

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Akreditasi merupakan salah satu bentuk penilaian atau evaluasi kelayakan dan mutu perguruan tinggi atau program studi yang dilakukan oleh organisasi atau badan mandiri di luar perguruan tinggi, dimana salah satu indikator penilaiannya adalah persentase kelulusan tepat waktu (BAN-PT, 2019). Oleh karena itu, persentase kelulusan tepat waktu mahasiswa di perguruan tinggi merupakan indikator penting dalam mengevaluasi kualitas pendidikan dan efektivitas sistem pendidikan. Namun, pada praktiknya masih banyak mahasiswa yang tidak dapat menyelesaikan studinya secara tepat waktu. Berdasarkan Laporan Audit Mutu Internal Akademik Undiksha Tahun 2020 menunjukkan bahwa beberapa prodi di Universitas Pendidikan Ganesha memiliki angka kelulusan mahasiswa tepat waktu yang berada dibawah standar yang telah ditetapkan (Lembaga Pengembangan Pembelajaran dan Penjamin Mutu Undiksha, 2020).

Terdapat banyak faktor yang mempengaruhi lama masa studi mahasiswa itu sendiri, diantaranya faktor akademik, faktor internal, maupun faktor eksternal (Winalia *dkk.*, 2019). Untuk dapat membantu pihak perguruan tinggi dalam memahami faktor-faktor yang mempengaruhi waktu kelulusan mahasiswa, maka salah satu pendekatan analisis data yang dapat digunakan adalah dengan melakukan *clustering*. *Clustering* adalah sebuah teknik analisis data yang digunakan untuk

mengelompokkan objek-objek data yang serupa ke dalam kelompok-kelompok yang disebut *cluster*, *clustering* mampu melakukan pemisahan, pemecahan, maupun segmentasi kumpulan data ke dalam kelompok-kelompok (*cluster*) berdasarkan kesamaan ciri, sifat, dan pola data tersebut (Rusdiana *dkk.*, 2009).

Dengan melakukan *clustering*, data mahasiswa akan dikelompokkan berdasarkan kesamaan karakteristik yang dimilikinya sehingga dengan analisis yang lebih lanjut maka akan dapat diidentifikasi karakteristik data mahasiswa tersebut berdasarkan kelompok *cluster* yang terbentuk. Hasil dari analisis ini dapat memberikan wawasan kepada perguruan tinggi untuk mengambil tindakan yang sesuai dengan karakteristik data pada kelompok tersebut, seperti mengidentifikasi mahasiswa yang memerlukan dukungan tambahan, mengimplementasikan program-program khusus untuk meningkatkan tingkat kelulusan, atau melakukan perubahan dalam kurikulum untuk meningkatkan efisiensi studi. Selain itu, dengan melakukan *clustering* data mahasiswa, kita dapat memahami berbagai faktor yang mempengaruhi kelulusan, seperti tingkat dukungan akademik yang diberikan, tingkat keberlanjutan finansial mahasiswa, dan faktor sosial lainnya.

Terdapat beberapa algoritma yang populer digunakan dalam *clustering* yaitu K-Means, K-Medoids, *Fuzzy C-Means*, DBSCAN, *Hierarchical Clustering*, dan *Mixture of Gaussians*. K-Means dan DSCAN merupakan algoritma *clustering* yang dipilih untuk diterapkan pada penelitian ini. Pemilihan penggunaan kedua algoritma tersebut didasari beberapa alasan, diantaranya adalah perbedaan karakteristik yang dimiliki kedua algoritma tersebut, dimana K-Means merupakan suatu algoritma pengelompokan non-hierarki (*partitioning*) yang dilakukan dengan cara mempartisi data ke dalam beberapa *cluster* berdasarkan kedekatan jarak antar data ke pusat

*cluster*. Sedangkan *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN) merupakan suatu algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data. Selain itu DBSCAN memiliki kemampuan yang berbeda dalam melakukan pengelompokan data jika dibandingkan dengan algoritma K-Means dan K-Medoids, dimana DBSCAN dapat digunakan untuk mendeteksi *outliers* atau *noise* pada data (Ester dkk., 1996).

Selain karena perbedaan karakteristik K-Means dan DBSCAN diatas, alasan pemilihan kedua algoritma ini juga didasari oleh perbedaan hasil yang didapatkan pada penelitian yang telah dilakukan oleh peneliti sebelumnya, diantaranya yaitu penelitian oleh Mohamed Nafuri dkk. (2022) yang berjudul *Clustering Analysis for Classifying Student Academic Performance in Higher Education*, yang melakukan penelitian dengan tujuan untuk melakukan klasifikasi mahasiswa B40 (tipe mahasiswa dengan penghasilan rendah) berdasarkan performa, perilaku, dan prestasi mereka pada HEIs (*Higher Education Institutions*). Pada penelitian disebutkan bahwa K-Means menghasilkan *cluster* terbaik jika dibandingkan dengan BIRCH dan DBSCAN. Sedangkan penelitian oleh Valarmathy dkk. (2019) dengan judul *Performance Evaluation and Comparison of Clustering Algorithms Used in Educational Data Mining*, disebutkan bahwa DBSCAN merupakan metode yang memiliki performansi terbaik dalam melakukan *clustering* jika dibandingkan dengan EM, CLOPE, DBSCAN, *Filtered Cluster*, *Farthest First*, COWEB, K-Means *Clustering* dan CLARA.

