

Penerapan Algoritma K-Means Untuk Pengelompokan Riset Grup Dosen Program Studi Teknik Informatika Universitas Bina Darma

Yurenza¹, R.M Nasrul Halim²

^{1,2}Teknik Informatika, Universitas Bina Darma, Palembang, Indonesia

Article Info

Article history:

Diterima 5 September 2024
Revisi 5 Oktober 2024
Diterbitkan 7 Oktober 2024

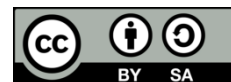
Keywords:

Algoritma K-Means
Klasifikasi
Riset Grup
Cluster

ABSTRAK

UBD atau yang dikenal Universitas bina darma adalah salah satu universitas yang terletak di kota Palembang provinsi sumsel yang bergerak di bidang ilmu komputer dan berbagai bidang studi lain nya. Universitas bina darma sudah sering melakukan publikasi karya ilmiah nya di laman jurnal nasional maupun internasional serta menghasilkan penelitian yang berkualitas di berbagai bidang program studi, misalnya yang dilaksanakan oleh program studi teknik informatika, telah banyak mahasiswa-mahasiswi maupun dosen program studi teknik informatika yang mempublikasikan karya ilmiah nya di berbagai jurnal nasional dengan berbagai penelitian berdasarkan grup riset nya masing-masing. Banyak nya penelitian yang di lakukan oleh dosen dan juga mahasiswa di berbagai bidang program studi teknik informatika membuat dosen perlu melakukan proses klasifikasi atau pengelompokan publikasi dosen berdasarkan bidang penelitian bertujuan untuk mengidentifikasi kecenderungan masing-masing dosen dibidang teknik informatika terhadap bidang penelitian tertentu. Pengelompokan riset grup menggunakan metode tertentu di program studi teknik informatika masih belum terlaksana. Dengan menerapkan metode k-means diharapkan dapat pengelompokan riset grup dosen program studi teknik informatika berdasarkan bidang penelitian guna mengetahui kecenderungan setiap dosen teknik informatika terhadap bidang penelitian tertentu serta dosen dapat mengetahui minat riset sehingga dosen bisa merencanakan segala sesuatu dengan mudah tentang bidang penelitian nya. Hasil analisis ini terdiri dari dua bagian. Bagian pertama berdasarkan dosen dikelompokan 3 cluster berdasarkan 29 data dosen yaitu cluster 0 merupakan cluster dengan penelitian terendah, cluster1 dengan penelitian sedang, dan cluster 2 dengan penelitian tertinggi. Bagian kedua berdasarkan judul penelitian dikelompokan 7 cluster dengan jumlah 729 data, didapatkan cluster 0 dengan jumlah 11 anggota, kemudian cluster 1 berjumlah 101 anggota, kemudian cluster 2 berjumlah 3 anggota, kemudian cluster 3 berjumlah 447 anggota, kemudian cluster 4 berjumlah 164 anggota, selanjutnya cluster 5 berjumlah 2 anggota, terakhir cluster 6 berjumlah 1 anggota.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Yurenza,
Universitas Bina Darma, Jl. Jendral Ahmad Yani No.3, 9/10 Ulu, Kec. Seberang Ulu I, Palembang,
Sumatera Selatan, 30111, Indonesia
Email: yurenza0009@gmail.com

1. PENDAHULUAN

Universitas Bina Darma adalah salah satu universitas swasta yang ada di Palembang yang bergerak di bidang ilmu komputer dan berbagai bidang studi lain nya. Universitas bina darma telah banyak melahirkan lulusan terbaik di berbagai bidang komputer dan sistem informasi, yang tentunya sudah sering melakukan publikasi karya ilmiah nya di laman jurnal nasional maupun internasional serta menghasilkan penelitian yang berkualitas di berbagai bidang program studi. Hal ini sesuai dengan salah satu dari misi yang hendak di tuju

Universitas Bina Darma Palembang yakni melakukan penelitian yang berstandar internasional. Berdasarkan hal tersebut seluruh Civitas Akademika dituntut untuk lebih memaksimalkan menjalankan fungsi dan juga tugas nya masing-masing agar mencapai misi Universitas, dalam hal ini dosen-dosen memiliki kewajiban untuk melakukan penelitian serta berkontribusi dalam melaksanakan publikasi karya ilmiah sesuai dengan grup risetnya. misalnya yang dilaksanakan oleh Program Studi Teknik Informatika, telah banyak mahasiswa-mahasiswi maupun dosen Program Studi Teknik Informatika yang mempublikasikan karya ilmiahnya di berbagai jurnal nasional dengan berbagai penelitian berdasarkan grup riset nya masing-masing.

Research Group (Riset Grup) merupakan kumpulan beberapa peneliti yang memiliki pengalaman, latar belakang, dan inti penelitian yang sama. Kelompok riset grup dapat ditemukan di berbagai instansi seperti instansi Keuangan sampai instansi Perguruan Tinggi[1]. Mengingat banyak nya penelitian yang di lakukan oleh Dosen serta Mahasiswa-Mahasiswi Program Studi Teknik Informatika di perlukannya pengelompokan riset grup tersebut sehingga dapat dengan mudah di kelola dan mengidentifikasi kecenderungan masing-masing dosen teknik informatika terhadap area penelitian tertentu, dosen juga dapat merencanakan segala sesuatu nya dengan mudah mengenai bidang penelitiannya.

