

Metode Clustering K-Means dalam Memprediksi Prestasi Siswa

Nur Anisah

Universitas Pembangunan PancaBudi Medan
Email : nuranisahasriel123@gmail.com

Abstrak

Prestasi merupakan hasil dari kesungguhan belajar setiap siswa dengan cari disiplin ilmu. Ketentuan sebagai siswa/siswi yang berprestasi disekolah ditentukan berdasarkan nilai raport sekolah yang terdiri dari nilai-nilai mata pelajaran di SMP Negeri 22 Medan tahun ajaran 2017/2018 yang akan digunakan untuk mendapatkan informasi siswa yang berprestasi dengan menggunakan metode K-Means Clustering dimana data yang digunakan adalah 305 data siswa. Hasil yang didapatkan dengan menggunakan metode tersebut bahwa siswa yang berprestasi menghasilkan informasi yang sangat signifikan daripada siswa yang tidak yang tidak berprestasi. Dimana frekuensi pelajar berprestasi sangat sedikit hasilnya.

Kata kunci: K-Means, Clustering, Data mining, Prestasi

Abstract

Achievement is the result of every student's seriousness in learning by looking for scientific disciplines. Provisions as students who excel in school are determined based on the school report card grades which consist of subject values. Another provision for making these students an outstanding student is discipline. Data mining is a technique for analyzing high-achieving student data. There are 305 student data of SMP Negeri 22 Medan for the 2017/2018 school year which will be used to obtain information on outstanding students using the K-Means Clustering method. The results obtained by using this method, students who excel produce information that is very significant for students who are not achieving. Where is the range of outstanding students has few results.

Keywords: K-Means, Clustering, Data mining, Prestasi

1. Pendahuluan

Pendidikan suatu hal kemajuan dan kemandirian suatu bangsa. Jika pendidikan makin maju suatu bangsa, maka akan semakin maju dan mandiri bangsa tersebut. Dengan adanya pendidikan para generasi penerus bangsa dibentuk kualitasnya.

Pendidikan nasional memiliki fungsi sebagai pengembang diri dan kemampuan untuk membentuk watak dan peradaban bangsa yang bermartabat sebagai langkah untuk mencerdaskan kehidupan bangsa sehingga berpotensi untuk mengembangkan peserta didik agar menjadi manusia yang beriman dan bertakwa kepada Tuhan Yang Maha Esa yang berakhlak mulia, memiliki ilmu, cakap dan kreatif, mandiri serta menjadi warga negara yang demokratis serta bertanggung jawab (Depdiknas.2003,p.4).

Prestasi suatu hal pencapaian yang diperoleh dari kesungguhan dalam belajar serta disiplin ilmu. Ada beberapa hal yang menjadi ketentuan pada pelajar yang dinyatakan sebagai pelajar yang berprestasi diantaranya adalah ketentuan dari nilai raport serta nilai kedisiplin selama menimba ilmu di sekolah tersebut.

Untuk mendapatkan hasil informasi tentang pelajar yang memiliki prestasi perlu dilakukan suatu pengelompokan data dan memberikan daftar data pelajar yang dijadikan suatu informasi bagi pihak sekolah. Adapun metode pengelompokan *K-Means* digunakan untuk mengelompokkan data-data yang memiliki ciri yang sama dan mengelompokkan data kedalam sebuah *cluster*.

Maka dari itu menggunakan salah satu metode data mining , dapat dikelompokkan pelajar yang berprestasi berdasarkan metode *K-Means Clustering*.

2. Metode Penelitian

2.1. Dataset

Untuk mengklasifikasikan dengan metode *K-Means* tersebut maka data yang digunakan bersumber dari data pelajar SMP Negeri 22 Medan tahun ajaran 2017/2018 sebanyak 305 data pelajar.

Dari 305 data yang didapat diperoleh 1 atribut spesial dan 6 regular atribut yang dimana terdiri dari Nilai raport pada semester 2 dari kelas 1 sampai kelas 3 pada setiap pelajar serta jumlah pelanggaran yang dilakukan pelajar tersebut selama 3 tahun menimba ilmu disekolah tersebut.

