

# PENERAPAN OPTIMASI JUMLAH KLUSTER PADA KMEANS UNTUK PENGELOMPOKAN KELAS MATA KULIAH KOSENTRASI MAHASISWA SEMESTER AKHIR

Suhardi Rustam

Fakultas Ilmu Komputer , Universitas Ichsan Gorontalo

Jl. Achmad Nadjamuddin No. 17 Telp. 0435-829975 Fax. 0435-829976 Kota Gorontalo  
suhardirstm@gmail.com

*Abstrak, trending universitas sebagai institusi pendidikan memiliki peranan penting atas kemampuan lulusan, besarnya kapasitas data yang dimiliki universitas diperlukan optimasi dalam pengolahan data. Data-data tersebut diantaranya tentang data akademik mahasiswa. jumlah data akademik mahasiswa semester akhir juga sangat melimpah, melimpahnya data tersebut sampai sekarang menimbulkan permasalahan tentang pengelompokan kelas kompetensi mahasiswa sesuai bidang minat dan keahlian. Melimpahnya data namun tidak memberikan pengetahuan apapun sehingga tidak bermanfaat bagi universitas terutama fakultas kecuali pengetahuan administratif. Ribuan mahasiswa yang aktif pada Universitas ichsan gorontalo disertai dengan jumlah lulusan yang masih kurang ideal setiap periode lulusan, dengan pemodelan optimasi jumlah kluster pada kmeans maka akan menghasilkan kluster kelompok untuk kuliah konsentrasi dari masing-masing kluster yang dihasilkan, setelah dilakukan pengklasteran maka hasil yang akan didapat akan di visualisasikan dalam bentuk grafik chart, Data akademik yang akan digunakan yaitu data mahasiswa angkatan 2016-2017 yang telah mengambil kelas mata kuliah konsentrasi. Penerapan optimasi jumlah kluster pada algoritma KMeans dimana  $K=5$  menghasilkan kluster untuk mengelompokkan Kelas Mata Kuliah Kosentrasi mahasiswa semester akhir dan masing-masing kluster tersebut memiliki nilai optimasi kluster serta faktor penyebab paling banyak pemilihan kelas kosentrasi adalah nilai mata kuliah utama program studi*

**Kata Kunci :** Akademik Data Mining, K-Means, optimasi jumlah kluster)

## I. PENDAHULUAN

Permasalahan kematangan kemampuan lulusan Universitas sebagai institusi pendidikan memegang peranan penting dalam menghasilkan lulusan dengan berbagai kemampuan sesuai visi misi dari institusi tersebut. Hal ini diperlukan usaha yang terukur dan konsisten

dalam system pengajaran serta tata kelola yang baik sehingga sesuai dengan pedoman universitas. Banyaknya rekaman unit-unit data niversitas ichsan gorontalo menyimpan kumpulan data yang banyak.Tsunami data mengindikasikan bahwa data-data ini sangat melimpah namun tidak memberikan pengetahuan apapun sehingga tidak bermanfaat bagi universitas terutama fakultas kecuali pengetahuan administratif. Supaya data ini bermanfaat maka data harus diolah terlebih dahulu menjadi pengetahuan yang memiliki nilai[1]. Universitas ichsan gorontalo dengan jumlah mahasiswa mencapai ribuan orang yang disertai dengan jumlah lulusan yang masih kurang ideal setiap periode lulusan, maka perlu diterapkan pola penentuan kelas kosentrasi mata kuliah yang efektif bagi pencapaian kemampuan mahasiswa. Perlu digali informasi yang bisa digunakan untuk lebih meningkatkan ketepatan kelas kosentrasi yang salah satunya bisa jadi ukuran untuk menfokuskan kemampuan mahasiswa[2].

Penyediaan data berupa Data akademik yang akan digunakan yaitu data mahasiswa angkatan 2016-2017 yang telah mengambil kelas mata kuliah kosentrasi. Dan data ini juga sebagai data testing yang akan dicoba kedalam model algoritma klustering dan klasifikasi. Ada banyaknya algoritma yang digunakan untuk menyelesaikan persoalan pada data mining. Pada masalah universitas ini digunakan algoritma data mining yaitu algoritma klustering dan klasifikasi serta gabungan kedua algoritma tersebut. Algoritma klustering yang digunakan adalah K-Means klustering [3], algoritma ini memiliki kelebihan yang dapat diterapkan dan dijalankan yang dapat diterapkan dan dijalankan untuk pengelompokan kelas mata kuliah kosentrasi yang ingin dirancang, relative cepat untuk di adaptasi, dan paling banyak dipraktekkan dalam data mining, algoritma ini merupakan algoritma sangat penting dalam data mining. Algoritma K-Means merupakan metode untuk mengelompokan objek atau data yang mempunyai atribut dan mempunyai jumlah data yang banyak kedalam bentuk satu atau lebih kelompok sehingga

data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster*/kelompok yang sama dan mirip [4].

