Algoritma K-Means Untuk Pengelompokan Topik Skripsi Mahasiswa

By Muhammad Rafi Muttaqin

WORD COUNT 3548
TIME SUBMITTED 14-AUG-2020 09:02PM
PAPER ID 62177193

ILKOM Jurnal Ilmiah
Vol 12, No. 2, Agustus 2020, pp. xx-xx

E-ISSN: 2548 - 7779

Algoritma K-Means Untuk Pengelompokan Topik Skripsi Mahasiswa

Muhammad Rafi Muttaqin [a,1,*], Meriska Defriani [a,2]

[a] *Teknik Informatika, STT Wastukancana, Jalan Cikopak No.53,Kabupaten Purwakarta, 41151, Indonesia*
[1] *rafi@stt-wastukancana.ac.id; [2]meriska@stt-wastukancana.ac.id*
**corresponding author*

INFORMASI ARTIKEL

Diterima :
Direvisi :
Diterbitkan :

Kata Kunci:
Data Mining
Knowledge Discovery in Database
Clustering
K-Means
Rapidminer

ABSTRAK

Dalam membantu mengembangkan teknologi di bidang pendidikan serta membawa suatu perubahan besar dalam daya saing antar individu maupun kelompok, untuk dapat melakukannya membutuhkan suatu informasi serta data yang cukup agar dapat dianalisis lebih lanjut. Dalam hal ini STT Wastukancana Purwakarta berada dibawah naungan Yayasan Bunga Bangsa melihat bahwa mahasiswa STT Wastukancana Purwakarta mempunyai beberapa kendala dalam tugas akhir, salah satunya yaitu sulit dalam menentukan topik judul skripsi yang akan dibuatnya sehingga terkadang topik judul skripsi yang diambil tidak sesuai dengan kemampuan masing-masing mahasiswa. Masalah ini dapat diatasi dengan menerapkan metode clustering. Metode analisis yang digunakan adalah Knowledge Discovery in Database (KDD). Metode pengelompokan mahasiswa menggunakan metode clustering dan algoritma K-Means sebagai perhitungan clusteringnya dimana Clustering ini bertujuan untuk membagi mahasiswa ke dalam cluster berdasarkan nilai yang diperolehnya dari semester 1 s/d 7, sehingga dapat menghasilkan rekomendasi mahasiswa dalam mengambil topik skripsi. Alat yang digunakan untuk mengimplementasi algoritma tersebut yaitu Rapidminer. Hasil dari penelitian ini adalah pengelompokan mahasiswa sesuai dengan keahliannya, yang didapat berdasarkan cluster yang memiliki nilai yang paling tinggi dan didominasi pada mata kuliah yang paling banyak sesuai dengan mata kuliah yang sudah dikelompokan masing-masing keahlian. Sehingga hasil cluster ini yang menjadi acuan sebagai rekomendasi mahasiswa dalam mengambil topik judul skripsi.

Keywords:
Data Mining
Knowledge Discovery in Database
Clustering
K-Means
Rapidminer

ABSTRACT

In helping to develop technology in the field of education as well as bringing about a major change in competitiveness between individuals and groups, to be able to do so requires sufficient information and data to be analyzed further. In this case STT Wastukancana Purwakarta is under the auspices of Bunga Bangsa Foundation, seeing that STT Wastukancana Purwakarta students have several obstacles in their final project, one of which is difficult in determining the topic of the thesis title to be made so that sometimes the topic of the thesis title taken is not in accordance with their abilities each student. This problem can be overcome by applying the clustering method. The analytical method used is Knowledge Discovery in Database (KDD). The method of grouping students uses the clustering method and the K-Means algorithm as a clustering calculation where the Clustering aims to divide students into clusters based on grades obtained from semester 1 to 7, so as to produce student recommendations in taking thesis topics. The tool used to implement the algorithm is Rapidminer. The results of this study are grouping students according to their expertise, which is obtained based on the cluster that has the highest score and is dominated by the most subjects according to the subjects that have been grouped by each expertise. So the results of this cluster are used as a reference for students to take the thesis title topic.

