JOURNAL INNOVATIONS COMPUTER SCIENCE
E-ISSN: 2961-970X, P-ISSN: 2961-9718
Volume 1 Number 2, November 2022, pp 94-108.
DOI: https://doi.org/10.56347/jics.v1i2.72

Optimasi *Support Vector Mechine* (SVM) Menggunakan *K-Means* dan *K-Medoids* untuk Klasterisasi Tema Tugas Akhir

Gaffy Patria [a*]
[a*] Sekolah Tinggi Manajemen Informatika Dan Komputer (STMIK) Abulyatama Aceh, Kota Banda Aceh, Provinsi Aceh, Indonesia.

ABSTRACT
The large amount of final project document data from study programs at the Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Abulyatama can make a major contribution to the difficulty of the process of grouping a student's final project theme. The clustering process that has been carried out manually so far has been very ineffective and inefficient, so a data mining application is needed to manage the data, especially for clustering the data. The goal to be achieved from writing this thesis is to implement the Support Vector Machine with K-Means and K-Medoids to optimize the final assignment clustering. the results of the Optimization Support Vector Machine (SVM) analysis using K-Means and K-Medoids for Grouping Student Final Project Themes can be concluded in a number of ways, namely; with the K-Means Clustering method it can be seen that there are 23 data mining, 10 networks, 26 artificial intelligence, and 21 websites, and website 11 items.

ARTICLE HISTORY
Received 12 November 2022
Accepted 25 November 2022
Published 30 November 2022

KEYWORDS
Support Vector Machines; K-Means; K-Medoids, Final Project; Rapidminer.

ABSTRAK
Banyaknya jumlah data dokumen tugas akhir dari program studi di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Abulyatama dapat memberi kontribusi besar dalam sulitnya proses mengelompokkan suatu tema tugas akhir mahasiswa. Proses klasterisasi yang dilaksanakan secara manual selama ini sangat tidak efektif dan efesien, sehingga diperlukan suatu penerapan data mining untuk mengelola data tersebut khususnya untuk klasterisasi data tersebut. Adapun tujuan yang ingin dicapai dari penulisan skripsi ini adalah untuk penerapan Support Vector Mechine dengan K-Means dan K-Medoids untuk mengoptimalkan klasterisasi tugas akhir. hasil analisis Optimasi Support Vector Machine (SVM) Menggunakan K-Means dan K-Medoids Untuk Pengelompokan Tema Tugas Akhir Mahasiswa dapat disimpulkan beberapa hal yaitu; dengan metode K-Means Clustering dapat diketahui bahwa data mining sebanyak 23, jaringan sebanyak 10, Kecerdasan buatan 26, dan website sebanyak 21. Sedangkan dengan metode K-Medoids Clustering dapat diketahui bahwa data mining 18 items, jaringan 32 items, kecerdasan buatan 19 items, dan website 11 items.

KATA KUNCI
Support Vector Mechine; K-Means; K-Medoids, Tugas Akhir; Rapidminer.

CONTACT Gaffy Patria gaffypatria@gmail.com Sekolah Tinggi Manajemen Informatika Dan Komputer (STMIK) Abulyatama Aceh.

1. Pendahuluan

Banyaknya jumlah data dokumen tugas akhir dari program studi di Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Abulyatama dapat memberi kontribusi besar dalam sulitnya proses mengelompokkan suatu tema tugas akhir mahasiswa. Banyaknnya mahasiswa yang mengambil mata kuliah tugas akhir ini membuat petugas akademik susah mengelola dan mengelompokkan data tersebut. Hal tersebut disebabkan karena banyaknya data mahasiswa yang mengambil mata kuliah tugas akhir dan pemilihan tema tugas akhir yang sangat beragam antara satu dengan lainnya sesuai dengan penelitian yang dilaksanakan. Proses klasterisasi yang dilaksanakan secara manual selama ini sangat tidak efektif dan efesien, sehingga diperlukan suatu penerapan data mining untuk mengelola data tersebut khususnya untuk klasterisasi data tersebut.

Data mining itu sendiri merupakan adalah ekstraksi pola yang menarik dari data dalam jumlah besar yang belum diketahui sebelumnya dan berguna [1][2]. Proses penemuan pola terbukti dapat membantu proses pengelompokan berbagai topik skripsi yang ada sehingga diperoleh informasi yang bermakna dalam menentukan tren penelitian Universitas dari tahun ke tahun [3][4]. Dengan adanya pengelompokan dokumen, maka tidak harus membuka halaman terlalu banyak, karena dokumen hasil pencarian telah dikelompokkan berdasarkan kategori yang dapat menggambarkan isi dari suatu dokumen, hal tersebut tentunya dapat mempermudah dalam menemukan beberapa dokumen yang diinginkan [5], oleh karenanya sebelum proses tersebut dilakukan maka, proses tersebut perlu dianalisis sebelum benar-benar diaplikasikan ke dalam program yang sifatnya aplikatif [6]. Oleh karena itu, diperlukan metode pengelompokan (*clustering*) yang nantinya dapat dipastikan keberhasilan pengelompokan suatu dokumen tugas akhir dengan baik. Metode *clustering* dapat digunakan dalam pengkategorian atau pengelompokan dokumen [7][8]. Caranya adalah dengan mengelompokan dokumen-dokumen ke dalam *cluster*s berdasarkan kedekatan atau kemiripan (*similarity*) antar dokumen. Sehingga dokumen yang berhubungan dengan suatu tema tertentu secara otomatis ditempatkan pada *cluster* yang sama [9]. Saat ini ada beberapa algoritma *clustering* diantaranya *K-Means Clustering* dan *K-Medoids*. Adapun tujuan yang ingin dicapai dari penelitian ini adalah untuk penerapan *Support Vector Mechine* dengan *K-Means* dan *K-Medoids* untuk mengoptimalkan klasterisasi tugas akhir.

