

PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PENENTUAN JURUSAN SISWA SMA YANG BERBASIS WEB (STUDI KASUS SMA 1 CISARUA BOGOR)

Mulyati¹, Halimah Tus Sadiyah², Andre Jan Febry³

¹Program Studi Ilmu Komputer, FMIPA Universitas Pakuan, Bogor 16143

E-mail : mulyati@unpak.ac.id

²Program Studi Sistem Informasi Program Diploma Universitas Pakuan, Bogor 16143

E-mail : h5.tussadiyah47@gmail.com

ABSTRAK

SMAN 1 Cisarua Bogor is a high school education that prioritizes the preparation of students to continue their education to a higher level with specialization. The manifestation of this specialization is in the form of majors that include majors in Natural Sciences (IPA) and Social Sciences (IPS). However, in the process of determining these majors the school still uses the manual method in processing student academic data, so it takes a long time to find out the results of student majors. Therefore we need the application of data mining technology. One of them is clustering using the K-Means Clustering algorithm. The k-means algorithm has a high accuracy so it is very efficient in processing large amounts of data. The results of the study were obtained from 324 students who were clustered based on the value of psychological tests and majors' interests, 183 students entered the science department and 141 students entered the social studies department with an accuracy rate of 71.3%.

Kata Kunci: Data Mining, K-Means clustering, SMA 1 Cisarua

1. PENDAHULUAN

SMAN 1 Cisarua Bogor merupakan jenjang pendidikan menengah atas yang mengutamakan penyiapan siswa untuk melanjutkan pendidikan kejenjang yang lebih tinggi dengan pengkhususan. Perwujudan pengkhususan tersebut berupa diselenggarakannya penjurusan yang dimulai dari kelas X (sepuluh), yakni penjurusan Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Penentuan jurusan siswa pada SMA tersebut dilihat berdasarkan hasil ujian Psikotes dan juga peminatan jurusan dari siswa sendiri. Namun, dalam proses penentuan jurusan tersebut masih menggunakan sistem manual sehingga membutuhkan waktu yang lama untuk mengetahui hasil penjurusan siswa. Oleh karena itu diperlukan penerapan teknologi data mining, salah satunya pengklusteran dengan menggunakan algoritma K-means.

Algoritma K-means merupakan salah satu metode dalam data mining. Algoritma ini mempartisi data ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan yang memiliki karakteristik yang berbeda dikelompokkan pada cluster yang lain [1]. Metode K-means sangat mudah digunakan dan memiliki kompleksitas waktunya linear [2]. serta memiliki ketelitian yang cukup tinggi sehingga sangat efisien dalam pengolahan data dalam jumlah yang besar [3]. Beberapa penelitian yang sudah dilakukan mengenai penggunaan algoritma K-means, diantaranya oleh Lase & Erwin [4], yang berjudul "Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru. Hasil penelitian adalah aplikasi dapat digunakan untuk membantu proses pemilihan jurusan pada siswa SMK Swasta Harapan Baru secara otomatis. Penelitian lainnya oleh Nurhayati [5] dengan judul Penerapan Algoritma K-means dalam Data Mining untuk Peminatan Jurusan Bagi Siswa Kelas X memiliki tingkat akurasi sebesar 75% terhadap 253 data sampel.

Berdasarkan permasalahan dan penelitian terkait, maka dalam penentuan jurusan bagi siswa SMA 1 Cisarua Bogor digunakan metode K-Means Clustering. Diharapkan dengan menggunakan metode ini dapat memudahkan dalam penentuan jurusan siswa.

1.1 Tujuan Penelitian

Adapun tujuan penelitian ini adalah menerapkan algoritma K-Means Clustering Pada penentuan Jurusan Siswa SMA pada studi kasus di SMA 1 Cisarua Bogor dengan berbasis web.

