

## PENILAIAN KINERJA TIM EVALUASI PERIJINAN PEMBUKAAN PROGRAM STUDI DENGAN TEKNIK CLUSTERING K-MEANS

Ibnu Akil

Fakultas Teknologi Informatika  
Universitas Bina Sarana Informatika  
[www.bsi.ac.id](http://www.bsi.ac.id)  
ibnu.ial@bsi.ac.id

**Abstract**—The Ministry of Research, Technology and Higher Education considers the process of evaluating study program permits to be long, in turns it will affect the development of education in colleges. Therefore, an effective and efficient licensing assessment team is needed, so it is necessary to classify which assessment team performs well and which does not perform well. Clustering is an effective data mining technique for grouping. One of the clustering techniques that is quite empirically proven is K-Means. The purpose of this research is to group the high-performing assessment team with the average one using the K-Means clustering technique, in order to improve or speed up the evaluation process of the study program permit proposal at the Ministry of Education and Culture.

**Keyword:** clustering, data mining, k-means

**Abstrak**—Proses evaluasi perijinan program studi dianggap lama oleh Menteri Kemenristekdikti. Hal ini dapat menyebabkan tertundanya perkembangan pendidikan di kampus-kampus. Karena itu diperlukan tim penilai perijinan yang efektif dan efisien, maka perlu dikelompokkan mana tim penilai yang kinerjanya baik dan yang kurang baik. Clustering merupakan Teknik data mining yang efektif untuk melakukan pengelompokan. Salah satu Teknik clustering yang cukup terbukti secara empiris adalah K-Means. Tujuan dari penelitian ini adalah melakukan pengelompokan tim penilai yang berkinerja tinggi dengan yang biasa saja dengan menggunakan Teknik clustering K-Means, guna meningkatkan atau mempercepat proses evaluasi usulan perijinan Program Studi pada Kementerian Pendidikan dan Kebudayaan.

**Kata kunci:** pengelompokan, data mining, k-means

### PENDAHULUAN

Penilaian kinerja merupakan kekuatan dari manajemen kinerja, yang pada gilirannya mempengaruhi kinerja organisasi (Akinbowale et al., 2014). Mengelompokkan karyawan atau staff mana yang memiliki kinerja yang baik dengan yang kurang baik bisa jadi hal yang krusial. Evaluasi perijinan program studi memerlukan kinerja tim penilai yang tinggi. Proses perijinan pembukaan program studi memerlukan waktu yang lama sebagaimana disampaikan oleh Menteri Kemenristekdikti (LPM-UMI, 2019). Karena itu diperlukan tim penilai perijinan yang efektif dan efisien, maka perlu dikelompokkan mana tim penilai yang kinerjanya baik dan yang kurang baik.

Clustering merupakan Teknik data mining yang efektif untuk melakukan pengelompokan. *Clustering* atau pengelompokan telah menjadi cara yang efektif didalam *artificial intelligent* untuk algoritma *unsupervised learning*, dimana *clustering* mampu memperlihatkan keterkaitan data dalam satu kelompok atau lebih. Menurut Tan dan kawan-kawan dalam

(Irwansyah, 2017), *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu kelompok memiliki tingkat kemiripan yang maksimum. Dalam dunia data sains kita dapat menggunakan analisa *clustering* untuk mendapatkan suatu pencarian nilai dari data yang kita miliki dengan melihat dataset terbagi menjadi kelompok-kelompok apa saja (Seif, 2018).

Selain itu *clustering* adalah teknik yang secara luas digunakan di dunia industri. Bahkan sebenarnya *clustering* digunakan pada hampir setiap domain mulai dari perbankan, mesin rekomendasi, *clustering* dokumen sampai pada segmentasi citra (Pulkit, n.d.). Dalam kasus ini peneliti mencoba menerapkan teknik *clustering* di dalam mengelompokkan tim penilai yang memiliki kinerja tinggi atau rendah dari evaluasi perijinan program studi di Kementerian Pendidikan dan Kebudayaan. Tujuannya adalah untuk memilih tim penilai atau evaluator yang akan secara aktif ditugaskan untuk mengevaluasi proposal perijinan program studi.