Pada penelitian sebelumnya telah dilakukan oleh Bakti dan Indriyatno (2017) dengan judul penerapan text mining untuk klasifikasi tema tugas akhir. Berdasarkan hasil penelitian, Algoritma *Clustering* dapat digunakan dalam pengkategorian atau pengelompokan dokumen. Salah satu penggunaan algoritma *clustering* yaitu dengan menerapkan metode *K-Means*. Pengklasteran dokumen abstrak tugas akhir berbahasa Indonesia dengan menerapkan Algoritma *K-Means* menunjukan klaster yang dihasilkan cukup baik, sehingga dapat dijadikan rekomendasi bahwa metode *K-Means* klastering cukup baik jika diterapkan dalam penerapan sistem temu kembali, dengan indikator jarak antar klaster yang dihasilkan sangat dekat yaitu sebesar 0,001 ketika dihitung dengan metode Davies Bouldin Index [10]. Pada penelitian yang dilakukan oleh Ramdani, Chrisnanto, & Maspupah (2018) juga melakukan penelitian untuk mengklasifikasi jenis tema penelitian berdasarkan abstrak menggunakan *Vector Space Model* (VSM) dan *K-Nearest Neighbor* (K-NN). Jumlah abstrak yang digunakan dalam penelitian ini sebanyak 75 data abstrak sebagai data training dan 25 data sebagai data uji. Jumlah kelas pada proses KNN ditentukan sebanyak 15 kelas yang merepresentasikan jumlah tema penelitian, dengan nilai K ideal ditentukan sebanyak 1, 3, 5, 7 dan 9. Berdasarkan hasil penelitian diperoleh bahwa pengklasifikasian tema penelitian dapat dilakukan dengan akurasi yang baik, dimana algoritma KNN mampu mengelompokan data abstrak ke dalam 15 kelas [11]. Selanjutnya penelitian yang dilakukan oleh Sankoh *dkk* (2015), pada penelitian ini bertujuan untuk

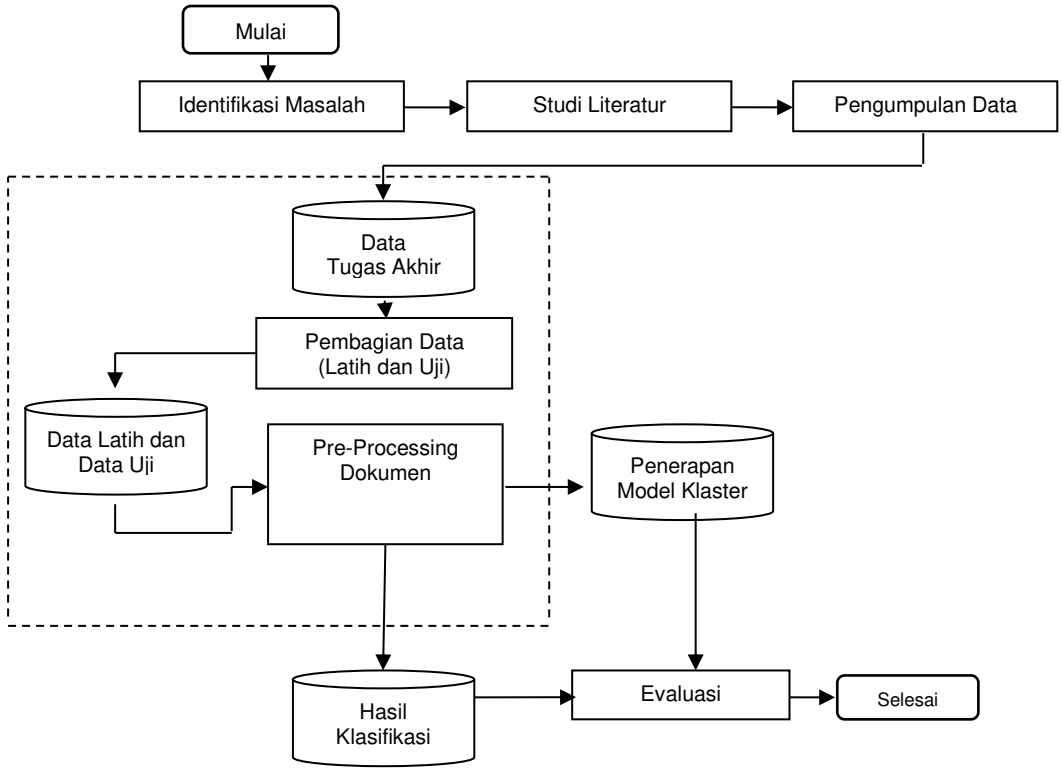
mengembangkan metode pengelompokan *file* berupa kombinasi pendekatan *pre-processing*, *neural network*, *K-Means*, dan *particle swarm optimization* dengan masukan berupa file audio sehingga diperoleh keluaran berupa kumpulan file audio yang telah terkelompok berdasarkan jenis musik. Hasil dari penelitian ini yaitu berupa suatu sistem dengan pengembangan metode dalam pengelompokan file audio berdasarkan jenis musik. Metode pada tahap pre-processing memiliki hasil terbaik pada pendekatan berdasarkan analisa spectrum melodi gitar dan bass, di mana memiliki nilai precission 95%, recall 100% dan F-Measure 97,44% [12].

Selain itu, penelitian oleh Raharjo dan Winarko (2014) dengan judul klasterisasi, klasifikasi dan peringkasan teks berbahasa Indonesia. Penelitian ini akan melakukan survei paper penelitian data mining teks berbahasa Indonesia. Dari paper yang didapatkan terlihat bahwa sebagian besar topik penelitian data mining bertujuan adalah untuk melakukan pengelompokan suatu berita baik online maupun cetak berdasar atas acuan tertentu, penelitian lain ditujukan untuk mengolah teks di media sosial seperti twitter. Artikel ini akan memperlihatkan metode yang digunakan dan tujuan dari paper dalam bidang klasterisasi, klasifikasi dan peringakasan dokumen berbahasa Indonesia. Jumlah dokumen latih dan uji bervarisi di masing-masing paper, namun secara umum cukup banyak yang menggunakan dokumen latih dan uji dengan perbandingan 90% dan 10% dari total dokumen. Survei juga menunjukkan bahwa cukup banyak paper yang kurang memperhatikan masalah penulisan dan pengacuan daftar pustaka. Dari hasil survei tersebut maka diperlukan penelitian dalam bidang klasifikasi, klasterisasi dan peringkasan Bahasa Indonesia dengan metode yang lebih beragam [13].

