

Implementasi Algoritma *K-Means Clustering* Berbasis Jarak Manhattan Untuk Klasterisasi Konsentrasi Bidang Mahasiswa

Riska Fitri Nur Alifah¹, Abd. Charis Fauzan^{2,*}

Program Studi Ilmu Komputer, Universitas Nahdlatul Ulama Blitar, Indonesia

¹riskafitrinur2401@gmail.com; ²abdcharis@unublitar.ac.id

*corresponding author

INFO ARTIKEL

Sejarah Artikel

Diterima: 5 Januari 2023

Direvisi: 25 April 2023

Diterbitkan: 29 April 2023

Kata Kunci

Davies-Bouldin Index

Jarak Manhattan

Konsentrasi Bidang

K-Means Clustering

ABSTRAK

Konsentrasi bidang studi merupakan pengkhususan studi atau jurusan mahasiswa terhadap suatu bidang studi yang sesuai dengan minat atau kemampuan mahasiswa. Sesuai dengan keputusan Ketua Program Studi mahasiswa diwajibkan untuk memilih konsentrasi bidang, konsentrasi bidang yang telah dipilih harus ditekuni mahasiswa sampai akhir semester. Dalam penentuan pememilihan konsentrasi bidang adalah salah satu hal yang sangat penting bagi mahasiswa karena berpengaruh terhadap kegiatan akademik mahasiswa selanjutnya. Tujuan penelitian ini adalah mengimplementasikan algoritma *K-Means Clustering* dalam mengetahui konsentrasi bidang mahasiswa Ilmu Komputer Universitas Nahdlatul Ulama Blitar. Dengan demikian pengelompokan konsentrasi bidang mahasiswa dapat diimplementasikan dalam algoritma *K-Means Clustering* dalam mengetahui konsentrasi bidang mahasiswa Ilmu Komputer Universitas Nahdlatul Ulama Blitar. Salah satu metode yang dapat digunakan dalam mengelompokkan mahasiswa sesuai dengan konsentrasi bidangnya yaitu metode *K-Means Clustering*. *K-Means Clustering* merupakan salah satu metode *cluster analysis* yang berusaha mempartisi data yang ada kedalam bentuk *cluster* atau kelompok berdasarkan karakteristiknya. Sehingga data yang memiliki kemiripan akan berada pada satu *cluster* sedangkan data yang memiliki ketidaksamaan akan berada pada *cluster* lain. Pada pengelompokan data konsentrasi bidang mahasiswa sebanyak 62 data menggunakan metode *K-Means Clustering* menunjukkan hasil berupa 4 *cluster* yaitu Sistem Informasi, Sistem Cerdas, Net-Centric Computing, dan Game Animasi. Berdasarkan dari perancangan, implementasi, pengujian dan pembahasan maka diperoleh kesimpulan bahwa metode K-Means dengan menggunakan jarak Manhattan dapat digunakan untuk mengklaster konsentrasi bidang mahasiswa. Tingkat kevalitan dihitung dengan metode *Davies-Bouldin Index* (DBI) menghasilkan nilai sebesar 0.27 yang merupakan baik karena semakin mendekati 0 maka tingkat kevalitan semakin baik.

PENDAHULUAN

Konsentrasi bidang studi pengkhususan studi atau jurusan mahasiswa terhadap suatu bidang studi yang sesuai dengan minat atau kemampuan mahasiswa. Dengan demikian konsentrasi bidang juga memiliki tujuan agar mahasiswa dapat lebih mudah dalam mengambil topik tugas akhir [1]. Sesuai dengan keputusan Ketua Program Studi mahasiswa diwajibkan untuk memilih konsentrasi bidang, konsentrasi bidang yang telah dipilih harus ditekuni mahasiswa sampai akhir semester. Dalam penentuan pememilihan konsentrasi bidang, merupakan salah satu hal yang sangat vital bagi mahasiswa karena berpengaruh

terhadap kegiatan akademik mahasiswa selanjutnya. Sesuai keputusan Program Studi, konsentrasi bidang yang ada pada Program Studi Ilmu Komputer terbagi menjadi 4 pilihan yaitu, Sistem Informasi, Sistem Cerdas, Net-Centric Computing, dan Game Animasi.