Pengelompokan riset grup sebelum nya menerapkan metode manual yang memungkinkan terjadi nya kesalahan teknis maupun kekeliruan, solusi yang dapat di berikan yakni menerapkan metode algoritma k-means. K-means adalah salah satu metode pengelompokan data yang berupaya mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok (*cluster*) berdasarkan kemiripan karakteristik data. Metode k-means memisahkan data ke dalam beberapa kelompok sejenisnya supaya data yang memiliki karakteristik sama dapat di kelompokkan ke dalam satu kelompok yang sama serta data yang memiliki karakteristik berbeda dapat di kelompokkan ke dalam kelompok yang lainnya[2]. Hal tersebut di laksanakan secara bertahap sampai diperoleh kelompok yang tetap. Dengan menerapkan metode k-means dosen dapat mengetahui kemiripan judul-judul penelitian berdasarkan kategori yang telah banyak dilakukan pada penelitian sebelumnya secara lebih mudah.

1.1 Studi Literatur

1.1.1 Data Mining

Data mining dapat didefinisikan sebagai serangkaian teknik yang secara otomatis diterapkan untuk memeriksa dan mengungkap kaitan yang kompleks pada kumpulan data besar. Kumpulan data yang dimaksud adalah berupa data tabular yang sering digunakan dalam teknologi manajemen basis data rasional. Tetapi, teknik data mining juga dapat diterapkan pada domain data spasial, teks, dan multimedia[3].

Data mining adalah tahapan yang menggunakan teknik statistika, matematika, kecerdasan buatan dan *machine learning* untuk memperoleh informasi yang bermanfaat juga pengetahuan yang tersusun di database besar. Berdasarkan uraian tersebut, data mining melibatkan pengolahan pengetahuan yang ada didalam database kemudian di olah untuk menemukan pola dan menerapkan teknik statistik matematika, kecerdasan buatan, serta *machine learning* guna memisahkan dan mendeskripsikan informasi pengetahuan dari *database*[4]

1.1.2 Operasi Data Mining

Operasi dalam *data mining* terbagi menjadi 2 kategori yakni metode deskriptif dan metode prediktif. Metode prediktif bertujuan untuk memperkirakan nilai dari suatu variabel berdasarkan nilai variabel lain, di mana hasilnya digunakan untuk verifikasi *hipotesis*, pelaporan dan *querying*. Sementara itu, metode deksriptif bertujuan untuk menemukan pola atau anomali serta hubungan dalam data yang mudah di mengerti oleh manusia, hasil yang di dapat pada metode ini digunakan untuk menganalisis data, *database*, eksplorasi, segmentasi, deteksi deviasi dan analisis keterkatitan[5]

1.1.3 Knowledge Discovery in Database (KDD)

Definisi KDD menurut Han dan Kamber., (2006) dalam[6], istilah *Knowledge Discovery in Database* (KDD) dan *data mining* digunakan secara bergantian untuk memberikan deskripsi tentang metode penyelidikan informasi yang tersimpan di dalam suatu kumpulan data besar. Meskipun kedua istilah itu memiliki konsep yang berbeda, mereka saling terkait, salah satu tahap dalam proses KDD adalah *data mining*.

Data mining, dikenal sebagai penemuan pengetahuan dalam database (KDD) merupakan proses pengumpulan dan penggunaan data historis untuk mengidentifikasi pola, hubungan, atau keteraturan dalam sejumlah besar data[7]

1.1.4 Research Group

Research Group (Riset Grup) ialah suatu kumpulan dari beberapa orang yang melakukan penelitian yang dimana mereka sudah punya pengalaman, latar belakang yang mumpuni, serta banyak memiliki inti penelitian serupa. Kelompok penelitian dapat dijumpai dalam berbagai instansi seperti instansi Keuangan hingga instansi Perguruan Tingg[1]. *Research group* (riset grup) adalah sekelompok peneliti di perguruan tinggi yang telah menghasilkan karya-karya penelitian yang mencakup penelitian mendasar, proving,

pengembangan metode baru dan bersifat lintas serta multidisipliner. Sifat utama dari riset grup adalah banyaknya karya ilmiah di jurnal baik nasional maupun internasional yang dituliskan oleh para dosen dan mahasiswa di suatu lembaga, keberadaan *research group* bisa berada di Program Studi, Fakultas dan Universitas[8]

1.1.5 Clustering

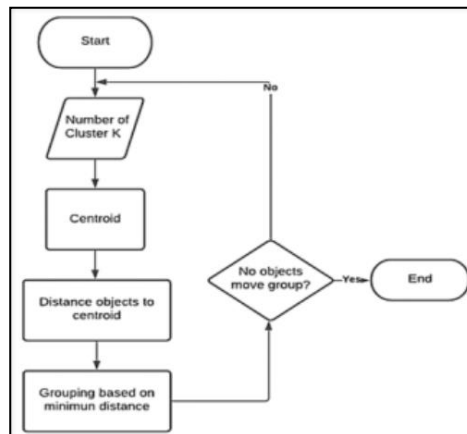
Clustering atau pengelompokan data adalah proses penting yang digunakan untuk menemukan kesamaan dalam data dan mengelompokkan data serupa ke dalam kelompok-kelompok. *Clustering* dianggap sebagai salah satu metode pembelajaran tanpa pengawasan yang paling signifikan, di mana tujuan utamanya adalah menemukan pola dalam kumpulan data yang tidak berlabel. *Clustering* membagi kumpulan data menjadi beberapa kelompok (*cluster*), dengan tingkat kesamaan yang lebih tinggi di dalam kelompok tertentu dibandingkan dengan kelompok lainnya [9].