Dengan perbedaan karakteristik algoritma yang cukup kontras, serta perbedaan hasil yang didapatkan pada penelitian sebelumnya maka peneliti tertarik untuk mengangkat judul penelitian yaitu “Klasterisasi Data Mahasiswa

Undiksha Menggunakan Algoritma K-Means dan DBSCAN serta Analisis Peluang Kelulusan Mahasiswa pada Suatu *Cluster*”, yang bertujuan untuk mengelompokan data mahasiswa Undiksha berdasarkan kemiripan karakteristik data yang dimilikinya ke dalam beberapa *cluster* dengan menggunakan algoritma K-Means dan DBSCAN, melakukan analisis peluang kelulusan mahasiswa pada setiap *cluster* yang terbentuk, serta mengetahui algoritma manakah yang menghasilkan kualitas *cluster* terbaik.

## 1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut maka rumusan masalah yang akan dikaji sebagai berikut:

1. Bagaimanakah proses klasterisasi dengan menggunakan algoritma K-Means dan DBSCAN dalam melakukan pengelompokan data mahasiswa Undiksha?
2. Bagaimanakah visualisasi dan hasil interpretasi *cluster* yang telah dibentuk dengan algoritma K-Means dan DBSCAN?
3. Bagaimanakah analisis probabilitas kelulusan mahasiswa pada suatu *cluster*?
4. Berdasarkan analisis validasi *cluster* dengan menggunakan *silhouette coefficient*, algoritma manakah yang menghasilkan kualitas *cluster* terbaik?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dari penelitian adalah sebagai berikut:

1. Untuk mengetahui proses klusterisasi dengan menggunakan algoritma K-Means dan DBSCAN dalam melakukan pengelompokan data mahasiswa Undiksha.
2. Untuk mengetahui visualisasi dan hasil interpretasi *cluster* yang telah dibentuk dengan algoritma K-Means dan DBSCAN.
3. Untuk dapat mengetahui probabilitas kelulusan mahasiswa pada suatu *cluster*.
4. Untuk mengetahui algoritma manakah yang menghasilkan kualitas *cluster* terbaik.

#### 1.4 Batasan Masalah

Agar penyusunan dan pembahasan penelitian ini lebih terarah serta tidak keluar dari inti permasalahan, maka dibuat batasan masalah dalam penelitian ini, yaitu:

1. Model akan dibuat menggunakan *Python 3.10.0*.
2. *Dataset* yang digunakan merupakan data mahasiswa Universitas Pendidikan Ganesha (Undiksha) angkatan tahun 2015 s.d. 2016.
3. Metode yang digunakan untuk *clustering* data akademik dan non-akademik siswa adalah metode K-Means dan DBSCAN.
4. Evaluasi model menggunakan metode *Silhouette Coefficient*.
5. Hasil dari penelitian ini hanya terbatas pada hasil *clustering* data mahasiswa Undiksha, analisis *cluster* dan probabilitas kelulusan mahasiswa, serta perbandingan kualitas *cluster* K-Means dan DBSCAN, serta visualisasi dan analisis *cluster* yang terbentuk.

6. Probabilitas kelulusan yang dimaksud adalah hasil analisis besarnya peluang label lulus tepat waktu atau lulus tidak tepat waktu pada suatu *cluster*. Nilai peluang ini didapatkan dengan melakukan perhitungan menggunakan metode peluang sederhana dan *Softmax*.

## 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini dapat dibagi menjadi dua jenis, yaitu manfaat teoritis dan manfaat secara praktis sebagai berikut:

### 1.5.1 Manfaat Teoritis

Penelitian ini dapat menjadi sumber informasi, referensi dan materi acuan bagi peneliti lain yang tertarik untuk mengeksplorasi dan melakukan penelitian pada topik yang serupa di masa mendatang.

### 1.5.2 Manfaat Secara Praktis

#### a. Bagi Mahasiswa

Adapun manfaat yang dapat diperoleh penulis dengan melakukan penelitian ini adalah untuk memenuhi tagihan tugas akhir atau skripsi dan juga meningkatkan wawasan berpikir ilmiah serta kemampuan menganalisis suatu permasalahan khususnya pada bidang *data science*. Selain itu dengan dilakukannya penelitian ini diharapkan dapat menjadi langkah awal dalam mengembangkan penelitian-penelitian selanjutnya terkait dengan *clustering* data mahasiswa.

#### b. Bagi Civitas Akademika Undiksha

Hasil penelitian ini dapat menjadi referensi berharga bagi institusi pendidikan, terutama bagi civitas akademika Undiksha, dalam mengoptimalkan manajemen data mahasiswa. Penerapan algoritma *clustering* dapat membantu pihak

universitas dalam mengelompokkan mahasiswa berdasarkan karakteristik tertentu, sehingga memungkinkan universitas untuk memberikan layanan yang lebih terfokus dan efektif sesuai dengan kebutuhan setiap kelompok mahasiswa. Selain itu dengan eksplorasi lebih lanjut, hasil dari penelitian ini dapat digunakan untuk melakukan identifikasi dini terhadap mahasiswa yang memiliki potensi lulus tidak tepat waktu, sehingga pihak perguruan tinggi dapat membuat perencanaan studi dan bimbingan secara khusus sesegera mungkin terhadap mahasiswa yang terindikasi memiliki potensi untuk lulus tidak tepat waktu.

c. Bagi Masyarakat

Adapun manfaat penulisan proposal ini bagi masyarakat yakni dapat memberikan tambahan ilmu pengetahuan dan wawasan yang lebih luas kepada masyarakat, terutama terkait dengan penerapan algoritma K-Means dan DBSCAN dalam permasalahan *clustering* data mahasiswa. Selain itu, kedepannya penelitian ini dapat dijadikan sebagai acuan dalam penyusunan laporan penelitian terkait.

