

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data Mining atau disebut juga *Knowledge Discovery in Database (KDD)* adalah proses pengekstrasian pengetahuan dari data yang berukuran besar untuk menemukan pola-pola baru sehingga mendapatkan sebuah pengetahuan dan informasi baru. Teknologi *Data Mining* digunakan untuk menggali pengetahuan yang ada didalam database. Salah satu metode yang digunakan dalam *Data Mining* adalah metode *clustering* dengan algoritma *K-Means*.

Algoritma *K-Means Clustering* adalah teknik dalam *Data Mining* yang mengelompokkan item-item data yang mirip kedalam satu kelompok atau *cluster*. Algoritma *K-Means* melakukan dua proses yaitu proses penentuan pusat *cluster* (*centroid*) dan proses pencarian anggota dari tiap-tiap *cluster*. Hasanah, et al. (2019) dalam penelitian *Clustering* dengan algoritma *K-Means* melakukan pengelompokan terhadap data nilai rapor siswa dengan data sampel sebanyak 23 siswa yang menghasilkan 3 *cluster* tingkat kompetensi siswa untuk pemilihan peserta Olimpiade Sains Nasional.

K-Means pada penelitian terdahulu lainnya digunakan sebagai metode pengelompokan siswa ke dalam beberapa kelas dalam suatu mata kuliah berdasarkan nilai tugas, nilai ujian tengah semester, nilai ujian akhir semester, dan indeks prestasi kumulatif (IPK) semester mahasiswa. Kemudian diperoleh jumlah *cluster* atau kelas dan jumlah data yang digunakan mempengaruhi kualitas *cluster* yang dibentuk (Pradnyana dan Permana, 2018). Penelitian lain dilakukan oleh Mardalius (2018), dimana metode *Clustering* dengan algoritma *K-Means* digunakan untuk mengelompokkan data siswa berdasarkan nilai ujian semester mata pelajaran Ujian Nasional. Metode ini melakukan pengelompokan untuk belajar tambahan siswa yang akan mengikuti Ujian Nasional.

Penelitian dengan mengambil data dari UCI (University California Irvine) dengan metode *K-Means Clustering and Multiple Linear Regressions (MLR)*, dimana

set data didasarkan dari nilai matematika siswa, nilai ulangan, kuis, dan tugas siswa sebagai faktor utama dalam memprediksi kinerja akademik siswa (Omolewa, et al. 2019). Sedangkan Amirullah (2019) dengan Algoritma *Clustering K-Means* dan *Algoritma Greedy*, dengan data 37 siswa menghasilkan kelompok yang terbagi pada 5 *cluster* kualitas siswa, sehingga dari *cluster* tersebut terbentuk kelompok kerja siswa untuk pemerataan kualitas siswa.

Pasina, et al. (2019) mengklasifikasikan mahasiswa ke dalam kelompok sesuai dengan preferensi gaya belajar mereka. *The Felder dan Silverman's Index Learning Style (ILS)* digunakan sebagai metode untuk mengevaluasi preferensi gaya belajar siswa. Instruktur di pendidikan tinggi dapat mengadopsi pendekatan pengajaran yang sesuai untuk meningkatkan kinerja akademik dengan mengetahui gaya belajar siswa.

Sekolah merupakan lembaga pendidikan paling strategis untuk mewujudkan tujuan pendidikan nasional sebagaimana tertera dalam Undang-Undang Republik Indonesia Nomor 20 Tahun 2003. Kualitas sekolah dan peserta didik dapat diukur dari kualitas pembelajaran yang dilaksanakan oleh guru. Sekolah dengan latar belakang siswa yang beragam perlu melakukan langkah-langkah strategis dalam mengelola pembelajaran untuk mendapatkan hasil belajar yang optimal.

Desain dan strategi pembelajaran yang tepat dapat meningkatkan motivasi belajar siswa. Peranan guru sangat penting dalam mengelola pembelajaran dan menciptakan suasana belajar yang kondusif. Metode *card sort* digunakan untuk melihat peningkatan keaktifan siswa dari rata-rata presentase seluruh indikator keaktifan siswa dalam pembelajaran (Hanifah dan Wulandari,2018).

Implementasi *Data Mining* dengan metode *Algoritma K-Means Clustering* diatas memiliki kesamaan dengan masalah yang akan penulis bahas. Klasterisasi siswa di SMK Negeri 2 Padang Panjang selama ini dilakukan secara manual oleh guru Bimbingan dan Konseling (BK). Algoritma *K-Means* dapat melakukan klasterisasi siswa secara sistematis. Penulis mengangkat judul penelitian tentang **“Klasterisasi Penempatan Siswa yang Optimal untuk Meningkatkan Nilai Rata-Rata Kelas Menggunakan *K-Means*”**.

