

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perguruan tinggi merupakan lembaga pendidikan tertinggi yang semakin banyak peminatnya di Indonesia. Menurut Kementerian Pendidikan Dan Kebudayaan (2015), perkembangan data statistik Perguruan Tinggi menunjukkan bahwa jumlah lembaga Perguruan Tinggi, mahasiswa baru, mahasiswa lulusan, dan dosen menunjukkan perkembangan meningkat tiap tahunnya.

Kelulusan mahasiswa yang tepat waktu akan menguntungkan pihak mahasiswa dan perguruan tinggi. Untuk mahasiswa, semakin cepat lulus maka kesempatan untuk mengikuti berbagai seleksi dalam mencari pekerjaan semakin banyak. Sedangkan bagi perguruan tinggi, kelulusan mahasiswa dengan tepat waktu dapat memajukan kualitas, meningkatkan reputasi, dan juga berpengaruh pada akreditasi perguruan tinggi tersebut. (Premashanti, Gita. 2015) Pada penelitian ini, digunakan teknik data *mining* untuk menemukan pola kelulusan mahasiswa yang sudah lulus, kemudian dijadikan dasar untuk memprediksi kelulusan mahasiswa pada tahun berikutnya.

*Data mining* adalah proses menemukan hubungan dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipahami sehingga hubungan tersebut dapat menjadi dasar klasifikasi data. Untuk mengetahui tingkat kelulusan mahasiswa dalam satu tahun ajaran dapat dilakukan suatu prediksi berdasarkan data-data mahasiswa pada tingkat atau tahun ajaran pertama. Ada beberapa faktor yang mempengaruhi prediksi kelulusan mahasiswa yang sesuai dengan waktu studi, diantaranya : rata-rata IPK terakhir, jumlah SKS.(Kondo, Ferry.2015) Dalam menentukan prediksi kinerja mahasiswa dengan akurasi yang tinggi bermanfaat untuk mengidentifikasi siswa dengan prestasi akademik (Suhartinah dan Emastuti, 2010).

Untuk menyelesaikan permasalahan pada penjelasan diatas adalah dengan pemanfaatan algoritma *K-Means Clustering*. Algoritma *K-Means* adalah algoritma klastering yang paling sederhana dibanding algoritma klastering yang lain. Algoritma ini mempunyai kelebihan mudah diterapkan dan dijalankan, relatif cepat, mudah untuk diadaptasi, dan paling banyak dipraktekkan dalam tugas data mining. *Clustering* merupakan suatu metode untuk pengelompokan dokumen dimana dokumen dikelompokan dengan konten untuk mengurangi ruang pencarian yang diperlukan dalam merespon suatu *query* (Natalius, 2010).

Algoritma *K-Means* merupakan algoritma yang membutuhkan parameter input sebanyak  $k$  dan membagi sekumpulan  $n$  objek kedalam *cluster* sehingga tingkat kemiripan antar anggota dalam satu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah. Kemiripan anggota terhadap *cluster* diukur dengan kedekatan objek terhadap nilai *mean* pada *cluster* atau disebut sebagai *centroid cluster* (Rismawan dan Kusumadewi, 2008). Data nilai mahasiswa hanya akan menjadi sekumpulan data yang tidak berguna jika tidak dilakukan penggalian data terhadapnya. Banyak informasi terpendam yang dapat diambil dari sekumpulan data tersebut sehingga dapat memberikan suatu pengetahuan untuk penentuan kebijakan. Penggalian data dapat dilakukan dengan cara pengelompokan data nilai mahasiswa menjadi beberapa kelompok, kelompok nilai baik dan nilai buruk. Karena itu, pada penelitian ini akan dikembangkan suatu sistem yang dapat mengklastering kelulusan mahasiswa menggunakan algoritma *K-Means* dengan melihat pola kelulusan mahasiswa beberapa periode sebelumnya.

## 1.2 Rumusan Masalah

1. Bagaimana hasil prediksi sistem dalam menentukan kelulusan mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember?
2. Seberapa besar tingkat akurasi algoritma *K-Means* dalam memprediksi ketepatan kelulusan mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember?

## 1.3 Tujuan

1. Mampu memprediksi sistem mengenai kelulusan mahasiswa Jurusan Teknik Informatika di Universitas Muhammadiyah Jember, apakah bernilai lulus cepat, lulus tepat dan lulus terlambat.
2. Mengetahui tingkat akurasi algoritma *K-Means* dalam memprediksi ketepatan kelulusan mahasiswa di Universitas Muhammadiyah Jember.

## 1.4 Batasan masalah

1. *Dataset* berupa data mahasiswa Jurusan Teknik Informatika angkatan 2011.
2. Kriteria mahasiswa yang digunakan yaitu :
  - a. Rata-rata IPK terakhir
  - b. Jumlah SKS
3. Menggunakan bahasa pemrograman PHP dan *database* MySQL.
4. Hasil Klasifikasi yaitu lulus cepat, lulus tepat dan lulus terlambat untuk angkatan 2011.
5. *Ceteroid* telah diketahui sebagai tolak ukur dalam penentuan klasifikasi dan apabila mendekati atau sama dengan *ceteroid* tersebut maka proses berhenti, berikut ini adalah *threshold* (batasan) dari *ceteroid* :

<i>Ceteroid</i>	IPK	SKS
<i>Centroid 1</i>	3.44	142.25
<i>Centroid 2</i>	3.14	130.25
<i>Centroid 3</i>	2.86	115.00

### **1.5 Manfaat**

1. Membantu administrasi perguruan tinggi untuk memberikan peringatan dini dan pembimbingan awal bagi mahasiswa yang kemungkinan tidak lulus tepat waktu.
2. Membantu perguruan tinggi dalam membuat kebijakan untuk bisa meningkatkan kelulusan mahasiswa.