Algoritma *K-Medoids Clustering* adalah salah satu algoritma yang digunakan untuk klasifikasi atau pengelompokan data. Untuk melakukan *clustering* dengan metode partisi dapat menggunakan *K-Means* dan *K-Medoids*. *K-Means* merupakan suatu algoritma peng*cluster*an yang cukup sederhana yang mempartisi *dataset* kedalam beberapa *cluster* k [14]. Algortima *K-Medoids*, juga dikenal sebagai *partitioning around medoids*, adalah varian dari metode *K-Means* [15][16]. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan mean yang dimiliki oleh setiap *cluster*, dengan tujuan mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang ada dalam *dataset*. Algoritma *K-Medoids* hadir untuk mengatasi kelemahan Algoritma *K-Means* yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [17]. *K-Means* dan *K-Medoids* adalah contoh dari metode *partitioning*. Algoritma *K-Means* sensitif terhadap *outlier* karena objek dengan nilai yang sangat besar dapat secara substansial mendistorsi distribusi data [18]. Untuk mengambil nilai rata-rata dari objek dalam sebuah *cluster* sebagai titik acuan, medoid dapat digunakan, yang merupakan objek dalam sebuah *cluster* yang paling terpusat. Strategi dasar dari algoritma *clustering K-Medoids* adalah untuk menemukan k *cluster* dalam n objek dengan pertama kali secara arbitrarily menemukan wakil dari objek (medoid) untuk tiap-tiap *cluster*. Seperti metode partisi *clustering* yang lainnya, metode *Partitioning* Around *Medoids* (PAM) atau disebut *K-Medoids* juga digunakan untuk pengelompokkan dokumen, dalam metode PAM ini setiap *cluster* dipresentasikan dari sebuah objek di dalam *cluster* yang disebut dengan medoid, Tujuannya adalah menemukan kelompok k-*cluster* (jumlah *cluster*) diantara semua objek data di dalam sebuah kelompok data. *Cluster*nya dibangun dari hasil mencocokkan setiap objek data yang paling dekat dengan *cluster* yang dianggap sebagai medoid sementara.

2. Metodelogi Penelitian

Suatu penelitian dikatakan berhasil apabila telah tercapainya tujuan penelitian tersebut. Namun sebelum mencapai tujuannya harus melalui banyak proses atau tahapan-tahapan. Adapun tahapan-tahapan tersebut penulis tuangkan dalam diagram penelitian berikut ini.



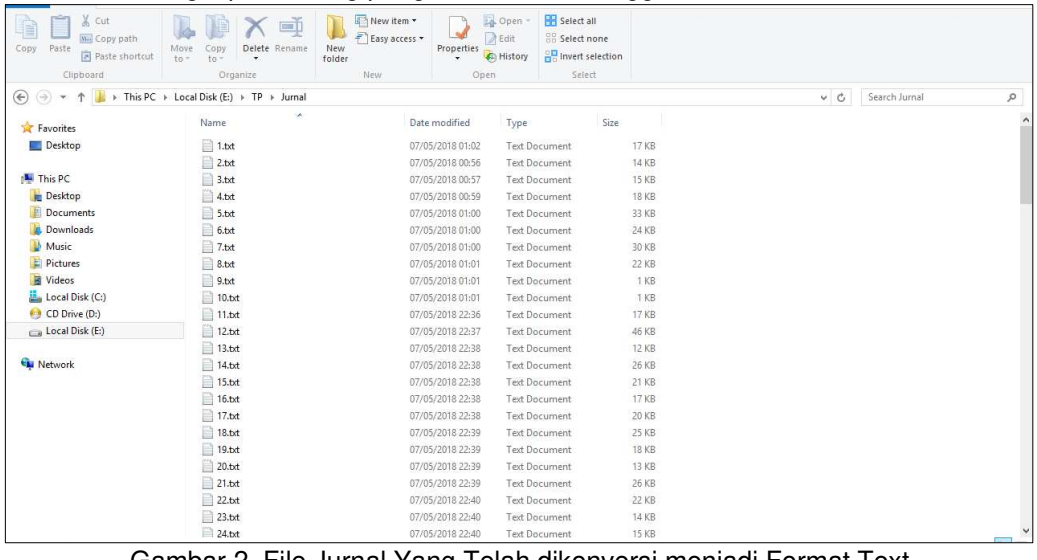
Gambar 1. Tahapan Penelitian

Berdasarkan diagram alir penelitian diatas dapat dijelaskan bahwa untuk mencapai tujuan penelitian yang diajukan pada skripsi ini terdapat beberapa proses atau tahapan yang harus dilalui. Pertama mulai selanjutnya indentifikasi masalah, lalu studi literatur dan pengumpulan data, data tugas akhir mahasiswa sistem informasi disimpan dalam *database*, setelah data terkumpul data akan di bagi menjadi 2 (dua) kelompok yaitu data latih dan data uji. Hasil pengelompokan data latih dan uji pun disimpan dalam *database.* Adapun data penelitian ini dikumpulkan dari *internet* dalam bentuk jurnal penelitian jurusan komputer dari berbagai kampus dan penerbit jurnal di Indonesia yang telah dipublikasi, diantaranya adalah Universitas Indonesia, Politeknik Harapan Bersama Tegal, Universitas Ahmad Dahlan Yogjakarta, ITB Bandung, IPB Bogor dan lain-lain. Semua jurnal ini akan dikelompokkan dalam 4 (empat) pengelompokan tema, yaitu tema data mining, jaringan, kecerdasan buatan, dan website. Suatu penelitian akan memerlukan perangkat untuk membantu tercapainya suatu tujuan penelitian. Adapun alat dan bahan yang diperlukan dalam penelitian ini berupa *hardware* (perangkat keras) dan *software* (perangkat lunak). Berikut ini perangkat keras yang digunakan yaitu:

Tabel 1. Spesifikasi *Hardware* dan *Software*

Spesifikasi		Type/ Versi
Software	Microsoft Office Excell	Office 2016
	Rapidminer	8.1 64 Bit
	OS Windows 8.1	8.1
Hardware	Processor	Intel Core I5 1.7 Ghz
	RAM	4 Gb
	VGA	Nvidia Geforce 820M
	Hardisk	500 Gb Seagate Barracuda

Dataset diperoleh dari data berbagai sumber jurnal di Indonesia pada tahun terbit antara 2015 sampai dengan 2017, sebanyak 80 jurnal dengan berbagai macam jenis tema terdiri dari data mining, jaringan, kecerdasan buatan, dan website. Alat atau tools yang digunakan dalam penelitian ini adalah dengan menggunakan Software Rapid Miner 8.1, sebagai pendukung pengolahan data menggunakan Microsoft Excel 2016.