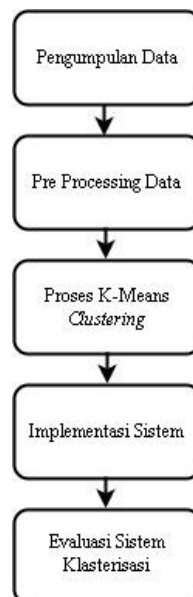
Proses pemilihan konsentrasi bidang secara manual dapat dilakukan dengan mahasiswa memilih sesuai dengan minat masing-masing mahasiswa. Akan tetapi apabila salah dalam memilih konsentrasi bidang dapat mempersulit mahasiswa. Dengan demikian pemilihan atau penentuan konsentrasi bidang dapat dilakukan dengan pengelompokan menggunakan metode *K Means Clustering*. Salah satu metode yang dapat digunakan dalam mengelompokkan mahasiswa sesuai dengan konsentrasi bidangnya yaitu metode *K-Means Clustering*. *K-Means Clustering* merupakan salah satu metode yang berusaha mempartisi data yang ada kedalam bentuk *cluster* atau kelompok [2]. Dalam penelitian ini dilakukan klasifikasi konsentrasi bidang mahasiswa dengan metode perhitungan *K-Means Clustering* dengan menggunakan Jarak Manhattan. Seperti halnya dalam penelitian yang dilakukan oleh Dinata 2020 yaitu tentang klasterisasi transportasi bus menyatakan bahwa Jarak Manhattan memiliki keakuratan lebih daripada Jarak Eclidean. Sehingga jarak Manhattan dapat bekerja dengan baik dalam memberikan rekomendasi untuk klasterisasi [3].

Beberapa penelitian yang dijadikan rujukan dalam penelitian ini antara lain: Jurnal hasil penelitian yang dilakukan oleh Klarinda dkk pada tahun 2018 tentang aplikasi penentuan konsentrasi mahasiswa Teknik Sipil pada Universitas Swadaya Gunung Djati dengan menggunakan metode *K-Means Clustering*. Penelitian ini membahas tentang membuat aplikasi pembantu penentuan konsentrasi jurusan mahasiswa. Tujuan dari penelitian ini adalah untuk membantu mahasiswa dalam menentukan minat konsentrasi sesuai dengan nilai mahasiswa dengan menggunakan metode *K-Means Clustering*. Data yang digunakan berupa NIM Mahasiswa dan Nilai mata kuliah Minat Studi dari semester 1 sampai 5. Hasil dari penelitian adalah menghasilkan sistem penentuan konsentrasi jurusan dengan metode *K-Means Clustering* sesuai kriteria pengelompokan. Ada tiga konsentrasi jurusan yang ada yaitu Struktur Bangunan, Pengendalian Air, dan Struktur Tanah [4]. Dalam penelitian Ayu & Saryanti pada tahun 2019 tentang penerapan penerapan teknik *clustering* untuk pengelompokan konsentrasi mahasiswa dengan metode *K-Means Clustering*. Penelitian tersebut membahas mengenai pengelompokan nilai dari mata kuliah konsentrasi mahasiswa untuk mengetahui seberapa kelompok yang terbentuk dari data yang ada. Data yang digunakan adalah data nilai mahasiswa STIKOM Bali Program Studi Sistem Komputer angkatan 2016 matakuliah Jaringan Komputer, Pemrograman Berorientasi Obyek, Fisika, dan Organisasi Komputer. Tujuan dari penelitian ini adalah mahasiswa dapat memilih bidang konsentrasi dan mahasiswa dapat mengetahui kemampuan akademik yang dimiliki dengan demikian mahasiswa dapat lebih mudah dalam pengambilan tema judul skripsi [5]. Penelitian yang dilakukan oleh Rahayu 2019 tentang Penerapan *K-Means Clustering* Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa data yang digunakan berupa IPK mahasiswa, jarak tempuh mahasiswa, jumlah kehadiran, dan pendapatan orang tua. Metode yang digunakan adalah metode *K Means clustering*. Hasil yang diperoleh dari pengelompokan adalah Algoritma *K-Means* bisa membantu mengklasterisasi calon penerima beasiswa dalam 4 cluster yaitu sangat layak, kurang layak, dipertimbangkan, dan tidak layak menerima beasiswa bidikmisi [6]. Penelitian yang dilakukan oleh Ramadhan 2020 tentang kualitas suhu luar ruangan yang baik untuk melakukan kegiatan. Metode yang digunakan adalah metode *K Means Clustering*. Data yang digunakan berupa data angka dari dataset cuaca BMKG yang memberikan info tentang suhu minimum, maksimum, kelembaban rata-rata, hingga kecepatan angin maksimum di wilayah masing-masing. Hasil

dari penelitian ini adalah 3 *cluster* yaitu kualitas suhu baik, kualitas suhu sedang dan kualitas suhu buruk [7]. Berdasarkan penelitian tersebut, sistem pendukung penentuan konsentrasi bidang pada Program Studi Ilmu Komputer Universitas Nahdlatul Ulama Blitar akan diimplementasikan dengan metode *K-Means Clustering* menggunakan jarak Manhattan.