Teknik clustering cukup banyak, namun pada dasarnya terbagi menjadi dua kategori utama, yaitu; *partitioning* dan *hierarchical* (Soni & Ganatra, 2012).

*Hierarchical clustering* menghasilkan urutan yang bersarang dari sebuah partisi, dengan kelompok tunggal inklusif dibagian atas dan kelompok-kelompok tunggal titik individu berada dibawahnya, dimana masing-masing tingkat menengah dianggap sebagai menggabungkan dua kelompok dari tingkat yang lebih rendah berikutnya atau memisahkan satu kelompok dari yang lebih tinggi berikutnya (Grover, 2014). Sementara Steinbach membagi *hierarchical* yang disebut sebagai *agglomerative hierarchical* menjadi dua yaitu; *intra cluster similarity*, dan *centroid similarity* (Steinbach et al., 2000).

Teknik *Partitioning*

*Partitioning clustering* menempatkan satu set titik-titik data kedalam K-cluster dengan proses perulangan. Fungsi kriteria (J) yang telah ditentukan menempatkan data ke dalam set K yang ke-i. Sebagai hasilnya nilai di dalam set K (perhitungan maksimalisasi dan minimisasi), pengelompokan dapat dilakukan (Kutbay, 2018). *Partitioning* adalah metode clustering yang digunakan untuk mengklasifikasikan observasi-observasi di dalam suatu data set menjadi banyak kelompok berdasarkan kemiripan mereka. Algoritma ini membutuhkan si penganalisa untuk menentukan jumlah kelompok yang akan dibuat (Kassambara, 2021). "K-means merupakan salah satu teknik *partitioning clustering* yang memisahkan data ke K daerah bagian yang terpisah" (Kusuma & Ellyana, 2018). Ada juga yang membagi Teknik *clustering* menjadi *Soft Clustering* atau *Overlapping Clustering* dan *Hard Clustering* atau *Exclusive Clustering* (Bora & Gupta, 2014). Teknik-teknik berbeda yang diajukan tentunya memiliki fitur-fitur yang berbeda serta memiliki kelebihan dan kekurangan, tentunya akan menjadi sulit untuk memilih mana teknik yang tepat untuk kasus ini.

Beberapa peneliti sebelumnya telah melakukan komparasi antara algoritma-algoritma *clustering* diantaranya Neha R. Soni dari

India (Soni & Ganatra, 2012). Soni menilai algoritma-algoritma *clustering* tersebut berdasarkan; kompleksitas, tipe data, besaran data, bentuk cluster, dan input parameter. Dari hasil penelitian tersebut peneliti mempertimbangkan pemilihan algoritma pada kompleksitas dan besaran data saja. Maka disini akan digunakan algoritma K-means dimana K-means memiliki kelebihan diantaranya: kemudahan implementasi, kesederhanaan, efisiensi, dan teruji secara empiris. Sedangkan

kelemahannya adalah: *scalability*, kelompok yang tidak seimbang, tidak cocok untuk kelompok-kelompok dari bentuk *non convex*, dan sensitif terhadap noise (Soni & Ganatra, 2012).

## BAHAN DAN METODE

Penelitian ini dilakukan beberapa tahapan dimulai dari mengumpulkan data terkait teori K-Means, kemudian contoh-contoh implementasi K-Means dalam berbagai kasus. Menganalisa beberapa algoritma *clustering* yang sejenis, serta pemilihan algoritma K-Means berdasarkan hasil analisis. Kemudian perancangan implementasi dalam kasus penilaian kinerja Tim Evaluasi yang meliputi, pengumpulan data dasar untuk penilaian, kemudian implementasi algoritma K-Means dengan menggunakan software RapidMiner dan menganalisis hasilnya. Untuk lebih jelasnya perhatikan gambar 1 yang merupakan proses rancangan penelitian.



