

K-MEANS CLUSTERING DALAM PENGELOMPOKAN RELEVANSI PEKERJAAN S1 INFORMATIKA (STUDI KASUS JURUSAN TEKNIK INFORMATIKA UMM MALANG)

Dikky Cahyo Hariyanto^{*1}, Sri Harini², Totok Chamidy³

1. Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang
2. Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Magister Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang

Article Info

Kata Kunci: K-Means; Clustering; Data mining; Relevansi

Keywords: *K-Means; Clustering; Data Mining; Relevance*

Article history:

Received 3 March 2024
Revised 17 March 2024
Accepted 31 March 2024
Available online 1 June 2024

DOI :

<https://doi.org/10.29100/jipi.v9i2.5507>

* Corresponding author.

Dikky Cahyo Hariyanto

E-mail address:

dikky037@gmail.com

ABSTRAK

Perguruan tinggi merupakan salah satu tingkatan dalam menuntut ilmu diharapkan untuk menciptakan lulusan yang mampu dan berkompeten sesuai dengan bidang ilmu yang diambil. Banyaknya jumlah lulusan dari perguruan tinggi dengan jumlah lapangan pekerjaan yang aa menyebabkan banyaknya lulusan bekerja tidak sesuai dengan jurusnya, sehingga perlu adanya evaluasi akan tingkat keberhasilan capaian pembelajaran lulusan yang ada di perguruan tinggi. Tujuan dari penelitian ini adalah untuk menganalisa data relevansi pekerjaan para lulusan S1 Teknik Informatika dengan apa yang mereka pelajari sesuai dengan capaian pembelajaran yang ada di program studi Teknik Informatika Universitas Muhammadiyah Malang menggunakan K-Means clustering. Menggunakan data hasil kuisisioner pengukuran capaian pembelajaran para lulusan dan pengukuran kesesuaian perkerjaan sebanyak 137 responden yang telah diuji validitas, reabilitas dan multikolinieritas, hasil penelitian ini menunjukkan dta dibentuk menjadi tiga cluster dengan analisa bahwa 29.92% lulusan Teknik Informatika UMM mampu memenuhi capaian pembelajaran lulusan dan mendapatkan pekerjaan yang relevan denga apa yang mereka pelajari, 49.63% lulusan lainnya juga mendapat pekerjaan yang relevan dengan jurusannya walaupun kurang menguasai keterampilan khusus yang diukur pada capaian pembelajara lulusan, dan 20.45% lulusan lainnya mendapatkan pekerjaan yang kurang relevan dengan bidang teknik Informatika.

ABSTRACT

Higher education is one of the levels of studying that is expected to produce graduates who are capable and competent in accordance with the field of knowledge taken. The large number of graduates from tertiary institutions with a large number of job opportunities causes many graduates to work in ways that do not match their majors, so there is a need for an evaluation of the level of success of graduates' learning achievements in tertiary institutions. The aim of this research is to analyze data on the relevance of the work of undergraduate graduates in Informatics Engineering to what they have learned in accordance with the learning outcomes in the Informatics Engineering study program at the University of Muhammadiyah Malang using K-Means clustering. Using data from questionnaires measuring graduate learning outcomes and measuring job suitability for 137 respondents who had been tested for validity, reliability and multicollinearity, the results of this research showed that the data was formed into three clusters with the analysis that 29.92% of UMM Informatics Engineering graduates were able to meet graduate learning outcomes and obtain jobs that are relevant to what they studied, 49.63% of other graduates also got jobs that were relevant to their major even though they lacked mastery of specific skills as measured by graduate learning outcomes, and 20.45% of other graduates got jobs that were less relevant to the field of Informatics engineering.

I. PENDAHULUAN

PERGURUAN tinggi merupakan salah satu tingkatan dalam menuntut ilmu yang selalu diharapkan masyarakat dapat menciptakan lulusan yang mampu serta kompeten dengan bidang ilmunya. Tingginya minat untuk menuntut ilmu ke perguruan tinggi membuat banyak sekali berdirinya institusi perguruan tinggi. Data yang dihimpun dari Badan Pusat Statistik (BPS) pada tahun 2021 menunjukkan ada 3115 Universitas yang ada di Indonesia.