METODE

Penelitian akan dilakukan berdasarkan rancangan peneliti seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Diagram alir penelitian

Metode yang digunakan pada pembuatan sistem pengelompokan ini terdiri dari :

Pengumpulan Data

Penelitian ini dilaksanakan pada program studi Ilmu Komputer Fakultas Ilmu Eksakta Universitas Nahdlatul Ulama Blitar. Waktu penelitian dilaksanakan pada perkuliahan semester ganjil periode 2020/2021. Data berupa data “nilai mahasiswa angkatan 2017” dan “mata kuliah kurikulum 2016” berbentuk excel .csv. Data nilai dan mata kuliah mahasiswa diperoleh dari Pusat Teknologi Informasi dan Pangkalan Data Universitas Nahdlatul Ulama Blitar. Sedangkan konsentrasi bidang ditentukan oleh program studi. Data yang diperlukan dalam penelitian yaitu :

- NIM dan nama mahasiswa aktif Ilmu Komputer Universitas Nahdlatul Ulama Blitar angkatan 2017.
- Nilai mahasiswa dari mata kuliah yang sesuai dengan konsentrasi bidang yaitu Sistem Informasi Enterprise, Pengolahan Citra, Jaringan Komputer dan Komunikasi Data, Rekayasa Perangkat Lunak, Kecerdasan Baitan, Sistem Temu Kembali Informasi, Sistem Terdistribusi, Interaksi Manusia dan Komputer, Sistem Informasi Geografis, Tata Kelola Teknologi Informasi, Robotika, Keamanan Jaringan dan Grafika Komputer .

Pre-processing Data

Dalam pre-processing data akan dijelaskan bagaimana data setelah di dapatkan dari Pusat Teknologi Informasi dan Pangkalan Data Universitas Nahdlatul Ulama Blitar hingga data siap diolah dalam *K-Means Clustering*:

1. Data yang didapat berupa 2 file excel ber ekstensi .csv. File 1 berupa data kurikulum mata kuliah tahun 2016 beserta atributnya dan file 2 berisikan NIM mahasiswa Ilmu Komputer tahun 2017, nilai dari seluruh mata kuliah yang telah ditempuh dan atribut lainnya.
2. Setelah data didapatkan, data dipilah yang diperlukan saja. Kedua file digabungkan dan ditambahkan nama mahasiswa. Data yang diperlukan berupa NIM, nama mahasiswa, nama mata kuliah dan nilai sesuai dengan ketentuan konsentrasi bidang yang telah ditentukan oleh program studi.
3. Data yang sudah diubah dan dikelompokkan berdasarkan konsentrasi bidang yang meliputi 13 mata kuliah, kemudian dicari rata-rata nilainya.

Proses *K-Means Clustering* Berbasis Manhattan

Algoritma *K-Means Clustering* merupakan salah satu metode yang mengelompokkan data ke dalam kelompok-kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kelompok dan data dengan karakteristik berbeda akan dikelompokkan dalam kelompok lain. Algoritma *K-Means Clustering* merupakan algoritma yang mengelompokkan data berdasarkan kemiripan pada titik pusat. Tujuan dari algoritma *K-Means Clustering* adalah mengelompokkan data dengan memaksimalkan kemiripan data ke dalam satu *cluster* dan meminimalkan kemiripan pada *cluster* lain. Ukuran kemiripan yang digunakan dalam *cluster* adalah fungsi jarak. Sehingga kemiripan data didapat dari jarak terpendek antar data terhadap titik pusat (centroid) [9]. Tahapan algoritma *K-Means Clustering* adalah sebagai berikut [10]. Tentukan (k) sebagai jumlah *cluster* yang akan dibentuk

1. Tentukan titik pusat (*cluster*) awal secara acak
2. Hitung jarak dari setiap objek ke masing-masing centroid dari masing-masing *cluster* dengan jarak Euclidian, sebagaimana tertera pada Persamaan (1).

$$d(x, y) = |x, y|$$

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

Keterangan :

x_i =objek x ke-i

y_i =data y ke-i

n = banyaknya objek

Karena pada penelitian ini menggunakan jarak Manhattan maka Persamaan yang digunakan adalah Persamaan (2).