Gambar 2. File Jurnal Yang Telah dikonversi menjadi Format Text

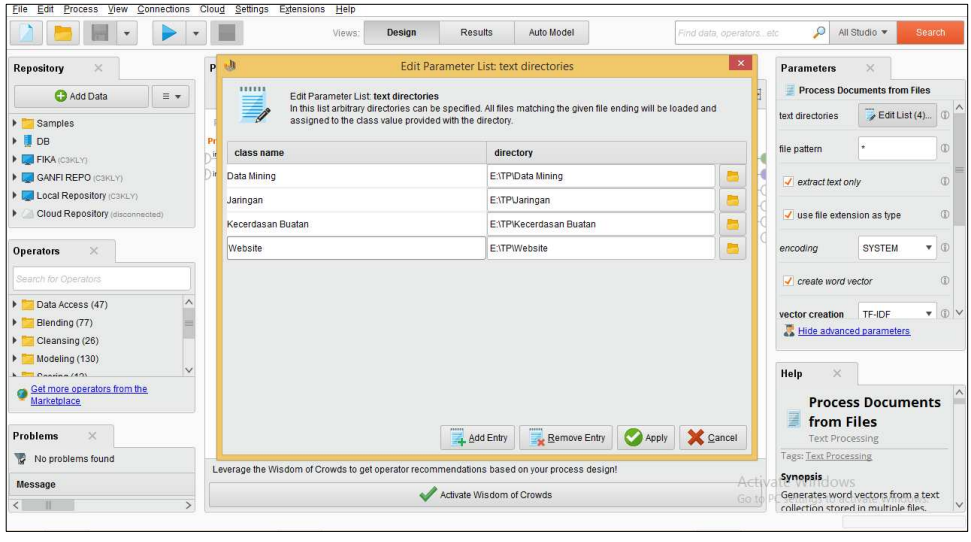
Data jurnal berjumlah 80 jika dirincikan seperti pada tabel berikut 2.

Tabel 2. Pengelompokkan Manual Jurnal

No	Tema Jurnal	Jumlah Jurnal
1	Data Mining	20
2	Jaringan	20
3	Kecerdasan Buatan	20
4	Website	20
Total		80

Pada *processing text* maka akan dilakukan tahap seperti *tokenizing*, *stopwords*, dan *transform case*. Adapun langkah-langkah sebagai pengujian *processing text* akan dilakukan tahap sebagai berikut:

1) *Process Documents from File*
   Pada proses ini, bertujuan untuk memproses dokumen agar mendapatkan daftar kata-kata yang berbeda dan jumlah individu dari setiap dokumen. Pada bagian ini juga nantinya akan memilih file jurnal yang telah dikonversi sebelumnya dan memilih folder target jurnal.



Gambar 3. Parameter *List Document*

Ketika telah memilih data jurnal maka pada tahap ini *processing text* ini dilakukan beberapa subproses agar dokumen dapat dipakai untuk melakukan proses pengelompokan.

2) Sub *Process Documents from File*

Pada *sub process documents from file* diantaranya

a) *Tokenize*

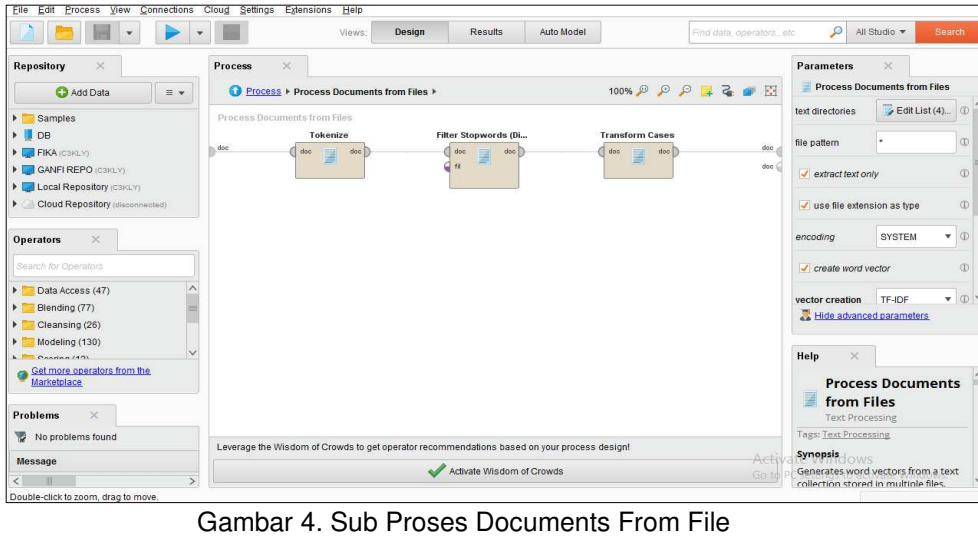
Proses yang bertujuan untuk memisah teks menjadi beberapa token berdasarkan pembatas berupa spasi atau tanda baca.

b) *Transform Case*

Proses yang bertujuan untuk mengubah semua karakter dalam dokumen menjadi huruf kecil atau huruf besar menjadi satu bagian masing-masing.

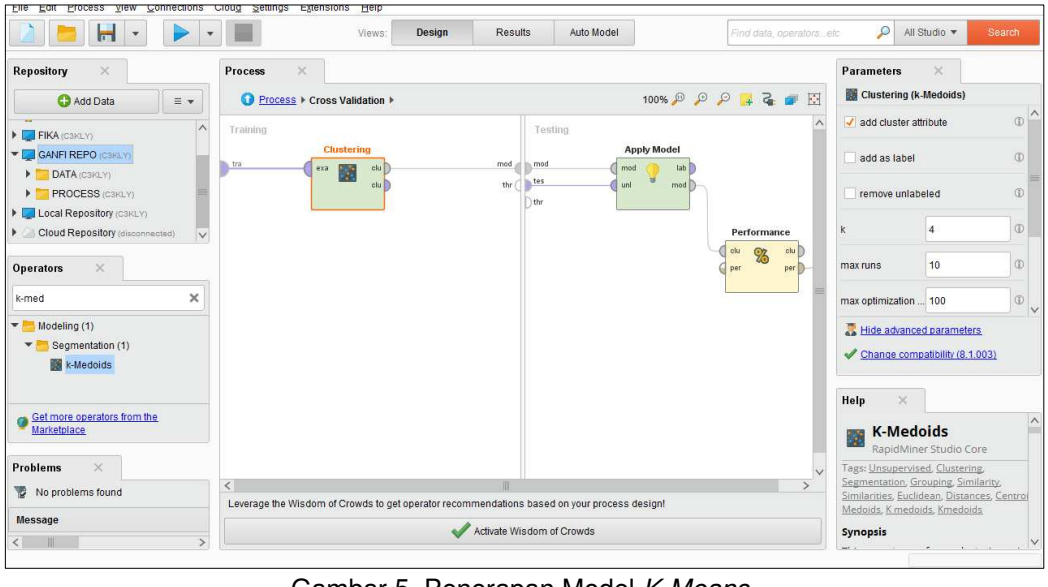
c) *Filter Stopwords*

Memfilter *stopwords* dari dokumen dengan menghapus setiap token yang sama dengan kata dari daftar kata bawaan. *Stop words* adalah kata umum (*common words*) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna.