I. Pendahuluan

Tugas akhir atau skripsi merupakan salah satu syarat yang harus dilaksanakan oleh mahasiswa di sebuah perguruan tinggi untuk dapat lulus menjadi seorang sarjana. Dalam pengambilan topik skripsi, program studi Teknik Infomatika STT XYZ Purwakarta memberikan beberapa pilihan topik atau peminatan yang dapat dipilih oleh mahasiswa. Pemilihan topik atau peminatan tersebut akan lebih baik jika tidak hanya sesuai dengan minat tapi juga sesuai dengan kemampuan masing-masing mahasiswa. Kemampuan tersebut dapat dilihat dari data akademik mahasiswa, yaitu dari hasil studi selama proses perkulihan dari semester satu sampai semester tujuh. Hampir seluruh matakuliah yang diselenggarakan memiliki korelasi dengan topik atau peminatan yang dapat dipilih. Oleh karena itu, dapat dilakukan suatu analisis pada data akademik mahasiswa yang hasilnya dapat membantu untuk menentukan topik skripsi yang sesuai dengan minat dan kemampuan. Dengan topik dan peminatan yang sesuai, mahasiswa dapat memaksimalkan proses pengerjaan skripsi sehingga dapat menyelesaikan skripsi tepat waktu. Penyelesaian skripsi yang tepat waktu dapat memberikan keuntungan untuk program studi terutama untuk kebutuhan akreditasi.

Pada penelitian ini dilakukan pengelompokan dengan menggunakan data mining. Data mining merupakan teknik untuk menemukan suatu pola di dalam sejumlah besar data. Algoritma yang digunakan adalah K-Means. Algoritma K-Means merupakan algoritma yang mudah untuk diimplementasikan, diadaptasi, dan membutuhkan waktu pembelajaran yang relatif cepat. Algoritma K-Means sebelumnya pernah digunakan oleh Bakti dan Indriyatno [1] untuk mengelompokkan dokumen tugas akhir yang kemudian hasilnya akan digunakan untuk analisis sistem temu kembali. Selain itu, Solehudin, dkk [2] untuk mengelompokan dokumen skripsi berdasarkan tema, objek, dan metode penelitian.

II. Metode

A. *Data Mining*

Data mining merupakan sekumpulan teknik untuk menemukan pengetahuan yang sebelumnya tidak diketahui dalam basis data yang besar. Pola yang ditemukan tersebut dapat digunakan untuk membantu pengambilan sebuah keputusan[3]. Data mining tidak hanya dapat digunakan dalam menemukan pengetahuan atau fenomena baru, melainkan juga untuk meningkatkan pemahaman kita mengenai apa yang kita ketahui [4]. Data mining juga didefinisikan sebagai proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidenfikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar atau data warehouse[5].

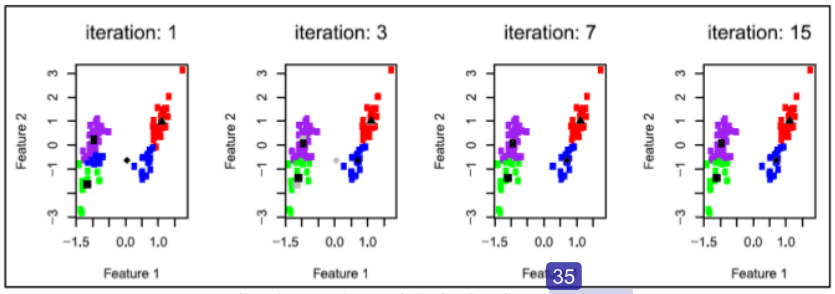
B. *Clustering*

Clustering adalah teknik pembelajaran tanpa pengawasan yang tidak memerlukan dataset berlabel. Clustering didefinisikan sebagai pengelompokan sekumpulan objek yang mirip ke dalam kelas atau cluster. Selama proses analisis, data dikelompokkan ke dalam kelas atau cluster sehingga objek di dalam sebuah cluster memiliki kemiripan yang tinggi satu sama lain tetapi memiliki perbedaan yang tinggi dibandingkan dengan objek di cluster lain[3]. Clustering adalah metode yang tidak diawasi yang digunakan untuk memisahkan data menjadi kelompok sehingga objek yang dimiliki oleh satu kelompok serupa dan berbeda dari objek di kelompok lain[6][7].

C. *K-Means*

Algoritma K-Means merupakan algoritma klasterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya[8]. K-Means juga merupakan metode pengelompokan paling sederhana [9] yang mengelompokkan data ke dalam k kelompok berdasar centroid masing-masing kelompok. Hanya saja hasil dari K-Means sangat dipengaruhi parameter k dan inisialisasi centroid. Pada umumnya K-Means menginisialisasi centroid secara acak [10]. Langkah-langkah dari algoritma K-Means menurut Witten,dkk [11] adalah sebagai berikut :

1. Menentukan satu objek secara acak sebagai pusat cluster awal.
2. Menetapkan setiap objek ke kluster yang objeknya memiliki kemiripan berdasarkan nilai rata-rata objek di dalam cluster.
3. Memperbaharui rata-rata nilai cluster dengan menghitung rata-rata objek di setiap cluster.
4. Mengulangi langkah ke 2 dan 3 hingga anggota cluster tidak berubah.