Sesuai karakteristik algoritma K-means dalam hal memilih bobot nilai[5]dengan kemiripan nilai[6]dan bebas memilih pusat kluster data[7]karena pusat klusternya bebas untuk dipilih[8]yang sering digunakan dalam penelitian untuk pengelompokan [9]hal ini perlu di optimalkan hasilnya dengan jumlah kluster, hasil K-means juga akan mengoptimalkan K yang sifatnya bias nilai[10].

Dalam penelitian ini tentang bagaimana menerapkan optimasi jumlah kluster pada Algoritma K-Means dalam mengelompokkan Kelas Mata Kuliah Kosentrasi Mahasiswa Semester Akhir dan bagaimana memvisualisasikan faktor dari hasil kluster sehingga lebih akurat dan berkualitas

## II. LANDASAN TEORI

### A. Data Mining

Data terkini yang juga memiliki batasan waktu data tak berguna, data yang disimpan dengan jarak terpisah dari data operasional sebagai data warehouse (gudang data). Sehingga jenis data ini dapat kategorisasi dalam manajemen data dibagi dua, adalah OLTP (*On Line Transaction Processing*) yaitu manajemen data operasional, kemudian data yang dimanfaatkan merupakan data terkini. Untuk akan datang.

Sebuah *Data* dalam gudang datang merupakan kumpulan database akan dipakai demi kebutuhan membantu kebijakan serta diatur dalam bentuk yang terpisah dari database operasional [11]. Data warehouse adalah bersifat (1) *Subject oriented*, tentang diorganisasi atas dasar subject misal data pelanggan, data produk, dsb (2) *Terintegrasi*, yaitu dibangun lebih dari satu sumber data yang beragam misal database relasional, file, juga bentuk data lain sehingga memerlukan teknik *data cleaning* dan *Data Integration* dapat memastikan konsistensi penamaan, pengukuran atribut, dsb.(3) *Time-variant*, horizon data sangat lebih lama daripada data operasional, seperti data-data dalam kurung waktu 5-10 tahun. (4) *Non volatile*, dikarena non perubahan data.

### B. Algoritma K-Means

Algoritma K-Means [12] *clustering* adalah sebagai berikut

1. Pilih K buah titik *centroid* secara acak
2. Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap *cluster* merupakan titik *centroid* yang telah dipilih sebelumnya
3. Perbaharui nilai titik *centroid*
4. Ulangi langkah 2 dan 3 sampai nilai dari titik *centroid* tidak lagi berubah

Distance space digunakan untuk menghitung jarak antara data dan centroid. Adapun persamaan yang dapat digunakan salah satunya yaitu *Euclidean Distance Space*. *Euclidean distance space* sering digunakan dalam perhitungan jarak, hal ini dikarenakan hasil yang diperoleh merupakan jarak terpendek antara dua titik yang

diperhitungkan. Adapun persamaannya adalah sebagai berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2}$$

dimana

$d_{ij}$  = Jarak objek antara objek i dan j

P = Dimensi data

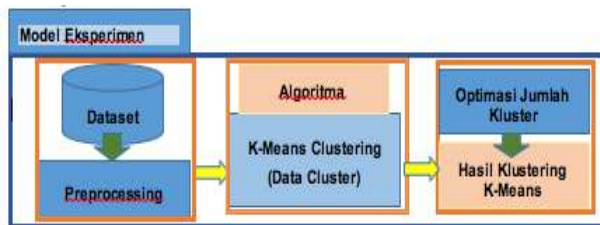
$X_{ik}$  = Koordinat dari obyek i pada dimensi k