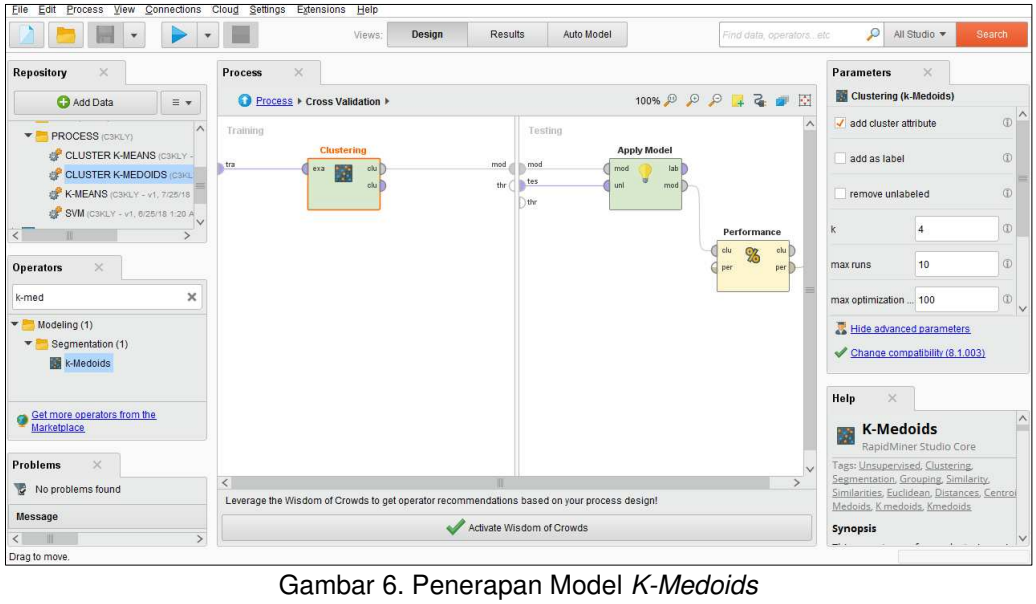


Gambar 4. Sub Proses Documents From File

Proses penerapan model dilakukan terhadap hasil sebelum proses yang merupakan representasi data dalam bentuk model ruang *vector*. Metode pertama ialah penerapan model dokumen yang ada dengan *K-Means* dan *K-Medoids*.

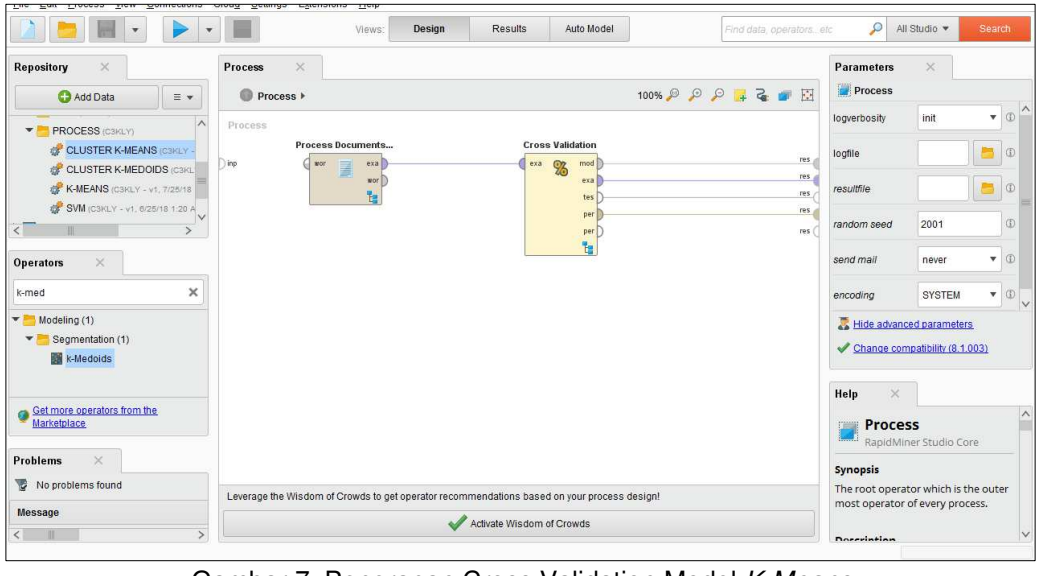


Gambar 5. Penerapan Model *K-Means*

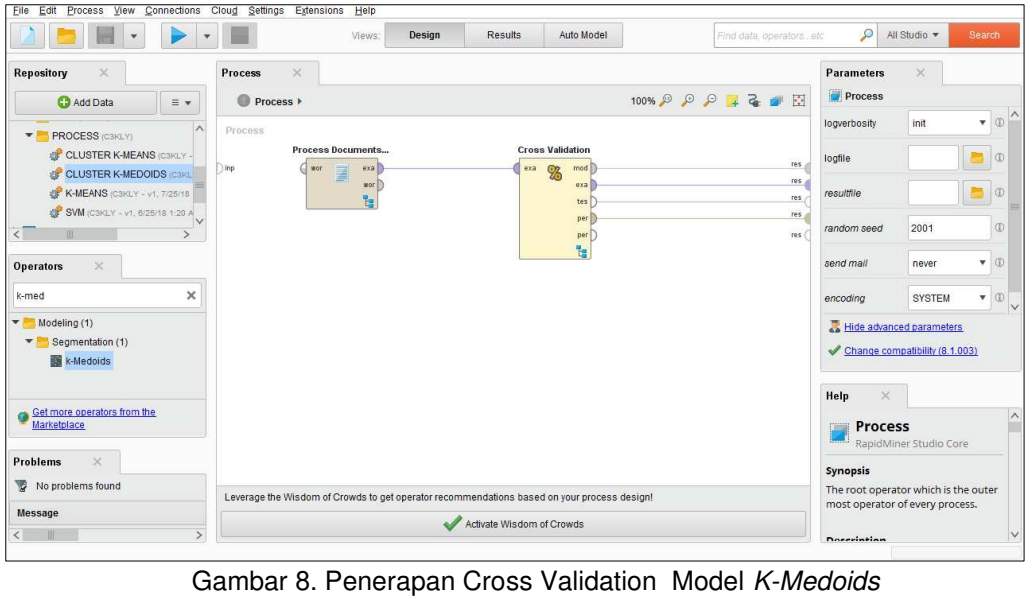


Gambar 6. Penerapan Model *K-Medoids*

Sebagai evaluasi dari model yang diusulkan, yaitu dengan menggunakan *Kfolds Cross Validations* untuk mencari nilai akurasi yang kemudian hasil dari akurasi tersebut dievaluasi dengan cara membandingkan tingkat akurasi yang dihasilkan oleh model *K-Means* dan *K-Medoids*.