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

Keterangan :

d = jarak antar x dan y

x = data pusat *cluster*

y = data pada atribut

3. Kelompokkan masing-masing objek ke dalam centroid terdekat
4. Lakukan iterasi, kemudian hitung pusat *cluster* dengan Persamaan (3).

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

5. Ulangi langkah 3-4 hingga data tidak lagi berpindah ke *cluster* yang lain.

Implementasi Sistem

Implementasi sistem adalah tahap pengembangan dari hasil perancangan menjadi kode kode program untuk menjadi sistem berbasis komputer. Dari data yang telah tersusun kemudian diaplikasikan ke dalam program klasterisasi data konsentrasi bidang mahasiswa. Sistem yang dibuat merupakan sistem berbasis aplikasi dekstop yang akan dibuat menggunakan bahasa pemrograman Java dengan Integrated Davelopment Environment (IDE) Netbeans versi 8.2.

Evaluasi Sistem Klasterisasi

Davies-Bouldin Index (DBI) adalah metode yang digunakan untuk mengevaluasi hasil *cluster*. Validasi yang dilakukan adalah seberapa baik *clustering* yang sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari data set. Jumlah klaster terbaik ditunjukkan dengan nilai DB yang semakin kecil [11]. Berikut langkah-langkah perhitungan *Davies-Bouldin Index* [12].

1. *Sum Of Square Within-Cluster* (SSW). Perhitungan *Sum Of Square Within-Cluster* (SSW) bertujuan untuk mengetahui seberapa keterikatan atau kemiripan dalam anggota satu *cluster*. Semakin kecil nilai yang di dapat semakin bagus karena semakin mirip. Berikut persamaan SSW tertera pada Persamaan (4) :

$$SSW_i = \frac{1}{m_1} \sum_{j=i}^{m_1} d(x_j, c_j) \quad (4)$$

Keterangan :

m_1 : jumlah data dalam *cluster* data ke-i

c_j : centroid *cluster* ke-i

$d(x_j, c_j)$: jarak setiap data ke yang dihitung menggunakan jarak ecluidean

2. *Sum of Square Between-Cluster* (SSB). Perhitungan *Sum of Square Between-Cluster* (SSB) bertujuan untuk mengetahui separasi antar *cluster* atau seberapa besar perbedaan antar *cluster* sehingga terpisah ke dalam klompok lain. Semakin besar nilainya maka semakin bagus. Persamaan yang digunakan sebagaimana tertera pada Persamaan (5).

$$SSB_{ij} = d(x_i, x_j) \quad (5)$$

Keterangan :

$d(x_i, x_j)$: jarak antara data ke-i dan data ke-j di *cluster* lain.

3. Ratio (Rasio). Perhitungan rasio bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke-i dan *cluster* ke-j untuk menghitung rasio yang dimiliki masing-masing *cluster*. Untuk menentukan rasio menggunakan Persamaan (6).

$$R_{ij, \dots, n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_{ij} + \dots + SSB_{ni, nj}} \quad (6)$$

Keterangan :

$(SSW)_i$: *Sum Of Square Within-Cluster* pada centroid i

$(SSB)_{ij}$: *Sum of Square Between Cluster* data ke i dengan j pada *cluster* yang berbeda.

4. *Davies Bouldin Index* (DBI). Setelah mendapatkan nilai rasio kemudian menghitung DBI. Semakin kecil nilai DBI yang diperoleh (non negatif ≥ 0) maka *cluster* tersebut semakin baik. Berikut DBI pada Persamaan (7).

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j \dots k}) \quad (7)$$

Keterangan :

$R_{(i,j)}$: ratio dari nilai SSW dan SSB

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai hasil yang diperoleh dari pengujian. Data yang telah diperoleh, akan diproses untuk mendapatkan hasil pengelompokan yaitu, Sistem Informasi, Sistem Cerdas, NCC, dan Game Animasi.

Hasil Pre-processing Data

Dalam pemilihan data yang dijadikan objek penelitian adalah nilai mahasiswa ilmu komputer angkatan 2017. Dari data yang telah didapatkan dipilih untuk data yang diperlukan berupa beberapa nilai dari mata kuliah yang ditempuh sebagaimana tertera pada Tabel 1.