1.1.6 K-Means Clustering

Menurut[10], K-Means adalah salah satu metode dalam data mining yang digunakan untuk mengelompokkan data atau melakukan *Clustering*, dimana data dengan karakteristik serupa akan dikelompokkan kedalam satu *cluster* sementara data dengan karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* berbeda.

Menurut[11], *Clustering* adalah proses membagi data ke dalam beberapa kelompok, dimana setiap kelompok berisi objek dengan karakteristik yang mirip satu sama lain.

Tahapan *clustering* dengan metode k-means dapat dilihat pada gambar 1 sebagai berikut :



Gambar 1. Algoritma K-Means

1. Menentukan jumlah kluster k sebagai jumlah kluster yang ingin dibentuk.
2. Inisialisasi k pusat kluster dapat dilakukan dengan beberapa cara, tetapi cara yang paling umum digunakan adalah dengan mengambil data yang ada secara acak.
3. Menghitung jarak setiap data input ke masing-masing centroid menggunakan rumus jarak *Euclidean* (*Euclidean Distance*) untuk menentukan jarak yang paling dekat antara setiap data dengan centroid. Berikut adalah persamaan untuk menghitung *Euclidean Distance* [12] :

$$d(x - y) = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (1)$$

Keterangan :

d = titik dokumen

x= pusat *cluster*

y = data

4. Mengklasifikasikan setiap data berdasarkan kedekatannya dengan centroid (jarak terkecil).
5. Memperbarui nilai centroid. Nilai centroid baru di peroleh dari rata-rata cluster yang berkaitan dengan menggunakan rumus :

$$C_k = \left(\frac{1}{n_k}\right) \sum d_1 \quad (2)$$

Keterangan :

N_k = jumlah dokumen pada *cluster*

C_k dan d_1 = dokumen pada *cluster* k

6. Mengulangi langkah 2 hingga beberapa langkah sampai tidak ada lagi perubahan pada anggota setiap cluster.

1.1.7 Rapidminer

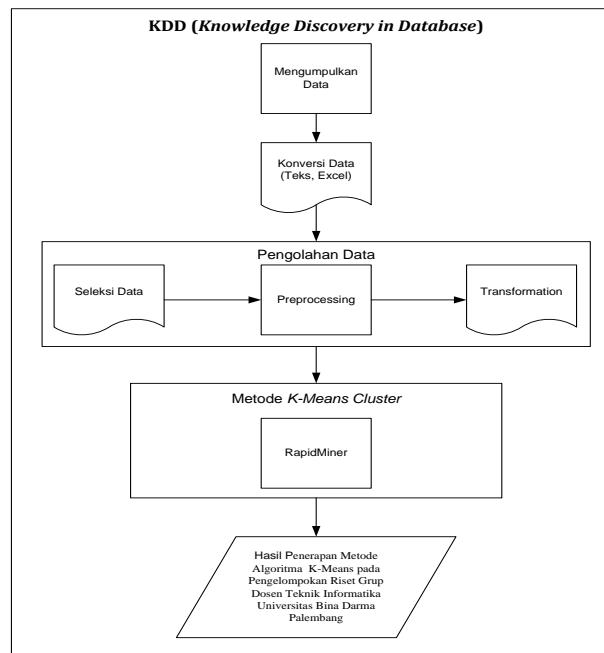
RapidMiner adalah perangkat lunak yang menawarkan tempat terintegrasi untuk *machine learning*, *text mining*, *predictive analytics*, dan *business analytics*[13].

RapidMiner digunakan dalam berbagai aplikasi bisnis, komersial serta untuk penelitian, pendidikan, pelatihan, pembuatan prototipe dengan cepat, dan pengembangan aplikasi. Platform ini mendukung semua tahap dalam proses pembelajaran mesin, persiapan data, visualisasi hasil, validasi dan pengoptimalan[14].

RapidMiner dikembangkan pada open core model, dengan menggunakan RapidMiner Basic Edition pengguna dapat mengunduhnya dibawah lisensi AGPL[13]

2. METODE

Kerangka penelitian ini menggunakan algoritma K-Means untuk melakukan pengelompokan riset grup dosen[15] Proses gambaran kerja penerapan metode algoritma k-means pada penelitian ini dapat dilihat pada gambar berikut :



Gambar 2. Tahapan Penelitian

Alur tahapan pada penelitian ini adalah sebagai berikut :

1. Data Selection

Dalam penelitian ini yang mana datanya berasal dari data penelitian dosen Universitas Bina Darma Palembang, data ini dibagi menjadi tiga bagian yang pertama yaitu data *Research Group* Program Studi Teknik Informatika, kedua data nama-nama dosen UBD Program Studi Teknik Informatika didalam Pangkalan Data Pendidikan Tinggi (PDDikti), dan yang ketiga merupakan data penelitian yang dilakukan oleh dosen Universitas Bina Darma Program Studi Teknik Informatika pada tahun 2019-2022.

