



Sistem Rekomendasi Pemilihan Peminatan Menggunakan *Density Canopy* *K-Means*

Ridho Ananda¹, Muhammad Zidny Naf'an², Amalia Beladonna Arifa³, Auliya Burhanuddin⁴

¹Teknik Industri, Fakultas Rekayasa Industri dan Desain, Institut Teknologi Telkom Purwokerto

^{2,3,4}Teknik Informatika, Fakultas Informatika, Institut Teknologi Telkom Purwokerto

¹ridho@ittelkom-pwt.ac.id, ²zidny@ittelkom-pwt.ac.id, ³amalia@ittelkom-pwt.ac.id, ⁴auliya@ittelkom-pwt.ac.id

Abstract

The carelessly selection of specialization course leaves some students with difficulty. Therefore, it is needed a recommendation system to solve it. Several approaches could be used to build the system, one of them was *K-Means*. *K-Means* required the number of initial centroid at random, so its result was not yet optimal. To determine the optimal initial centroid, *Density Canopy* (DC) algorithms had been proposed. In this research, DC and *K-Means* (DCKM) was implemented to build the recommendation system in the problem. The alpha criterion was also proposed to improve the performance of DCKM. The academic quality dataset in the 2018 informatics programs students of ITTP was used. There were three main stages in the system, namely determination of the weight of the course in dataset, implementation of DCKM, and determination of specialization recommendations. The results showed that the system by using DCKM has good quality based on the *Silhouette* results (at least 0.655). The system also used standar valuation scale in ITTP and *silhouette index* in the process of system. The results showed that 176 (65.91%) students were recommended in IT specialization, 25 (9.36%) students were recommended in MM specialization and 66 (24.7%) students were recommended in SC specialization.

Keywords: Recommendation System, *Density Canopy*, *K-Means*, *Silhouette Index*, Selection of Specialization

Abstrak

Adanya mahasiswa yang kesulitan saat menjalani mata kuliah peminatan merupakan akibat dari pemilihan peminatan tanpa didasarkan pada data. Oleh karena itu dibutuhkan sistem rekomendasi pemilihan peminatan yang didasarkan pada data. Beberapa pendekatan telah dilakukan salah satunya dengan *Clustering K-Means*. *K-Means* membutuhkan jumlah dan letak *centroid* awal secara acak sehingga hasil *clustering* belum tentu optimal. Beberapa ilmuwan telah mengusulkan algoritme dalam penentuan *centroid* awal salah satunya adalah algoritme *Density Canopy*. Paper ini mengimplementasikan algoritme *Density Canopy* dan *K-Means* (DCKM) dalam membangun sistem rekomendasi pemilihan peminatan. Penambahan kriteria α (*alpha*) diusulkan pada proses DCKM untuk meningkatkan kinerja algoritme tersebut. Data yang digunakan adalah data mutu akademik semester 1 dan 2 mahasiswa Prodi Informatika Angkatan 2018 ITTP. Sistem rekomendasi yang dibangun memiliki tiga tahapan utama yaitu (1) penentuan bobot mata kuliah pada dataset, (2) implementasi DCKM dan (3) penentuan rekomendasi peminatan. Hasil penelitian menunjukkan bahwa sistem yang dibangun memiliki kualitas yang baik berdasarkan perolehan hasil *Silhouette* minimal 0.6552. Skala penilaian institusi dan derajat keanggotaan *index Silhouette* $s(i)$ juga dilibatkan sehingga mengurangi unsur subjektivitas. Hasil sistem rekomendasi menunjukkan 176 (65.91%) mahasiswa direkomendasikan pada peminatan TI, 25 (9.36%) mahasiswa direkomendasikan pada peminatan MM dan 66 (24.7%) mahasiswa direkomendasikan pada peminatan SC.

Kata kunci: Sistem Rekomendasi, *Density Canopy*, *K-Means*, *Silhouette Index*, Pemilihan Peminatan

1. Pendahuluan

Institut Teknologi Telkom Purwokerto (ITTP) merupakan salah satu perguruan tinggi di Purwokerto yang telah memiliki 2 fakultas dan 8 program studi di tahun 2019. Salah satu program studi yang perkembangannya cukup pesat adalah informatika. Informatika memiliki tiga peminatan yang wajib dipilih oleh mahasiswa saat semester 4, yaitu teknologi informasi (TI), pengolahan informasi multimedia (MM), dan sistem cerdas (SC).

Sebagian besar pemilihan peminatan yang dilakukan oleh mahasiswa di ITTP belum didasarkan pada data dikarenakan belum adanya sistem yang dibangun untuk membantu mengambil keputusan pemilihan peminatan berdasarkan data. Akibatnya, ada beberapa mahasiswa yang kesulitan saat menjalani mata kuliah peminatan. Ada juga mahasiswa yang mengambil topik tugas akhir di luar peminatan yang diambil dikarenakan merasa kesulitan dengan peminatan tersebut. Perwalian mahasiswa oleh dosen tentang pemilihan peminatan juga seringkali hanya memberikan informasi terkait peminatan tersebut tanpa melihat kondisi dan kemampuan mahasiswa. Hal itu dikarenakan belum adanya sistem rekomendasi pemilihan peminatan bagi mahasiswa berdasarkan data mereka.