Tabel. 1 Hasil Pemilahan Data

NO	NIM	NAMA	SIG	TATA KELOLA	RPL	SIE	ROBOTIK A	KEC BUATAN	STKI	PENG CITRA	KEJAR	SISTER	JARKOM	GRAFIKA	IMK
1	1755201005	BIYADA CYNTIA IRJAYANI	3.70	4.00	3.00	3.50	2.70	3.70	3.70	3.70	3.50	3.50	3.70	3.50	3.70
2	1755201011	INDERA CAHYO WIBOWO	4.00	4.00	3.70	3.50	4.00	4.00	4.00	3.70	3.70	3.70	3.70	4.00	3.50
3	1755201012	KHARISMA SABBHATUL	3.70	4.00	3.00	3.50	3.70	3.70	4.00	3.70	4.00	3.50	3.70	3.50	3.50
4	1755201017	MOH. AFIF ROFIQI	3.70	4.00	3.50	3.50	3.70	3.70	4.00	3.50	3.70	3.50	3.70	3.50	3.70
5	1755201023	MOHAMMAD ZAINURROKHIM	4.00	4.00	3.00	3.50	3.50	3.50	1.70	3.50	3.50	3.50	3.70	3.70	3.50
...
...
62	1755201113	GUNAWAN	3.00	4.00	3.70	3.00	3.50	3.70	2.70	3.50	4.00	2.70	3.50	3.50	3.50

Hasil K-Means Clustering Berbasis Manhattan

Pada pengelompokan data konsentrasi bidang mahasiswa sebanyak 62 data menggunakan metode *K-Means Clustering* menunjukkan hasil berupa 4 *cluster* yaitu Sistem Informasi, Sistem Cerdas, Net-Centric Computing, dan Game Animasi. Adapun langkah-langkah dari pengelompokan konsentrasi bidang mahasiswa sebagai berikut:

1. Menentukan jumlah *cluster*. Dalam penelitian kali ini jumlah *cluster* yang dibuat adalah 4 *cluster* yaitu Sistem Informasi, Sistem Cerdas, Net-Centric Computing, dan Game Animasi.

2. Untuk menentukan titik pusat (centroid) awal dilakukan secara acak pada data yang ada. Centroid 1 diambil dari data ke-7, Centroid 2 diambil dari data ke-12, Centroid 3 diambil dari data ke- 20, dan Centroid 4 diambil dari data ke- 24 seperti yang tertera pada Tabel 2.

Tabel 2. Centroid awal

Data Ke-7	C1	3.85	3.78	3.80	3.60
Data Ke-12	C2	3.78	3.65	3.70	3.70
Data Ke-20	C3	3.68	3.73	3.70	3.85
Data Ke-24	C4	3.13	3.35	3.47	3.50

3. Menghitung jarak data ke centroid dengan menggunakan jarak manhattan. Setelah melakukan iterasi yang pertama kemudian hitung centroid baru untuk melakukan iterasi berikutnya. Hitung data dengan centroid baru dan ulangi sampai hasil dari pengelompokan *cluster* tidak berubah lagi. Berdasarkan proses iterasi yang telah dilakukan dalam penelitian ini telah dilakukan perhitungan sebanyak 8 iterasi berikut hasil dari pengelompokan seperti yang tertera pada Tabel 3. Dalam tabel tersebut data ke-1 dengan hasil perhitungan K Measn menggunakan jarak Manhattan menghasilkan nilai atribut C1 atau Sistem Informasi 4.24, nilai C2 atau Sistem Cerdas 0.80, nilai C3 atau Net-Centric Computing 0.42, dan nilai C4 atau Game Animasi 12.28. dari keempat nilai tersebut nilai yang paling mendekati adalah nilai C3 maka data ke-1 masuk ke dalam cluater 3 atau Net-Centric Computing.

Tabel 3 Hasil Pengelompokan

ke	C1	C2	C3	C4	CLUSTER
1	4.24	0.80	0.42	12.28	3
2	5.25	1.81	0.59	13.29	3
3	4.63	1.19	0.47	12.67	3
...
...
60	4.08	0.64	0.58	12.12	3
61	3.05	6.49	7.71	5.57	1
62	3.75	0.31	0.91	11.79	2

Hasil Implementasi Sistem

Aplikasi klasterisasi data konsentrasi bidang mahasiswa berdasarkan nilai mata kuliah yang telah ditempuh ini dibuat dengan menggunakan pemrograman java netbeans. Dalam tahap pembuatannya telah disesuaikan dengan perancangan sistem sebelumnya. Dimana sistem bekerja melalui tahap penginputan dari data yang akan diproses sesuai dengan tujuan masing-masing pemrosesan yang nantinya akan menghasilkan output. Rancangan aplikasi klasterisasi konsentrasi bidang mahasiswa berdasarkan nilai mata kuliah yang telah ditempuh menggunakan metode K-Means dapat diimplementasikan seperti Gambar 2.