2. Preprocessing

Pada tahap ini, dilakukan proses integrasi data pada data penelitian, diikuti dengan pembersihan data (data *cleaning*) untuk menghasilkan dataset yang bersih dan siap digunakan dalam tahap selanjutnya yaitu proses *mining*.

3. Transformation

Pada tahap ini adalah tahapan merubah data ke dalam bentuk yang sesuai untuk di-*mining*. Hasil data yang sudah siap untuk di data mining sebanyak 729 record untuk di cluster dalam penerapan metode *K-Means Clustering*.

Tabel 1. Transformasi Data

| No Dosen | No Research Group | | | | | | |
|----------|-------------------|----|----|---|---|---|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 0 | 21 | 24 | 0 | 0 | 0 | 19 |
| 2 | 0 | 9 | 11 | 0 | 0 | 0 | 9 |
| 3 | 0 | 10 | 8 | 0 | 0 | 0 | 3 |
| 4 | 0 | 10 | 24 | 0 | 0 | 0 | 7 |
| 5 | 0 | 15 | 22 | 0 | 0 | 0 | 4 |
| 6 | 0 | 8 | 10 | 0 | 0 | 0 | 3 |
| 7 | 0 | 6 | 26 | 1 | 0 | 0 | 4 |
| 8 | 0 | 0 | 7 | 0 | 0 | 0 | 3 |
| 9 | 2 | 9 | 14 | 1 | 0 | 0 | 5 |
| 10 | 0 | 5 | 8 | 0 | 0 | 0 | 0 |
| 11 | 0 | 4 | 21 | 0 | 0 | 0 | 5 |
| 12 | 0 | 4 | 30 | 2 | 0 | 0 | 7 |
| 13 | 0 | 9 | 10 | 0 | 0 | 0 | 2 |
| 14 | 0 | 7 | 14 | 0 | 0 | 0 | 1 |
| 15 | 0 | 1 | 7 | 0 | 0 | 0 | 1 |
| 16 | 0 | 1 | 10 | 0 | 0 | 0 | 5 |
| 17 | 0 | 0 | 13 | 2 | 0 | 0 | 0 |
| 18 | 0 | 4 | 12 | 2 | 0 | 0 | 1 |
| 19 | 0 | 2 | 12 | 0 | 0 | 0 | 2 |
| 20 | 0 | 2 | 16 | 2 | 0 | 0 | 4 |
| 21 | 0 | 1 | 12 | 0 | 0 | 0 | 2 |
| 22 | 0 | 2 | 14 | 0 | 0 | 0 | 3 |
| 23 | 0 | 5 | 33 | 0 | 0 | 0 | 4 |
| 24 | 0 | 3 | 13 | 0 | 0 | 0 | 1 |
| 25 | 0 | 14 | 34 | 0 | 0 | 0 | 3 |
| 26 | 0 | 7 | 11 | 1 | 0 | 0 | 1 |
| 27 | 0 | 7 | 26 | 0 | 0 | 0 | 2 |
| 28 | 0 | 0 | 3 | 0 | 0 | 0 | 0 |
| 29 | 0 | 0 | 5 | 0 | 0 | 0 | 0 |

4. Data Mining

Pada tahap data mining dalam penelitian ini, informasi penting yang tersembunyi dalam data yang ada dicari menggunakan metode K-Means *Clustering*. Sebelumnya, telah dilakukan beberapa tahapan untuk mempersiapkan data, yaitu proses *data selection*, *preprocessing*, *data integration* dan *data transformation*. Berdasarkan hasil data yang sudah melalui proses transformasi maka akan dilakukan perhitungan menggunakan metode K-Means *Clustering*.

Tabel 2. Hasil Pembentukan *Cluster* Iterasi -1

| No Dosen | C0 | C1 | C2 | Jarak | Cluster |
|----------|-------|-------|-------|-------|---------|
| 1 | 25,08 | 24,19 | 31,83 | 24,19 | 1 |
| 2 | 7,07 | 7,87 | 12,00 | 7,07 | 0 |
| 3 | 2,45 | 6,78 | 9,27 | 2,45 | 0 |
| 4 | 14,90 | 11,58 | 20,15 | 11,58 | 1 |
| 5 | 13,56 | 11,49 | 20,74 | 11,49 | 1 |
| 6 | 1,41 | 4,24 | 7,87 | 1,41 | 0 |
| 7 | 16,43 | 12,25 | 19,90 | 12,25 | 1 |
| 8 | 9,54 | 9,95 | 2,24 | 2,24 | 2 |
| 9 | 5,48 | 4,24 | 11,58 | 4,24 | 1 |
| 10 | 4,90 | 6,63 | 4,24 | 4,24 | 2 |
| 11 | 12,45 | 8,19 | 14,87 | 8,19 | 1 |
| 12 | 21,31 | 17,15 | 24,04 | 17,15 | 1 |
| 13 | 0,00 | 4,47 | 8,60 | 0,00 | 0 |
| 14 | 4,58 | 0,00 | 9,22 | 0,00 | 1 |
| 15 | 8,60 | 9,27 | 0,00 | 0,00 | 2 |
| 16 | 8,54 | 7,81 | 5,00 | 5,00 | 2 |
| 17 | 9,90 | 7,62 | 6,48 | 6,48 | 2 |
| 18 | 5,83 | 4,24 | 6,16 | 4,24 | 1 |
| 19 | 7,28 | 5,39 | 5,20 | 5,20 | 2 |

