

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Knowledge Discovery in Database (KDD) seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi yang tersembunyi dalam suatu basis data yang besar. Salah satu tahapan dalam *Knowledge Discovery in database* (KDD) adalah *Data Mining*. *Data Mining* adalah proses mencari pola atau informasi dalam basis data dengan menggunakan teknik atau metode tertentu (Tri Wahyudi, *et al.* 2017). Salah satu metode yang dapat digunakan dalam *Data Mining* adalah *metode Clustering* dengan algoritma *K-Means*. Algoritma *K-Means Clustering* adalah teknik dalam *Data Mining* yang mempartisi data yang ada ke dalam beberapa *cluster* sehingga data yang memiliki karakteristik yang sama akan dikelompokkan ke dalam satu *cluster* dan karakteristik yang berbeda di *cluster* lain (Anggoro Eko Wicaksono, 2016).

Data Mining dapat diimplementasikan dalam pengelompokan data-data di berbagai kasus, seperti pengelompokan data-data penyakit berdasarkan tingkatan keseriusan penyakitnya (Sulastri dan Gufroni, 2017). *Data Mining* dengan algoritma *K-Means Clustering* juga pernah dikembangkan untuk membantu menentukan calon penerima beasiswa bidik misi di Polbeng (Jaroji, *et al.* 2016). Menurut Panchoo, *et al.* (2016) pengelompokan fitur data berdimensi tinggi yang bertujuan untuk mengurangi

beban komputer berlebihan, juga bisa memanfaatkan algoritma *K-Means Clustering*. Hasil penelitian ini mampu mencapai tingkat keakuratan sebesar 80%. Algoritma *K-Means* juga dapat diimplementasikan dengan algoritma lain, salah satunya Algoritma Genetika. Perpaduan dua algoritma ini menunjukkan hasil bahwa algoritma *K-Means* yang diimplementasikan dengan algoritma genetika dapat meningkatkan tingkat keakuratan hasil klasifikasi data sebuah *dataset* (Husein, Asraa Abdullah, 2018)

Melimpahnya jumlah siswa dalam suatu institusi pendidikan khususnya Sekolah Menengah Atas (SMA) menyebabkan data-data siswa dalam sistem juga melimpah. Hal itu mengakibatkan pihak sekolah kesulitan dalam mengelola data yang banyak secara manual dan sering keputusan yang diambil kurang efektif. Pemilihan peserta Olimpiade Sains Nasional (OSN) juga membutuhkan proses seleksi akurat dari pihak sekolah ataupun guru agar hasil yang didapat juga memuaskan. Proses pemilihan calon peserta Olimpiade Sains Nasional (OSN) dilakukan dengan memanfaatkan data-data nilai siswa pada beberapa semester sebelumnya.

Penelitian ini dilakukan untuk mencari kelompok nilai akademik siswa dengan cara menghitung jarak terdekat antara data dengan titik tengah (*centroid*) suatu *cluster*. Sehingga pengelompokan tersebut akan menghasilkan beberapa klaster siswa yaitu siswa yang sangat berkompeten, berkompeten, dan kurang berkompeten untuk mengikuti Olimpiade Sains Nasional (OSN). Melalui penelitian ini dengan menggunakan algoritma *K-Means Clustering* diharapkan dapat membantu pihak sekolah dan guru dalam mengelompokkan siswa berdasarkan nilai akademik dengan hasil yang efektif dan efisien.

1.2. Perumusan Masalah

Sebelumnya pemilihan peserta Olimpiade Sains Nasional di SMA Negeri 5 Sijunjung dipilih secara manual, yang mana kemungkinan siswa yang terpilih kurang tepat sasaran lumayan besar. Maka dalam hal ini, dapat dirumuskan beberapa masalah sebagai berikut:

1. Bagaimana menerapkan algoritma *K-Means Clustering* mampu mengelompokkan siswa berdasarkan nilai rapor akademik agar kecenderungan terhadap kesalahan dalam pemilihan peserta Olimpiade Sains Nasional (OSN) dapat dikurangi?
2. Bagaimana menerapkan algoritma *K-Means Clustering* dengan menggunakan aplikasi *RapidMiner* mampu mengelompokkan siswa berdasarkan nilai rapor akademik agar kecenderungan terhadap kesalahan dalam pemilihan peserta Olimpiade Sains Nasional (OSN) dapat dikurangi?