Gambar 1. Ilustrasi dari Algoritma K-Means

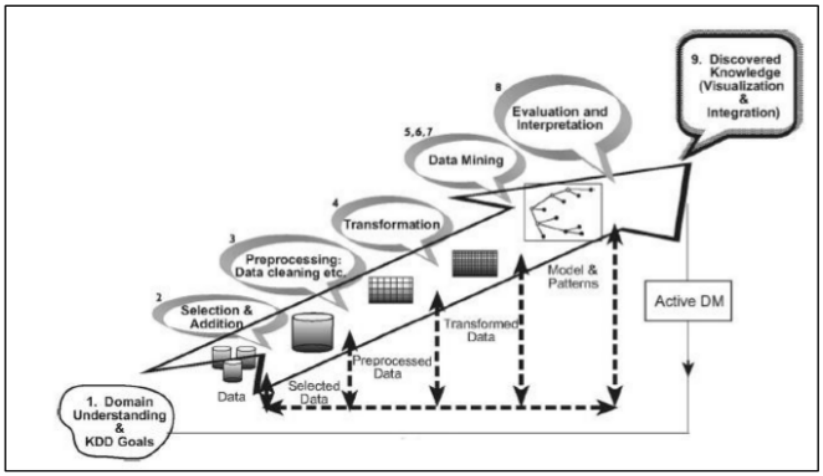
Gambar 1 merupakan ilustrasi dari Algoritma K-Means. Setiap plot menunjukan partisi yang diperoleh setelah iterasi tertentu. Centroid setiap cluster ditandai dengan warna hitam. Pada contoh tersebut terdapat empat cluster yang ditandai dengan warna yang berbeda-beda[12].

D. Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah Knowledge Discovery in Database (KDD). KDD merupakan sebuah proses komputasi yang didalamnya terdapat penggunaan algoritma-algoritma matematika yang berfungsi untuk mengekstraksi data dan melakukan perhitungan probabilitas kemungkinan tindakan di masa yang akan datang[3]. Hasil dari KDD berupa pengetahuan yang sebelumnya tidak diketahui, potensial, dan bermanfaat [13]. Data Mining merupakan salah satu langkah dari serangkaian proses iterative KDD[14]. Terdapat 9 tahapan dalam proses KDD [15] seperti yang dapat dilihat pada Gambar 1. Penjelasan dari setiap tahapan dalam proses KDD adalah sebagai berikut:

a. Domain Understanding and KDD Goals

Pada tahapan ini dilakukan pemahaman mengenai apa yang akan dilakukan dalam proses pemodelan data mining dan penentuan tujuan dilakukan data mining hingga implementasi ke lingkungan dimana proses penemuan pengetahuan akan berlangsung. Tahapan ini dapat direvisi ketika hasil pemodelan data mining tidak sesuai dengan tujuannya.



Gambar 2. Proses KDD

b. Selection and Addition

Pada tahapan ini dilakukan pengumpulan data. Data yang telah dikumpulkan kemudian dilakukan seleksi atribut dan hasil dari seleksi tersebut diintegrasikan menjadi sebuah dataset. Proses pembangunan dataset merupakan suatu proses yang penting karena proses pembelajaran data mining dan penemuan pola baru didasarkan pada dataset yang telah dibentuk.

c. Preprocessing and Data Cleaning

Pada tahapan ini dilakukan pembersihan data untuk meningkatkan keandalan data. Pembersihan data dilakukan dengan menangani nilai kosong, menangani baris data yang tidak relevan, dan menghilangkan noise atau outlier. Proses persiapan awal ini dapat melibatkan metode statistik yang kompleks atau menggunakan algoritma data mining yang spesifik.

d. Transformation

Pada tahapan ini dilakukan pengembangan data sehingga data dipersiapkan dengan lebih baik dan siap untuk dilakukan pemodelan data mining. Hal yang dapat dilakukan untuk mempersiapkan data menjadi lebih baik adalah melakukan reduksi dimensi seperti pemilihan fitur dan ekstrasi sampel data. Selain itu, dapat juga dilakukan transformasi atribut seperti mengubah atribut numerik menjadi atribut diskrit dan transformasi fungsional.