Gambar 2. Tampilan halaman aplikasi

Data	NIM	Nama	SIG	Tata	RPL	SIE	Robo...	Kec...	STKI	Peng...	KEJAR	SIST...	JARK...	GRAF...	IMK
1	1755...	BIYA...	3.7	4	3	3.5	2.7	3.7	3.7	3.7	3.5	3.5	3.7	3.5	3.7
2	1755...	INDE...	4	4	3.7	3.5	4	4	3.7	3.7	3.7	3.7	3.7	4	3.5
3	1755...	KHA...	3.7	4	3	3.5	3.7	3.7	4	3.7	4	3.5	3.7	3.5	3.5
4	1755...	MOH...	3.7	4	3.5	3.5	3.7	3.7	4	3.5	3.7	3.5	3.7	3.5	3.7
5	1755...	MOH...	4	4	3	3.5	3.5	3.5	1.7	3.5	3.5	3.5	3.7	3.7	3.5
6	1755...	NUR...	4	4	3	3.5	3.5	3.5	4	3.7	4	3.5	3.7	3.7	3.5
7	1755...	RISK...	4	4	3.7	3.7	3.7	3.7	4	3.7	4	3.7	3.7	3.5	3.7
8	1755...	AHM...	4	3	3.5	3.5	3.5	3.5	3.7	3.5	3.7	3.5	3.5	3.5	3.7
9	1755...	RAH...	3	4	3	3.7	2.7	3.5	3.7	3.5	3.7	3.5	3.7	3.5	3.7
10	1755...	EDO...	4	4	3	3.5	3.7	3.5	4	3.50	3.7	3.5	3.5	3.5	3.7
11	1755...	ACH...	4	4	3	3.5	3.7	3.5	4	3.5	3.5	3.5	3.7	3.7	3.7
12	1755...	FAD...	3.7	4	3.7	3.7	3.7	3.7	3.7	3.5	3.7	3.7	3.7	3.7	3.7
13	1755...	ETVL...	3.7	4	3.5	3.5	3.5	3.7	3.7	3	3.5	3.5	3.5	3.5	3.5
14	1755...	BINTL...	4	4	3.7	3.5	3	3.5	3.7	3.5	3.7	3.5	3.7	3.5	3.7
15	1755...	AHM...	4	4	3.5	3.5	3.5	3.7	3.7	3.5	3.5	3.7	3.7	3.7	3.7
16	1755...	AFWA...	4	4	3.7	3.7	3.7	4	3.7	3.7	3.7	3.7	3.7	3.7	3.7
17	1755...	AYU...	4	4	3.7	3.7	4	4	4	3.7	4	3.7	3.7	3.5	3.7
18	1755...	AHM...	4	4	3.5	3.5	3.5	3.5	4	3.5	3.5	3.5	3.7	3.7	3.7
19	1755...	MAR...	4	4	3.5	3.5	3.5	3.5	4	3.7	3.5	3.5	3.7	3.5	3.7
20	1755...	FIQIH...	3.7	4	3.5	3.5	3.5	3.7	4	3.7	3.7	3.7	3.7	3.7	4
21	1755...	ZAIN...	3.7	3	3	3	3	3.5	3.5	3.5	3.5	3	3.5	0	3.5
22	1755...	MOH...	3	3.5	3.5	2.7	3.5	3.7	3.5	4	3.7	3.5	4	3.7	3.5
23	1755...	NUVIL...	4	4	3.5	3.5	3.7	3.5	1.7	3.7	3.7	3.5	3.7	3.5	3.7
24	1755...	MUN...	2	4	3.5	3	2.7	3.5	3.7	3.5	3.7	3	3.7	3.5	3.5
25	1755...	MUH...	4	4	3	3.5	4	3.7	4	3.7	4	3.7	3.7	3.7	3.5

Gambar 3. Data set

Gambar 3 berisikan form data keseluruhan mahasiswa prodi ilmu komputer angkatan 2017 dan nilai-nilai yang akan digunakan sebagai perhitungan pengelompokan.