| | | | | | |
|----|-------|-------|-------|-------|---|
| 20 | 9,64 | 6,08 | 9,75 | 6,08 | 1 |
| 21 | 8,25 | 6,32 | 5,10 | 5,10 | 2 |
| 22 | 8,12 | 5,10 | 7,35 | 5,10 | 1 |
| 23 | 23,43 | 19,21 | 26,48 | 19,21 | 1 |
| 24 | 6,78 | 4,24 | 6,32 | 4,24 | 1 |
| 25 | 24,54 | 21,21 | 30,03 | 21,21 | 1 |
| 26 | 2,65 | 3,32 | 7,28 | 2,65 | 0 |
| 27 | 16,12 | 12,00 | 19,95 | 12,00 | 1 |
| 28 | 11,58 | 13,19 | 4,24 | 4,24 | 2 |
| 29 | 10,49 | 11,58 | 2,45 | 2,45 | 2 |

Tabel 3. Hasil Pembentukan *Cluster* Iterasi -2

| No Dosen | C0 | C1 | C2 | Jarak Terdekat | Cluster |
|----------|-------|-------|-------|----------------|---------|
| 1 | 13,77 | 24,19 | 20,61 | 13,77 | 0 |
| 2 | 22,85 | 7,87 | 22,57 | 7,87 | 1 |
| 3 | 26,00 | 6,78 | 25,38 | 6,78 | 1 |
| 4 | 11,23 | 11,58 | 9,50 | 9,50 | 2 |
| 5 | 11,62 | 11,49 | 12,93 | 11,49 | 1 |
| 6 | 24,82 | 4,24 | 23,37 | 4,24 | 1 |
| 7 | 13,79 | 12,25 | 7,89 | 7,89 | 2 |
| 8 | 31,15 | 9,95 | 27,76 | 9,95 | 1 |
| 9 | 20,47 | 4,24 | 19,37 | 4,24 | 1 |
| 10 | 28,58 | 6,63 | 26,06 | 6,63 | 1 |
| 11 | 17,85 | 8,19 | 13,14 | 8,19 | 1 |
| 12 | 13,68 | 17,15 | 6,26 | 6,26 | 2 |
| 13 | 24,72 | 4,47 | 23,48 | 4,47 | 1 |
| 14 | 22,37 | 0,00 | 19,79 | 0,00 | 1 |
| 15 | 31,05 | 9,27 | 27,70 | 9,27 | 1 |
| 16 | 27,85 | 7,81 | 24,51 | 7,81 | 1 |
| 17 | 27,35 | 7,62 | 22,74 | 7,62 | 1 |
| 18 | 25,48 | 4,24 | 22,25 | 4,24 | 1 |
| 19 | 26,28 | 5,39 | 22,53 | 5,39 | 1 |
| 20 | 22,86 | 6,08 | 18,65 | 6,08 | 1 |
| 21 | 26,87 | 6,32 | 22,85 | 6,32 | 1 |
| 22 | 24,52 | 5,10 | 20,52 | 5,10 | 1 |
| 23 | 13,25 | 19,21 | 3,97 | 3,97 | 2 |
| 24 | 25,22 | 4,24 | 21,48 | 4,24 | 1 |
| 25 | 7,00 | 21,21 | 5,77 | 5,77 | 2 |
| 26 | 24,87 | 3,32 | 22,73 | 3,32 | 1 |
| 27 | 13,86 | 12,00 | 8,14 | 8,14 | 2 |
| 28 | 35,10 | 13,19 | 31,93 | 13,19 | 1 |
| 29 | 33,45 | 11,58 | 30,04 | 11,58 | 1 |

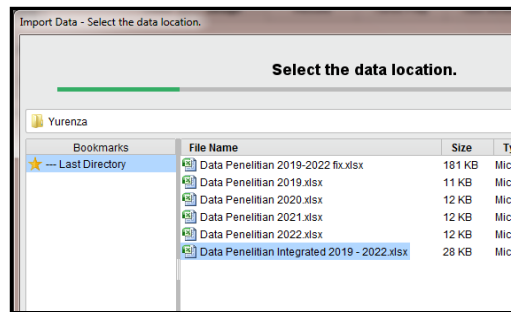
Berdasarkan perhitungan mulai dari iterasi ke 1 sampai iterasi ke 2, didapatkan persamaan hasil pembentukan *cluster*, maka tidak perlu dilakukan iterasi berikutnya dikarenakan hasil *clustering* telah mencapai stabil dan *konvegen*. Untuk hasil akhir mengenai pola data mining dijelaskan pada hasil dan pembahasan dalam bentuk analisa melalui perangkat lunak Rapidminer.

3. HASIL DAN PEMBAHASAN

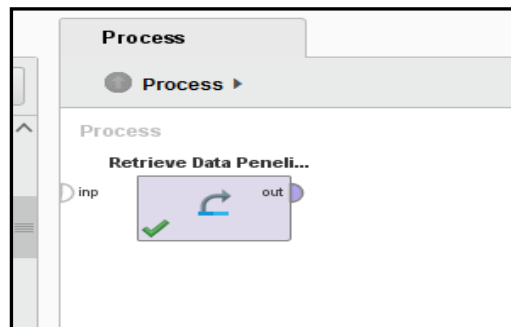
Hasil dan pembahasan mengenai penelitian diuraikan berikut ini :

3.1. Data

Data yang telah melalui proses transformasi akan dimasukkan ke dalam *tools* rapidminer.