e. Data Mining

Tahapan data mining terdiri dari tiga tahapan, yaitu pemilihan model data mining, pemilihan algoritma data mining, dan penggunaan data mining. Model data mining yang dipilih adalah model yang sesuai dengan kebutuhan, yaitu klasifikasi, regresi, atau pengelompokan. Algoritma yang dipilih disesuaikan dengan model data mining yang dipilih dengan mempertimbangkan kelebihan dan kekurangan dari algoritma tersebut. Setelah dilakukan pemilihan model dan algoritma data mining, kemudian data mining digunakan untuk menemukan pola atau aturan yang baru. Algoritma data mining dimungkinkan untuk digunakan berulang kali hingga memperoleh hasil yang sesuai. Pada penggunaan data mining juga dilakukan pengaturan parameter kontrol algoritma.

f. Evaluation and Interpretation

Pada tahapan ini dilakukan evaluasi dan penafsiran pengetahuan yang berupa pola hasil penggunaan data mining. Selain itu juga dilakukan pendokumentasian pengetahuan yang ditemukan untuk penggunaan lebih lanjut.

g. Discovered Knowledge

Pada tahapan ini, pengetahuan yang ditemukan telah siap untuk diimplementasikan ke dalam sebuah sistem. Tahapan ini menentukan keaktifan keseluruhan proses[15].

III. Hasil dan Pembahasan

Berikut ini adalah pemaparan dari hasil pemodelan data mining dalam penelitian ini:

A. Domain Understanding and KDD Goals

Tugas akhir atau skripsi merupakan salah satu syarat yang harus dilaksanakan oleh mahasiswa di sebuah perguruan tinggi untuk dapat lulus menjadi seorang sarjana. Dalam pengambilan topik skripsi, program studi Teknik Infomatika STT XYZ Purwakarta memberikan beberapa pilihan topik atau peminatan yang dapat dipilih oleh mahasiswa. Pemilihan topik atau peminatan tersebut akan lebih baik jika tidak hanya sesuai dengan minat tapi juga sesuai dengan kemampuan masing-masing mahasiswa. Kemampuan tersebut dapat dilihat dari data akademik mahasiswa, yaitu dari hasil studi selama proses perkulihan dari semester satu sampai semester tujuh. Hampir seluruh matakuliah yang diselenggarakan memiliki korelasi dengan topik atau peminatan yang dapat dipilih. Oleh karena itu, dapat dilakukan suatu analisis pada data akademik mahasiswa yang hasilnya dapat membantu untuk menentukan topik skripsi yang sesuai dengan minat dan kemampuan. Analisis tersebut dapat dilakukan dengan data mining. Dengan topik dan peminatan yang sesuai, mahasiswa dapat memaksimalkan proses pengerjaan skripsi sehingga dapat menyelesaikan skripsi tepat waktu. Penyelesaian skripsi yang tepat waktu dapat memberikan keuntungan untuk program studi terutama untuk kebutuhan akreditasi.

B. Selection and Addition

Data yang digunakan dalam penelitian ini adalah data akademik mahasiswa angkatan 2012 Program Studi Teknik Informatika STT XYZ sejumlah 178 data mahasiswa. Data tersebut berupa transkrip nilai mahasiswa yang didapatkan dari Sistem Informasi Akademik (SIMAK) STT XYZ. Data tersebut terdiri dari atribut nim dan nama mahasiswa, kode, nama, sks, bobot, grade dan semester setiap matakuliah, IPS setiap semester, IPK, dan total SKS yang telah ditempuh. Namun tidak seluruh atribut tersebut digunakan untuk membangun dataset. Dilakukan seleksi data dengan menyeleksi atribut apa saja yang akan digunakan untuk membangun dataset. Atribut yang digunakan adalah Nim, Nama, SKS, Bobot, dan Semester setiap matakuliah.

Selain data akademik mahasiswa, data yang juga digunakan dalam penelitian ini adalah data pengelompokan matakuliah sesuai bidang keahlian yang tersedia di Prodi Informatika STT XYZ. Bidang keahlian tersebut adalah analisis, pemrograman, dan jaringan. Bidang keahlian jaringan terdiri dari matakuliah Dasar Kelistrikan dan Elektronika, Jaringan Komputer dan Komunikasi Data, Keamanan Komputer dan Jaringan, Sistem Digital, Workshop Instalasi Komputer, serta Workshop Jaringan Komputer. Bidang keahlian analisis terdiri dari matakuliah Aljabar Linier dan Matrik, Analisa dan Perancangan Sistem Informasi, Cyber Law, Interaksi Manusia dan Komputer, Kalkulus I, Kalkulus II, Konsep & Perancangan Basis Data, Logika Matematika, Manajemen Project Perangkat Lunak, Matematika Diskrit, Metode Numerik, Model dan Simulasi, Organisasi & Arsitektur Komputer, Organisasi Komputer, Sistem Berkas, Sistem Informasi Manajemen, Sistem Operasi, Statistik, Teori Bahasa & Otomata, Teori Graph, serta Sistem Pendukung Keputusan. Bidang keahlian pemrograman terdiri dari matakuliah Algoritma dan Pemrograman I, Algoritma dan Pemrograman II, Algoritma dan Pemrograman III, Pemrograman Berorientasi Objek, Pemrograman Web, Praktek Pemrograman I, Praktek Pemrograman II, Praktek Pemrograman III, Rekayasa Perangkat Lunak, Rekayasa Sistem Informasi, Struktur Data, Teknik Kompilasi, serta Sistem Pendukung Keputusan.