1.2 Batasan Masalah

Adapun Batasan Masalah dari penelitian ini adalah:

1. Aplikasi Penentuan Jurusan Siswa SMA 1 Cisarua Bogor kelas X dengan menggunakan Algoritma K-means Clustering
2. Data yang digunakan nilai psikotes dan minat jurusan, dengan jurusan yang disediakan adalah IPA dan IPS

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah dapat memudahkan pihak sekolah terutama SMA 1 Cisarua Bogor dalam penentuan jurusan IPA atau IPS pada Siswa kelas X (Sepuluh).

2. TINJAUAN PUSTAKA

2.1 K-means Clustering

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil [6]. Langkah – langkah melakukan clustering dengan algoritma K-means adalah sebagai berikut [7]:

1. Tentukan k sebagai jumlah cluster yang dibentuk.
2. Bangkitkan k Centroid (titik pusat cluster) awal secara random.
Penentuan centroid awal dilakukan secara random/acak dari objek – objek yang tersedia sebanyak k cluster, kemudian untuk menghitung centroid cluster ke- i berikutnya, digunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1,2,3, \dots, n \quad (1)$$

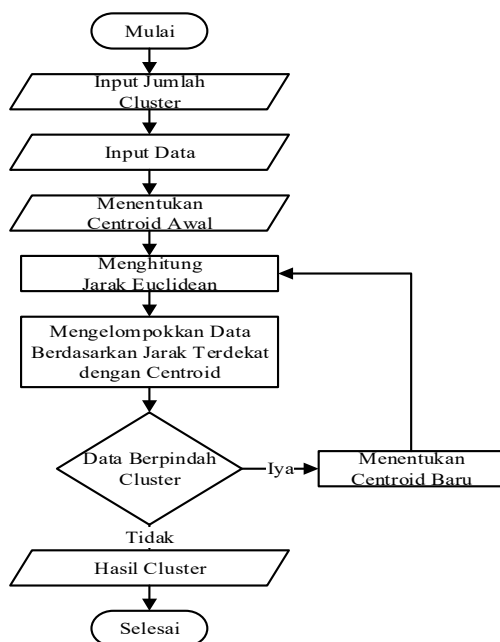
v : centroid pada cluster
 x_i : objek ke- i
 n : banyaknya jumlah objek yang menjadi anggota cluster
3. Hitung jarak setiap objek ke masing – masing centroid dari masing – masing cluster. Untuk menghitung jarak antara objek dengan centroid digunakan Euclidian Distance. Berikut rumus yang digunakan :

$$d(x, y) = ||x - y|| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1,2,3, \dots, n \quad (2)$$

x_i : objek x ke- i
 y_i : data y ke- i
 n : banyaknya objek

4. Alokasikan masing – masing objek ke dalam centroid yang paling terdekat.
Untuk melakukan pengalokasian objek kedalam masing – masing cluster pada saat iterasi secara umum dapat dilakukan dengan cara hard k-means, dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik pusat cluster tersebut.
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan (1).
6. Ulangi langkah 3 jika posisi centroid baru tidak sama.
Pengecekan konvergensi dilakukan dengan membandingkan matriks group assignment pada iterasi sebelumnya dengan matriks group assignment pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma k-means cluster sudah konvergen, tetapi jika berbeda maka belum konvergen sehingga perlu dilakukan iterasi berikutnya.

Berikut merupakan gambaran *flowchart* Algoritma K-means dapat dilihat pada Gambar 1:

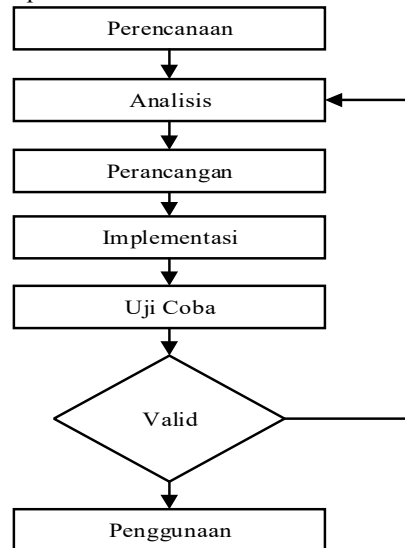


Gambar 1. *Flowchart* Algoritma K-means [5]

3. METODE PENELITIAN

Metode penelitian yang digunakan adalah Metode *Systems Development Life Cycle* (SDLC). Metode ini terdiri atas 6 tahap, yaitu: perencanaan, analisis, perancangan, implementasi, testing, dan

penggunaan. Tahap-tahap pendekatan SDLC dapat dilihat pada Gambar 2.