Gambar 1. Rancangan Proses Penelitian

### Data Penelitian

Data ini diambil dari sistem berjalan yaitu aplikasi <http://silemkerma.kemdikbud.go.id> mengenai proses penugasan evaluator dan hasil evaluasi dari evaluator tersebut. Data tersebut meliputi data evaluator, jumlah penugasan, jumlah penugasan yang selesai dan jumlah penugasan yang kadaluarsa.

### Analisis Data

Dari jumlah evaluator sebanyak 563, peneliti mengambil sample hanya 15 record untuk ditampilkan, namun dalam prosesnya semua data akan diproses. Perhatikan tabel 1 kolom Penugasan adalah jumlah seluruh penugasan yang telah diberikan, kolom "Dalam Proses" adalah usulan yang sedang dalam penilaian evaluator, kolom "Selesai" adalah jumlah seluruh penugasan yang telah selesai dikerjakan, kolom "Expired" adalah jumlah penugasan yang telah melebihi batas waktunya, dan kolom "Rerata Waktu Kerja" adalah rerata dari lama penilaian masing-masing penugasan.

Tabel 1 Sample Data Penelitian

ID	Nama Evaluator	Penugasan	Dalam Proses	Selesai	Expired	Rerata Waktu Kerja
1	Liliana Sugiharto, Dr.	463	0	457	2	7,31
2	Lukas, Dr., Ir., MAI, CISA	382	2	379	2	4,86
3	Achmad Benny Mutiara, Dr. Prof.	350	0	348	0	8,31
4	Euphrasia Susy Suhendra, Dr. Prof.	311	5	304	1	7,72
5	Suhendrik Hanwar	294	2	287	4	9,63
6	Abimanyu Dipo Nusantara, Ir M.P Dr	259	0	258	0	8,79
7	Rahmat Gunawan, M.Si Dr.	207	0	195	2	9,29
8	Faisal, SE., M.Si., MA., Dr.	197	0	196	1	5,40
9	Mohamad Adam, S.E., M.E., Dr., Prof.	174	0	174	1	3,87
10	Lipur Sugiyanta, Ph.D	165	16	146	2	7,42
11	Chandrasa Soekardi	163	0	160	1	6,49
12	Dr. Evi Gravitiani, SE, M.Si	155	1	152	2	10,92
13	Andi Sularso, MSM, Dr. Prof.	149	0	149	1	7,09
14	Yanti, Dr., MKeb	144	0	144	1	5,20
15	Dr. Heri Fathurahman	139	1	132	4	11,35

**Teknik Clustering K-Means**

Algoritma K-Means adalah algoritma yang bersifat perulangan yang mencoba membagi dataset menjadi kelompok belum terdefinisi yang terbedakan dan tidak tumpang tindih dimana setiap poin data hanya dimiliki oleh satu group (Dabbura, 2018).

Cara algoritma K-Means bekerja adalah sebagai berikut:

1. Tentukan jumlah kelaster K.
2. Inisialisasi centroids (titik pusat) dengan tahap pertama mengacak dataset dan secara acak memilih K poin data untuk setiap titik pusat tanpa menimpa.
3. Teruskan perulangan sampai tidak ada perubahan pada setiap titik pusat.
4. Hitung jumlah jarak kuadrat antara poin data - poin data dengan setiap titik pusat.
5. Tempatkan setiap poin data ke titik pusat terdekat.
6. Hitung setiap titik pusat tersebut untuk kelaster-kelaster dengan mendapatkan rerata dari semua poin data yang dimiliki oleh masing-masing kelompok.

