



## **CLUSTERING DATA MINING ALGORITMA K-MEANS DALAM PENGOLAHAN NILAI PAT PADAMATA PELAJARAN BAHASA INGGRIS, BAHASA INDONESIA DAN MATEMATIKA DENGAN PEMBELAJARAN VIRTUAL LEARNING**

**(Studi Kasus Siswa Kelas XI SMA AL Maghfirah Telajung Bekasi)**

**Sudirman<sup>1</sup>, Ibnu Sulaeman<sup>2</sup>**

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa

<sup>1</sup>sudirman@pelitabangsa.ac.id

### **Abstraksi**

Penelitian ini dilatar belakangi oleh permasalahan mengenai nilai PAT (penilaian akhir tahun) pada mata pelajaran Bahasa Inggris, Bahasa Indonesia dan Matematika siswa kelas XI SMA AL Maghfirah Telajung Cikarang Barat Bekasi yang belum dikelompokkan. Penelitian ini bertujuan untuk mengetahui hasil penggunaan data mining *clustering* algoritma *k-means* dalam pengolahan data PAT (Penilaian Akhir Tahun) pada mata pelajaran Bahasa Inggris, Bahasa Indonesia dan Matematika dengan pembelajaran *virtual learning* pada siswa kelas XI SMA AL Maghfirah Telajung Bekasi. Subjek penelitian ini adalah data nilai PAT (penilaian akhir tahun) kelas XI SMA AL Maghfirah sebanyak 37 siswa. Teknik *pengclusteran* data dengan menggunakan metode *clustering* algoritma *k-means* melalui software *Rapidminer Studio*. Hasil penelitian menunjukkan bahwa metode tersebut dapat mengelompokkan data nilai PAT (penilaian akhir tahun) berdasarkan jarak antar datanya. Adapun hasil pengelompokan data tersebut adalah sebagai berikut : *cluster 0* (rendah) sebanyak 18 siswa dan *cluster 1* (tinggi) sebanyak 19 siswa. Dari penelitian tersebut dapat disimpulkan bahwa teknik *pengclusteran* data menggunakan metode *clustering* algoritma *k-means* melalui aplikasi *Rapidminer Studio* dapat dikatakan berhasil dalam mengclusterkan data nilai PAT (penilaian akhir tahun) pada SMA AL Maghfirah Telajung Cikarang Barat Bekasi.

**Kata kunci :** Nilai PAT, *Clustering*, *K-Means* dan *Rapidminer Studio*

### **Abstract**

*This research was motivated by problems regarding the PAT (year-end assessment) scores in the subjects of English, Indonesian and Mathematics for class XI students of SMA AL Maghfirah Telajung Cikarang Barat Bekasi which have not been grouped. This study aims to determine the results of using k-means algorithm data mining clustering in processing PAT (Year-End Assessment) data on English, Indonesian and Mathematics subjects with virtual learning in class XI SMA AL Maghfirah Telajung Bekasi. The subjects of this study were 37 students of PAT (year-end assessment) grade XI SMA AL Maghfirah data. The data clustering technique uses the k-means algorithm clustering method through the Rapidminer Studio software. The results showed that this method can classify the PAT value data (year-end assessment) based on the distance between the data. The results of data grouping are as follows: cluster 0 (low) of 18 students and cluster 1 (high) of 19 students. From this research it can be concluded that the data clustering*

*technique using the k-means algorithm clustering method through the Rapidminer Studio application can be said to be successful in clustering PAT value data (year-end assessment) at SMA AL Maghfirah Telajung Cikarang Jawa Barat*

**Keywords:** *PAT Value, Clustering, K-Means and Rapidminer Studio*

### **1. Pendahuluan**

Pendidikan merupakan salah satu metode dalam meningkatkan kualitas kehidupan. Pendidikan dapat ditempuh dari jalur-jalur yang telah ditentukan oleh pemerintah baik pendidikan formal maupun pendidikan informal. Pendidikan formal meliputi jenjang pendidikan mulai dari jenjang sekolah dasar hingga perguruan tinggi. Setiap jenjang satuan pendidikan berdasarkan UU No 20 Tahun 2003 Pasal 58 Tentang Sistem Pendidikan Nasional menyebutkan bahwa terdapat dua jenis evaluasi yaitu evaluasi internal yang dilakukan oleh pendidik terhadap hasil belajar peserta

didik dan evaluasi eksternal yang dilakukan oleh lembaga Negaraterhadap peserta didik.