C. Preprocessing and Data Cleaning

Dataset yang telah dibangun pada tahapan sebelumnya kemudian lakukan pembersihan untuk mendapatkan dataset yang lebih baik. Pada penelitian ini dilakukan penghapusan baris data yang tidak relevan dengan tujuan KDD. Baris data yang mengandung matakuliah yang tidak termasuk dalam bidang keahlian seperti Bahasa Inggris I, Bahasa Inggris II, Konsep Teknologi, Pengantar Teknologi Informasi, Pendidikan Agama I. Pendidikan Agama II, Bahasa Indonesia, Kewarganegaraan, Ilmu Sosial Dasar, Kewirausahaan, Metode Penelitian, Praktek Industri, Pengantar Intelegensi Buatan, Sistem Multimedia dan Jaringan, Skripsi, dihapus dari dataset. Selain itu, juga dilakukan penghapusan baris data dengan Nim dan matakuliah yang sama. Contohnya, terdapat beberapa matakuliah yang sama namun hanya cara penulisan nama matakuliahnya saja yang berbeda, seperti matakuliah Bahasa Inggris 2 dan Bahasa Inggris II. Baris data yang nilai atribut bobotnya bernilai 0 juga dilakukan penghapusan. Setelah proses pembersihan data, jumlah dataset sebanyak 116 data mahasiswa.

D. Transformation

Tahapan selanjutnya adalah transformasi data. Transformasi data yang dilakukan pada penelitian ini adalah membuat atribut baru yang didapatkan dari hasil perkalian SKS dan Bobot yang kemudian dikalikan dengan 100. Hal tersebut dilakukan agar setiap baris data memiliki perbedaan nilai yang cukup signifikan. Selain itu nama matakuliah diinisialkan untuk memudahkan proses pemodelan data mining. Inisial setiap matakuliah dapat dilihat pada Tabel 1. Contoh dataset yang telah siap untuk dilakukan data mining dapat dilihat pada Tabel 2.

Tabel 1. Inisialisasi Nama Matakuliah

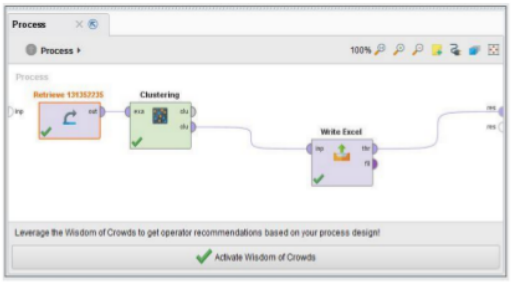
Nama Matakuliah	No. Matakuliah
Algoritma & Pemrograman I	1
Algoritma & Pemrograman II	2
Algoritma & Pemrograman III	3
Aljabar Linier dan Matrik	4
Analisa dan Perancangan Sistem Informasi	5
Cyber Law	6
Dasar Kelistrikan & Elektronika	7
Interaksi Manusia dan Komputer	8
Jaringan Komputer & Komunikasi Data	9
Kalkulus I	10
Kalkulus II	11
Keamanan Komputer & Jaringan	12
Konsep & Perancangan Basis Data	13
Logika Matematika	14
Manajemen Project Perangkat Lunak	15
Matematika Diskrit	16
Metode Numerik	17
Model dan Simulasi	18
Organisasi & Arsitektur Komputer	19
Organisasi Komputer	20
Pemrograman Berorientasi Objek	21
Pemrograman WEB	22
Praktek Pemrograman I	23
Praktek Pemrograman II	24
Praktek Pemrograman III	25
Rekayasa Perangkat Lunak	26
Rekayasa Sistem Informasi	27
Sistem Berkas	28
Sistem Digital	29
Sistem Informasi Manajemen	30
Sistem Operasi	31
Sistem Pendukung Keputusan	32
Statistik	33
Struktur Data	34
Teknik Kompilasi	35
Teori Bahasa & Otomata	36
Teori Graph	37
Workshop Instalasi Komputer	38
Workshop Jaringan Komputer	39