Gambar 7. Penerapan Cross Validation Model *K-Means*



Gambar 8. Penerapan Cross Validation Model *K-Medoids*

3. Hasil dan Pembahasan

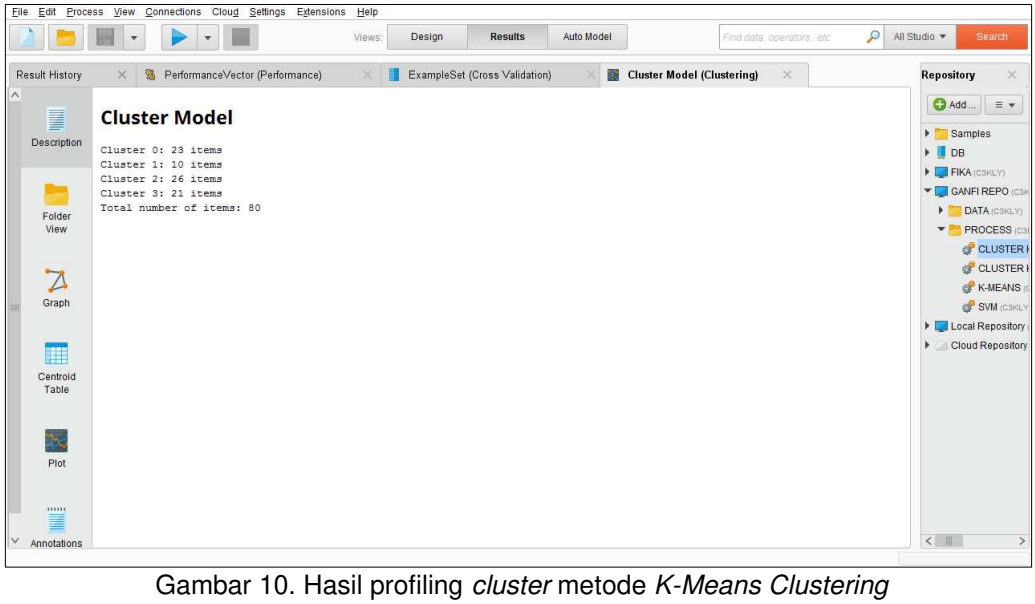
3.1 Hasil

Masih sulitnya dalam menentukan pengelompokan tema tugas akhir mahasiswa sering dialami oleh setiap perguruan tinggi. Tujuan penelitian ini adalah memberikan sebuah penunjang keputusan bagi pengambil kebijakan di program study khususnya pada STMIK Abulyatama agar setiap judul tugas akhir mahasiswa yang diajukan sesuai dengan tema yang telah ditentukan. Penggunaan algoritma *K-Means* dan *K-Medoids* yang digunakan akan membuktikan tingkat klaterisasi judul tugas akhir yang lebih baik. Untuk lebih rinci tahapan perancangan tersebut dapat dilihat pada penjelasan pada bagian ini. Dalam kasus ini, pengklasteran terhadap tema tugas akhir dibagi menjadi empat tema, yaitu data mining, jaringan, kecerdasan buatan, dan website. Untuk pengklasteran pertama, digunakan metode *K-Means Clustering* dengan hasil

pengklasteran menggunakan software rapidminer seperti pada gambar berikut:

Gambar 9. Hasil pengklasteran metode *K-Means Clustering*

Gambar 9 menunjukkan keanggotaan klaster dan profiling *cluster* yang digunakan untuk melihat nilai rata-rata dari anggota masing-masing variabel pada setiap klaster sehingga diperoleh karakteristik setiap klaster. Adapun hasil profiling *cluster* dengan metode *K-Means Clustering* ditunjukkan pada gambar 10 berikut.



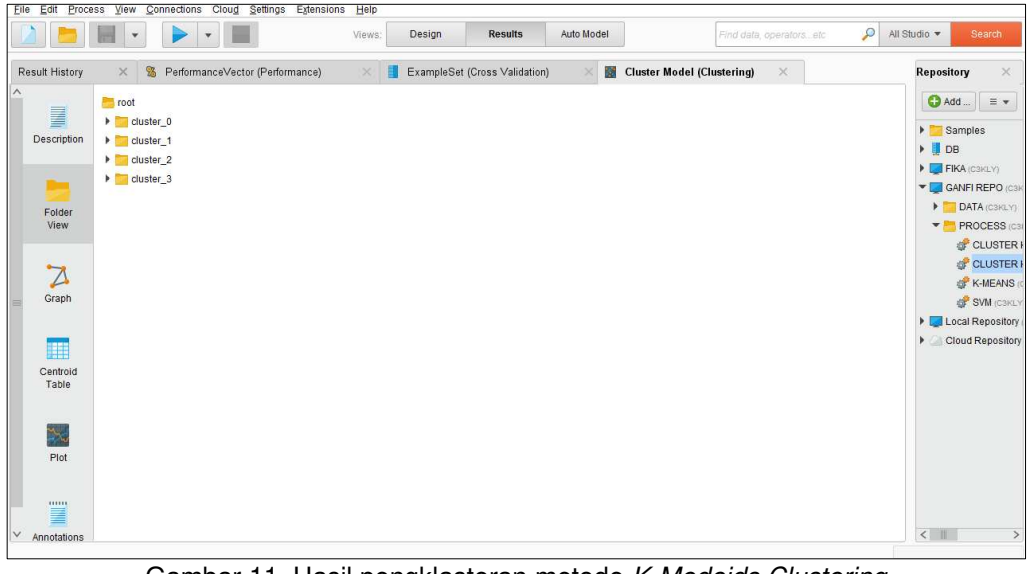
Gambar 10. Hasil profiling *cluster* metode *K-Means Clustering*

Dari gambar 10 dapat dilihat bahwa jurnal atau tugas akhir dalam *Cluster* 0: 23 items, *Cluster* 1: 10 items, *Cluster* 2: 26 items, *Cluster* 3: 21 items, dari total 80. Adapun rincian dari klister tersebut adalah:

Tabel 3. Hasil *profiling cluster* dengan metode *K-Means Clustering*

	*Cluster*_0	*Cluster*_1	*Cluster*_2	*Cluster*_3
Jurnal	2	21	1	5
	3	26	8	6
	4	27	9	22
	7	29	10	23
	14	31	11	24
	18	34	12	25
	44	38	13	28
	45	41	15	32
	46	76	16	33
	48	77	17	35
	49		19	36
	50		20	37
	53		30	39
	54		42	40
	56		43	51
	64		47	52
	65		57	55
	67		58	62
	69		59	68
	71		60	74
	72		61	75
	73		63	
	80		66	
			70	
			78	
			79	

Sedangkan hasil pengklasteran metode *K-Medoids Clustering* menggunakan software rapidminier ditunjukkan pada gambar 11 berikut.



