

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG PENELITIAN

Pendidikan adalah hal yang penting dalam kehidupan seseorang. Dengan adanya pendidikan, seseorang akan mendapatkan pemahaman tentang ilmu pengetahuan yang terus menerus berkembang serta mencerdaskan generasi penerus bangsa yang nantinya akan menjadi penggerak Indonesia. Kota Semarang memiliki 47 SMP Negeri dan 170 SMP Swasta sebagai pilihan ketika siswa-siswi SD (Sekolah Dasar) telah menyelesaikan ujian dan dinyatakan lulus di SD dimana mereka bersekolah saat ini. Dengan banyaknya pilihan SMP di kota Semarang, diperlukan sebuah cara untuk menentukan predikat sebuah sekolah. Salah satunya adalah meninjau dari nilai Ujian Nasional (UN).

UN menurut Departemen Pendidikan Nasional (Depdiknas) adalah sistem evaluasi standar pendidikan dasar dan menengah secara nasional dan persamaan mutu tingkat pendidikan antar daerah yang dilakukan oleh Pusat Penilaian Pendidikan di Indonesia berdasarkan Undang-Undang Republik Indonesia nomor 20 tahun 2003. Dengan adanya UN, pemerintah dapat mengukur pencapaian hasil belajar peserta didik, mengukur mutu pendidikan di tingkat nasional dan mempertanggungjawabkan

penyelenggaraan pendidikan kepada masyarakat sesuai dengan yang tercantum pada Keputusan Menteri Pendidikan Nasional Nomor 153/U/2003 tentang Ujian Akhir Nasional.

Data mining adalah gabungan beberapa disiplin ilmu komputer (ACM 2006) yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan data yang meliputi metode-metode yang merupakan bagian dari *Artificial Intelligence*, *Machine Learning*, statistika dan sistem basis data (ACM 2006). Data mining juga dapat didefinisikan sebagai teknik yang cepat dan mudah dalam menemukan pengetahuan, pola dan relasi antar data secara otomatis (Suyanto, 2018). Sebagai contoh, data mining dapat digunakan untuk menemukan pola penggunaan internet di suatu negara dalam rentang waktu tertentu. Tentunya jumlah data akses internet akan sangat banyak dan akan menghabiskan waktu yang sangat lama jika analisa dilakukan secara konvensional.

Data mining memiliki berbagai jenis tugas, beberapa diantaranya adalah *Classification* dan *Clustering*. *Clustering* adalah metode partisi satu set objek data ke dalam suatu himpunan bagian yang disebut dengan cluster tanpa perlu diberitahu label kelasnya. *Clustering* juga dapat digunakan untuk himpunan data yang sudah diketahui kelasnya. *Clustering* terbagi menjadi beberapa algoritma salah satunya adalah *Hierarchical Agglomerative Clustering*. *Hierarchical Agglomerative Clustering* adalah sebuah metode clustering secara hirarki berdasarkan

tingkatan tertentu sehingga menyerupai bentuk pohon dimana proses pengelompokannya dilakukan secara bertingkat atau bertahap yang menggunakan strategi *agglomerative*. Strategi *Agglomerative* atau *bottom-up* adalah strategi pengelompokan data yang dimulai dari sebuah data tunggal yang dijadikan cluster lalu digabungkan menjadi cluster yang lebih besar

Data mining dapat menemukan pola antar data untuk mendapatkan informasi secara cepat sehingga data mining dapat digunakan untuk mendapatkan informasi pengelompokan SMP sekota Semarang dari rerata nilai UN. Data rerata nilai UN yang meliputi 4 mata pelajaran yaitu Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA akan dianalisa dengan menggunakan algoritma *Hierarchical Agglomerative Clustering* dengan 2 metode yang berbeda yaitu *Complete Linkage* dan *Average Linkage* kemudian hasilnya akan dibandingkan.