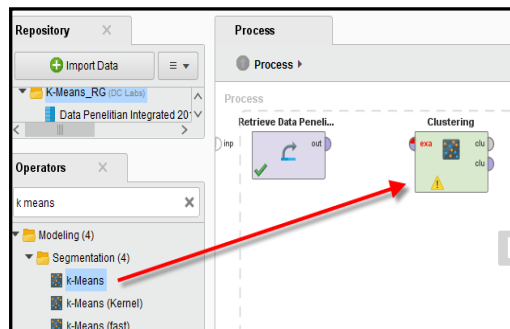
Gambar 3. Import Data Hasil Transformasi



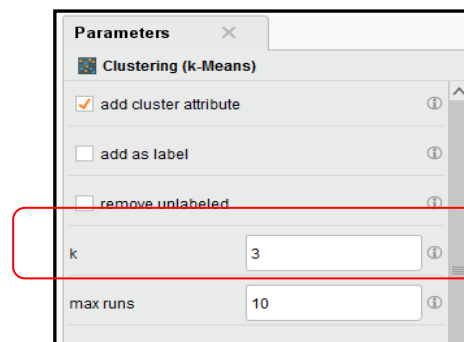
Gambar 4. Data Hasil Setelah Import

3.2. Pembentukan Model K-Means Clustering

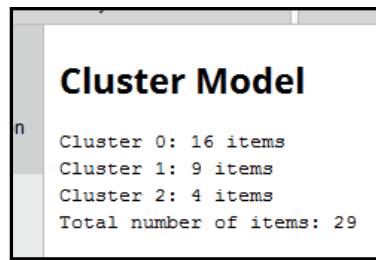
Pada tahap ini akan dilakukan proses pembentukan klusterisasi dengan algoritma k-means. Berikut adalah pembuatan model pada rapidminer menggunakan k-means clustering.



Gambar 5. Pembentukan Model Klusterisasi



Gambar 6. Mengatur Parameter K-Means Clustering



Gambar 7. Model Cluster K-Means

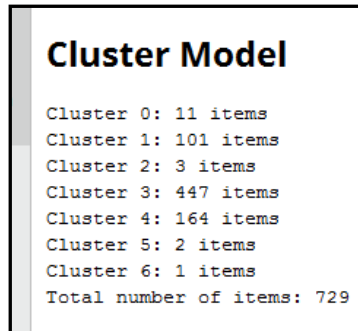
Tabel 4. Hasil Clustering K-Means Dosen

| Cluster | No Dosen | No Research Group | | | | | | |
|---------|----------|-------------------|------|------|-----|-----|-----|------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| C0 | 8 | 0.0 | 0.0 | 7.0 | 0.0 | 0.0 | 0.0 | 3.0 |
| C0 | 10 | 0.0 | 5.0 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| C0 | 13 | 0.0 | 9.0 | 10.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| C0 | 14 | 0.0 | 7.0 | 14.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| C0 | 15 | 0.0 | 1.0 | 7.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| C0 | 16 | 0.0 | 1.0 | 10.0 | 0.0 | 0.0 | 0.0 | 5.0 |
| C0 | 17 | 0.0 | 0.0 | 13.0 | 2.0 | 0.0 | 0.0 | 0.0 |
| C0 | 18 | 0.0 | 4.0 | 12.0 | 2.0 | 0.0 | 0.0 | 1.0 |
| C0 | 19 | 0.0 | 2.0 | 12.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| C0 | 20 | 0.0 | 2.0 | 16.0 | 2.0 | 0.0 | 0.0 | 4.0 |
| C0 | 21 | 0.0 | 1.0 | 12.0 | 0.0 | 0.0 | 0.0 | 2.0 |
| C0 | 22 | 0.0 | 2.0 | 14.0 | 0.0 | 0.0 | 0.0 | 3.0 |
| C0 | 24 | 0.0 | 3.0 | 13.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| C0 | 26 | 0.0 | 7.0 | 11.0 | 1.0 | 0.0 | 0.0 | 1.0 |
| C0 | 28 | 0.0 | 0.0 | 3.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| C0 | 29 | 0.0 | 0.0 | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| C1 | 1 | 0.0 | 21.0 | 24.0 | 0.0 | 0.0 | 0.0 | 19.0 |
| C1 | 2 | 0.0 | 9.0 | 11.0 | 0.0 | 0.0 | 0.0 | 9.0 |
| C1 | 3 | 0.0 | 10.0 | 8.0 | 0.0 | 0.0 | 0.0 | 3.0 |
| C1 | 4 | 0.0 | 10.0 | 24.0 | 0.0 | 0.0 | 0.0 | 7.0 |
| C1 | 5 | 0.0 | 15.0 | 22.0 | 0.0 | 0.0 | 0.0 | 4.0 |
| C1 | 6 | 0.0 | 8.0 | 10.0 | 0.0 | 0.0 | 0.0 | 3.0 |
| C1 | 7 | 0.0 | 6.0 | 26.0 | 1.0 | 0.0 | 0.0 | 4.0 |
| C1 | 9 | 2.0 | 9.0 | 14.0 | 1.0 | 0.0 | 0.0 | 5.0 |
| C1 | 11 | 0.0 | 4.0 | 21.0 | 0.0 | 0.0 | 0.0 | 5.0 |
| C2 | 12 | 0.0 | 4.0 | 30.0 | 2.0 | 0.0 | 0.0 | 7.0 |
| C2 | 23 | 0.0 | 5.0 | 33.0 | 0.0 | 0.0 | 0.0 | 4.0 |
| C2 | 25 | 0.0 | 14.0 | 34.0 | 0.0 | 0.0 | 0.0 | 3.0 |
| C2 | 27 | 0.0 | 7.0 | 26.0 | 0.0 | 0.0 | 0.0 | 2.0 |