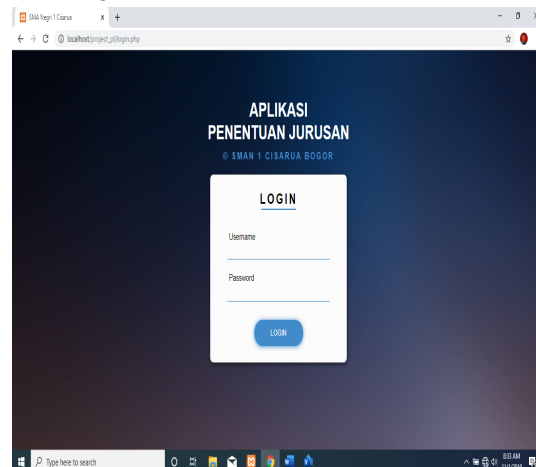
Gambar 2. Tahapan Metode SDLC [8]

4. HASIL DAN PEMBAHASAN

Berikut ini merupakan hasil dari tampilan Aplikasi Penentuan Jurusan Siswa SMA Dengan Algoritma K-means clustering beserta uraian mengenai halaman dari aplikasi tersebut yang akan dijelaskan dibawah ini:

1. Tampilan Form Login

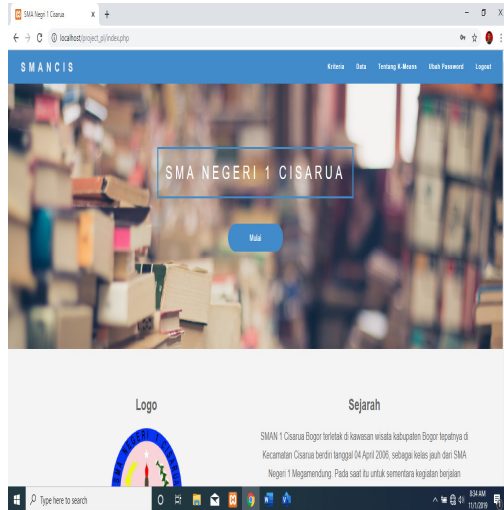
Merupakan tampilan utama yang keluar dari program. *User* memasukan *username* dan *password* kemudian klik tombol *login*, dapat dilihat pada Gambar 3.



Gambar 3. Tampilan Form Login

2. Tampilan Menu Utama

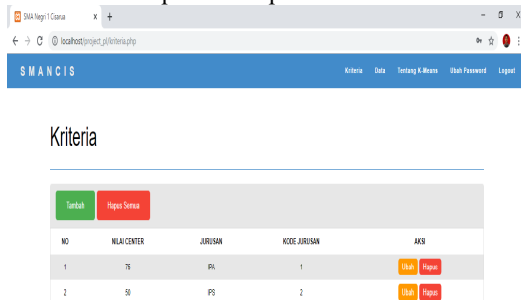
Merupakan tampilan menu utama yang berisi profil sekolah SMAN 1 Cisarua, dapat dilihat pada Gambar 4.



Gambar 4. Tampilan Menu Utama

3. Tampilan Data Kriteria

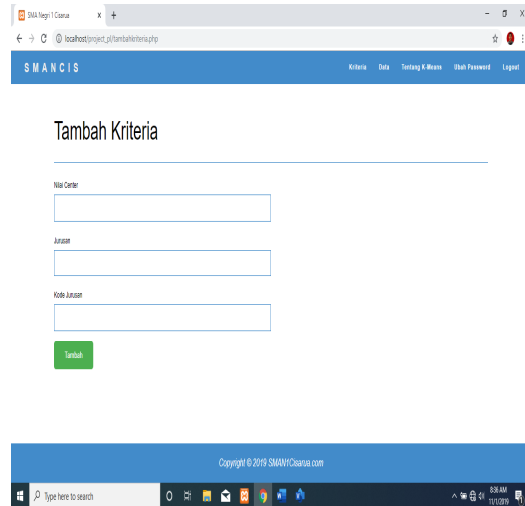
Merupakan tampilan daftar data kriteria (centroid awal), berfungsi untuk mengubah dan menghapus data kriteria. Dapat dilihat pada Gambar 5.