Tabel 2. Contoh Dataset yang Siap Untuk Pemodelan Data Mining

No Mata Kuliah	NIM Mahasiswa			
	121351001	*121351002*	*...*	*121351017*
1	600	400	...	600
2	600	600	...	900
3	1200	600	...	600
4	900	900	...	900
5	900	900	...	1200
6	900	900	...	900
7	600	400	...	400
8	600	400	...	400
9	600	900	...	600
10	900	900	...	900
11	900	900	...	900
12	900	900	...	900
13	900	900	...	900
14	600	600	...	400
15	600	600	...	800
16	600	600	...	300
17	900	900	...	900
18	1200	600	...	900
19	900	900	...	900
20	900	900	...	900
21	600	600	...	900
22	900	900	...	900
23	800	400	...	400
24	900	1200	...	900
25	900	900	...	900
26	1200	900	...	900
27	900	900	...	900
28	600	800	...	600
29	400	400	...	400
30	900	600	...	900
31	600	600	...	900
32	1200	1200	...	1200
33	900	600	...	600
34	600	900	...	900
35	900	900	...	900
36	1200	1200	...	1200
37	1200	1200	...	1200
38	600	400	...	600
39	600	600	...	900

E. Data Mining

Model data mining yang digunakan dalam penelitian ini adalah model pengelompokan atau clustering. Algoritma yang digunakan adalah algoritma K-Means. Data mining diimplementasikan dengan menggunakan Rapidminer. Proses pemodelan dengan Rapidminer dapat dilihat pada Gambar 2. Operator yang digunakan adalah Retrieve, Clustering, dan Write Excel. O*perator Retrieve* yang digunakan untuk menambahkan masukan. O*perator clustering* digunakan untuk mengelompokan data. Data dikelompokan menjadi tiga kelompok sesuai dengan jumlah bidang keahlian mahasiswa. O*perator Write Excel* digunakan untuk menuliskan hasil data mining dalam file *excel* agar lebih mudah dalam pembacaan pola hasil pemodelan. Hasil pemodelan data mining pada penelitian ini dapat dilihat pada Tabel 3.



Gambar 1 Proses Model *K-Means Cluster*ing Dengan *Rapidminer*

Tabel 3. Hasil Pengelompokan Dataset

No Matakuliah	Cluster	Nilai Cluster
1	cluster_2	600,0
2	cluster_2	600,0
3	cluster_1	1200,0
4	cluster_0	900,0
5	cluster_0	900,0
6	cluster_0	900,0
7	cluster_2	600,0
8	cluster_2	600,0
9	cluster_2	600,0
10	cluster_0	900,0
11	cluster_0	900,0
12	cluster_0	900,0
13	cluster_0	900,0
14	cluster_2	600,0
15	cluster_2	600,0
16	cluster_2	600,0
17	cluster_0	900,0
18	cluster_1	1200,0
19	cluster_0	900,0
20	cluster_0	900,0
21	cluster_2	600,0
22	cluster_0	900,0
23	cluster_0	800,0
24	cluster_0	900,0
25	cluster_0	900,0
26	cluster_1	1200,0
27	cluster_0	900,0
28	cluster_2	600,0
29	cluster_2	400,0
30	cluster_0	900,0
31	cluster_2	600,0
32	cluster_1	1200,0
33	cluster_0	900,0
34	cluster_2	600,0

35	cluster_0	900,0
36	cluster_1	1200,0
37	cluster_1	1200,0
38	cluster_2	600,0
39	cluster_2	600,0

F. Evaluation and Interpretation

Hasil pemodelan pada tahapan sebelumnya kemudian dilakukan evaluasi. Kelompok dengan nilai terbesar dipilih untuk dianalsis matakuliah apa saja yang terdapat di dalam kelompok tersebut. Pada Tabel 2, kelompok yang memiliki nilai terbesar adalah kelompok atau cluster 1. Cluster 1 memiliki anggota sebanyak 19 anggota, yaitu matakuliah dengan kode 2,5,6,9,12,13,17,21,24, 26, 27, 30, 31, 32, 34, 35, 36, 37, 39. Pada cluster tersebut terdapat 7 matakuliah yang masuk dalam bidang keahlian pemrograman, 9 matakuliah yang masuk dalam bidang keahlian analisa, 2 matakuliah yang masuk ke dalam bidang jaringan, dan 1 matakuliah yang dapat masuk ke dalam bidang keahlian analisa dan pemrograman. Jumlah matakuliah pada setiap kelompok bidang keahlian menunjukan kemampuan dari mahasiswa. Jumlah matakuliah pada kelompok bidang pemrograman memiliki jumlah paling banyak dibandingkan dengan kelompok yang lain menunjukan bahwa mahasiswa memiliki kemampuan yang lebih baik pada bidang pemrograman sehingga dapat disimpulkan bahwa mahasiswa direkomendasikan untuk mengambil topik skripsi yang berkaitan dengan pemrograman.

