

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pendidikan tinggi merupakan salah satu pilar penting dalam pembangunan sumber daya manusia yang unggul dan kompetitif. Di era globalisasi dan revolusi industri 4.0, perguruan tinggi dituntut untuk tidak hanya mencetak lulusan yang kompeten secara akademik, tetapi juga mampu bersaing di dunia kerja yang semakin kompleks[1]. Namun, kenyataan di lapangan menunjukkan bahwa tidak semua mahasiswa yang diterima di perguruan tinggi mampu menyelesaikan studinya dengan baik. Fenomena mahasiswa yang mengalami putus studi atau *drop out* (DO) menjadi tantangan serius yang dihadapi oleh banyak perguruan tinggi, baik negeri maupun swasta.

*Drop out* merupakan kondisi di mana mahasiswa menghentikan proses studinya sebelum menyelesaikan program pendidikan yang ditempuh[2]. Penyebab dari fenomena ini sangat beragam, mulai dari faktor ekonomi, masalah pribadi, motivasi belajar yang rendah, hingga prestasi akademik yang kurang memadai. Tingginya angka *drop out* tidak hanya merugikan mahasiswa secara individu, tetapi juga berdampak negatif bagi institusi pendidikan. Citra perguruan tinggi dapat menurun, efisiensi penggunaan sumber daya menjadi rendah, dan perencanaan akademik menjadi terganggu. Oleh karena itu, upaya untuk mendeteksi dan mencegah risiko *drop out* sejak dini menjadi kebutuhan mendesak dalam sistem manajemen pendidikan tinggi.

Salah satu indikator awal yang dapat digunakan untuk memprediksi risiko *drop out* adalah tingkat kehadiran dan prestasi akademik mahasiswa, seperti nilai Indeks Prestasi Kumulatif (IPK). Mahasiswa yang sering tidak hadir dalam perkuliahan atau memiliki IPK di bawah standar umumnya memiliki risiko lebih tinggi untuk gagal menyelesaikan studinya. Data kehadiran dan nilai mahasiswa sesungguhnya menyimpan informasi yang sangat berharga jika dianalisis dengan tepat. Di sinilah peran teknologi dan analisis data sangat

dibutuhkan.

Perkembangan teknologi informasi dan data *science* memungkinkan institusi pendidikan untuk memanfaatkan data historis mahasiswa guna mendeteksi pola-pola yang mengarah pada risiko *drop out*. Salah satu metode yang dapat digunakan untuk mengelompokkan mahasiswa berdasarkan karakteristik tertentu adalah teknik *clustering*. *Clustering* merupakan bagian dari data mining yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok (*cluster*) yang memiliki kemiripan tertentu. Teknik ini sangat berguna dalam proses pengambilan keputusan, karena dapat membantu pihak kampus memahami kelompok-kelompok mahasiswa yang perlu mendapat perhatian khusus.

Salah satu algoritma *clustering* yang paling populer dan banyak digunakan adalah *K-Means*. Algoritma ini bekerja dengan mengelompokkan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya berdasarkan kedekatan jarak antar data. Keunggulan dari *K-Means* adalah kemampuannya dalam menangani dataset dalam jumlah besar dengan proses yang relatif cepat dan efisien. Dalam konteks pendidikan, *K-Means* dapat digunakan untuk mengelompokkan mahasiswa berdasarkan IPK dan tingkat kehadiran, sehingga pihak kampus dapat mengetahui kelompok mahasiswa yang berisiko mengalami *drop out*.