Gambar 5. Tampilan Data Kriteria

4. Tampilan Form Tambah Kriteria

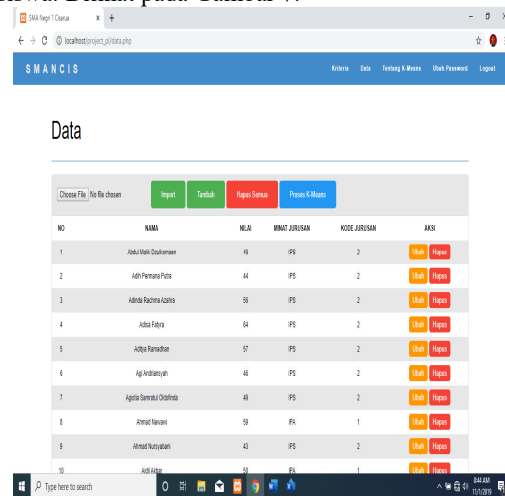
Merupakan tampilan untuk menambahkan kriteria (centroid awal), bisa dilihat pada Gambar 6.



Gambar 6. Tampilan Form Tambah Kriteria

5. Tampilan Data Siswa

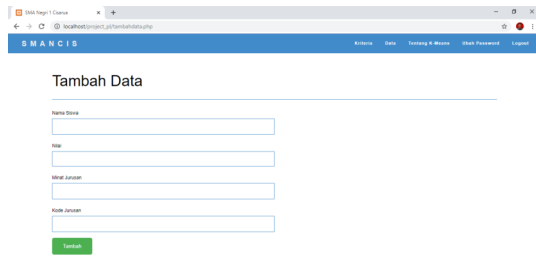
Merupakan tampilan daftar data siswa, berfungsi untuk menambah, mengubah dan menghapus data siswa. Dilihat pada Gambar 7.



Gambar 7. Tampilan Data Siswa

6. Tampilan Form Tambah Data Siswa

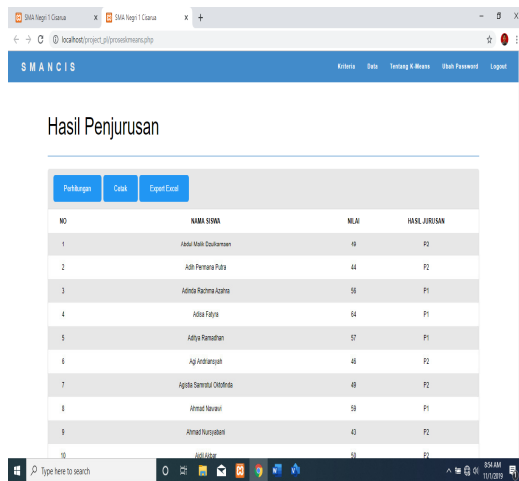
Tampilan Form Tambah Data Siswa merupakan tampilan untuk menambahkan data siswa berupa nama, nilai, minat jurusan, dan kode jurusan. Dilihat pada Gambar 8.