Berdasarkan Tabel 4.diatas didapatkan C0 memiliki 4 anggota yaitu dosen dengan nomor **8,10,13,14,15,16,17,18,19,20,21,22,24,26,28** dan **29**. Kemudian C1 memiliki 9 anggota yaitu dosen dengan nomor **1,2,3,4,5,6,7,9**, dan **11**, terakhir C2 memiliki 4 anggota yaitu dosen dengan nomor **12,23,25** dan **27**. Berdasarkan data ini C0 merupakan klaster dengan penelitian terbanyak, C1 merupakan klaster dengan penelitian sedang, dan C2 merupakan klaster dengan penelitian terendah atau sedikit.

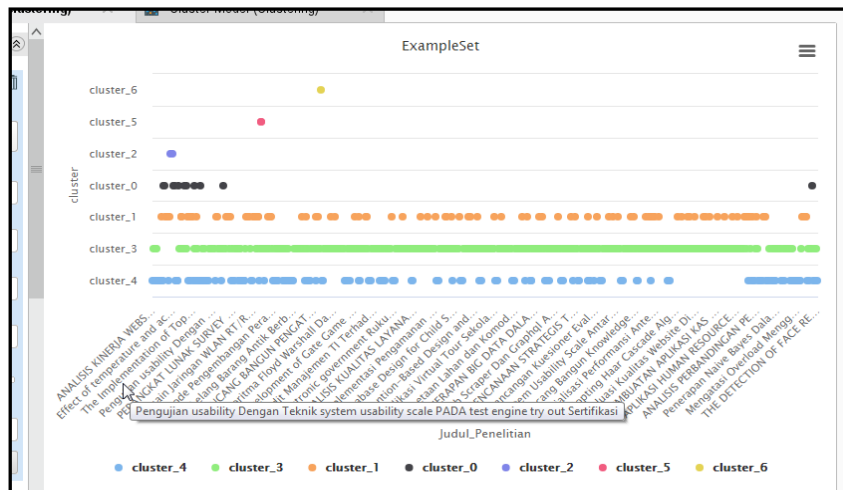
3.2.1. Pembentukan Model K-Means Clustering Judul Penelitian

Peneliti mengatur jumlah *clustering* yang akan digunakan yaitu sebanyak 7 *cluster*.



Gambar 8. Hasil *Clustering* K-Means

Dari Gambar 8 di atas, terlihat *cluster* 0 memiliki anggota 11 items, kemudian *cluster* 1 memiliki anggota 101 items, kemudian *cluster* 2 memiliki anggota 3 items, *cluster* 3 memiliki anggota 447 items, *cluster* 4 memiliki anggota 164 items, *cluster* 5 memiliki anggota 2 items dan *cluster* 6 memiliki anggota 1 items. Total keseluruhan anggota adalah 729 items. Kemudian visualisasi dari hasil klasterisasi dapat dilihat dibawah ini.



Gambar 9. Visualisasi *Cluster* K-Means

Gambar 9 di atas, merupakan hasil pembentukan visualisasi *cluster* dimana sumbu x yaitu atribut judul_penelitian, kemudian untuk kolom yaitu *research_group*.

Tabel 5. Klasterisasi Jumlah Penelitian

| No Research Group | Nama Research Group | Jumlah Penelitian |
|-------------------|--|-------------------|
| 1 | Mathematic and Applied Industrial Analytic | 3 |
| 2 | Software Engineering, Mobile Application and Cloud Computing | 164 |
| 3 | Big Data, Data Analytic and Business Intelligence | 447 |
| 4 | Artificial Intelligence and Soft Computing | 11 |
| 5 | IT Governance, IT Risk and IT Audit | 1 |
| 6 | Multimedia, Visual Informatics and Game | 2 |
| 7 | Network Technology, Cyber Security, IoT and Embeded System. | 101 |

Dari Tabel 5. diatas dapat diketahui bahwa judul penelitian yang paling banyak yaitu pada group *Big Data, Data Analitic and Business Intelligence* sedangkan yang terendah yaitu pada group *IT Governance, IT Risk and IT Audit*.