Beberapa penelitian sebelumnya telah berhasil mengimplementasikan *K-Means* untuk berbagai tujuan di dunia pendidikan, seperti pengelompokan berdasarkan minat, prestasi akademik, dan perilaku belajar mahasiswa. Namun, belum banyak studi yang secara khusus memanfaatkan algoritma ini untuk mendeteksi risiko *drop out* mahasiswa dengan fokus pada parameter IPK dan kehadiran. Padahal, kedua parameter tersebut merupakan indikator awal yang dapat dengan mudah diakses dan dimanfaatkan oleh pihak akademik.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means* dalam proses pengelompokan risiko *drop out* mahasiswa berdasarkan data IPK dan tingkat kehadiran. Penelitian dilakukan menggunakan data sekunder dari rekap absensi dan nilai

IPK mahasiswa di salah satu perguruan tinggi swasta. Melalui pendekatan ini, diharapkan pihak kampus dapat memperoleh informasi yang lebih akurat dan strategis dalam mengambil kebijakan preventif terhadap potensi putus studi. Dengan demikian, kualitas pendidikan dan efisiensi manajemen akademik dapat terus ditingkatkan.

Penelitian ini tidak hanya berkontribusi dalam konteks peningkatan mutu pendidikan di tingkat institusi, tetapi juga memperkaya khasanah literatur akademik di bidang penerapan teknologi data mining dalam dunia pendidikan. Diharapkan hasil dari penelitian ini dapat menjadi referensi dan acuan bagi institusi lain yang ingin menerapkan pendekatan serupa dalam mendeteksi dan mengelola risiko *drop out* mahasiswa. Berdasarkan uraian latar belakang di atas maka penelitian ini mengambil judul **implementasi algoritma *k-means* untuk pengelompokan risiko *drop out* mahasiswa.**

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana implementasi algoritma *K-Means* dalam mengelompokkan mahasiswa KIP STMIK YMI Tegal angkatan 2022 berdasarkan IPK dan kehadiran?
2. Bagaimana hasil pengelompokan risiko *drop out* yang dihasilkan oleh algoritma *K-Means*?

## 1.3 Pembatasan Masalah

Agar penelitian lebih terarah, maka pembatasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan terbatas pada mahasiswa penerima KIP STMIK YMI Tegal angkatan 2022.
2. Variabel yang digunakan dalam proses *clustering* adalah IPK terakhir yang diperoleh dan persentase kehadiran mahasiswa.

## 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Menerapkan algoritma *K-Means* untuk mengelompokkan mahasiswa

berdasarkan Indeks Prestasi Kumulatif (IPK), kehadiran (absensi) mahasiswa, dan Satuan Kredit Semester (SKS).

2. Menghasilkan pengelompokan risiko drop out mahasiswa dalam beberapa klaster risiko berdasarkan data yang tersedia.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Manfaat untuk Penulis
  - a. Menambah wawasan dan pemahaman penulis tentang implementasi algoritma *K-Means* dalam pengolahan data pendidikan.
  - b. Melatih keterampilan teknis dalam menggunakan tools seperti Microsoft Excel dan *Python* (Google Colab) untuk keperluan analisis data dan visualisasi hasil *clustering*.
  - c. Memberikan pengalaman nyata dalam menyusun penelitian ilmiah dari tahap perumusan masalah hingga analisis hasil.
  - d. Meningkatkan kemampuan dalam melakukan analisis data secara kuantitatif berbasis data mining.
  - e. Menjadi portofolio akademik yang dapat mendukung pengembangan karier penulis di bidang teknologi informasi dan analisis data.
2. Manfaat untuk STMIK YMI Tegal
  - a. Memberikan informasi berbasis data mengenai tingkat risiko drop out mahasiswa penerima KIP angkatan 2022.
  - b. Menjadi acuan bagi pihak kampus untuk melakukan deteksi dini dan intervensi terhadap mahasiswa yang berpotensi mengalami penurunan akademik.
  - c. Mendukung pengambilan kebijakan akademik yang lebih tepat sasaran melalui segmentasi mahasiswa berdasarkan tingkat risiko.
  - d. Meningkatkan efektivitas pengelolaan pendidikan dan kualitas layanan akademik di lingkungan STMIK YMI Tegal.
  - e. Menjadi referensi awal enerapan teknologi data mining untuk kepentingan evaluasi dan perencanaan akademik di masa mendatang.