Centroid	Sistem I...	Sistem ...	NCC	Game A...
Centro...	3.5	3.55	3.56666...	3.6
Centro...	3.675	3.475	3.5	3.5
Centro...	3.175	3.375	3.33333...	1.75
Centro...	3.625	3.84999...	3.80000...	3.6

Gambar 4. Penentuan centroid awal

Gambar 4 merupakan form penentuan data centroid ini berfungsi untuk menentukan

centroid awal dari tiap-tiap kelas. Centroid awal diinputkan oleh user.

Gambar 5. Perhitungan iterasi

Pada Gambar 5 merupakan form perhitungan iterasi ini berfungsi untuk memproses perhitungan data sesuai dengan centroid yang telah ditentukan oleh user. Disini data akan dihitung menggunakan metode K-Means sampai dengan hasil iterasi tidak berubah lagi.

Evaluasi dengan *Davies-Bouldin Index (DBI)*.

Metode pengujian hasil *clustering* tergolong baik atau kurang baik dapat dinilai dari salah satu metode validasi *cluster*. Pada penelitian ini metode validasi yang digunakan untuk menguji hasil *cluster* adalah metode *Davies-Bouldin Index (DBI)*. Berikut hasil pengujian dengan DBI.

1. Menghitung *Sum Of Square Within-Cluster (SSW)*. Langkah pertama yang dilakukan adalah menghitung SSW lakukan pada semua *cluster*. Semakin kecil nilai yang dihasilkan maka semakin bagus *cluster* yang didapat seperti yang tertera pada Tabel 4 [13]. Dalam Tabel 4 fitur dari setiap *cluster* dihitung jaraknya menggunakan jarak euclidean dengan centroid terakhir. Csi merupakan centroid Sistem Informasi, Csc merupakan centroid Sistem Cerdas, Cncc merupakan centroid Net-Centric Computing, dan Cga merupakan centroid Game Animasi. Dan dari hasil perhitungan jarak tersebut menghasilkan nilai SSW sebesar 1.39 dan hasil tersebut tergolong baik karena nilai yang dihasilkan kecil.

Tabel 4. Hasil SSW

Fitur				Cluster yang diikuti	Centroid				Jarak data ke cluster (euclidean)	SSW
C1	C2	C3	C4		Csi	Csc	Cncc	Cga		
3.50	2.50	2.33	3.60	1	2.51	2.3	2.59	2.53	1.49	1.39
2.30	2.43	3.23	3.60	1					1.27	
3.35	2.78	2.90	1.85	1					1.22	
2.63	1.93	2.07	1.75	1					1.01	
0.75	1.88	2.40	1.85	1					1.94	

2. Menghitung *Sum of Square Between-Cluster (SSB)*. Setelah melakukan perhitungan SSW kemudian menghitung SSB. Disini ketika semakin besar nilai yang di dapatkan maka semakin bagus pula hasil perhitungan sebagaimana tertera pada Tabel 5 [13].

Perhitungan SSB ini menggunakan Persamaan (5) dimana SSB ini menghitung jarak antar kelompok.

Tabel 5. Hasil SSB

		Centroid				
		SSB	1	2	3	4
Centroid	1	0.00	1.73	2.34	4.07	
	2	1.73	0.00	0.62	5.78	
	3	2.34	0.62	0.00	6.38	
	4	4.07	5.78	6.38	0.00	

3. Menghitung Rasio. Setelah nilai SSB didapatkan selanjutnya menghitung nilai rasio seperti yang tertera pada Tabel 6. Rasio bertujuan untuk menghitung perbandingan yang dimiliki oleh masing-masing *cluster*.

Tabel 6. Hasil Rasio

		Centroid				
		RASIO	1	2	3	4
Centroid	1	0.00	1.08	0.70	0.61	
	2	1.08	0.00	1.16	0.27	
	3	0.70	1.16	0.00	0.21	
	4	0.61	0.27	0.21	0.00	

4. Menghitung DBI. Setelah menghitung SSW, SSB, dan Rasio kemudian didapat hasil DBI. Dari perhitungan data pengelompokan konsentrasi bidang mahasiswa didapat evaluasi K-Means sebesar 0,27, dimana nilai semakin mendekati 0 (non negative) hasil yang didapat semakin bagus. berikut hasil DBI tertera pada Tabel 7 [13].