$X_{jk}$  = Koordinat dari obyek j pada dimensi k

## III. METODE

Metodel penelitian ini, mempergunakan cara penelitian eksperimen, untuk tahap penelitian dimulai dari pengumpulan dataset akademik mahasiswa sampai dengan mendapatkan hasil evaluasi dan hasil untuk mendapatkan tujuan yang akan dicapai, adapun tahapannya adalah, di mulai dari penyiapan dataset kemudian dilakukan langkah preprocessing terhadap data[12] untuk menghilangkan outliernya, selanjutnya adalah menormalisasikan data kemudian menggunakan *tool data mining orange* untuk mengelola data tersebut dengan menggunakan metode optimasi jumlah kluster pada K-Means [13]. Model yang penelitian ini adalah sebagai berikut :

### a. Eksperimen



Gambar 1 Model Eksperimen Optimasi Jumlah Kluster pada KMeans

### Pengumpulan Data

Pada penelitian tersebut, yang data digunakan ini, juga diperoleh dari penelitian yang dilakukan pada data universitas ichsan Gorontalo yang berkaitan langsung dengan bagian pusat pengelolaan data kemahasiswaan universitas (BAAK Universitas). Identifikasi variable dalam penelitian menentukan cara pengolahan data dan alat yang digunakan untuk menganalisis data, adapun variable dalam penelitian ini adalah sebagai berikut :

- a. Indeks prestasi kumulatif, adalah penghitungan IP dengan menggabungkan semua mata kuliah yang telah ditempuh sampai suatu semester tertentu
- b. Indeks prestasi semester, adalah nilai rata rata dari seluruh matakuliah yang telah diambil oleh mahasiswa
- c. Nilai mata kuliah prasyarat konsentrasi, adalah nilai mata kuliah yang pelengkap untuk mata kuliah konsentrasi

d. Jumlah kehadiran dari mata kuliah prasyarat konsentrasi, merupakan nilai kehadiran mahasiswa dalam mengikuti mata kuliah tersebut

**Pengolahan Data Awal**

Penelitian ini dibatasi hanya pada tahapan ini melakukan *preprocessing* terhadap data untuk menghilangkan outlier[7], merapikan setiap variable yang terkait sehingga siap untuk ke tahap penggunaan selanjutnya

**Eksperimen**

Eksperimen pada tahapan ini adalah menjalan seluruh prosedur yang telah ditentukan/disusun dalam sebuah kerangka pemikiran untuk menyelesaikan masalah, adapun tahapannya adalah, di mulai dari fase pertama yaitu penyiapan dataset kemudian dilakukan langkah preprocessing terhadap data untuk menghilangkan outliernya, selanjutnya adalah menormalisasikan data kemudian menggunakan tools orange data mining untuk mengelola data tersebut dengan menggunakan metode *k-means clustering* pada tahapan kelastering ini menggunakan jumlah kluster sebanyak  $k=5$ , selanjutnya pada tahapan fase ini juga yaitu penggunaan optimasi jumlah kluster dengan nilai input algoritma K-Means sehingga akan menghasilkan data klustering.

Fase pertama, Secara umum algoritma dasar dari K-Means Clustering adalah sebagai berikut :

1. Tentukan jumlah cluster
2. Alokasikan data ke dalam cluster secara random
3. Hitung centroid/rata-rata dari data yang ada di masing-masing cluster
4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan

Fase ini juga dengan  $K=5$ , secara umum algoritma kmeans dioptimalkan dengan nilai terbaik yang hasilkan oleh optimasil jumlah kluster sesuai jumlah K

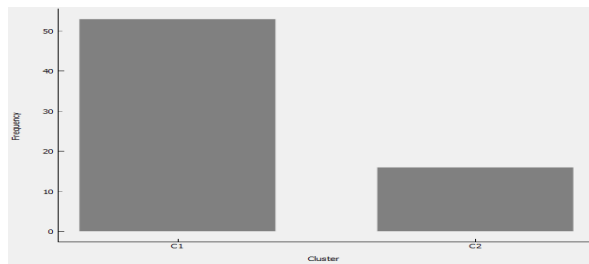
**Evaluasi**

Pengujian eksperimen yang dilakukan berupa eksperimen klustering pada algoritma K-Means dan eksperimen nilai akurasi pada setiap nilai K yang dimiliki saat optimalisasi jumlah kluster yang disebut dengan *Silhouette Scores*