1.2. PERUMUSAN MASALAH

Berdasarkan latar belakang di atas dapat dirumuskan permasalahan dalam penelitian ini adalah :

Menganalisa rerata nilai UN SMP sekota Semarang dengan menggunakan algoritma *Hierarchical Agglomerative Clustering*

Adapun batasan masalah yang dibahas pada skripsi ini adalah sebagai berikut :

1. Data yang digunakan adalah rerata nilai UN dari 217 SMP sekota Semarang tahun 2018/2019 dari 4 mata pelajaran yaitu bahasa Indonesia, bahasa Inggris, Matematika dan IPA
2. Algoritma yang digunakan adalah *Hierarchical Agglomerative Clustering* dengan 2 metode yaitu *Complete Linkage* dan *Average Linkage* sebagai perbandingan
3. Tools atau aplikasi yang digunakan adalah R-Studio

1.3. TUJUAN DAN MANFAAT PENELITIAN

1.3.1. TUJUAN

1. Menganalisa rerata nilai UN SMP sekota Semarang tahun ajaran 2018/2019 dengan menggunakan algoritma *Hierarchical Agglomerative Clustering*.
2. Membandingkan hasil metode *Hierarchical Agglomerative Clustering Complete Linkage* dan *Average Linkage*.

1.3.2. MANFAAT

Penelitian ini diharapkan dapat memberi manfaat baik bagi siswa, penulis, maupun Kampus Universitas Stikubank (UNISBANK) Semarang.

1. Bagi Siswa

Membantu para orang tua memilih sekolah dari 217 SMP yang ada di kota Semarang.

2. Bagi Penulis

- a. Memenuhi kewajiban kelulusan penulis dengan pembuatan tugas akhir / skripsi.
- b. Dapat menjadi referensi untuk penelitian-penelitian selanjutnya.

3. Bagi Kampus Universitas STIKUBANK (UNISBANK) Semarang

- a. Melihat sejauh mana penulis dapat menerapkan teori yang sudah di dapat dalam bangku kuliah.
- b. Sebagai referensi untuk membantu mahasiswa lain yang sedang menyelesaikan skripsi.

1.4. METODE PENELITIAN

1.4.1. PERUMUSAN OBYEK PENELITIAN

Data yang akan dianalisa adalah nilai rerata UN dari 217 SMP sekota Semarang. Nilai rerata tersebut meliputi 4 mata pelajaran yang diujikan yaitu:

1. Matematika
2. IPA (Ilmu pengetahuan Alam)
3. Bahasa Inggris
4. Bahasa Indonesia

Dari data-data tersebut akan dibagi menjadi 3 cluster dikarenakan cluster ini mendekati keadaan sesungguhnya. Berikut pembagian clusternya:

1. Unggul
2. Menengah / Sedang
3. Rendah

1.4.2. METODE PENGUMPULAN DATA

Data nilai UN dari 217 SMP se-kota Semarang didapatkan dari web <https://puspendik.kemdikbud.go.id/hasil-un/> dimana pada web tersebut terdapat data nilai-nilai UN se-Indonesia yang dapat didownload secara gratis. Data didownload pada 19 Mei 2020.

1.4.3. METODE ANALISIS

Clustering adalah teknik data mining dimana data dikelompokkan menjadi sejumlah kelompok dimana tiap objek dalam kluster memiliki kemiripan yang tinggi. Dalam penelitian ini, algoritma yang digunakan adalah *Hierarchical Clustering*. Algoritma tersebut bekerja dengan cara melakukan pengelompokan objek kedalam sebuah hirarki yang nantinya akan divisualisasikan. Algoritma ini memiliki 3 metode yaitu *Complete Linkage*, *Single Linkage* dan *Average Linkage*. *Complete linkage* adalah metode pengelompokan cluster dengan jarak terjauh sedangkan *Average Linkage* adalah metode pengelompokan cluster dengan jarak rata-rata antar anggotanya

1.5.SISTEMATIKA PENULISAN

Bab I Pendahuluan

- 1.1.Latar Belakang Penelitian
- 1.2.Perumusan Masalah
- 1.3.Tujuan dan Manfaat Penelitian
- 1.4.Metode Penelitian
- 1.5.Sistematika Penulisan

Bab II Tinjauan Pustaka

- 2.1. Data Mining
- 2.2. Clustering
- 2.3. Penelitian yang terkait
- 2.4. Perbedaan dengan penelitian terdahulu

Bab III Metode Penelitian

- 3.1. Penentuan Variabel
- 3.2. Teknik pengumpulan dan analisa data
- 3.3. Pengukuran jarak
- 3.4. Data Cleansing
- 3.5. Metode Clustering

Bab IV Hasil Penelitian dan Pembahasan

- 4.1. Clustering untuk SMP Sekota Semarang
- 4.2. Clustering untuk SMP Swasta
- 4.3. Clustering untuk SMP Negeri
- 4.4. Clustering untuk 4 nilai mata pelajaran

Bab V Kesimpulan dan Saran

5.1. Kesimpulan

5.2. Saran

Daftar Pustaka