4. KESIMPULAN

Berikut kesimpulan yang peneliti dapatkan :

1. Penerapan k-means *clustering* dilakukan terhadap judul penelitian Riset Grup Dosen Teknik Informatika Universitas Bina Darma tahun 2019 sampai tahun 2022 dengan atribut judul penelitian, nama dosen dan tahun.
2. Hasil *clustering* Riset Grup Dosen Teknik Informatika terdiri dari dua bagian clustering. Bagian pertama berdasarkan dosen dikelompokan 3 *cluster* berdasarkan 29 data dosen yaitu *cluster 0* merupakan *cluster* dengan penelitian terendah, *cluster1* dengan penelitian sedang, dan *cluster 2* dengan penelitian tertinggi. Bagian kedua berdasarkan judul penelitian dikelompokan 7 *cluster* dengan jumlah 729 data, didapatkan *cluster 0* merupakan *cluster* dengan *research group Artificial Intelligence and Soft Computing (AISRG)* dengan jumlah 11 anggota, kemudian *cluster 1 Network Technology, Cyber Security, IoT and Embeded System (NCIERG)* dengan jumlah 101 anggota, kemudian *cluster 2 Mathematic and Applied Industrial Analytic (MAIARG)* dengan jumlah 3 anggota, kemudian *cluster 3 Big Data, Data Analitic and Business Intelligence (BDBIRG)* dengan jumlah 447 anggota, kemudian *cluster 4 Software Engineering, Mobile Application and Cloud Computing (SEMCRG)* dengan jumlah 164 anggota, selanjutna *cluster 5 Multimedia, Visual Informatics and Game (MAIARG)* dengan jumlah 2 anggota, terakhir *cluster 6 IT Governance, IT Risk and IT Audit (IGRARG)* dengan jumlah 1 anggota.
3. Kecenderungan minat penelitian dosen program studi teknik informatika pada bidang *Big Data, Data Analitic and Business Intelligence* sebanyak 447 Penelitian dan *Software Engineering, Mobile Application and Cloud Computing* sebanyak 164 penelitian.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Prodi Teknik Informatika di Universitas Bina Darma Palembang atas penyediaan data yang diperlukan dan kepada semua pihak yang telah memberikan dalam penelitian ini.

REFERENSI

- [1] Fitri, "Workshop Optimalisasi Pengelolaan Kelompok Penelitian atau Research Group (RG)," LPPM UAD.
- [2] N. Anggraini and L. Zahrotun, "PENGELOMPOKAN JUDUL PENELITIAN DOSEN MENGGUNAKAN METODE K-MEANS DENGAN COSINE SIMILARITY," *Jurnal Sarjana Teknik Informatika Fakultas Teknologi Industri Universitas Ahmad Dahlan*, pp. 2–2, Aug. 2019.
- [3] Mulyawan, G. Dwilestari, A. Bahtiar, F. M. Basysyar, and N. Suarna, "Classification of human development index using particle swarm optimization based on support vector machine algorithm," *IOP Conf Ser Mater Sci Eng*, vol. 1088, no. 1, p. 012033, Feb. 2021, doi: 10.1088/1757-899x/1088/1/012033.
- [4] D. Putro Utomo and Mesran, "Analisis Komparasi Metode Klasifikasi Data Mning dan Reduksi Atribut Pada Data Set Penyakit Jantung," *Jurnal Media Informatika Budidarma*, vol. 4, no. 2, pp. 437–444, Apr. 2020.
- [5] S. Adinugroho and Y. A. Sari, *Implementasi Data Mining Menggunakan WEKA*. Malang: Universitas Brawijaya Press, 2018, 2018.
- [6] I. Kadek, J. Arta, G. Indrawan, and R. Dantes, "DATA MINING REKOMENDASI CALON MAHASISWA BERPRESTASI DI STMIK DENPASAR MENGGUNAKAN METODE TECHNIQUE FOR OTHERS REFERENCE BY SIMILARITY TO IDEAL SOLUTION," *Jurnal Ilmu Komputer Indonesia (JIKI)*, vol. 4, no. 1, 2019.
- [7] S. Pujiono, R. Astuti, and F. Muhamad Basysyar, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN POLA PENJUALAN PRODUK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 615–620, Feb. 2024, doi: 10.36040/jati.v8i1.8360.
- [8] S. Rochmiyati, S. A. Widodo, D. Supriadi, I. Ghozali, and S. Y. Erlangga, *Buku Pedoman Penelitian*. Yogyakarta: Universitas Sarjanawiyata Tamansiswa, 2023.

-
- [9] I. Kamila, U. Khairunnisa, and Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019, doi: <http://dx.doi.org/10.24014/rmsi.v5i1.7381>.
- [10] R. Kesuma Dinata, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," 2020.
- [11] E. Juliana and dan Vivi Nur Aleyda, "PENERAPAN METODE CLUSTERING K-MENS UNTUK MEMBANTU MENENTUKAN TINGKATAN STATUS DAERAH DAMPAK COVID 19," 2021. [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/index>
- [12] S. Irma Yuniarti, "PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN USIA CALON PENERIMA VAKSIN DI KAB. NGAWI," *JurnalInformasidanKomputer*, vol. 9, no. 2, 2021.
- [13] J. Arunadevi, S. Ramya, and M. R. Raja, "A study of classification algorithms using Rapidminer," 2018. [Online]. Available: <http://www.ijpam.eu>
- [14] H. Rizqifaluthi and M. A. Yaqin, "Process Mining Akademik Sekolah menggunakan RapidMiner," *MATICS*, vol. 10, no. 2, p. 47, Mar. 2019, doi: 10.18860/mat.v10i2.5158.
- [15] W. P. Utami and K. Handoko, "PENERAPAN K-MEANS DALAM PENGELOMPOKAN DATA POLIS ASURANSI KENDARAAN BERMOTOR DI PT JASARAHARJA PUTERA," *JURNAL COMASIE*, 2020.