**IV.HASIL DAN PEMBAHASAN**

**Model preprocessing**

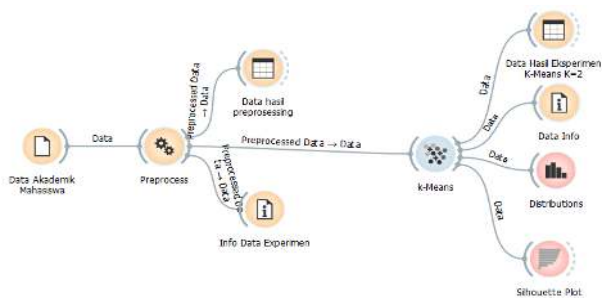
Pada model ini, diawali pada tahapan preprocessing, dimana seluruh data harus menghilangkan outlier/missing value yang tidak memiliki nilai variable dalam dataset



Gambar 2 tahapan preprocessing

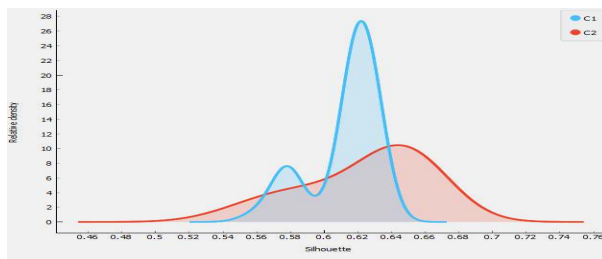
**Algoritma K-Means Model KMeans eksperimen**

Dalam model eksperimen ini algoritma K-means ditempatkan setelah data preprocessing, adapun gambarannya adalah sebagai berikut :

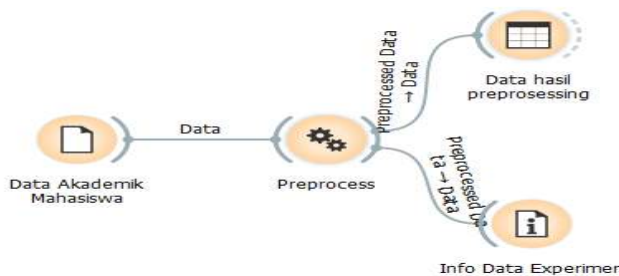


Gambar 3 Model K-Means

Dalam memproses dataset dengan menggunakan model seperti pada gambar 3 yaitu menggunakan model K-Means menghasilkan kluster yaitu Kluster C1 dan Kluster C2 dengan grafik frekuensi kluster pada Gambar 4.



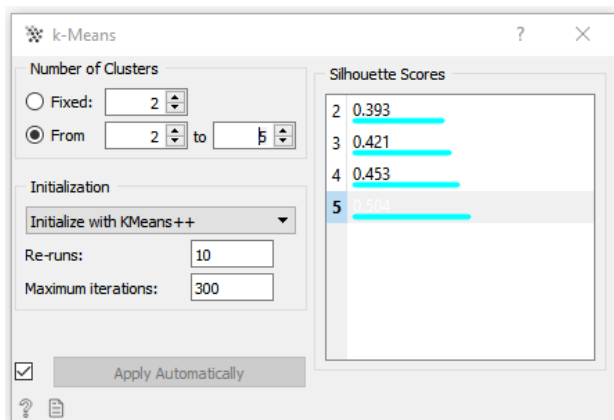
Gambar 4 Grafik frekuensi kluster



Gambar 5 Grafik Silhouette Kluster

Disamping itu juga ditampilkan grafik silhouette kluster yang menggambarkan kerapatan data kluster (relative density) dengan jarak antara garis/jarak dari setiap pinggiran kluster seperti Gambar 5.

Hasil eksperimen akademik data akademik menggunakan algoritma k-means dengan optimasi jumlah kluster adalah sebagai berikut :



Gambar 6 Nilai Optimasi Jumlah Kluaster

Tabel 1 Hasil Eksperimen Optimasi Jumlah Kluster pada K-Means

No	Algoritma	Kluster	Silhouette Scores
1.	K-Means dengan dataset 69 record	K=2	0.393
		K=3	0.421
		K=4	0.453
		K=5	0.504

Dari hasil eksperimen dengan optimasi jumlah kluster pada algoritma kmeans dengan dataset 69 record, maka untuk kluster k=2, nilai optimasinya 0.393, k=3, nilai optimasinya 0.421, k=4, nilai optimasinya 0.453 dan k=5, nilai optimasinya 0.504