Gambar 11. Hasil pengklasteran metode *K-Medoids Clustering*

Gambar 11 menunjukkan keanggotaan klaster dan profiling *cluster* yang digunakan untuk melihat nilai rata-rata dari anggota masing-masing variabel pada setiap klaster sehingga diperoleh karakteristik setiap klaster. Adapun hasil profiling *cluster* dengan metode *K-Medoids Clustering* ditunjukkan pada gambar 12 berikut.

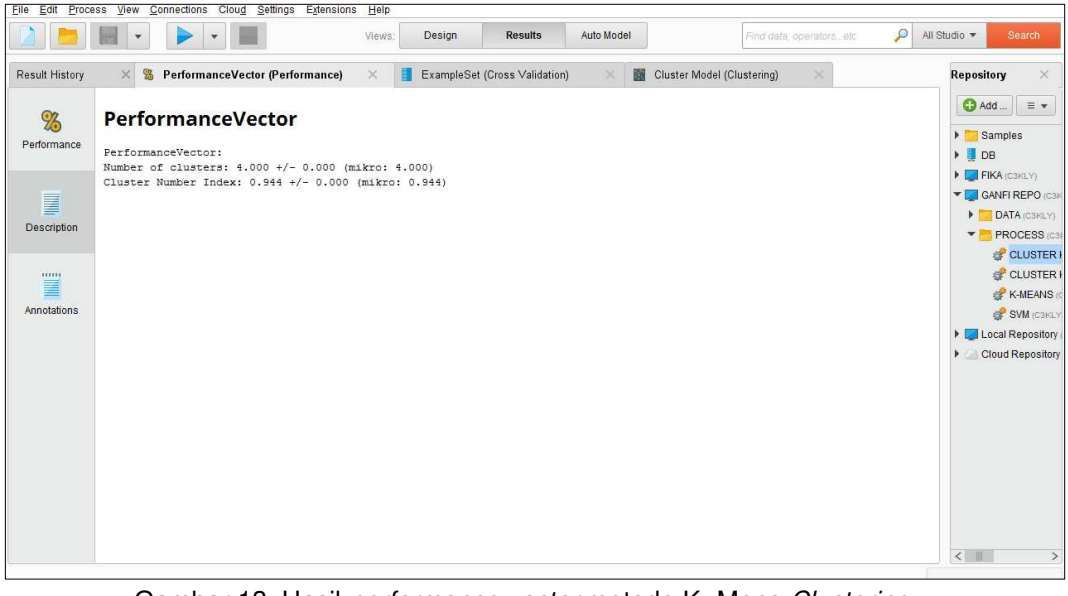
Gambar 12. Hasil profiling *cluster* metode K- *Medoids Clustering*

Dari gambar 12 dapat dilihat bahwa jurnal atau tugas akhir dalam *Cluster* 0: 18 items, *Cluster* 1: 32 items, *Cluster* 2: 19 items, *Cluster* 3: 11 items, dari total 80. Adapun rincian dari klister tersebut adalah:

Tabel 4. Hasil *Profiling Cluster* dengan Metode K- *Medoids Clustering*

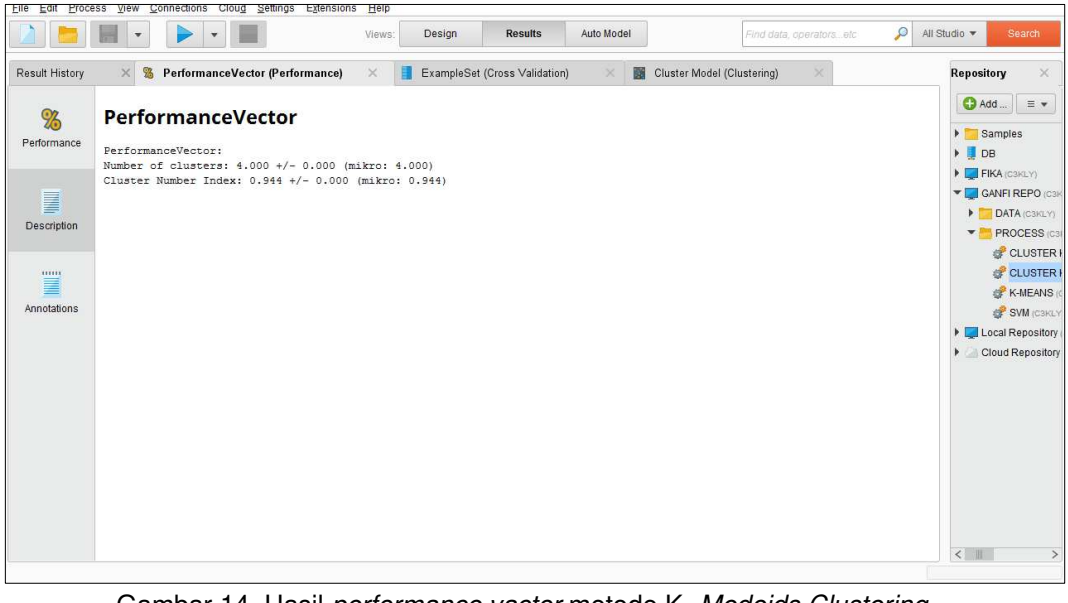
	*Cluster*_0	*Cluster*_1	*Cluster*_2	*Cluster*_3
Jurnal	8	1	11	7
	22	2	18	9
	28	3	32	21
	30	4	36	23
	35	5	37	24
	40	6	38	27
	41	10	42	31
	46	12	48	39
	49	13	52	76
	51	14	53	77
	54	15	58	80
	56	16	59	
	62	17	60	
	65	19	61	
	68	20	66	
	71	25	67	
	72	26	70	
	75	29	73	
		33	78	
		34		
		43		
		44		
		45		
		47		
		50		
		55		
		57		
		63		
		64		
		69		
		74		
		79		

Dari hasil *cross validation* maka didapatkan hasil *performance vector* dari penggunaan model *K-Means* dan *K-Medoids* sebagai berikut:



Gambar 13. Hasil *performance vactor* metode K- Mens *Clustering*

Dari PerformanceVector metode K- Mens *Clustering* maka didapatkan hasil Number of *cluster*s: 4.000 +/- 0.000 (mikro: 4.000) dan *Cluster* Number Index: 0.944 +/- 0.000 (mikro: 0.944).



