

BAB I

PENDAHULUAN

Bab pendahuluan ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, serta ruang lingkup dalam pembahasan penelitian tentang penggunaan metode *Hierarchical Clustering* dalam pengelompokan data untuk mengukur kemampuan belajar siswa dalam proses belajar di sekolah.

1.1. Latar Belakang

Perkembangan teknologi membantu aktivitas kehidupan manusia dalam berbagai bidang, khususnya dalam bidang pendidikan. e-Rapor merupakan salah satu inovasi dari perkembangan teknologi di bidang pendidikan. Teknologi ini merupakan bentuk dari rapor yang disajikan secara digital dan memiliki suatu pemrosesan otomatis. e-Rapor mempermudah guru dalam melakukan pengolahan nilai siswa maupun dalam penyimpanannya yang dilakukan secara digital dan pemrosesannya secara otomatis, sehingga dapat mempersingkat waktu yang dibutuhkan. Hal tersebut sejalan dengan pendapat Putri & Hartanto (2013) yang menyatakan bahwa dengan adanya e-Rapor maka proses melakukan pemasukan data dan nilai menjadi lebih efisien waktu dan biaya. Selain itu, e-Rapor turut mempermudah pelaporan hasil belajar siswa kepada para wali murid. Dengan mudahnya pelaporan hasil belajar ini menurut Aprilianti, Rhahmadiahti, Setianingrum, Pradana, Wulandari, dan Kurniawan (2023) bahwa e-Rapor membuat orang tua lebih mudah memantau perkembangan belajar anaknya dalam satu semester di sekolah. Oleh karena itu e-Rapor merupakan suatu kemajuan teknologi di bidang pendidikan yang masih terus dikembangkan agar dapat berfungsi dengan lebih baik dalam membantu guru memproses hasil belajar siswa secara lebih akurat. Salah satunya adalah dengan pemutakhiran proses pengelompokan siswa sesuai dengan hasil belajarnya.

Proses pengelompokan suatu data dapat dilakukan secara digital dengan menggunakan suatu metode atau algoritma *clustering* dengan kelebihan dan kekurangannya masing-masing. Penelitian tentang pengelompokan siswa berdasar nilai akademik pada siswa SMP yang telah dilakukan oleh Amalia, Supianto,

Setiawan, Zilvan, Yuliani, dan Ramdan (2021) dengan menggunakan metode *K-Means Clustering* menyebutkan bahwa *clustering* dengan menggunakan metode ini dapat memudahkan guru untuk mengelompokkan siswa sesuai dengan kesamaan karakteristiknya. Data yang dipakai dalam penelitian tersebut merupakan data numerik dari nilai ulangan harian, nilai verbal, nilai keterampilan, nilai ujian tengah semester, dan nilai ujian akhir semester yang digunakan sebagai data untuk membuat kelompok-kelompok belajar yang sesuai. Pemilihan jumlah kluster dapat dilakukan pada metode ini dapat membantu menentukan nilai yang optimal sehingga menghasilkan sebaran *dataset* untuk setiap kluster. Pada penelitian lainnya yang dilakukan oleh Lestari, Kartika, dan Budiman (2022) yang dilakukan untuk klasterisasi siswa agar terbentuk kelompok belajar yang sesuai dan dari penelitian ini metode *K-Means Clustering* menghasilkan *Silhouette Index* sebesar 0,758. Pada penelitian lain yang menggunakan metode *K-Means Clustering*, BIRCH, dan DBSCAN yang dilakukan oleh Nafuri, Sani, Zainudin, Rahman, dan Aliff (2022) membandingkan ketiga metode tersebut dalam klasterisasi data hasil belajar siswa yang menunjukkan bahwa hasil *Silhouette Index* dari metode *K-Means Clustering* mendapatkan hasil yang lebih baik dari dua metode yang lain. Disebutkan bahwa metode *K-Means Clustering* merupakan metode yang banyak digunakan dalam melakukan *clustering*. Penjelasan lain dalam penelitian tersebut bahwa metode BIRCH sangat baik bila dilakukan untuk *dataset* yang besar, namun dalam hasil penelitian tersebut masih ada data siswa yang belum diletakkan sesuai dengan karakteristiknya. Dilanjutkan dengan penjelasan DBSCAN yang merupakan metode yang bagus dalam menangani *noise*, namun metode tersebut kurang dapat memisahkan data-data sehingga terjadi tumpang tindih antar data sehingga metode ini kurang baik dalam mengelompokkan data antar siswa. Penelitian dengan menggunakan *dataset* yang lain pernah dilakukan oleh Sufairoh, Rani, Amalia, dan Rolliawati (2023) membandingkan metode *K-Means Clustering* dengan metode *Hierarchical Clustering* dan DBSCAN pada *dataset Marketplace Electronic Phone* menghasilkan nilai *Silhouette Index* sebesar 0,944473 pada metode *Hierarchical Clustering* yang lebih tinggi dibanding dua metode lainnya yaitu *K-Means Clustering* sebesar 0,690601 dan DBSCAN sebesar 0,283192. Disebutkan dalam penelitian yang dilakukan oleh Gupta, Sharma, dan Akhtar (2021) bahwa metode *Hierarchical Clustering* memiliki akurasi yang lebih baik dibanding dengan metode *K-Means Clustering*. Setiap metode *clustering* memiliki kelebihan dan kekurangannya masing-