Fungsi tujuan dari K-Means adalah:

$$J = \sum_{i=1}^m \sum_{k=1}^k w_{ik} || x^i - \mu_k ||^2 \dots\dots\dots(1)$$

Dimana  $w_{ik} = 1$  untuk setiap poin data  $x^i$  jika ia dimiliki oleh cluster  $k$ ; selain itu  $w_{ik} = 0$ , dan juga  $\mu_k$  adalah titik pusat (centroid) dari kelompok  $x^i$ .

**Proses Pengolahan Data**

Untuk memproses data penelitian dengan menggunakan Teknik K-Means akan lebih baik hasilnya jika menggunakan software data mining yang cukup handal, baik dari sisi fitur dan kemudahan di dalam penggunaannya. Rapid Miner adalah system yang mendukung desain dan dokumentasi dari keseluruhan proses data mining

(Hofmann & Klinkenberg, 2014). RapidMiner termasuk salah satu software yang memiliki kelebihan diantaranya lengkapnya pemodelan dengan banyak algoritma atau teknik analisis yang disebut operator.

**HASIL DAN PEMBAHASAN**

**Persiapan Data Awal**

Data yang didapat berupa raw data dari database langsung dengan menggunakan query. Ada dua table yang menjadi sumber data, pertama table penugasan dan yang kedua table hasil evaluasi. Field "Penugasan", "Dalam Proses", "Selesai" dan "Expired" merupakan summary dari seluruh data yang bersesuaian dengan field tersebut. Sedangkan field "Rerata Waktu Kerja" didapat dari menghitung jumlah hari antara tanggal penugasan dengan tanggal upload hasil penugasan keseluruhan, kemudian diambil nilai rerata dari hasilnya. Informasi-informasi ini dibutuhkan untuk melihat berapa banyak seorang evaluator telah menerima penugasan, kemudian jumlah penugasan yang sedang dia evaluasi, jumlah penugasan yang telah expire, jumlah penugasan yang telah selesai dan rerata berapa lama dia mengerjakan penugasan tersebut.

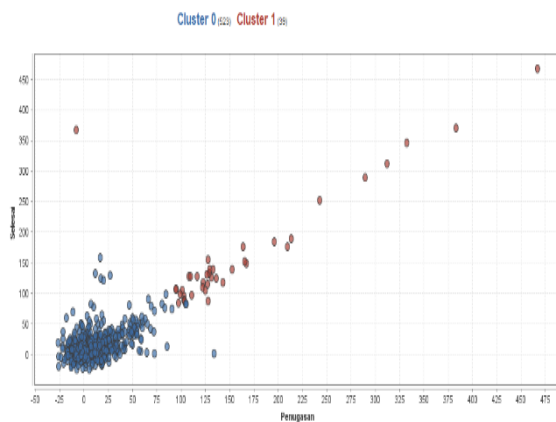
**Implementasi K-Means**

Setelah data disesuaikan dengan kebutuhan untuk diproses lebih lanjut dengan metode K-Means, maka dapat dilihat hasil olah K-Means dari RapidMiner. Yang pertama adalah *centroid table* seperti table 2 berikut:

Tabel 2. Centroid Table

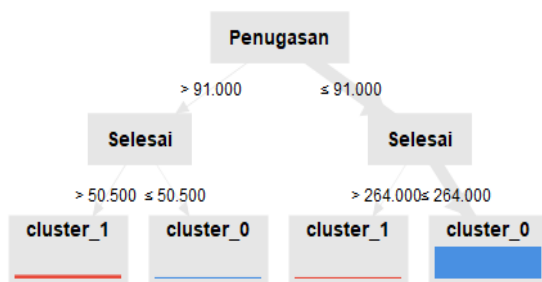
Cluster	Dalam Proses	Expired	Penugasan	Selesai	Waktu Kerja
Cluster 0	0.597	1.486	13.895	15.400	9.831
Cluster 1	1.103	1.385	158.872	165.103	6.823

Perhatikan gambar 2, dapat dilihat hasil centroid table untuk setiap kolom dari masing-masing kelompok (cluster).