Tabel 7. Hasil DBI

		Centroid				R Max	DBI
		RASIO	1	2	3	4	
Centroid	1	0.00	1.08	0.70	0.61	1.08	0.27
	2	1.08	0.00	1.16	0.27	1.16	
	3	0.70	1.16	0.00	0.21	1.16	
	4	0.61	0.27	0.21	0.00	0.61	

KESIMPULAN

Berdasarkan dari perancangan, implementasi, pengujian dan pembahasan maka diperoleh kesimpulan bahwa metode K-Means dapat digunakan untuk mengklaster konsentrasi bidang mahasiswa. *Cluster* yang didapatkan sebanyak 4 *cluster*, yaitu Sistem Informasi, Sistem Cerdas, Net-Centric Computing, dan Game Animasi. Hasil dari perhitungan sebanyak 8 iterasi dan kelompok yang didapat Sistem informasi sebanyak 5 anggota, Sistem Cerdas sebanyak 21 anggota, NCC sebanyak 32 anggota, dan Game Animasi sebanyak 4 anggota. Metode K-Means dapat diaplikasikan dengan jarak Manhattan dengan hasil keakuratan yang dihitung menggunakan perhitungan *Davies-Bouldin Index* (DBI) sebesar 0.27 yang merupakan hasil yang baik. Metode K-Means dapat diaplikasikan

dengan Java Netbeans. Hasil dari sistem pengklasteran konsentrasi bidang mahasiswa ini terdapat saran yang perlu diperhatikan dalam pengembangan berikutnya yaitu: 1) Dataset yang perlu dikembangkan disertakan pengelompokan sesuai konsentrasi bidang. 2) Menerapkan metode K Means *Clustering* dengan Jarak yang lain. 3) Evaluasi yang perlu dikembangkan dengan metode Evaluasi yang lain.

REFERENSI

- [1] N. L. G. P. Suwirmayanti, "Penerapan Teknik *Clustering* Untuk Pengelompokan Konsentrasi Mahasiswa Dengan Metode Self Organizing Map," *J. Ilm.*, vol. 2, no. 1, pp. 11–20, 2020.
- [2] S. Rustam, "Penerapan Optimasi Jumlah Kluster Pada K-means Untuk Pengelompokan Kelas Mata Kuliah Konsentrasi Mahasiswa Semester Akhir," *J. Sist. Inf. DAN Tek. Komput.*, vol. 5, no. 1, pp. 1–4, 2020.
- [3] R. K. Dinata, H. Akbar, and N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 104–111, 2020, doi: 10.33096/ilkom.v12i2.539.104-111.
- [4] C. Klarinda, "Aplikasi Penentuan Konsentrasi Mahasiswa Teknik Sipil Pada Universitas Swadaya Gunung Djati Menggunakan Metode *K-Means Clustering*," pp. 1–10, 2018.
- [5] I. G. Ayu and D. Saryanti, "PENERAPAN TEKNIK *CLUSTERING* UNTUK PENGELOMPOKAN KONSETRASI MAHASISWA DENGAN METODE K-MEANS," pp. 519–526, 2019.
- [6] A. E. Rahayu, K. Hikmah, N. Y. Ningsih, and A. C. Fauzan, "Penerapan *K-Means Clustering* Untuk Penentuan Klasterisasi Beasiswa Bidikmisi Mahasiswa," vol. 1, no. 2, pp. 82–86, 2019.
- [7] M. R. Ramadhan, A. C. Fauzan, N. Aziz, and T. Wahyuni, "K - Means *Clustering* for Determining Quality of Outdoor Temperature Based on BMKG Datasets," vol. 4, no. May, pp. 12–17, 2020.
- [8] 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, doi: 10.21009/pinter.1.2.10.
- [9] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma *K-Means Clustering* untuk Memprediksi Waktu Kelulusan," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 1, pp. 62–66, 2019.
- [10] A. M. H. P. S. Ramadani, I. Ambarita, "Metode K-Means Untuk Pengelompokan Masyarakat Miskin Dengan Menggunakan Jarak Kedekatan Manhattan City Dan Euclidean (Studi Kasus Kota Binjai)," *Inf. Syst. Dev.*, vol. 04, pp. 15–29, 2019.
- [11] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, "Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA," *Pros. Semin. Nas. Mat.*, vol. 2, pp. 161–170, 2019.
- [12] S. Butsianto and N. Saepudin, "Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 51–59, 2020, doi: 10.32672/jnkti.v3i1.2008.
- [13] E. PRASETYO, *Data Mining/ Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI, 2014.