Penelitian tentang sistem rekomendasi telah dilakukan di beberapa bidang dengan menggunakan berbagai algoritme, salah satunya algoritme *clustering*. Di bidang kuliner, metode *K-Means* dan *Simple Additive Weighting* telah dilakukan untuk memberikan rekomendasi lokasi wisata kuliner [1]. Di bidang pustaka, *K-Means Clustering* digunakan dalam sistem rekomendasi buku bagi pengunjung [2]. Dalam bidang perdagangan, *K-Means Clustering* dan metode *Decision Tree* digunakan dalam sistem rekomendasi *customer* pada *e-commerce* MTC karebosi [3]. Dalam pendidikan yang berkaitan dengan pemilihan mata kuliah peminatan telah dilakukan dengan menggunakan algoritme *K-Means* dan *Apriori* [4].

K-Means merupakan salah satu algoritme dalam *data mining* yang cukup populer dan masuk dalam kategori *unsupervised learning*. *K-Means* digunakan untuk mengumpulkan objek-objek yang memiliki kemiripan tinggi masuk dalam satu kelompok, serta memisahkan antara objek-objek yang memiliki kemiripan sangat rendah berdasarkan ukuran kuantitatif. Proses awal dari algoritme tersebut adalah penentuan jumlah dan letak *centroid* awal yang seringkali dilakukan secara acak, sehingga hasil *clustering* tidak bersifat tunggal apabila dilakukan perulangan pada proses *K-Means*. Untuk memperoleh hasil *K-Means* yang optimal, beberapa ilmuwan telah melakukan penelitian untuk mengoptimalkan hasil dari *K-Means* dengan mengusulkan algoritme *preprocessing* yang dapat memberikan *centroid* awal untuk proses *K-Means*. Penentuan *centroid* awal dengan pendekatan hiarki

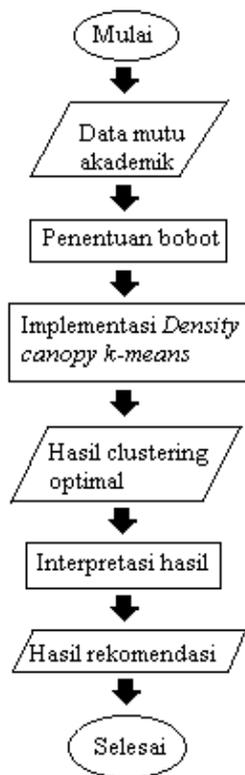
telah dilakukan dalam mengoptimalkan hasil *K-Means* pada pengelompokan provinsi di Indonesia [5]. Penentuan *centroid* awal dengan *Bayesian Information Criterion* (BIC) juga telah dilakukan dalam menganalisis mutu pendidikan SMA di Jawa Tengah [6]. Algoritme lain juga diusulkan dalam penentuan *centroid* awal untuk *K-Means*, salah satunya dengan algoritme *canopy* [7]. Selanjutnya algoritme *canopy* dikembangkan menjadi algoritme *Density Canopy*, dimana algoritme tersebut dapat memberikan banyak dan letak *centroid* awal secara optimal serta robust terhadap data outlier [8].

Berdasarkan uraian di atas, pada paper ini disusun suatu sistem rekomendasi pemilihan peminatan mahasiswa dengan menggunakan algoritme *K-Means*. Penentuan *centroid* awal dan letaknya untuk proses *K-Means* menggunakan algoritme *Density Canopy* karena keunggulannya yaitu *robust* terhadap *outlier*. Paper ini juga memberikan sedikit modifikasi pada algoritme *Density Canopy* untuk memperoleh hasil yang lebih optimal berdasarkan ukuran *Silhouette*. Dalam penentuan rekomendasi, ukuran *Silhouette* juga dilibatkan. Data yang digunakan dalam penelitian ini adalah data mutu akademik mahasiswa angkatan 2018 pada semester 1 dan 2 dimana pada awal dua semester itu merupakan pondasi dari mata kuliah peminatan yang akan dipelajari. Penelitian ini diharapkan dapat meningkatkan performa dari algoritme *Density Canopy K-Means* serta mengaplikasikannya untuk membangun sistem rekomendasi pemilihan peminatan mahasiswa di ITTP.

2. Metode Penelitian

2. 1. Alur penelitian

Alur penelitian dari paper ini memiliki 3 tahapan utama, yaitu tahap perencanaan, implementasi algoritme, dan interpretasi hasil. Tahap perencanaan meliputi perhitungan bobot dari mata kuliah di semester 1 dan 2 terhadap mata kuliah peminatan berdasarkan syarat lulus dan syarat ambil dari mata kuliah peminatan tertentu. Tahap implementasi algoritme adalah proses *clustering* dari data untuk masing-masing peminatan dengan *Density Canopy k-means* sehingga diperoleh *cluster* dengan jumlah yang optimal. Setiap peminatan mungkin akan membentuk *cluster* dengan jumlah yang berbeda, sehingga digunakan skala penilaian yang mengacu pada buku panduan akademik ITTP untuk menstandarisasi hasil *cluster* setiap peminatan. Tahapan interpretasi hasil berkaitan dengan rekomendasi pemilihan peminatan untuk setiap mahasiswa berdasarkan hasil *cluster* yang diperoleh dan ukuran keanggotaan dari index *Silhouette*. Secara sederhana, alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2. 2. Data mutu akademik mahasiswa