Banyaknya jumlah universitas ini berdampak pada banyaknya lulusan yang dihasilkan. Dari fenomena yang ada terdapat banyak sekali lulusan perguruan tinggi yang mendapatkan pekerjaan yang tidak sesuai dengan apa yang mereka pelajari selama kuliah. Tujuan dari penelitian ini adalah untuk melakukan analisa data relevansi pekerjaan para lulusan S1 Teknik Informatika dengan apa yang mereka pelajari sesuai dengan capaian pembelajaran yang ada di program studi Teknik Informatika Universitas Muhammadiyah Malang menggunakan K-Means.

Algoritma K-means merupakan metode yang banyak digunakan di bidang analisis kluster dan penambangan data karena kesederhanaan dan efisiensi komputasinya. Selain itu algoritma ini relatif dapat digunakan untuk data yang berskala besar sehingga sangat mudah diterapkan pada berbagai studi kasus di kehidupan nyata. Disisi lain algoritma ini juga secara umum sering diterapkan ke cluster dengan berbagai bentuk dan ukuran yang berbeda seperti cluster dengan model elips. Algoritma ini bertujuan untuk mempartisi kumpulan data tertentu menjadi sejumlah cluster yang telah ditentukan, dengan setiap titik data ditugaskan ke cluster dengan mean terdekat. Proses ini membantu mengidentifikasi pengelompokan dalam data sangat bermanfaat untuk pengenalan pola dan interpretasi data [1].

Algoritma ini memiliki beberapa langkah dalam implementasinya dimana :

1. langkah pertama adalah menentukan nilai K (*cluster*) yang ingin dibuat.
2. Menentukan pusat *centroid* secara acak.
3. Menghitung jarak antar *centroid* menggunakan persamaan *Euclidean distance* dimana metode ini menjadi metode yang paling efektif dalam perhitungan jarak cluster [2].
$$[(x,y), (a,b)] = \sqrt{(x-a)^2 + (y-b)^2} \dots \dots \dots (1)$$
4. Menggabungkan data berdasarkan jarak terpendek dengan *centroid*
5. Menentukan *centroid* baru berdasarkan keanggotaan *cluster*
6. Mengulang langkah ke-3 sampai ke-5 hingga keanggotaan setiap cluster tidak ada perubahan.

Algoritma K-means telah banyak dibahas dalam literatur, dan memiliki beberapa kelemahan. Salah satu kelemahan utama adalah sensitivitasnya terhadap pemilihan awal pusat cluster, yang dapat berdampak signifikan terhadap hasil akhir pengelompokan [3]. Ketergantungan pada nilai awal sering kali menyebabkan kualitas hasil pengelompokan yang buruk, terutama ketika *centroid* awal tidak dipilih dengan tepat.

K-means dikenal luas pada bidang data mining dandiakui sebagai salah satu dari 10 algoritma data mining teratas [4]. Algoritma ini digunakan untuk mengelompokkan dan mengelompokkan data di berbagai domain, seperti analisis data perkotaan [5], evaluasi risiko utang pemerintah [6], penilaian prestasi siswa [7], keuangan penambangan strategi pasar [8], pengelompokan profil kesehatan [9], analisis tingkat kejahatan [10] serta pengelompokkan kategori rumah sakit berdasarkan asilitas dan ketersediaan dokter [11]. Algoritma ini dikenal dengan proses yang tidak memerlukan pengawasan dan kemampuannya mengelompokkan data menggunakan sistem partisi [9]. Disisi lain. algoritma ini juga telah diterapkan dalam model pemantauan, seperti pemantauan getaran turbin angin berdasarkan data SCADA (*Supervisory Control and Data Acquisition*) [12].

