BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Salah satu cara untuk mengetahui faktor nilai *cumlaude* mahasiswa Fakultas Teknik Universitas Muhammadiyah Yogyakarta adalah dengan menerapkan metode *clustering* dengan algoritma *K-Means* untuk penelitiannya.

Berdasarkan beberapa kajian penulis terhadap penelitian terdahulu yang membahas mengenai analisis faktor nilai *cumlaude* mahasiswa Fakultas Teknik Universitas Muhammadiyah Yogyakarta sejauh pemahaman penulis belum pernah diteliti. Namun, ada beberapa penelitian yang membahas tentang analisa faktor-faktor tingginya nilai indeks prestasi mahasiwa yang sudah pernah dilakukan, diantaranya adalah:

Daruyani dkk. (2013) dengan judul penelitian, Faktor-faktor yang mempengaruhi indeks prestasi mahasiswa FSM Universitas Diponegoro semester pertama dengan metode regresif *logistik biner*. Dalam penelitian ini di jelaskan bahwa untuk mengetahui faktor-faktor apa saja yang mempengaruhi indeks prestasi mahasiswa dapat menggunakan analisis regresi logistik biner karena variable respon yang diamati terdiri dari satu variabel. Estimasi parameter model menggunakan fungsi maksimum likelihood. Untuk menguji signifikansi dari parameter-parameter menggunakan uji rasio likelihood dan uji wald. Setelah dilakukan pengujian secara keseluruhan terhadap variabel prediktor nilai rapor, nilai UN, jalur masuk pilihan jurusan, tempat tinggal, metode belajar, biaya hidup

perbulan, hubungan mahasiswa dengan teman, hubungan mahasiswa dengan keluarga serta motivasi belajar semua variabel ini signifikan mempengaruhi indeks prestasi mahasiswa. Setelah pengujian secara individu variabel nilai UN dan hubungan mahasiswa dengan teman signifikan mempengaruhi indeks prestasi mahasiswa.

Daely (2013) dengan judul penelitian, Analisis stastistik faktor-faktor yang mempengaruhi indeks prestasi mahasiswa. Dalam penelitian ini menjelaskan Terdapat beberapa faktor yang mempengaruhi indeks prestasi, khususnya di prodi S1 Matematika FMIPA USU. Dengan metode analisis faktor diperoleh empat faktor yang mempengaruhi indeks prestasi mahasiswa S1 Matematika FMIPA USU yaitu, Faktor Lingkungan dan Pengawasan Orang Tua, Faktor Kondisi Finansial dan Motivasi Belajar, Faktor Kualitas Belajar dan Pembagian Waktu Belajar, dan Faktor Kualitas Pengajaran Dosen dan Kesehatan Mahasiswa.

Nugroho dkk. (2013) dengan judul penelitian, Penerapan Algoritma C4.5 untuk klasifikasi predikat kelulusan mahasiswa fakultas komunikasi dan informatika Universitas Muhammadiyah Surakarta. Penelitian data *mining* dilakukan untuk memanfaatkan data yang melimpah sebagai sumber daya strategis untuk Fakultas dan departemen untuk mengklasifikasikan 'Tingkat keunggulan menggunakan data *mining techniques*. Mahasiswa unggulan itu diklasifikasikan menggunakan algoritma C4.5. Jumlah sampel ditentukan dengan menggunakan persamaan Slovin. Ada 341 Data mahasiswa yang diambil dari total 2.358 mahasiswa FKI yang telah lulus sebagai data harus diklasifikasikan. Pengolahan data dilakukan pada pemisahan atribut yang dibutuhkan untuk proses data *mining*,

standarisasi data (preprocessing), dan konversi data real ke data nominal. Atribut yang digunakan terdiri dari sekolah utama (setara dengan SMA), jenis kelamin, rumah sekolah, rata-rata jumlah sks per semester, dan peran asisten yang dianggap penting dalam mempengaruhi tingkat siswa unggulan. Hasil penelitian menunjukkan bahwa yang variabel mempengaruhi tngginya nilai mahasiswa adalah partisipasi mereka sebagai asisten dengan akurasi 73,91%. Itu Hasil penelitian menunjukkan bahwa variabel yang digunakan sebagai pertimbangan untuk fakultas mendapatkan maksimum tingkat keunggulan partisipasi siswa menjadi asisten.