Data mutu akademik mahasiswa yang digunakan adalah data hasil belajar mata kuliah semester 1 dan 2 mahasiswa angkatan 2018 Prodi Informatika ITTP. Data tersebut memuat 267 sebagai objek mahasiswa dan 16 mata kuliah sebagai indikator, data berbentuk matriks berukuran 267×16 . Tabel 1 memberikan informasi mata kuliah pada semester 1 dan 2.

Tabel 1. Daftar Mata Kuliah Semester 1 Dan 2

No	Mata kuliah	Kode
1	Fisika	Y1
2	Kalkulus	Y2
3	Algoritme Pemrograman	Y3
4	Praktikum Algoritme Pemrograman	Y4
5	Berpikir komputasi	Y5
6	Kepemimpinan dan Keterampilan Komunikasi	Y6
7	Bahasa Indonesia	Y7
8	Pengantar Teknologi Informasi dan Komunikasi	Y8
9	Agama	Y9
10	Kalkulus Lanjut	Y10
11	Matematika Diskrit	Y11
12	Arsitektur dan Organisasi Komputer	Y12
13	Sistem Digital	Y13
14	Praktikum Sistem Digital	Y14
15	Algoritme dan Struktur Data	Y15
16	Praktikum Algoritme dan Struktur Data	Y16

Tabel 2. Daftar Mata Kuliah Peminatan Teknologi Informasi

No	Mata kuliah	Kode
1	Komputasi awan	TI1
2	Kriptografi	TI2
3	Sistem terdistribusi	TI3
4	keamanan jaringan dan sistem	TI4
5	sistem basis data	TI5
6	tata kelola teknologi informasi	TI6
7	digital forensic	TI7
8	Sistem tertanam dan robotika	TI8
9	pemrograman parallel	TI9
10	Jaringan komputer lanjut	TI10

Tabel 3. Daftar Mata Kuliah Peminatan Informasi Multimedia

No	Mata kuliah	Kode
1	Dasar multimedia	MM1
2	Multimedia interaktif	MM2
3	grafika komputer	MM3
4	pengolahan citra	MM4
5	pemrosesan bahasa alami	MM5
6	Pengenalan suara	MM6
7	Computer vision	MM7
8	Basis data multimedia	MM8
9	Pemodelan animasi dan 3-dimensi	MM9
10	Temu kembali informasi	MM10

Tabel 4. Daftar Mata Kuliah Peminatan Sistem Cerdas

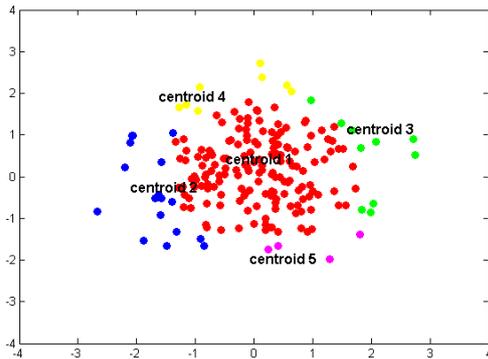
No	Mata kuliah	Kode
1	Pembelajaran mesin	SC1
2	Bioinformatika	SC2
3	metode stokastik	SC3
4	data mining	SC4
5	pengenalan pola	SC5
6	Data Science dan Analisis	SC6
7	Big data	SC7
8	Logika fuzzy	SC8
9	Game intelligent	SC9
10	Agroinformatika	SC10

Terdapat tiga peminatan yang ada pada Prodi Informatika ITTP, yaitu teknologi informasi (TI), informasi multimedia (MM), dan sistem cerdas (SC). Tabel 2 memberikan informasi mata kuliah pada peminatan TI, Tabel 3 memberikan informasi mata kuliah pada peminatan MM dan Tabel 4 memberikan informasi mata kuliah pada peminatan SC.

2. 3. Algoritme *Density Canopy K-Means*

Density Canopy K-Means merupakan algoritme *clustering* yang terdiri dari dua proses utama yaitu penentuan *centroid* (pusat *cluster*) awal dan lokasinya dengan algoritme *Density Canopy* serta proses *clustering* dengan algoritme *K-Means*. Pada

pelaksanaannya, *algoritme Density Canopy* akan memilih *centroid* yang memiliki banyak tetangga berdasarkan ukuran rata-rata jarak antarobjek pada dataset. Pada data dengan distribusi tertentu, ternyata hal itu dapat menyebabkan perolehan *centroid* yang terlalu banyak (tidak optimal) dikarenakan *centroid* pertama kemungkinan tidak jauh dari pusat data. Sedangkan *centroid* berikutnya berdasarkan sisa-sisa data pada pinggiran yang bukan tetangga dari *centroid* sebelumnya. Gambar 2 memberikan ilustrasi dari proses pemilihan *centroid* pada algoritme *Density Canopy*.