Algoritma K-means telah dibandingkan dengan metode lain, seperti dalam studi algoritma pengelompokan lalu lintas, yang mana algoritma ini dijadikan sebagai tolak ukur perbandingan [5]. Selain itu, algoritma ini telah digunakan untuk klasifikasi ucapan dan ekstraksi fitur audio, yang menunjukkan keserbagunaannya dalam berbagai tugas analisis data [13]. Algoritma tersebut juga telah digunakan untuk analisis kinerja di lembaga pendidikan, dimana data yang dikumpulkan dari universitas swasta dikelompokkan menggunakan metode K-means [14]. Aplikasi ini menyoroti kemampuan adaptasi dan efektivitas algoritma di berbagai bidang penelitian.

K-means telah banyak digunakan dalam berbagai aplikasi dalam data mining, menunjukkan signifikansi dan fleksibilitasnya dalam mengelompokkan dan menganalisis data di berbagai domain. Kemudahan dalam penerapan dan tingkat efektivitasnya inilah yang menjadikan penulis memilih untuk menggunakan K-Means dalam melakukan pengelompokan data dan melakukan analisis data relevansi pekerjaan lulusan perguruan tinggi khususnya program studi teknik Informatika Universitas Muhammadiyah Malang sebagai bahan evaluasi pihak terkait untuk keberlanjutan akademik dengan mengukur ketercapain para lulusan akan melalulu unsur variabel IPK, Unsur capaian pembelajaran lulusan (CPL) yang terdiri dari unsur Sikap (rx1), Keterampilan Umum (rx2),

Pengetahuan (rx3), Keterampilan Khusus (rx4), dan unsur kesesuaian pekerjaan yang didapat (ry).

II. METODE PENELITIAN

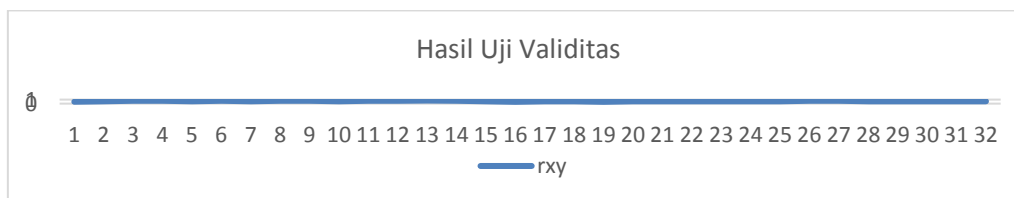
Metode pengumpulan data yang digunakan dalam penelitian ini melalui kuisioner pengukuran capaian pembelajaran para lulusan dan pengukuran kesesuaian pekerjaan yang mereka dapatkan. Instrumen pertanyaan kuisioner mengdopsi dari kuisioner yang dibuat oleh [15]. Kuisioner disebarikan kepada para lulusan Program S1 Teknik Informatika Universitas Muhammadiyah Malang (UMM) dari tahun 2018 sampai tahun 2021 dengan total lulusan sebesar 836 lulusan. Data hasil pengisian kuisioner sebanyak 137 responden dengan 32 item pertanyaan yang mewakili unsur sikap, pengetahuan, keterampilan umum, dan keterampilan khusus ini diolah menggunakan proses Knowledges Discovery in Database serta dilakukan uji validitas, reliabilitas dan multikolinierias sebelum masuk ke komputasi dan diterapkan algoritma K-Means serta dilakukan perbaikan untuk dianalisa hasil pengelompokan datanya. Karakteristik responden pengisi kuisioner dapat dilihat pada Tabel 1.