Dessy Purnama Sari dkk. (2014) dengan judul ”Analsis cluster menggunakan algoritma *K-Means* untuk mengelompokan siswa kelas IV sekolah dasar Brawijaya *smart school* Malang”. Analisis *cluster* merupakan metode pengelompokan multivariat dengan tujuan utama yaitu mengelompokan objek atau subjek berdasarkan kemiripan karakteristik yang dimiliki. Analisis *cluster* memiliki homogenitas (kesamaan) yang tinggi antar anggota dalam satu kelompok (*within cluster*) dan heterogenitas (perbedaan) yang tinggi antar kelompok satu dengan kelompok lain (*between cluster*) (Hair dkk., 2010). Metode analisis *cluster* menggunakan algoritma *K-Means* adalah: menentukan jumlah kelompok yang akan dibentuk sebanyak 2 kelompok, menentukan titik pusat awal kelompok, menghitung jarak setiap objek pada setiap pusat kelompok dengan menggunakan jarak *Mahalanobis*, mengelompokan objek berdasarkan jarak terdekat dengan pusat kelompok, menentukan pusat kelompok baru dengan menghitung rata-rata pada setiap kelompok, menghitung kembali jarak setiap objek pada pusat kelompok dan

mengelompokannya hingga tidak ada objek yang berpindah dari kelompok, dan melakukan interpretasi karakteristik pada masing-masing kelompok.

Penulis juga mengambil makalah yang di tulis oleh Narwati (2014) dengan judul "Pengelompokan mahasiswa menggunakan algoritma *K-Means*". *K-means* merupakan salah satu metode data *clustering non hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster / kelompok. Metode ini mempartisi ke dalam cluster / kelompok sehingga data yang memiliki karakteristik yang sama (*High intra class similarity*) dikelompokkan ke dalam satu cluster yang sama dan yang memiliki karakteristik yang berbeda (*Law inter class similarity*) dikelompokkan pada kelompok yang lain [3]. Proses *clustering* dimulai dengan mengidentifikasi data yang akan dikluster, ***Xij*** (i=1,...,n; j=1,...,m) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap kluster ditetapkan secara bebas (sembarang), ***Ckj*** (k=1,...,k; j=1,...,m). Kemudian dihitung jarak antara setiap data dengan setiap pusat kluster. Untuk melakukan penghitungan jarak data ke-I (xi) pada pusat kluster ke-k (ck), diberi nama (***dik***), dapat diperoleh berdasarkan data penerimaan mahasiswa baru dan data akademik yang dapat diperoleh dari server Unisbank. Berdasarkan proses data *mining* dengan teknik *clustering* menggunakan algoritma *K-Means* yang diterapkan pada data akademik mahasiswa, diperoleh informasi dari pengelompokan atau pengklusteran nilaiTes mahasiswa saat masuk dari sejumlah 936 mahasiswa adalah sejumlah 116 mahasiswa atau sebesar 12,393% masuk kluster 1, 363 (38,782%) mahasiswa masuk kluster 2 dan 457 (48,825%) mahasiswa masuk kluster 3. Hal ini berarti hampir sebagian besar

kemampuan mahasiswa saat masuk kuliah adalah masuk kluster 3, atau berada pada kemampuan paling atas

2.2 Landasan Teori

2.2.1 Data *Mining*

Data *mining* adalah proses yang menggunakan teknik statistik,matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Istilah data *mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah hak untuk menemukan,menggali,atau menambang pengetahuan dari data atau informasi yang kita miliki. *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database (*KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan , pola atau hubungan dalam set data berukuran besar [6].

Menurut Fayyad dalam buku (Kusrini, 2009) Istilah data *mining* dan knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data *mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut: (Narwati, 2010)

1. Data *selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data dari hasil

il seleksi yang akan digunakan untuk proses data *mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing/Cleaning*

Sebelum proses data *mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkosisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). juga dilakukan proses *enrichement*, yaitu proses "memperkaya" data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

3. *Transformation*

Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data *mining*. Proses coding dalam KDD merupakan proses kreatif dan sangat bergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