Gambar 2. Sebaran Centroid Awal Pada Data Dengan Distribusi Tertentu

Untuk mengatasi hal tersebut, maka dilakukan sedikit modifikasi pada algoritme *Density Canopy* yaitu penambahan kriteria α (*alpha*) pada penentuan jarak tetangganya. Algoritme dari *Density Canopy* dengan penambahan kriteria tersebut adalah:

Step 1. Misalkan \mathbf{X} dataset berukuran $n \times p$ maka kali pertama dilakukan perhitungan jarak tetangga terjauh (d) dengan persamaan (1).

$$d = \frac{2\alpha}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n d_E(\mathbf{x}_i, \mathbf{x}_j) \quad (1)$$

Dimana \mathbf{x}_i merupakan vektor baris dari objek ke- i . $d_E(\mathbf{x}_i, \mathbf{x}_j)$ merupakan jarak Euclidean antara objek ke- i dengan objek ke- j . n merupakan banyak objek. α merupakan kriteria modifikasi yang dilakukan pada paper ini dan memiliki rentan nilai pada $(0, \infty)$. Apabila $\alpha \in (0, 1)$ berarti nilai d lebih kecil dari jarak tetangga terjauh pada *Density Canopy*. Apabila $\alpha \in (1, \infty)$ berarti nilai d lebih besar dari jarak tetangga terjauh pada *Density Canopy*. Sedangkan apabila $\alpha = 1$ maka nilai d sama seperti jarak tetangga terjauh pada *Density Canopy*.

Step 2. Menghitung banyaknya tetangga yang dimiliki oleh masing-masing objek dengan persamaan (2).

$$\rho(i) = \sum_{j=1}^n f(d_E(\mathbf{x}_i, \mathbf{x}_j) - d) \quad (2)$$

dengan ketentuan:

$$f(x) = \begin{cases} 1, & x < 0 \\ 0, & x \geq 0 \end{cases}$$

dimana $\rho(i)$ adalah banyak tetangga dari objek ke- i .

Step 3. Pilih objek ke- i dengan nilai $\rho(i)$ maksimum sebagai *centroid* pertama, selanjutnya objek ke- i dan objek lain yang masuk dalam kategori tetangga objek ke- i dihapuskan dari dataset.

Step 4. Pada dataset sisa, dilakukan perhitungan $\rho(i)$ dengan menggunakan persamaan (2). Menghitung rata-rata jarak objek ke- i dengan tetangganya atau $a(i)$ dengan persamaan (3). Menghitung kepadatan lokal atau $q(i)$ dengan persamaan (4).

$$a(i) = \frac{1}{k} \sum_{j=1}^k d_E(\mathbf{x}_i, \mathbf{x}_j) \quad (3)$$

dimana k adalah banyak tetangga dari objek ke- i . $d_E(\mathbf{x}_i, \mathbf{x}_j)$ merupakan jarak Euclidean antara objek ke- i dengan objek ke- j yang merupakan tetangganya.

$$q(i) = \begin{cases} \min d_E(\mathbf{x}_i, \mathbf{x}_j) & \text{jika } \exists j, \rho(i) < \rho(j) \\ \max d_E(\mathbf{x}_i, \mathbf{x}_j) & \text{jika } \forall j, \rho(i) > \rho(j) \end{cases} \quad (4)$$

Secara sederhana apabila ada tetangga dari objek ke- i , misal $A = \{1, 2, \dots, m\}$, dengan $p(j) > p(i)$ untuk setiap $j \in A$ maka $a(i)$ dipilih dari minimum $d_E(\mathbf{x}_i, \mathbf{x}_j)$ dimana $j \in A$. Sedangkan apabila semua tetangga dari objek ke- i , misal $B = \{1, 2, \dots, k\}$, dengan $\rho(j) < \rho(i)$ untuk setiap $j \in B$ maka $a(i)$ dipilih dari maksimum $d_E(\mathbf{x}_i, \mathbf{x}_j)$ dimana $j \in B$.

Step 5. Melakukan perhitungan bobot dari tiap objek dengan persamaan (5).

$$w(i) = \rho(i) \times \frac{1}{a(i)} \times q(i) \quad (5)$$

dimana $w(i)$ merupakan bobot dari objek ke- i .

Step 6. Objek ke- i yang memiliki nilai $w(i)$ terbesar dijadikan sebagai *centroid* berikutnya. *Centroid* yang baru dan objek yang masuk dalam kategori tetangganya dihapuskan dalam dataset.

Step 7. Proses kembali pada **step 4** hingga tidak ada lagi objek yang tersisa pada dataset.

Centroid-centroid yang telah diperoleh, akan digunakan sebagai *centroid* awal pada proses pengelompokan dengan *K-Means*.

K-Means merupakan algoritme *clustering* yang cukup populer dan termasuk dalam *unsupervised learning*. *K-Means* telah banyak digunakan untuk menyelesaikan beberapa masalah seperti permasalahan segmentasi marketing [9], pelayanan kesehatan masyarakat [10], dan masalah sistem rekomendasi [11]. Dalam

prosesnya, *K-Means* memasukkan objek pada *centroid* terdekat kemudian mengkoreksi *centroid* dengan rata-rata objek-objek pada *centroid*. Proses berulang secara iteratif hingga tidak ada objek yang berpindah ke *centroid* yang lain.