Gambar 2. Grafik Scatter Plot

Perhatikan Gambar 2, kita bisa lihat titik-titik poin data pada masing-masing kelompok. Cluster 0 berwarna biru dan cluster 1 berwarna merah, yang menggambarkan pemisahan centroid dari cluster 0 dan 1.



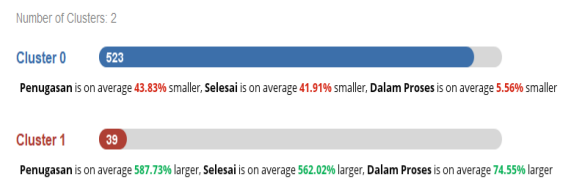
Gambar 3. Cluster Tree

Gambar 3 merupakan decision tree dari output data. Tujuan dari decision tree adalah untuk mengidentifikasi objek kelas tertentu. Decision tree menggunakan atribut objek yang berbeda untuk mengklasifikasikan subset objek yang berbeda dan tidak menggunakan hanya satu atribut (Križanić, 2020). Perhatikan bahwa kelompok yang Selesai  $\leq 264.000$  masuk ke dalam cluster 0. Sedangkan  $>$  dari 264.000 masuk ke dalam cluster 1.

Attributes	Dalam ...	Expired	Penuga...	Selesai	Waktu ...
Dalam Proses	1	0.044	0.141	0.100	0.004
Expired	0.044	1	0.020	-0.003	0.040
Penugasan	0.141	0.020	1	0.880	-0.049
Selesai	0.100	-0.003	0.880	1	-0.071
Waktu Kerja	0.004	0.040	-0.049	-0.071	1

Gambar 4. Tabel Correlation Matrix

Untuk melihat hubungan antar kolom atau atribut, dapat menggunakan Correlation Matrix. Statistic yang kompleks dan relasi dari angka-angka perlu di sederhanakan dengan menggunakan correlation matrix (McGrew, 2021). Perhatikan gambar 4, atribut yang memiliki correlation yang kecil ( $<0.0$ ) dengan atribut lain adalah atribut Waktu Kerja, maka dapat dikatakan atribut Waktu Kerja tidak memiliki hubungan yang signifikan dengan atribut lainnya, karenanya atribut Waktu Kerja dikeluarkan dari summary, perhatikan gambar 5.



Gambar 5. Summary

Dari 563 evaluator setelah melalui proses clustering dengan K-Means yang terbagi menjadi 2 kelompok, yaitu cluster 0 dan cluster 1, perhatikan Gambar 5. Cluster 0 sebanyak 523 dimana rerata Penugasan  $< 43,83\%$ , selesai  $< 41,91\%$ , dan Dalam Proses  $< 5,56\%$ . sedangkan cluster 1 rerata Penugasan  $> 587,73\%$ , Selesai  $> 562,02\%$ , Dalam Proses  $> 74,55\%$ . Maka dari summary ini bisa diketahui bahwa evaluator yang masuk ke dalam cluster 1 adalah evaluator yang Penugasan, Selesai dan Dalam Proses-nya lebih besar dari evaluator yang masuk ke dalam cluster 0. Maka dari sini didapatkan 39 evaluator yang memiliki kinerja dan prestasi yang baik. Perhatikan tabel 3 berikut adalah data evaluator yang masuk ke dalam cluster 1.

Tabel 3. Data Evaluator Cluster 1

Penugasan	Dalam Proses	Selesai	Expired	Waktu Kerja	cluster	id
463,0	0,0	457,0	2,0	7,3	cluster_1	1,0
382,0	2,0	379,0	2,0	4,9	cluster_1	2,0
350,0	0,0	348,0	0,0	8,3	cluster_1	3,0
311,0	5,0	304,0	1,0	7,7	cluster_1	4,0
294,0	2,0	287,0	4,0	9,6	cluster_1	5,0
259,0	0,0	258,0	0,0	8,8	cluster_1	6,0