G. Discovered Knowledge

Pengetahuan baru yang telah didapatkan pada tahapan sebelumnya dapat digunakan untuk memberikan rekomendasi topik skripsi pada mahasiswa Prodi Teknik Informatika STT XYZ sehingga harapannya mahasiswa dapat menyelesaikan skripsinya dengan baik.

IV. Kesimpulan

Berdasarkan pemaparan mengenai pemodelan data mining, dapat disimpulkan bahwa hasil pengelompokan dengan nilai cluster yang paling tinggi dapat menunjukan kemampuan mahasiswa pada tiap kelompok bidang keahlian. Hasil pengelompokan dianalisis dengan melihat jumlah matakuliah pada setiap kelompok bidang keahlian. Jumlah matakuliah yang paling banyak pada suatu kelompok bidang keahlian menandakan bahwa mahasiswa memiliki kemampuan lebih baik pada bidang tersebut sehingga direkomendasikan topik skripsi yang sesuai dengan kelompok bidang keahlian tersebut. Dengan pengambilan topik skripsi yang sesuai dengan bidang keahlian, harapannya mahasiswa dapat menyelesaikan skripsinya dengan lebih baik dan tepat waktunya.

Ucapan Terima Kasih

Terima kasih kepada Ketua, Bagian Akademik, dan para dosen STT XYZ yang telah banyak membantu menyelesaikan penelitian ini hingga dapat menjadi sebuah karya ilmiah yang diterbitkan.

Daftar Pustaka

[1] V. K. Bakti and J. Indriyatno, "Klasterisasi Dokum Tugas Akhir Menggunakan K-Means Clustering sebagai Analisa Penerapan Sistem Temu Kembali," *J. Ilm. Manaj. Inform. dan Komput.*, vol. 1, no. 1, pp. 31–34, 2017.

[2] M. Sholehhudin, M. Fauzi Ali, and S. Adinugroho, "Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus : Universitas Brawijaya)," vol. 2, no. 11, pp. 5518–5524, 2018.

[3] P. Bhatia, *Data Mining and Data Warehousing : Principles and Practical Techniques*. United Kingdom: Cambridge University Pres, 2019.

[4] M. Feng *et al.*, "Big Data Analytics and Mining for Effective Visualization and Trends Forecasting of Crime Data," *IEEE Access*, vol. 7, pp. 106111–106123, 2019.

[5] E. Turban, R. E. Sharda, and D. Delen, *Decision Support Systems and Intelligent Systems (9th Edition)*. Prentice Hall, 2010.

[6] S. Sharma and ShikhaRai, "Genetic K-Means Algorithm – Implementation and Analysis," *Int. J. Recent Technol. Eng.*, vol. 1, no. 2, pp. 117–120, 2012.

[7] C.-P. Wei and I.-T. Chiu, "Approach, Turning telecommunications call details to churn prediction: a data mining," *Expert Syst. Appl.*, vol. 23, no. 2, pp. 103–112, 2002.

[8] Suyanto, *Data Mining Untuk Klasifikasi dan Klasterisasi*. bandung: Informatika, 2017.

[9] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *Multidisciplany Sci. J.*, vol. 16, no. 2, pp. 226–235, 2019.

[10] O. Somantri, S. Wiyono, and Dairoh, "Metode K-Means untuk Optimasi Klasifikasi Tema Tugas Akhir Mahasiswa Menggunakan Support Vector Machine (SVM)," *Sci. J. Informatics*, vol. 3, no. 1, pp. 34–45, 2016.

[11] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco: Elsevier Inc, 2005.

[12] M. Z. Rodriguez, C. H. Comin, D. C. O. M. Bruno, D. R. Amancio, and L. da F. C. F. A. Rodrigues, "Clustering algorithms: A comparative approach," *PLoS One*, vol. 14, no. 1, pp. 1–34, 2019.

[13] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques Third Edition*. USA: Elsevier Inc, 2012.

[14] V. Gupta and G. Lehal, "A Survey of Text Mining Techniques and Applications," *J. Emerg. Technol. Web Intell.*, vol. 1, no. 1, pp. 60–76, 2009.

[15] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications 2nd Ed*. World Scientific Publishing Co., 2015.