1.2 Perumusan Masalah

Berdasarkan permasalahan yang ada, supaya tesis ini sesuai dengan tujuan yang ingin dicapai, maka penulis merumuskan beberapa permasalahan sebagai berikut :

1. Bagaimana cara melakukan klasterisasi penempatan siswa dengan metode *K-Means*?
2. Bagaimana menerapkan Algoritma *K-Means* untuk klasterisasi penempatan siswa yang optimal?
3. Bagaimana implementasi Algoritma *K-Means* dan *software RapidMiner versi 9.2* untuk penempatan siswa secara optimal guna mendapatkan komposisi kelas yang seimbang.

1.3 Batasan Masalah

Agar pembahasannya pada penelitian ini tidak menyimpang maka penulis membatasi ruang lingkup objek penelitian. Adapun ruang lingkup penelitian antara lain:

1. Penelitian dilakukan di SMK Negeri 2 Padang Panjang dimana yang akan diteliti adalah cara klasterisasi penempatan siswa yang optimal untuk menaikkan nilai rata-rata kelas. Data yang diolah adalah data 90 siswa Kompetensi Keahlian Teknik Komputer Jaringan kelas XI.
2. Sistem yang digunakan adalah *Data Mining* dengan menggunakan metode klasterisasi *K-Means*.
3. Implementasi dari Algoritma *K-Means* ini akan dapat dilihat dengan *Software RapidMiner versi 9.2*.

1.4 Tujuan Penelitian

Tujuan yang ingin diperoleh dari penelitian ini agar lebih bermanfaat kedepannya adalah:

1. Guru dapat melakukan klasterisasi penempatan siswa dengan metode *K-Means* untuk mendapatkan komposisi kelas yang seimbang guna meningkatkan nilai rata-rata kelas.
2. Menerapkan cara kerja Algoritma *K-Means* untuk mengambil keputusan yang tepat dalam klasterisasi penempatan siswa.
3. Menguji metoda Algoritma *K-Means* untuk klasterisasi siswa dengan menggunakan *software RapidMiner versi 9.2*.
4. Untuk pembuktian metoda Algoritma *K-Means* untuk klasterisasi penempatan siswa dilakukan perancangan program dengan bahasa pemrograman PHP.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat kedepannya, yang beberapa diantaranya adalah:

1. Sebagai acuan dan pedoman bagi guru dalam melakukan klasterisasi kelompok belajar siswa.
2. Mempermudah guru dan pihak sekolah dalam menentukan kelompok belajar siswa.
3. Bagi peneliti sendiri, bisa menambah pengetahuan yang bermanfaat untuk dapat diaplikasikan dalam proses pembelajaran dikelas.

1.6 Sistematika Penulisan

Sistematika yang digunakan dalam penyusunan thesis ini adalah sebagai berikut:

Bab I PENDAHULUAN

Bab ini berisikan tentang latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

Bab II LANDASAN TEORI

Bab ini berisikan tentang penerapan *Knowledge Discovery in Database (KDD)*, *Data Mining*, Algoritma *K-Means Clustering* dan sekilas tentang nilai rata-rata siswa.

Bab III METODE PENELITIAN

Bab ini berisi kerangka kerja, mendefinisikan ruang lingkup masalah, analisa masalah, menentukan tujuan, mempelajari literatur, mengumpulkan data dan informasi, menganalisa dan menerapkan *Data Mining*, mengimplementasikan *Algoritma K-Means Clustering*, menguji data dan menarik kesimpulan.

Bab IV ANALISA DAN PERANCANGAN

Bab ini berisi tentang analisa permasalahan, analisa kebutuhan sistem, menerapkan sistem *Algoritma K-Means Clustering*, menentukan variabel dalam penghitungan hasil untuk penentuan klasterisasi penempatan siswa.

Bab V IMPLEMENTASI SISTEM

Bab ini menjelaskan tentang implementasi *Algoritma K-Means Clustering*, kebutuhan perangkat keras dan lunak, pengujian dengan *Data Testing*, untuk klasterisasi penempatan siswa.

Bab VI KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan tentang penerapan algoritma *K-Means Clustering*, menggambarkan kekurangan dan kelemahan metode yang digunakan.