masing. Selaras dengan penelitian yang dilakukan Abdalla (2022) yang membandingkan antara metode *Hierarchical Clustering* dan metode *K-Means Clustering*, kedua metode memiliki kinerja yang sama tetapi memiliki tempat keunggulannya masing-masing dan menurutnya pada setiap kenaikan *cluster* menyebabkan kinerja yang semakin tinggi.

Clustering merupakan salah satu dari metode *data mining* yang digunakan untuk mengelompokkan suatu *dataset* sesuai dengan karakteristik yang sama dalam setiap kelompoknya. Senada dengan pernyataan Sammut dan Webb (2017) bahwa *clustering* merupakan pemisahan sekumpulan objek ke dalam suatu kelompok yang memiliki karakteristik yang sama dan berbeda dengan karakteristik kelompok yang lain. Pengelompokan data tersebut dapat dilakukan dengan menggunakan suatu metode atau algoritma dalam *clustering* yang merupakan suatu kumpulan aturan yang digunakan untuk mengelompokkan data berdasarkan karakteristiknya, beberapa metode yang umum digunakan untuk *clustering* ialah *K-Means Clustering*, *Hierarchical Clustering*, DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), maupun *Gaussian Mixture Model*. Menurut Munshi (2022) pemilihan metode untuk *clustering* bergantung pada kompleksitas *dataset* yang digunakan, hal ini dikarenakan setiap metode memiliki kelebihan dan kekurangannya masing-masing dalam menangani permasalahan.

Pemilihan metode *clustering* yang diterapkan dapat disesuaikan dengan karakteristik *dataset* yang digunakan. Dalam hal ini metode *Hierarchical Clustering* dapat digunakan sebagai metode dalam proses penentuan kelompok belajar siswa. Menurut hasil penelitian yang dipaparkan oleh Kaushik dan Mathur (2014) bahwa metode *Hierarchical Clustering* kurang sensitif terhadap *noise* serta memiliki kualitas dan performa yang bagus dibanding dengan metode *K-Means Clustering*. Menurut Cheng, Zhu, Huang, Wu, dan Yang (2019) *noise* merupakan suatu data yang dapat ditemukan secara acak berada di luar kumpulan data yang dapat mengganggu hasil pengelompokan data. Penggunaan teknik *clustering* yang tepat dapat mempermudah dalam penentuan kelompok belajar siswa berdasarkan hasil belajarnya.

1.2. Rumusan Masalah

Bagaimana cara mengolah data e-Rapor hasil belajar siswa pada SD XYZ Kecamatan Srumbung dengan menggunakan metode *Hierarchical Clustering*?

1.3. Tujuan & Manfaat

Tujuan yang ingin dicapai dari penelitian ini adalah melakukan proses pengolahan data e-Rapor dengan menggunakan metode *Hierarchical Clustering* untuk menghasilkan kelompok-kelompok siswa yang sesuai dengan data hasil belajarnya di sekolah.

Manfaat yang dapat diperoleh dari penelitian ini adalah mengetahui tingkat prestasi siswa dalam belajar di sekolah yang dapat dijadikan sebagai bahan tinjauan atau evaluasi bagi guru dalam mengajar, serta bagi orang tua siswa untuk dapat lebih memahami anaknya dalam belajar.

1.4. Ruang Lingkup Penelitian

Ruang lingkup dari penelitian *clustering* hasil belajar siswa SD XYZ Kecamatan Srumbung adalah sebagai berikut

1. Dikembangkan dengan menggunakan bahasa pemrograman Python dan *database* MySQL untuk melengkapi proses yang ada pada aplikasi e-Rapor SD XYZ Kecamatan Srumbung.
2. Menggunakan *dataset* hasil belajar siswa yang diperoleh selama proses pembelajaran, sebuah kelompok data merupakan kumpulan hasil belajar siswa pada satu tingkatan kelas.
3. Menghasilkan kelompok yang sesuai dengan tingkat nilai metrik evaluasi *Silhouette Index*.