Penugasan	Dalam Proses	Selesai	Expired	Waktu Kerja	cluster	id
207,0	0,0	195,0	2,0	9,3	cluster_1	7,0
197,0	0,0	196,0	1,0	5,4	cluster_1	8,0
174,0	0,0	174,0	1,0	3,9	cluster_1	9,0
165,0	16,0	146,0	2,0	7,4	cluster_1	10,0
163,0	0,0	160,0	1,0	6,5	cluster_1	11,0
155,0	1,0	152,0	2,0	10,9	cluster_1	12,0
149,0	0,0	149,0	1,0	7,1	cluster_1	13,0
144,0	0,0	144,0	1,0	5,2	cluster_1	14,0
139,0	1,0	132,0	4,0	11,3	cluster_1	15,0
135,0	1,0	134,0	1,0	3,7	cluster_1	16,0
134,0	0,0	130,0	1,0	9,3	cluster_1	17,0
134,0	0,0	132,0	1,0	6,8	cluster_1	18,0
133,0	1,0	124,0	2,0	4,3	cluster_1	19,0
132,0	2,0	9,0	1,0	6,4	cluster_0	20,0
131,0	2,0	122,0	3,0	10,6	cluster_1	21,0
131,0	4,0	127,0	1,0	8,2	cluster_1	22,0
127,0	0,0	126,0	1,0	3,3	cluster_1	23,0
121,0	0,0	120,0	1,0	2,1	cluster_1	24,0
117,0	0,0	115,0	1,0	9,7	cluster_1	25,0
116,0	0,0	115,0	1,0	10,1	cluster_1	26,0
115,0	0,0	115,0	1,0	3,2	cluster_1	27,0
114,0	0,0	111,0	1,0	4,8	cluster_1	28,0
113,0	0,0	109,0	1,0	5,9	cluster_1	29,0
112,0	0,0	107,0	2,0	3,8	cluster_1	30,0
110,0	1,0	109,0	1,0	5,4	cluster_1	31,0
106,0	4,0	101,0	0,0	6,1	cluster_1	32,0
105,0	0,0	105,0	1,0	2,8	cluster_1	33,0
102,0	0,0	99,0	1,0	3,1	cluster_1	34,0
99,0	0,0	99,0	1,0	4,8	cluster_1	35,0
98,0	0,0	98,0	1,0	5,8	cluster_1	36,0
98,0	2,0	92,0	3,0	24,2	cluster_1	37,0
98,0	1,0	96,0	2,0	9,3	cluster_1	38,0
94,0	0,0	93,0	1,0	5,1	cluster_1	39,0

**KESIMPULAN**

Dari hasil pengolahan data dengan Teknik clustering K-Means dengan menggunakan RapidMiner dapat diketahui bahwa 39 evaluator masuk ke dalam cluster atau kelompok 1. Cluster 1 adalah kelompok yang memiliki tingkat kecenderungan data Penugasan, Selesai dan Dalam Proses terbanyak. Sedangkan cluster 0 adalah sebaliknya. Maka dapat dikatakan evaluator yang masuk dalam cluster 1 adalah evaluator yang akan terus ditugaskan secara aktif atau menjadi evaluator utama, sedangkan evaluator yang masuk ke dalam cluster 0 akan dijadikan sebagai evaluator cadangan. Meskipun demikian pada kenyataannya evaluator yang mendapatkan sedikit penugasan bukan berarti tidak memiliki kinerja yang baik. Hal ini dikarenakan penugasan tersebut didasarkan kepada bidang ilmu evaluator harus sesuai dengan penugasan atau usulan yang akan dievaluasi. K-Means memang dapat mengelompokkan dengan baik, namun field Rerata Waktu Kerja tidak dimasukkan ke dalam summary, padahal Rerata

Waktu Kerja ini menjadi bahan pertimbangan yang lebih berbobot untuk ikut dikalkulasi, karena semakin kecil Rerata Waktu Kerja maka semakin baik kinerja dari evaluator tersebut. Sayangnya hal ini luput dari perhatian K-Means. Dapat disimpulkan bahwa nilai Rerata Waktu Kerja tidak berkorelasi dengan atribut lainnya sehingga dikeluarkan dari summary. Untuk penelitian lebih lanjut perlu dikembangkan model yang lebih baik untuk menganalisa clustering dengan kasus yang mirip dimana ada atribut yang tidak berkorelasi tetapi memiliki value yang penting.