2. 4. Silhouette Index

Silhouette Index merupakan salah satu validasi hasil *clustering* dan termasuk dalam kelompok validasi internal [12]. Telah dilakukan penelitian terkait validasi pada *clustering*, dengan kesimpulan bahwa validasi internal lebih baik dalam menilai hasil *clustering* [13]. Dilakukan juga investigasi untuk mendeteksi adanya outlier dengan menggunakan validasi internal *clustering* dengan *Calinski-Harabasz Index*, *Davies-Boundin Index*, *Silhouette Index*, *Dunn Index*, *R-Squared Index*, diperoleh kesimpulan *Silhouette Index* masuk dalam kelompok validasi internal terbaik bersama dengan *R-Squared Index* dalam mendeteksi adanya outlier dibandingkan validasi internal lainnya [14]. Hal itu yang menjadi alasan penggunaan *Silhouette Index* pada paper ini. Alasan yang lain adalah *Silhouette Index* memiliki rentang nilai untuk ketepatan objek tertentu dalam suatu *cluster* dimana hal itu akan dimanfaatkan pada proses rekomendasi. Nilai dari ketepatan objek tertentu dalam suatu *cluster* dinotasikan dengan $s(i)$ dan dihitung dengan menggunakan persamaan (6).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

dimana $b(i)$ merupakan rata-rata jarak objek ke- i dengan seluruh objek pada *cluster* terdekat (*nearest cluster*). $a(i)$ merupakan rata-rata jarak objek ke- i dengan objek lain dalam satu *cluster* (*intracluster*). Nilai $s(i)$ berada pada interval $[1, -1]$. Apabila $s(i) \approx 1$ maka objek- i berada pada *cluster* yang tepat. Apabila $s(i) \approx 0$ maka objek- i dapat masuk ke 2 *cluster* tertentu. Sedangkan apabila $s(i) \approx -1$ maka objek ke- i masuk dalam *cluster* yang tidak tepat [15]. *Silhouette Index* diperoleh dari rata-rata hasil $s(i)$ untuk setiap objek ke- i .

3. Hasil dan pembahasan

3.1. Penentuan Bobot

Langkah pertama dalam penelitian ini adalah penentuan bobot dari masing-masing variabel data berdasarkan peminatan tertentu. Bobot diperoleh berdasarkan korelasi antara mata kuliah peminatan tertentu terhadap mata kuliah pada dataset berdasarkan syarat lulus dan syarat ambil. Prosedur memperoleh bobot menggunakan persamaan (7).

$$w_{ij} = \frac{\sum_{k=1}^{10} \rho(\alpha_i, \beta_{jk})}{\sum_{i=1}^{16} \sum_{k=1}^{10} \rho(\alpha_i, \beta_{jk})} \quad (7)$$

dimana w_{ij} merupakan bobot dari mata kuliah ke- i terhadap peminatan ke- j . $\rho(\alpha_i, \beta_{jk})$ korelasi antara

mata kuliah ke- i dengan mata kuliah ke- k yang ada di peminatan ke- j . $\rho(\alpha_i, \beta_{jk}) = 1$ apabila ada korelasi antara mata kuliah ke- i dengan mata kuliah ke- k pada peminatan ke- j , dan $\rho(\alpha_i, \beta_{jk}) = 0$ untuk selainnya. Apabila dalam perhitungan bobot diperoleh $x_{ij} = 0$ maka diganti menjadi $w_{ij} = 0.5$. Nilai nol tersebut dikarenakan mata kuliah ke- i tidak ada korelasi terhadap setiap mata kuliah peminatan ke- j berdasarkan syarat lulus dan syarat ambil, namun dalam paper ini diasumsikan mata kuliah tersebut tetap memberikan pengaruh walaupun sedikit, sehingga diberi nilai secara sepihak. Mata kuliah yang memiliki $w_{ij} = 0.5$ kemungkinan pada mata kuliah umum seperti agama, kewarganegaraan, dan lain sebagainya.

Tabel 5. Bobot setiap mata kuliah pada peminatan tertentu

Kode	TI	MM	SC
Y1	0.026	0.038	0.027
Y2	0.051	0.189	0.164
Y3	0.051	0.038	0.110
Y4	0.051	0.038	0.110
Y5	0.026	0.019	0.055
Y6	0.026	0.019	0.014
Y7	0.026	0.019	0.014
Y8	0.359	0.151	0.014
Y9	0.026	0.019	0.014
Y10	0.051	0.189	0.164
Y11	0.051	0.151	0.164
Y12	0.051	0.019	0.014
Y13	0.051	0.019	0.014
Y14	0.051	0.019	0.014
Y15	0.051	0.038	0.055
Y16	0.051	0.038	0.055
Jumlah	1	1	1

Selanjutnya Tabel 5 memberikan informasi bobot masing-masing mata kuliah untuk peminatan tertentu. Pada tabel tersebut, diketahui bahwa mata kuliah yang memiliki bobot terbesar pada peminatan TI adalah Y8. Mata kuliah yang memiliki bobot terbesar pada peminatan MM dan SC adalah Y2 dan Y8.