Gambar 8. Tampilan Form Tambah Data Siswa

7. Tampilan Hasil Penentuan Jurusan

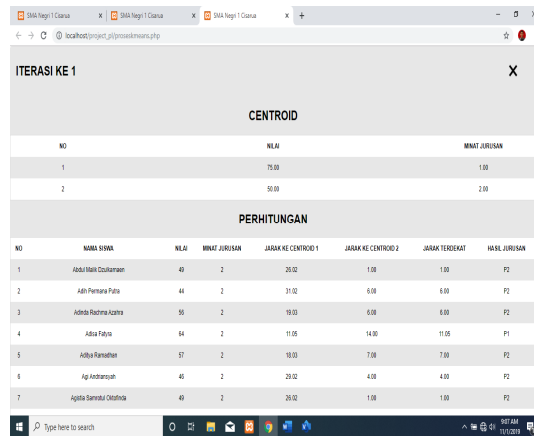
Tampilan Hasil Penentuan Jurusan merupakan tampilan data perhitungan secara menyeluruh algoritma K-means terhadap data yang diolah. Terdapat fitur untuk melihat detail perhitungan algoritma K-means dan mencetak hasil dari penentuan jurusan, dapat dilihat pada Gambar 9.



Gambar 9. Tampilan Hasil Penjurusan

8. Proses Pengujian dengan K-Means Clustering

Pada uji proses penentuan jurusan menggunakan algoritma K-Means Clustering menunjukkan bahwa data yang dihasilkan sesuai dengan rumus dan perhitungan manualnya. Dari 324 siswa yang diproses didapatkan hasil dari sistem yaitu 183 siswa masuk jurusan IPA dan 141 siswa masuk jurusan IPS. Untuk tingkat akurasi algoritma K-Means dilakukan perbandingan antara data minat jurusan siswa dan hasil yang sistem hasilkan. Dari 324 data, ada 231 data yang memiliki kesamaan, dan 93 data tidak memiliki kesamaan. Sehingga tingkat akurasi dengan menggunakan algoritma k-means yaitu 0,7130 atau 71,3% terhadap 324 data sampel. Dilihat pada Gambar 10.



Gambar 8. Tampilan Hasil Proses Algoritma K-means

5. KESIMPULAN

Penerapan algoritma K-Means Clustering pada penentuan jurusan di SMA 1 Cisarua Bogor dapat memudahkan pihak sekolah dalam pemilihan jurusan baik IPA ataupun IPS berdasarkan nilai Psikotes dan minat jurusan. Adapun Pada uji proses penentuan jurusan menggunakan algoritma K-Means dengan menggunakan aplikasi Web diperoleh bahwa dari 324 siswa yang yang dicluster berdasarkan nilai psikotes dan minat jurusan didapatkan hasil 183 siswa masuk jurusan IPA dan 141 siswa masuk jurusan IPS. Tingkat akurasi diperoleh sebesar 71,3% yang didapatkan dari hasil perbandingan data disistem dengan data minat siswa dari sekolah. Pengujian dilakukan sebanyak 324 data, dimana 231 data yang memiliki kesamaan dengan data disistem, dan 93 data tidak memiliki kesamaan.

REFERENSI

- [1] Alfina T, Santoso B, Barakbah AR. 2012. Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Teknik Industri ITS). Jurnal Teknik ITS. 1(1): 521-525. ISSN: 2301-9271
- [2] A. K. Jain, R. C. Dubes. 1988. *Algorithms for clustering data*. New Jersey: Prentice-Hall, Inc.
- [3] Sheih Al Syahdan, Anita Sinar. 2018. Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota, Jurnal Nasional Komputasi dan Teknologi Informasi. 1(2): 56-63.
- [4] Lase Y, Erwin P. 2019. Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan

- Jurusan Di SMK Swasta Harapan Baru. *Jurnal Penelitian Teknik Informatika*. 2(2): 43-47. e-ISSN: 2541-2019.
- [5] Nurhayati, & Luigi, A. P. 2015. Penerapan Algoritma K-Means dalam Data Mining untuk Peminatan Jurusan Bagi Siswa Kelas X (Studi Kasus SMA Negeri 29 Jakarta). *Prosiding Seminar Ilmiah Nasional Teknologi Komputer (SENATKOM 2015)*. Vol. 1: 9-13. ISSN: 2460 – 4690.
- [6] Agusta, Y. P. 2007. K-Means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*. Vol. 3: 47-60.
- [7] MacQueen, J. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press. 1: 281-297.
- [8] Nayan, B. R. 2010. Software Development Lifecycle Models. *Jurnal Hewlett-Packard Enterprise Services*. Vol. 35: 8-13.