Dalam dunia pendidikan, untuk memperoleh informasi pencapaian hasil belajar dan kemajuan belajar siswa maka digunakan evaluasi pembelajaran salah satunya yaitu PAT (Penilaian Akhir Tahun). Kegiatan tersebut bertujuan untuk mengetahui pencapaian hasil belajar siswa dan dapat mendorong pendidik untuk mengajar lebih baik dan meningkatkan kualitas, kinerja, atau produktifitas suatu lembaga dalam melaksanakan programnya. Berdasarkan hal tersebut maka setiap sekolah melakukan kegiatan evaluasi. Untuk mempermudah tenaga pendidik dalam mengolah data, diperlukannya suatu metode yang mempercepat dalam pengolahan data, salah satunya menggunakan data mining.

*Clustering* adalah salah satu metode dalam data mining yang mengelompokkan data secara otomatis tanpa perlu diberitahu label kelasnya [4]. Banyak metode clustering yang telah diusulkan oleh para ahli, salah satunya adalah *k-means*. Metode *k-means* merupakan algoritma klasterisasi yang paling banyak digunakan dalam aplikasi kecil hingga menengah karena kemudahan dalam mengaplikasikannya [14].

## 2. Landasan Pemikiran

### 2.1 Algoritma K-Menas

**Algoritma K-Menas** merupakan algoritma *clustering* yang berulang-ulang. Algoritma *k-means* dimulai dengan pemilihan secara acak  $k$ ,  $k$  disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai  $k$  secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan *centroid*, mean atau “means” [11].

Algoritma *k-means* bekerja dengan empat langkah, berikut ini adalah langkah-langkah dasar algoritma *k-means*: [14]

Cluster 1	18	48%
Cluster 2	19	52%
<b>Total</b>	<b>37</b>	<b>100%</b>

1. Penentuan *cluster* awal
2. Perhitungan jarak dengan pusat *cluster*
3. Pengelempokan data
4. Penentuan pusat *cluster* baru

### 2.2 Clustering

*Clustering* merupakan klasifikasi tanpa pengawasan dan merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma, yaitu dengan *Euclidean Distance*. [7]

Dalam pengertian lain *clustering* merupakan klasifikasi tanpa pengawasan dan merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Dari beberapa pendapat para ahli, peneliti menyimpulkan bahwa *clustering* adalah proses atau cara untuk mengelompokkan suatu data menjadi

beberapa kelompok tertentu atau *cluster* tertentu. [8].

### 2.3 Virtual Learning

*Virtual Learning* adalah sebuah platform berbasis website untuk pembelajaran dalam aspek digital yang biasa dipakai oleh beberapa institusi pendidikan. Penerapan *virtual learning* ditujukan untuk mengatasi masalah keterpisahan ruang dan waktu antar siswa dan pengajar melalui media komputer atau smartphone [9].

## 3. Metode Penelitian

Percobaan dilakukan dari parameter berikut :Jumlah

Cluster : 2

Jumlah Data : 37

Jumlah Atribut : 3 (Bahasa Inggris, Bahasa Indonesia dan Matematika)

Iterasi ke-1

1. Penentuan pusat cluster awal
2. Perhitungan jarak pusat cluster dengan rumus Euclidian Distance sebagai berikut:

Pengelompokan Data

Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data terhadap pusatcluster.

Penentuan pusat cluster baru

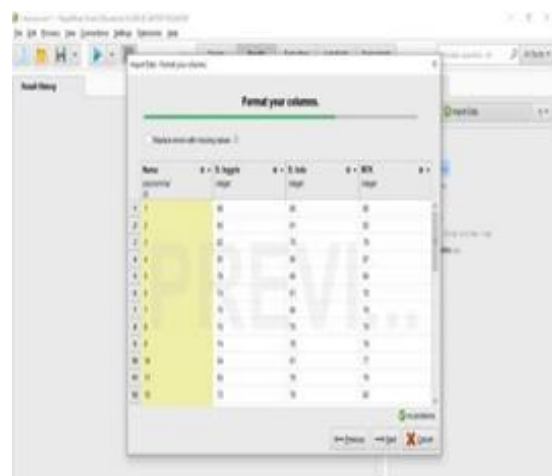
Lakukan hal yang sama pada C1 Bahasa Indonesia, C1 Matematika dan C2 Bahasa Indonesia, C2 Matematika.