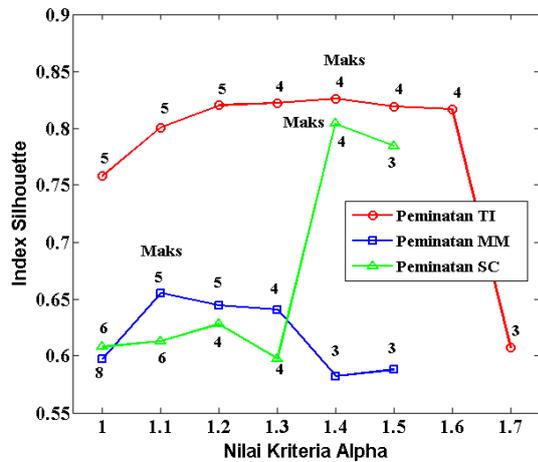
3.2. Hasil Clustering

Selanjutnya dilakukan proses *Clustering* menggunakan *Density Canopy K-Means* dengan ukuran jarak yang digunakan adalah jarak *Euclidean* terbobot, sebagaimana terlihat pada persamaan (8).

$$d_E(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\frac{\sum_{k=1}^p (x_{ik} - x_{jk})^2 w_{kl}}{\sum_{k=1}^p w_{kl}}} \quad (8)$$

dimana \mathbf{x}_i merupakan objek ke- i . x_{ik} merupakan nilai dari objek ke- i pada mata kuliah ke- k . w_{kl} merupakan bobot dari mata kuliah ke- k untuk peminatan ke- l . Sedangkan nilai α untuk *Density Canopy* dipilih $\alpha \geq 1$ dengan banyak *centroid* minimal sejumlah tiga. Pemilihan nilai α pada interval tersebut tujuannya untuk mendapatkan nilai d pada persamaan (1) lebih dari atau sama dengan tetangga terjauh dari *Density Canopy*. Nilai $\alpha < 1$ tidak digunakan dikarenakan akan menghasilkan *centroid* awal yang terlalu banyak

sehingga hasilnya tentu akan tidak optimal. Hasil validasi *Clustering* dengan beberapa nilai α (*alpha*) ditunjukkan pada Gambar 3.



Gambar 3. Hasil Validasi Clustering Untuk Setiap Peminatan Dengan Kriteria α Tertentu

Tabel 6. Sebaran Objek Di Setiap Peminatan

Cluster	TI	MM	SC
Cluster 1	230(86.14%)	195(73.03%)	237(88.76%)
Cluster 2	12(4.49%)	37(13.86%)	26(9.73%)
Cluster 3	24(8.98%)	30(11.23%)	1(0.37%)
Cluster 4	1(0.37%)	4(1.50%)	3(1.12%)
Cluster 5	-	1(0.37%)	-
Jumlah	267(100%)	267(100%)	267(100%)

Berdasarkan Gambar 3, maka hasil *clustering* terbaik untuk peminatan TI diperoleh dengan pemilihan $\alpha = 1.4$ sehingga terbentuk 4 *cluster* dengan nilai *Silhouette Index* 0.8258. Pada peminatan MM, hasil *clustering* terbaik pada pemilihan $\alpha = 1.1$ sehingga terbentuk 5 *cluster* dengan nilai *Silhouette Index* 0.6552. Sedangkan pada peminatan SC, hasil *clustering* terbaik pada pemilihan $\alpha = 1.4$ sehingga terbentuk 4 *cluster* dengan nilai *Silhouette Index* 0.8042. Tabel 6 memberikan informasi terkait sebaran objek pada setiap *cluster* di masing-masing peminatan.

Pada proses *Clustering* yang dilakukan, diketahui bahwa setiap objek akan masuk pada *cluster* tertentu di peminatan TI, MM, dan SC. Untuk memperoleh keputusan objek tersebut masuk dalam salah satu peminatan, maka perlu dianalisis profil dari masing-masing *cluster* yang terbentuk untuk setiap peminatan. Profil *cluster* untuk peminatan tertentu diperoleh dari rata-rata indikator dari *centroid cluster* tersebut, dengan menggunakan persamaan (9).

$$\mu_{ik} = \sum_{j=1}^p x_{ijk} * w_{jk} \quad (9)$$

dimana μ_{ik} profil dari *cluster* ke-*i* pada peminatan ke-*k*. x_{ijk} merupakan nilai data *cluster* ke-*i* pada mata kuliah ke-*j* dan peminatan ke-*k*. w_{jk} merupakan bobot dari mata kuliah ke-*j* pada peminatan ke-*k*. Selanjutnya dengan menggunakan skala penilaian berdasarkan buku petunjuk institusi tahun 2019/2020 seperti ditunjukkan pada Tabel 7.