#### IV.KESIMPULAN DAN SARAN

Sebagaimana dalam uraian pembahasan penelitian ini, dapat disimpulkan sebagai berikut, Penerapan algoritma KMeans dengan optimasi jumlah kluster dimana K=5 menghasilkan kluster untuk mengelompokkan Kelas Mata Kuliah Kosentrasi mahasiswa semester akhir dan masing-masing kluster tersebut maka untuk kluster k=2, nilai optimasinya 0.393, k=3, nilai optimasinya 0.421, k=4, nilai optimasinya 0.453 dan k=5, nilai optimasinya 0.504, serta nilai optimasi ini sebagai nilai terbaik, adapun saran dalam penelitian ini, untuk keperluan lebih lanjut data mining dalam pengolahan data akademik mahasiswa, maka disarankan untuk menguji coba disetiap iterasi dikarenakan dalam eksperimen ini fokus menggunakan k=5, penelitian semacam ini dapat dikembangkan dengan menggunakan optimasi di masing-masing algoritma.

#### DAFTAR PUSTAKA

[1] Elektro, J. T. (2015). *Penerapan Algoritma K-Means Dengan Optimasi Jumlah Cluster Untuk Pengelompokan Angk an Kerja Propinsi Jatim.*

[2] Rustam, S., & Annur, H. (2019). *AKADEMIK DATA MINING ( ADM ) K-MEANS DAN K-MEANS K-NN UNTUK MENGELOMPOKAN KELAS MATA KULIAH KOSENTRASI MAHASISWA SEMESTER AKHIR.* 11(28), 260–268.

[3] I. G. Ayu and D. Saryanti, “PENERAPAN TEKNIK CLUSTERING UNTUK PENGELOMPOKAN KONSETRASI MAHASISWA DENGAN METODE K-MEANS,” pp. 519–526, 2019.

[4] T. A. Munandar and W. O. Widyarto, “Clustering Data Nilai Mahasiswa Untuk Pengelompokan Konsentrasi Jurusan Menggunakan Fuzzy Cluster Means,” *Semin. Nas. Apl. Teknol. Inf.*, pp. 1907–5022, 2013.

[5] Widodo and D. Wahyuni, “Implementasi algoritma k-means clustering untuk mengetahui bidang skripsi mahasiswa multimedia pendidikan teknik informatika dan komputer universitas negeri jakarta,” *J. Pint.*, vol. 01, no. September, p. 11 pages, 2018.

[6] A. Rizky and F. Amiq, “Penerapan Metode Clustering Dengan Algoritma K- Means Untuk Rekomendasi Pemilihan Jalur Peminatan Sesuai Kemampuan Pada Program Studi,” 2013.

[7] P. S. Informatika, U. T. Yogyakarta, and S. P. Keputusan, “PADA UNIVERSITAS SWADAYA GUNUNG DJATI MENGGUNAKAN METODE K - MEANS CLUSTERING,” pp. 1–10.

[8] K. N. Sistem, G. A. Pradnyana, A. Aan, J. Permana, and U. P. Ganesha, “Perancangan Sistem Pembagian Kelas Kuliah Mahasiswa dengan Kombinasi Metode K-Means dan K-Nearest,” pp. 285–290, 2017.

[9] Rohman, A. (2015). Model Algoritma K-Nearest Neighbor (K-Nn) Untuk Prediksi Kelulusan Mahasiswa. *Neo Teknika*, 1(1). <https://doi.org/10.1017/CBO9781107415324.004>

[10] Rustam, S. (2018). Analisa Clustering Phising Dengan K-Means Dalam Meningkatkan Keamanan Komputer. *ILKOM Jurnal Ilmiah*, 10(2), 175. <https://doi.org/10.33096/ilkom.v10i2.309.175-181>

[11] Rustam, S., Santoso, H. A., & Supriyanto, C. (2018). Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang. *ILKOM Jurnal Ilmiah*, 10(3), 251. <https://doi.org/10.33096/ilkom.v10i3.342.251-259>

[12] J. Han, *Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*. 2011.

[13] G. I. Marthasari, “Implementasi Teknik Data Mining untuk Evaluasi Kinerja Mahasiswa Berdasarkan Data Akademik,” *Fountain Informatics J.*, vol. 2, no. 2, p. 20, 2017.