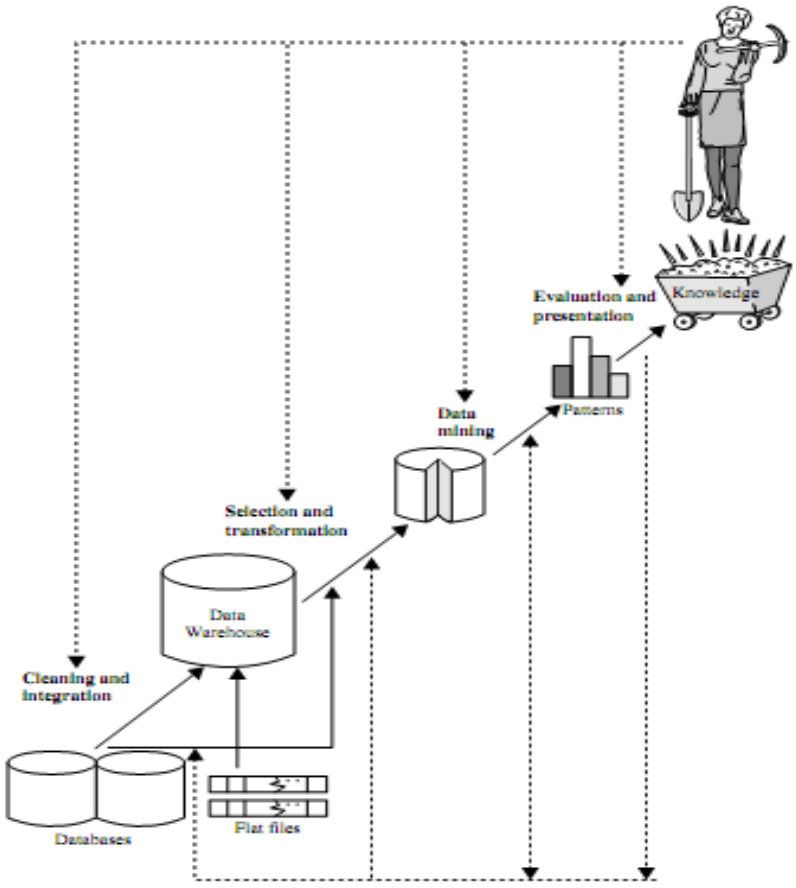
4. Data *mining*

Data *mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data *mining* sangat bervariasi. Pemilihan metode atau

algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation/Evalution*

Pola informasi yang dihasilkan dari proses data *mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.



Gambar 2.1 Proses data *mining* (Ridwan M. dkk 2013)

2.2.2 Clustering

Salah satu metode yang diterapkan dalam KDD adalah *clustering*. *Clustering* adalah membagi data ke dalam grup-grup yang mempunyai obyek yang karakteristiknya sama. Garcia-Molina [7] menyatakan *clustering* adalah mengelompokan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai suatu persamaan yang esensial.

Clustering memegang peranan penting dalam aplikasi data *mining*, misalnya eksplorasi data ilmu pengetahuan, pengaksesan informasi dan text *mining*, aplikasi basis data spesial, dan analisis weka. *Clustering* diterapkan dalam mesin pencari di internet. Web mesin pencariakan mencari ratusan dokumen yang cocok dengan kata kunci yang dimasukan. Dokumen-dokumen tersebut dikelompokan dalam cluster-cluster sesuai dengan kata-kata yang di gunakan (Anggriani S. 2007)

2.2.3 K-Means

K-Means merupakan salah satu metode pengelompokkan data nonhierarki yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang di *set* dalam suatu kelompok dan memaksimalkan variasi antar kelompok (Eko Prasetyo, 2012:178).

Pengertian dari *K-Means clustering* adalah *K-Means* dimaksudkan sebagai konstanta jumlah *cluster* yang diinginkan, *K-Means* dalam hal ini berarti nilai suatu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai *cluster*, sehingga *K-Means clustering* adalah suatu metode penganalisaan data atau metode data *mining* yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode *K-Means* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam suatu kelompok mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dasar algoritma *K-Means* adalah sebagai berikut:

1. Tentukan nilai *k* sebagai jumlah klaster yang ingin dibentuk.
2. Inisialisasi *k* sebagai *centroid* yang dapat dibangkitkan secara random.
3. Hitung jarak setiap data ke masing-masing *centroid* menggunakan persamaan *Euclidean Distance* yaitu sebagai berikut:

$$\boldsymbol{d(P,Q) = \sqrt{\sum_{j=1}^{p} (x_j(P) - x_j(Q))^2}}$$

4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan *centroid*-nya.
5. Tentukan posisi *centroid* baru (*k*).
6. Kembali ke langkah 3 jika posisi *centroid* baru dengan *centroid* lama tidak sama.

2.2.4 Weka

Weka adalah aplikasi data *mining open source* berbasis Java. Aplikasi ini dikembangkan pertama kali oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka terdiri dari koleksi algoritma machine learning yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data sampling. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data *mining* tetap terletak pada faktor pengetahuan manusia implementornya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan Weka mendukung beberapa file untuk *input*-nya, diantaranya:

- Comma Separated Values (CSV)
- Format C45
- Attribute-Relation File format (ARFF)

.Feri Sulianta dan Dominikus Juju (2010) *Data Mining: Meramalkan Bisnis Perusahaan*. Jakarta Elix Media Komputindo.