Tabel 7. Skala Penilaian Berdasarkan Buku Petunjuk Institusi

Nilai	Range
A	> 80,0
AB	70 – 79,99
B	65,00 – 69,99
BC	60,00 – 64,99
C	50,00 – 59,99
D	30,00 – 49,99
E	≤ 29,99

Tabel 8. Profil Masing-Masing Cluster Pada Peminatan Tertentu

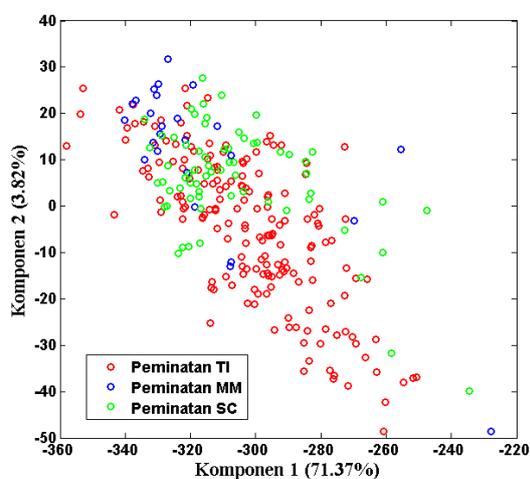
Cluster	TI	MM	SC
Cluster 1	73.00 (AB)	75.61 (AB)	78.21 (AB)
Cluster 2	61.30 (BC)	64.86 (BC)	66.64 (B)
Cluster 3	84.49 (A)	83.73 (A)	48.62 (C)
Cluster 4	43.87 (C)	66.12 (B)	66.37 (B)
Cluster 5	-	39.98 (C)	-

Berdasarkan skala penilaian tersebut, diperoleh profil dari masing-masing *cluster* pada peminatan tertentu ditunjukkan pada Tabel 8. Profil *cluster* akan melekat pada objek yang ada didalam *cluster* tersebut. Selanjutnya penentuan rekomendasi mahasiswa dilihat berdasarkan profil objek dan nilai keanggotaannya terhadap *cluster* berdasarkan nilai $s(i)$ sebagaimana beberapa datanya ditunjukkan pada Tabel 9.

Tabel 9. Profil Mahasiswa Dan Nilai Keanggotaan Terhadap Cluster Sebagai Penentu Rekomendasi Mata Kuliah Peminatan

Mahasiswa	TI		MM		SC		Rekomendasi
	$s(i)$	profil	$s(i)$	profil	$s(i)$	profil	
N1	0.85	AB	0.37	AB	0.79	AB	TI
N2	0.84	AB	0.31	AB	0.79	AB	TI
N3	0.90	A	0.13	AB	0.72	AB	TI
N4	0.80	AB	0.40	AB	0.80	AB	TI
N5	-0.15	AB	0.63	A	0.81	AB	MM
N6	0.78	AB	0.55	AB	0.56	AB	TI
N7	0.90	A	0.69	A	0.80	AB	TI
N8	0.81	AB	0.49	AB	0.74	AB	TI
N9	0.60	AB	0.35	AB	0.73	AB	SC
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
N267	0.83	AB	0.63	AB	0.61	AB	TI

Pada Tabel 9 diperoleh informasi profil dan ukuran keanggotaan dari masing-masing mahasiswa terhadap peminatan tertentu. Contoh prosedur rekomendasinya sebagai berikut: mahasiswa N1 masuk dalam profil AB untuk setiap peminatan TI, MM, dan SC. Untuk memutuskan mahasiswa tersebut direkomendasikan dalam peminatan tertentu, maka dipertimbangkan berdasarkan nilai keanggotaan dalam *cluster*. Nilai keanggotaan atau $s(i)$ pada peminatan TI, MM, dan SC dari mahasiswa N1 berturut-turut adalah 0.85, 0.37, dan 0.79. Berdasarkan nilai tersebut maka mahasiswa N1 direkomendasikan masuk dalam peminatan TI karena nilai keanggotaannya terbesar. Sedangkan pada mahasiswa N3 langsung direkomendasikan dalam peminatan TI dikarenakan profilnya dalam TI lebih baik dibandingkan dengan peminatan lainnya.



Gambar 4. Visualisasi sebaran rekomendasi mahasiswa

Dengan prosedur rekomendasi seperti itu maka diperoleh hasil 176 (65.91%) mahasiswa direkomendasikan pada peminatan TI, 25 (9.36%) mahasiswa direkomendasikan pada peminatan MM, dan 66 (24.7%) mahasiswa direkomendasikan pada peminatan SC. Gambar 4 memberikan visualisasi sebaran mahasiswa berdasarkan rekomendasi akhir yang diperoleh. Visualisasi tersebut menggunakan algoritme analisis komponen utama (PCA, *principal component analysis*) pada data asal dan tidak menyertakan bobot dari masing-masing peminatan.