Tabel 1. Daftar tabel Atribut

No	Nama Atribut	Nilai
1	Number	1,...,305
2	Nama	Nama Siswa
3	X1 raport semester 2 kelas 1	0 – 90
4	X2 raport semester 2 kelas 2	0 – 90
5	X3 raport semester 2 kelas 3	0 – 90
6	X4 daftar disiplin selama 3 tahun	0 – 50
7	Predikat	Baik , Kurang

Pada tahapan selanjutnya didalam data mining adalah melakukan preprosesing data yaitu data harus dibersihkan sebelum diproses, untuk menghindarkan terjadinya perulangan data. Teknik atau metode yang digunakan pada data preprocessing , diantaranya :

1. Normalisasi data
Yaitu melakukan proses standarisasi data sehingga menjadi data yang proporsional dengan menggunakan metode *Z- Transformation*.
2. Data Cleaning
Data Cleaning ialah menghilangkan nilai-nilai data yang salah, menghilangkan duplikasi data apabila adanya data tersebut terdapat data duplikasi. Adapun total dari data observasi pada dataset Nilai raport pelajar SMP Negeri 22 Medan sebanyak 305 record.
3. Missing Value
Missing Value pada atribut yang bernilai nominal akan digantikan dengan yang kemungkinan memiliki banyak dari atribut pada kolom yang sama.
4. Data Outlier
Outlier (pencilan) adalah data yang secara nyata berbeda dengan data-data yang lain.

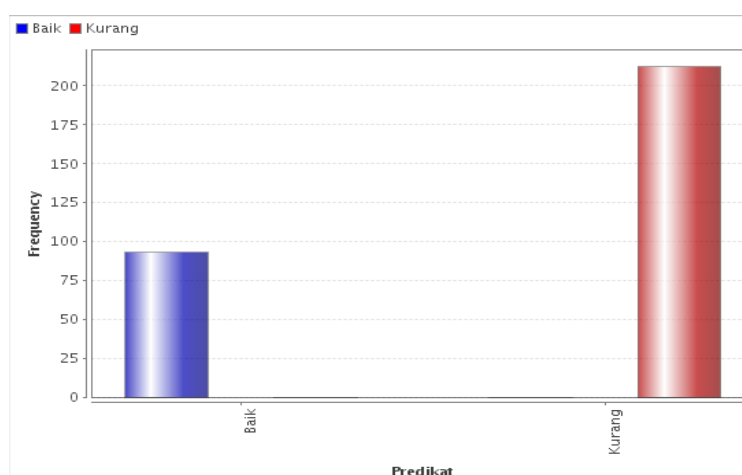
2.2. K-Means Clustering

Metode *K-Means* merupakan salah satu metode yang ada pada metode dalam fungsi *clustering* atau pengelompokan. Menurut (Larose, 2005) Clustering terpaku pada pengelompokkan data, observasi pada kasus mendasar kemiripan objek yang diteliti. Pengelompokan data dengan metode K-Means dilakukan dengan Algoritma:

1. Menentukan jumlah kelompok
2. Mengalokasikan data pada kelompok secara acak
3. Melakukan perhitungan pada *centroid* dari data yang ada pada setiap masing-masing kelompok diambil dari rata-rata (mean) disemua nilai data ke setiap fiturnya.
4. Mengalokasikan masing-masing data ke centroid terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, Yaitu dengan *Euclidean Distance* yaitu pengukuran jarak pada ruang jarak (*distance space*).

3. Hasil dan Pembahasan

Pada hasil yang didapat dengan menggunakan metode *K-Means* memberikan suatu hasil yang sangat signifikan dimana terdapat perbedaan yang tidak jauh antara pelajar yang berprestasi dengan pelajar yang tidak berprestasi disekolah tersebut. Pelajar yang berprestasi dimasukkan kedalam kategori “Baik” dan yang tidak prestasi “Kurang”. Berikut hasilnya dapat dilihat pada gambar berikut :



Gambar 1. Grafik Perbandingan Predikat

Ada 2 bagian yang terdapat pada gambar tersebut dibagian bagan biru, menampilkan sebagai predikat Biru yaitu sebagai “Baik” dan warna Merah “Kurang”. Pada Bagan biru menjelaskan bahwa prediksi “Baik” menunjukkan pada frequency mendekati 100 menurut hasil klasifikasi dengan metode *K-Means*.

4. Kesimpulan

Maka hasil analisis dan pembahasan mengenai metode K-Means dengan dataset yang digunakan adalah perbandingan yang signifikan dimana predikat Baik tidak terlalu jauh dari Predikat Kurang , sehingga memberikan informasi untuk indeks prestasi pada pelajar tersebut selama 3 tahun menimba ilmu disekolah. Adapun centroid distance yang dihasilkan dari dataset tersebut , dimana centroid cluster 0 nya = -3.747 dan centroid cluster 1 = -49.100 sehingga rata-rata centroid distance nya adalah -4.342 dengan menggunakan 305 dataset.

Daftar Pustaka

- [1] Depdiknas. (2003). Undang-Undang Republik Indonesia Nomor 20 Tahun 2003, tentang Sistem Pendidikan Nasional
- [2] Larose, & Daniel T. (2005). Discovering knowledge in data: an introduction to data mining. USA: John Wiley and Sons
- [3] Prasetyo, E., 2012, Data Mining : Konsep dan Aplikasi Menggunakan MATLAB, Andi, Yogyakarta