1.3. Batasan Masalah

Agar penulis tidak keluar dari permasalahan yang ada dan hasil penelitian lebih terarah dan tepat sasaran, maka penulis membatasi ruang lingkup permasalahan yang dapat diuraikan sebagai berikut:

1. Algoritma yang digunakan dalam melakukan *Clustering* data siswa adalah algoritma *K-Means*.
2. Data yang dijadikan rujukan dalam penelitian ini adalah data nilai rapor semester I, II kelas X dan nilai rapor semester I kelas XI di jurusan IPA SMA Negeri 5 Sijunjung

3. Pengelompokan (*Clustering*) yang digunakan dalam pemilihan peserta Olimpiade Sains Nasional (OSN) terdapat beberapa *cluster* yaitu (C0) Sangat Berkompeten; (C1) Berkompeten; (C2) Kurang Berkompeten.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk menerapkan algoritma *K-Means Clustering* dalam mengelompokkan siswa berdasarkan nilai rapor akademik sehingga mempermudah dalam memilih peserta Olimpiade Sains Nasional (OSN) secara tepat.
2. Untuk mengelompokkan siswa yang sangat berkompeten, berkompeten, dan kurang berkompeten untuk mengikuti Olimpiade Sains Nasional (OSN) berdasarkan nilai rapor akademik.
3. Untuk menerapkan algoritma *K-Means* dalam mengelompokkan nilai siswa dengan menggunakan aplikasi *RapidMiner*.
4. Untuk menguji sejauh mana *Data Mining* dengan algoritma *K-Means* ini mampu membantu pihak sekolah serta mengurangi kecenderungan kesalahan dalam pemilihan peserta Olimpiade Sains Nasional (OSN).

1.5. Manfaat Penelitian

Manfaat yang akan diperoleh dari penelitian ini adalah:

1. Dapat mempermudah pihak sekolah dan guru di SMA Negeri 5 Sijunjung dalam memilih siswa yang memenuhi kriteria untuk mengikuti Olimpiade Sains Nasional (OSN).

2. Dapat membantu mengelompokkan siswa yang berkompeten maupun yang kurang berkompeten secara efektif dan efisien.
3. Dapat menyelesaikan permasalahan pada sistem pemilihan peserta Olimpiade Sains Nasional (OSN) yang berjalan saat ini, sehingga kesalahan dalam pemilihan peserta dapat dikurangi dan hasil yang didapatkan maksimal.

4.1. Sistematika Penulisan

Sistematika dalam penulisan dalam penelitian ini dapat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini akan dijelaskan mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan pada penelitian ini.

BAB II LANDASAN TEORI

Bab ini berisi tentang kajian literatur dari jurnal-jurnal yang bertujuan untuk menjelaskan penerapan, prosedur, pengertian dan hal lainnya yang berhubungan dengan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang kerangka kerja atau tahapan-tahapan kerja yang digunakan selama penelitian, serta menjabarkan tahapan pengolahan data dengan metode *Clustering* untuk membantu mengelompokkan calon peserta olimpiade yang akan mewakili sekolah.

BAB IV ANALISA DAN PERANCANGAN

Bab ini menjelaskan bagaimana melakukan analisa menggunakan Data Mining dengan algoritma *K-Means* dan menjelaskan tahapan perhitungan sesuai dengan tahapan algoritma *K-Means* untuk mengklaster siswa ke dalam 3 klaster sehingga dapat membantu pihak sekolah dalam pemilihan siswa yang berpotensi untuk mengikuti Olimpiade Sains Nasional (OSN).

BAB V IMPLEMENTASI DAN HASIL

Bab ini berisi tentang penerapan algoritma *K-Means* serta pengujian dengan bantuan *software RapidMiner*.

BAB VI KESIMPULAN DAN SARAN

Bab ini berisi hasil dari penelitian berupa kesimpulan dan saran untuk pengembangan penelitian selanjutnya.