuk mengidentifikasi pola korosi dengan lebih baik (Almanza-Ortega, *et al.*, 2023).

*K-Means Clustering* adalah sebuah algoritma pengelompokan data yang bekerja dengan cara mengelompokkan data ke dalam beberapa *cluster* berdasarkan kedekatan jarak antara objek data dan pusat massa *cluster*. Algoritma ini memilih pusat *cluster* awal secara acak, kemudian menghitung jarak antara objek data dengan pusat massa *cluster*, dan objek data yang dekat dengan pusat massa akan menjadi bagian dari *cluster* tersebut. Langkah-langkah spesifik dari algoritma *K-Means* meliputi pemilihan sampel dengan nilai probabilitas maksimum sebagai pusat *cluster* berikutnya, dan mengulangi

langkah-langkah tersebut hingga K pusat *cluster* terpilih. Algoritma *K-Means Clustering* ini merupakan salah satu teknik yang paling umum digunakan dalam analisis data, dan telah banyak digunakan dalam berbagai bidang seperti pengenalan pola, pembelajaran mesin, pemrosesan gambar, dan penambangan data. Metode ini termasuk dalam kategori algoritma pengelompokan terpartisi yang membagi semua objek ke dalam kategori yang saling eksklusif, dengan tujuan meningkatkan kesamaan intra-*cluster* dan mengurangi kesamaan antar-*cluster* (Wang, *et al.*, 2023).

Mengklasifikasikan data akademik mahasiswa menjadi empat klaster, yaitu klaster mahasiswa berprestasi, berpotensi berprestasi, berpotensi bermasalah, dan klaster mahasiswa bermasalah (Fadrial, 2020).

*Dynamic K-means Clustering* merupakan pengembangan algoritma *k-means* untuk mengecek ulang kualitas *cluster* pada setiap iterasi, memungkinkan terjadinya perubahan dalam jumlah *cluster* untuk memenuhi keabsahan kualitas *cluster*. Algoritma klaster dinamis bertujuan untuk meningkatkan kualitas *cluster* sehingga menghasilkan angka yang optimal dari *cluster*. Algoritma *Dynamic K-means* memiliki kemampuan untuk mencari jumlah *cluster* ideal, namun terdapat kekurangan dalam penentuan titik *centroid* (pusat *cluster*) yang masih dipilih secara acak. Sehingga kesalahan penentuan *centroid* awal akan mempengaruhi jumlah proses iterasi dan waktu komputasi (Ariasa, *et al.*, 2020).

Data evaluasi pendidikan adalah data fakta yang diperoleh untuk efek pendidikan atau perkembangan peserta didik dalam segala aspek, dan evaluasi pendidikan adalah

proses penilaian nilai berdasarkan data tersebut. Data evaluasi siswa adalah salah satu bagian dari data besar pendidikan yang paling dikenal oleh para pendidik (Liu, 2022).

Kemajuan hasil belajar siswa harus dievaluasi dan dipantau. Hasil belajar siswa ditunjukkan dengan nilai ujian. Lembaga pendidikan dapat mengetahui kemampuan pengetahuan dan kompetensi siswa berdasarkan nilai ujian siswa. Algoritma *k-means* digunakan untuk mengevaluasi hasil belajar siswa. Hasil penerapan algoritma *k-means* dapat membantu perencana dan pimpinan akademik memantau hasil belajar siswa (Kurniawan, *et al.*, 2020).

Penelitian ini bertujuan untuk menganalisa data akademik mahasiswa, membangun pola pengelompokkan menggunakan metode Klastering *K-Means* sehingga dapat menghasilkan pengetahuan atau *knowledge*. Dan menguji data yang sudah diolah dengan menggunakan *tools RapidMiner*.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan di atas, maka dapat dirumuskan masalah dari penelitian ini adalah sebagai berikut :

1. Bagaimana penerapan metode *Clustering* Algoritma *K-Means* pada hasil evaluasi akademik mahasiswa?
2. Bagaimana kinerja algoritma *K-Means* dalam mengelompokkan mahasiswa berdasarkan hasil evaluasi akademik mereka?
3. Bagaimana hasil proses pengelompokkan (*clustering*) evaluasi akademik mahasiswa dapat diimplementasikan untuk mengembangkan kebijakan yang disesuaikan dengan klasifikasi mahasiswa?

### 1.3 Batasan Masalah

Agar penelitian ini lebih terarah maka penelitian diberi batasan-batasan sebagai berikut :

1. Data yang digunakan adalah data mahasiswa angkatan 2021 Program Studi Teknik Informatika
2. Pengelompokan dilakukan berdasarkan data informasi mahasiswa dari semester 1, 2, 3, dan 4 berupa Indeks Prestasi Semester (IPS)
3. *Tools* yang digunakan didalam pengolahan data adalah *RapidMiner*

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, maka terdapat tujuan yang ingin dicapai dari penelitian ini adalah :

1. Menerapkan algoritma *K-Means* dalam klusterisasi berdasarkan hasil evaluasi akademik mahasiswa
2. Menerapkan kinerja algoritma *K-Means* dalam mengelompokkan mahasiswa berdasarkan hasil evaluasi akademik
3. Mengoptimalkan penggunaan algoritma *K-Means* dalam klusterisasi evaluasi akademik mahasiswa untuk mendapatkan hasil yang lebih akurat dan relevan
4. Melakukan pengujian dari hasil evaluasi dengan Algoritma *K-Means* menggunakan aplikasi *RapidMiner*.

## **1.5 Manfaat Penelitian**

Sesuai dengan tujuan penelitian, maka manfaat yang diharapkan dengan dilakukannya penelitian ini adalah:

1. Memberikan informasi atau hasil secara tepat, cepat dan akurat dalam mengelompokan hasil evaluasi mahasiswa kepada pihak prodi yang merupakan salah satu basis untuk memantau perkembangan kinerja akademik mahasiswa.
2. Membantu pihak prodi akademik dalam mengembangkan strategi untuk meningkatkan pembelajaran dan prestasi akademik mahasiswa.
3. Sebagai bahan kajian bagi peneliti berikutnya yang berhubungan dengan penggunaan *Data Mining* pada Algoritma *K-Means* di perguruan tinggi.

## **1.6 Sistematika Penulisan**

Penelitian ini mencoba untuk membahas pokok permasalahan secara cermat dan sistematis. Untuk itu pembahasan dibuat dan disusun dengan sistematika penulisan sebagai berikut :

## **BAB I PENDAHULUAN**

Pada bab ini berisi pendahuluan yang mencakup latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

## **BAB II LANDASAN TEORI**

Pada bab ini berisi tentang tinjauan pustaka terhadap *Data Mining* khususnya metode klustering *K-Means* dan penerapannya dalam melakukan pencarian pengetahuan (*Knowledge*).

## **BAB III METODOLOGI PENELITIAN**

Pada bab ini menjelaskan tentang kerangka kerja (*frame work*) & metodologi yang akan dipakai di dalam melakukan penelitian.

## **BAB IV ANALISA DAN PERANCANGAN**

Pada bab ini akan dibahas tentang analisis pencarian pengetahuan (*Knowledge*) menggunakan metode Klustering *K-Means*.

## **BAB V IMPLEMENTASI DAN HASIL**

Bab ini membahas hasil implementasi dan pengujian dengan menggunakan *tools* yang sudah ada sebagai perbandingan

## **BAB VI KESIMPULAN DAN SARAN**

Bab ini berisikan tentang kesimpulan dari penelitian yang telah dilakukan serta keterbatasan dan saran yang digunakan dalam pengembangan penelitian selanjutnya dengan topik pembahasan yang sama.