**REFERENSI**

Akinbowale, M. A., Lourens, M. E., & Jinabhai, D. C. (2014). Employee performance measurement and performance appraisal policy in an organisation. *Mediterranean Journal of Social Sciences*, 5(9), 342–347. <https://doi.org/10.5901/mjss.2014.v5n9p342>

Bora, D. J., & Gupta, D. A. K. (2014). A Comparative study Between Fuzzy

- Clustering Algorithm and Hard Clustering Algorithm. *International Journal of Computer Trends and Technology*, 10(2), 108–113.  
<https://doi.org/10.14445/22312803/ijctt-v10p119>
- Dabbura, I. (2018). *K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks*. Towardsdatascience.Com.  
<https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>
- Grover, N. (2014). A study of various Fuzzy Clustering Algorithms. *International Journal of Engineering Research*, 3(3), 177–181.  
<https://doi.org/10.17950/ijer/v3s3/310>
- Hofmann, M., & Klinkenberg, R. (2014). RapidMiner Data Mining Use Cases and Business Analytics Applications. In *Taylor & Francis Group*. CRC Press.
- Irwansyah, E. (2017). *Clustering*. Socs.Binus.Ac.Id.  
<https://socs.binus.ac.id/2017/03/09/clustering/>
- Kassambara, A. (2021). *PARTITIONAL CLUSTERING IN R: THE ESSENTIALS*. Wwww.Datanovia.Com.  
<https://www.datanovia.com/en/courses/partitional-clustering-in-r-the-essentials/>
- Križanić, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study. *International Journal of Engineering Business Management*, 12, 1–9.  
<https://doi.org/10.1177/1847979020908675>
- Kusuma, A. W., & Ellyana, R. L. (2018). Penerapan Citra Terkompresi Pada Segmentasi Citra Menggunakan Algoritma K-Means. *Jurnal Terapan Teknologi Informasi*, 2(1), 65–74.  
<https://doi.org/10.21460/jutei.2018.21.65>
- Kutbay, U. (2018). Partitional Clustering. In *Recent Applications in Data Clustering: Vol. i* (Issue tourism, p. 13). IntechOpen Limited.  
<https://doi.org/10.5772/intechopen.75836>
- LPM-UMI. (2019). *Percepat Proses Izin Pembukaan Prodi serta Pendirian dan Perubahan PTS, dalam 15 Hari Kerja*.  
<https://lpm.umi.ac.id/news/detail/percepat-proses-izin-pembukaan-prodi-serta-pendirian-dan-perubahan-pts-dalam-15-hari-kerja>
- McGrew, K. (2021). *Be and see the WISC-V correlation matrix Kevin S. McGrew 05-05-21. May*.  
<https://doi.org/10.13140/RG.2.2.17324.18569>
- Pulkit, S. (n.d.). *The Most Comprehensive Guide to K-Means Clustering You'll Ever Need*.
- Seif, G. (2018). *The 5 Clustering Algorithms Data Scientists Need To Know*. Towardsdatascience.Com.  
<https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>
- Soni, N., & Ganatra, A. (2012). Comparative study of several Clustering Algorithms. *International Journal of Advanced Computer Research (ISSN (Print): 2249-7277 ISSN (Online): 2277-7970)*, 7(4), 2–6.
- Steinbach, M., Karypis, G., & Kumar, V. (2000). Comparison of Document Clustering Techniques. In *conservancy.umn.edu*.