Algoritma K-Means Untuk Pengelompokan Topik Skripsi Mahasiswa

ORIGINALITY REPORT

16%

SIMILARITY INDEX

PRIMARY SOURCES

1	jurnal.fikom.umi.ac.id Internet	73 words —	2%
2	www.scribd.com Internet	44 words —	1%
3	tip.ppj.unp.ac.id Internet	40 words —	1%
4	eprints.sinus.ac.id Internet	30 words —	1%
5	vokasi.uho.ac.id Internet	28 words —	1%
6	Ilias Gerostathopoulos, Stefan Kugele, Christoph Segler, Tomas Bures, Alois Knoll. "Automated Trainability Evaluation for Smart Software Functions", 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2019 Crossref	26 words —	1%
7	Suhardi Rustam, Haditsah Annur. "AKADEMIK DATA MINING (ADM) K-MEANS DAN K-MEANS K-NN UNTUK MENGELOMPOKAN KELAS MATA KULIAH KOSENTRASI MAHASISWA SEMESTER AKHIR", ILKOM Jurnal Ilmiah, 2019 Crossref	24 words —	1%
8	ecommons.luc.edu Internet	21 words —	1%

9	Suherman Suherman, Sunny Samsuni, Imam Lukman Hakim. "Sistem Rekomendasi Wisata Pantai menggunakan Metode Simple Additive Weighting", ILKOM Jurnal Ilmiah, 2020 Crossref	20 words —	1%
10	Fawaz Mohammed Jumaah, Sreedharan Baskara, Roslina Mohd Sidek, Fakhrul Zaman Rokhani. "PrimeTime web-based report analyzer (PTWRA) tool", 2015 6th Asia Symposium on Quality Electronic Design (ASQED), 2015 Crossref	19 words —	< 1%
11	arxiv.org Internet	18 words —	< 1%
12	eprints.fri.uni-lj.si Internet	17 words —	< 1%
13	ejournal.poltektegal.ac.id Internet	17 words —	< 1%
14	pt.scribd.com Internet	16 words —	< 1%
15	id.123dok.com Internet	15 words —	< 1%
16	repository.its.ac.id Internet	15 words —	< 1%
17	Hadi Santoso, Hilyah Magdalena. "Improved K-Means Algorithm on Home Industry Data Clustering in the Province of Bangka Belitung", 2020 International Conference on Smart Technology and Applications (ICoSTA), 2020 Crossref	14 words —	< 1%
18	Abdulrazzak Ali, Nurul A., Siti A., Amelia R.. "An Assessment of Open Data Sets Completeness", International Journal of Advanced Computer Science and Applications, 2019 Crossref	14 words —	< 1%

19	media.neliti.com Internet	13 words — < 1%
20	Praveen Kumar Kotturu, Abhishek Kumar. "Big Data based Adaptive Learning and Scope of Automation in Actionable Knowledge", 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020 Crossref	12 words — < 1%
21	journals.usm.ac.id Internet	12 words — < 1%
22	Yanshan Xiao, Feiqi Deng, Bo Liu, Shouqiang Liu, Dan Luo, Guohua Liang. "A Learning Process Using SVMs for Multi-agents Decision Classification", 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008 Crossref	12 words — < 1%
23	Achmad Solichin. "Comparison of Decision Tree, Naïve Bayes and K-Nearest Neighbors for Predicting Thesis Graduation", 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2019 Crossref	11 words — < 1%
24	docplayer.info Internet	11 words — < 1%
25	Antonius Rachmat Chrismanto, Willy Sudiarto Raharjo, Yuan Lukito. "Firefox Extension untuk Klasifikasi Komentar Spam pada Instagram Berbasis REST Services", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2019 Crossref	10 words — < 1%
26	docshare.tips Internet	10 words — < 1%
27	repository.usd.ac.id Internet	10 words — < 1%

28	Maryamah Maryamah, Made Agus Putra Subali, Lailly Qolby, Agus Zainal Arifin, Ali Fauzi. Jurnal Linguistik Komputasional (JLK), 2018 Crossref	9 words — < 1%
29	ml.scribd.com Internet	9 words — < 1%
30	Julius Rinaldi Simanungkalit, Haviluddin Haviluddin, Herman Santoso Pakpahan, Novianti Puspitasari, Masna Wati. "Algoritma Backpropagation Neural Network dalam Memprediksi Harga Komoditi Tanaman Karet", ILKOM Jurnal Ilmiah, 2020 Crossref	8 words — < 1%
31	id.scribd.com Internet	8 words — < 1%
32	elektrounud14.blogspot.com Internet	8 words — < 1%
33	www.kaskus.co.id Internet	8 words — < 1%
34	zulfiyahmusfiroh.blogspot.com Internet	8 words — < 1%
35	Hendro Priyatman, Fahmi Sajid, Dannis Haldivany. "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2019 Crossref	8 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	ON		