4. Kesimpulan

Paper ini telah mengimplementasikan algoritme DCKM sebagai sistem rekomendasi pemilihan peminatan bagi mahasiswa informatika angkatan 2018 dengan menggunakan data mutu akademik semester 1 dan 2. Pada prosesnya, diusulkan penambahan kriteria α untuk meningkatkan kinerja dari DCKM. Adanya kriteria tersebut menunjukkan peningkatan kemampuan algoritme DCKM berdasarkan validasi internal index Silhouette. Sistem rekomendasi yang dibangun pada penelitian ini memiliki kualitas yang baik berdasarkan perolehan hasil Silhouette pada peminatan TI sebesar

0.8258, peminatan MM sebesar 0.6552, dan peminatan SC sebesar 0.8042. Skala penilaian institusi dan derajat keanggotaan index Silhouette $s(i)$ juga dilibatkan dalam sistem rekomendasi sehingga mengurangi unsur subjektifitas dalam rekomendasi pemilihan peminatan. Hasil sistem rekomendasi menunjukkan 176 (65.91%) mahasiswa direkomendasikan pada peminatan TI, 25 (9.36%) mahasiswa direkomendasikan pada peminatan MM, dan 66 (24.7%) mahasiswa direkomendasikan pada peminatan SC.

Ucapan Terimakasih

Penelitian ini telah mendapatkan dukungan finansial dari LPPM Institut Teknologi Telkom Purwokerto berdasarkan hibah penelitian internal tahun 2019. Berdasarkan hal itu, kami mengucapkan terima kasih atas dukungan tersebut.

Daftar Rujukan

- [1] Saputa A., Mulyawan B., Sutrisno T., 2019. Rekomendasi Lokasi Wisata Kuliner di Jakarta Menggunakan Metode K-Means Clustering dan Simple Additive Weighting. *Jurnal Ilmu Komputer dan Sistem Informasi*, 7(1), pp. 14-21.
- [2] Putra I. M. A.W., Indrawa G., Aryanto K. Y. E., 2018. Sistem Rekomendasi Berdasarkan Data Transaksi Perpustakaan Daerah Tabanan dengan Menggunakan K-Means Clustering. *Jurnal Ilmu Komputer Indonesia*, 3(1), pp. 18-22.
- [3] Aprizal, Hasriani, Rauf A., 2018. Implementasi Sistem Rekomendasi Barang Customer pada E-Commerce MTC Karebosi Menggunakan Metode K-Means Clustering dan Metode Decision Tree, *Prosiding Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, Makassar, Indonesia. 7(2).
- [4] Afifuddin R. N., Nurjanah D., 2019. Sistem Rekomendasi Pemilihan Mata Kuliah Peminatan Menggunakan Algoritme K-means dan Apriori (Studi Kasus: Jurusan S1 Teknik Informatika Fakultas Informatika), *E-proceeding of Engineering*, 6(1), pp. 2359-2367.
- [5] Widiyanti, W., 2017. *Sentroid Awal Metode K-Means Dengan Pendekatan Metode Berhierarchy Dalam Pengelompokan Provinsi Di Indonesia*. Skripsi FMIPA: Institut Pertanian Bogor.
- [6] Ananda R., Burhanuddin A., 2019. Analisis Mutu Pendidikan Sekolah Menengah Atas Program Ilmu Alam Di Jawa Tengah Dengan Algoritme K-Means Terorganisir. *Journal of INISTA*, 2(1), pp. 065-072.
- [7] Kumar A., Ingle Y. S., Pande A., Dhule P., 2014. Canopy Clustering: A Review on Pre-Clustering Approach To K-Means Clustering. *IJIACS*, 3(5), pp. 22-29.
- [8] Zhang G., Zhang C., Zhang H., 2018. Improved K-Means Algorithm Based on Density Canopy. *Journal Knowledge-based Systems*, 145, pp. 289-297.
- [9] Abriyanti A., Damastuti N., 2019. Segmentasi Mahasiswa dengan Unsupervised Algoritme guna Membangun Strategi Marketing Penerimaan Mahasiswa. *Jurnal Insand Comtech*, 4(2), pp. 10-18.
- [10] Cahyana N. H., Aribowo A. S., 2018. Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat. *Seminar Nasional Informatika Medis*, pp. 24-31.
- [11] Listiani, L., Agustin, Y. H., Ramdhani, M. Z., 2019. Implementasi Algoritme K-Means Cluster Untuk Rekomendasi Pekerjaan Berdasarkan Pengelompokan Data

- Penduduk. *Seminar Nasional Sistem Informasi dan Teknik Informatika*, pp. 761-769.
- [12] Rendon E., Abundez I. M., Gutierrez C., Zagal S. D., Arizmendi A., Quiroz E.M., Arzate H. E., 2011. A Comparison of Internal and External Cluster Validation Indexes. in *Proceeding of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and Applications*.
- [13] Rendon E., Abundez I. M., Quiroz E. M., 2011. Internal Versus External Cluster Validation Indexes. *International Journal of Computers and Communications*, 5(1), pp. 27-34.
- [14] Anitha S., Metilda M., 2019. An Extensive Investigation of Outlier Detection by Cluster Validation Indices. *Ciencia e Tecnica Vitivinicola-A Science and Technology Journal*, 34(2),pp. 22-32.
- [15] Hassan S. I., Samad A., Ahmad O., Alam A., 2019. Partitioning and Hierarchical Based Clustering: A Comparative Empirical Assessment on Internal and External Indices, Accuracy, and Time. *Int. J. Inf. Tech.*, doi: <https://doi.org/10.1007/s41870-019-00406-7>