Gambar 14. Hasil *performance vactor* metode K- *Medoids Clustering*

Dari hasil PerformanceVector metode K- *Medoids Clustering* didapatkan Number of *cluster*s: 4.000 +/- 0.000 (mikro: 4.000) dan *Cluster* Number Index: 0.944 +/- 0.000 (mikro: 0.944).

3.2 Pembahasan

Dari hasil analisa dapat diketahui bahwa untuk hasil pengklasteran menggunakan metode *K-Means Clustering* lebih kecil dibandingkan dengan nilai hasil pengklasteran menggunakan metode *K-Medoids Clustering*, sehingga diperoleh metode terbaik untuk pengklasteran terhadap klaster adalah metode *K-Means Clustering*, dengan dengan hasil pengklasteran bahwa yang termasuk ke dalam tema data mining sebanyak 23, jaringan sebanyak 10, Kecerdasan buatan 26, dan website sebanyak 21. Sedangkan metode *K-Medoids Clustering* data mining 18 items, jaringan 32 items, kecerdasan buatan 19 items, dan website 11 items.

4. Kesimpulan

Berdasarkan hasil dan analisis yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut; 1) Dengan metode *K-Means Clustering* dapat diketahui bahwa data mining sebanyak 23, jaringan sebanyak 10, Kecerdasan buatan 26, dan website sebanyak 21, 2) Dengan metode *K-Medoids Clustering* dapat diketahui bahwa data mining 18 items, jaringan 32 items, kecerdasan buatan 19 items, dan website 11 items. Sedangkan saran untuk penelitian selanjutnya adalah Adanya pengaturan parameter berbeda dalam penerapan *K-Means* dan Metode *K-Medoids* sehingga didapatkan variasi nilai dan nantinya akan didapatkan hasil terbaik dengan tingkat pengelompokan yang lebih baik, dan Model yang diterapkan untuk penelitian selanjutnya dapat diimplementasikan kedalam sebuah aplikasi dengan menggunakan bahasa pemrograman tertentu.

Referensi

[1] Han, Jiawei *dan* Kamber, Micheline. (*2016*). *Data Mining : Concept and. Techniques Second Edition.* Morgan Kaufmann Publishers.

[2] Larose, Daniel T. *(*2015). *Discovering Knowledge in Data : An Introduction to Data Mining.* John Willey & Sons, Inc.

[3] Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, *34*(3), 1707-1720. DOI: https://doi.org/10.1016/j.eswa.2007.01.035.

[4] Kontostathis, A., Galitsky, L. M., Pottenger, W. M., Roy, S., & Phelps, D. J. (2004). A survey of emerging trend detection in textual data mining. *Survey of text mining*, 185-224. DOI: https://doi.org/10.1007/978-1-4757-4305-0_9.

[5] Chakrabarti, S. (2000). Data mining for hypertext: A tutorial survey. *ACM SIGKDD Explorations Newsletter*, *1*(2), 1-11. DOI: https://doi.org/10.1016/S0167-9236(99)00038-X.

[6] Ng, R. T., & Han, J. (2002). CLARANS: A method for clustering objects for spatial data mining. *IEEE transactions on knowledge and data engineering*, *14*(5), 1003-1016. DOI: https://doi.org/10.1109/TKDE.2002.1033770.

[7] Zendrato, F. S. G., Triayudi, A., & Endah Tri, E. (2022). Analisis Clustering Dokumen Tugas Akhir Mahasiswa Sistem Informasi Universitas Nasional menggunakan Metode K-Means Clustering. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(1), 70-76. DOI: https://doi.org/10.35870/jtik.v6i1.389.

[8] Salam, A., & Albahri, F. P. (2022). Sistem Rekomendasi Tugas Akhir Mahasiswa pada AMIK Indonesia untuk Mendukung Merdeka Belajar-Kampus Merdeka Menggunakan Metode Collaborative Filtering (CF). *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(2), 281-288. DOI: https://doi.org/10.35870/jtik.v6i2.420.

[9] Wali, M., & Safrizal, S. (2018). Similar text sebagai Pengkodean Aplikasi Plagiarisme. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *2*(1), 11-19. DOI: https://doi.org/10.35870/jtik.v2i1.43.

[10] Bakti, V. K., & Indriyatno, J. (2017). Klasterisasi Dokumen Tugas Akhir Menggunakan K-Means Clustering, Sebagai Analisa Penerapan Sistem Temu Kembali. *KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer*, *1*(1), 31-34. DOI: https://doi.org/10.32485/kopertip.v1i1.8.

[11] Ramdani, T., Chrisnanto, Y. H., & Maspupah, A. (2018, July). Pengklasifikasian Tema Penelitian Berdasarkan Abstrak Menggunakan Vector Space Model (VSM) dan K-Nearest Neighbor (K-NN). In *Seminar Nasional Teknologi Informasi* (Vol. 1, pp. 703-713).

[12] Sankoh, A. S., Musthafa, A. R., Rosadi, M. I., & Arifin, A. Z. (2015). Klasterisasi Jenis Musik Menggunakan Kombinasi Algoritma Neural Network, K-Means dan Particle Swarm Optimization. *Jurnal Buana Informatika*, *6*(3). DOI: https://doi.org/10.24002/jbi.v6i3.431.

[13] Raharjo, S., & Winarko, E. (2014). Klasterisasi, klasifikasi dan peringkasan teks berbahasa indonesia. *Prosiding KOMMIT*.

[14] Nainggolan, R., & Lumbantoruan, G. (2018). Optimasi performa cluster K-Means menggunakan Sum of Squared Error (SSE). *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, *2*(2), 103-108. DOI: https://doi.org/10.46880/jmika.Vol2No2.pp103-108.

[15] Ahmad, A., & Gata, W. (2022). Sentimen Analisis Masyarakat Indonesia di Twitter Terkait Metaverse dengan Algoritma Support Vector Machine. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(4), 548-555. DOI: https://doi.org/10.35870/jtik.v6i4.569.

[16] Nanja, M. (2022). indonesia Clustering Daerah Penyebaran Covid-19 Di Indonesia Menggunakan Algoritma K-Medoids. *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, *6*(4), 608-615. DOI: https://doi.org/10.35870/jtik.v6i4.608.

[17] Triyanto, W. A. (2015). Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran Produk. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, *6*(1), 183-188. DOI: https://doi.org/10.24176/simet.v6i1.254.

[18] Andini, A. D., & Arifin, T. (2020). Implementasi algoritma k-medoids untuk klasterisasi data penyakit pasien di rsud kota bandung. *Jurnal Responsif: Riset Sains dan Informatika*, *2*(2), 128-138. DOI: https://doi.org/10.51977/jti.v2i2.247.