Cluster baru untuk iterasi ke-2

C1	54,11	81,27	79,44
C2	76,68	81	79,63

Hasil Perhitungan Manual

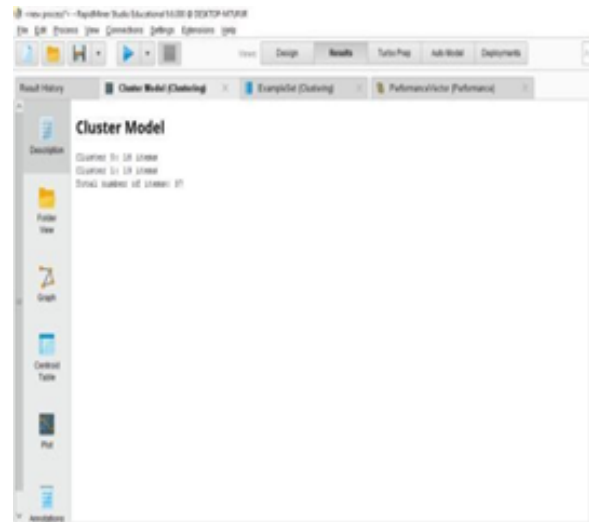
## 4. Hasil Penelitian dan Pembahasan



**Gambar 1.** Proses Penginputan Data Pada Aplikasi



Gambar 2. Proses Pengolahan Data Pada Aplikasi



Gambar 4. Hasil Pengolahan Data pada Aplikasi

Tabel Hasil Pengolahan Data Pada Aplikasi

Cluster 0	18	48%
Cluster 2	19	52%
<b>Total</b>	<b>37</b>	<b>100%</b>



Gambar 3. Parameter Pengolahan Data Pada Aplikasi

5. Penutup

1. Metode *Clustering* dengan algoritma *k-means* dapat digunakan untuk mengelompokkan nilai PAT (penilaian akhir tahun) pada mata pelajaran Bahasa Inggris, Bahasa Indonesia dan Matematika dengan pembelajaran *virtual learning* pada siswa kelas XI lembaga pendidikan SMA Al Maghfirah Telajung Bekasi dengan jumlah data sebanyak 37 data dan pembagian data menjadi 2 *cluster* yaitu *cluster 0* (rendah) dan *cluster 2* (tinggi).
2. Berdasarkan hasil perhitungan dari pengujian data *clustering* algoritma *k-means* yang dilakukan secara manual terdapat hasil sebagai berikut, *cluster 1* memiliki jumlah anggota sebanyak 18 dengan presentase 48% (rendah), *cluster 2* memiliki jumlah anggota sebanyak 19 dengan presentase 52% (tinggi) dengan nilai *Davies Bouldin Index* sebesar - 0,289. Sedangkan hasil perhitungan dari pengujian *clustering* algoritma *k-means* yang dilakukan menggunakan aplikasi *Rapidminer Studio* terdapat hasil sebagai berikut, *cluster 0* memiliki jumlah anggota sebanyak 18 dengan presentase 48% (rendah), *cluster 1* memiliki jumlah anggota.
3. Dari data diatas dapat disimpulkan bahwa nilai PAT (penilaian akhir tahun) pada mata pelajaran Bahasa Inggris, Bahasa Indonesia dan Matematika dapat dikelompokkan menjadi beberapa *cluster* berdasarkan jarak antar datanya.

**Daftar Pustaka**

- [1] F. Eka M dan dkk, "Implementasi Algoritma K-means Untuk menentukan Kelompok Pengayaan Materi Mata Pelajaran Ujian Nasional," *Jurnal Teknik Informatika*, vol. 8, no. 1, 2015.
- [2] A. P. Windarto, "Penerapan Data Mining Pada Ekspor Buah-Buahan Menurut Negara Tujuan Menggunakan K-Means Clustering" *Techno.com*. vol. 16, no. 4, pp. 348–357, 2017.
- [3] Venkateswarlu, B., & Raju, P. G. "Mine Blood Donors Information through Improved K-Means clustering," 1309.2597. 2013.
- [4] R. Tedy dan Kusumadewi, "Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index Dan Ukuran Kerangka," *Seminar Nasional Teknologi Informasi*, 2018
- [5] Suyanto, "Data Mining Untuk Klasifikasi dan Klasterisasi data", Informatika Bandung. 2017.
- [6] Siti, J. "Pemanfaatan Teknologi Informasi Dan Komunikasi Untuk Meningkatkan Kualitas Pembelajaran," 2015.