TABEL I
 KARAKTERSTIK DATA RESPONDEN

Karakteristik Jenis Kelamin	Frekuensi	Persentase
Laki-laki	103	75.18%
Perempuan	34	24.82%
Total	137	100%
Masa Tunggu Kerja		
<1 Bulan	42	30.65%
<3 Bulan	26	18.97%
<6 Bulan	24	17.51%
<12 Bulan	22	16.05%
>12 Bulan	23	16.82%
Total	137	100%

A. Uji Validitas dan Reliabilitas

Data dilakukan pengujian validitas dan reliabilitas untuk melihat kualitas data dan apakah instrument dalam kuisioner ini dapat digunakan untuk mengukur apa yang akan diukur. Instrumen dinyatakan berkualitas dan dapat dipertanggungjawabkan apabila telah terbukti validitas dan reliabilitasnya [16]. Berikut hasil uji kuisioner yang telah dilakukan:



Gambar. 1. Grafik Hasil Uji Validitas

Gambar 1 menunjukkan grafik hasil uji validitas dari ke 32 instrumen kuisioner yang digunakan pada penelitian ini dinyatakan valid dimana hasil r_{xy} saat dibandingkan dengan nilai r_{tabel} yang didapat pada daftar nilai r_{tabel} dengan tingkat signifikansi 0.05 dengan uji 2 sisi dimana nilainya adalah 0.1678. Dari hasil perbandingan tersebut maka instrumen dinyatakan valid.

TABEL II
 HASIL UJI RELIABILITAS

Jml Varian Butir	Varians Total	Koefisien Reliabilitas (r11)	Interpretasi
24.008	31.810	0.950	Sangat Reliabel

Tabel 2 menunjukkan bahwa hasil uji reabilitas nilai koefisien reabilitas (r11) yang menunjukkan angka 0.950 dan nilai tersebut sudah diatas 0.80 dimana dengan ini dapat disimpulkan bahwa hasil uji reabilitas *Cronbach Alfa* pada kuisioner dalam penelitian ini tingkat reabilitasnya sangat tinggi atau sangat reliabel.

B. Transformasi Data

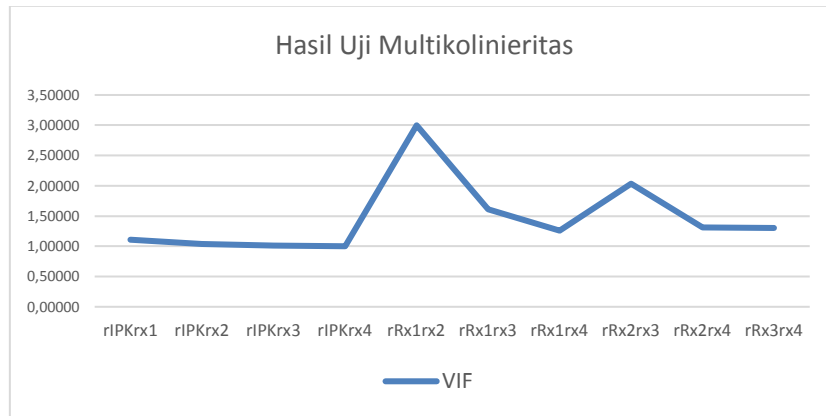
Data hasil uji validitas dan reliabilitas selanjutnya dilakukan pengolahan dimana data akan di integrasikan sesuai dengan unsur capaian pembelajaran menjadi atribut Rx1, Rx2, Rx3, Rx4 dan tingkat kesesuaian pekerjaan yang mereka dapatkan menjadi Ry serta ditambah satu atribut lagi berupa IPK yang diisikan oleh responden saat mengisi kuisioner seperti pada tabel berikut:

TABEL III
 DATASET YANG SETELAH PROSES DATA INTEGRATION

Data ke	IPK	Rx1	Rx2	Rx3	Rx4	Ry
1	3,2	75	71	75	53	1
2	3,7	75	75	50	50	8
3	3,8	81	100	90	81	7
4	3,4	81	75	85	50	6
5	3,5	69	54	65	61	3
....
137	3,6	75	67	40	39	4

C. Uji Multikolinieritas

Data hasil transformasi selanjutnya dilakukan uji multikolinieritas untuk melihat apakah seberapa kuat hubungan/korelasi antar variabel dengan cara mencari nilai tolerance (toleransi) dan VIF (*variance inflation factor*).

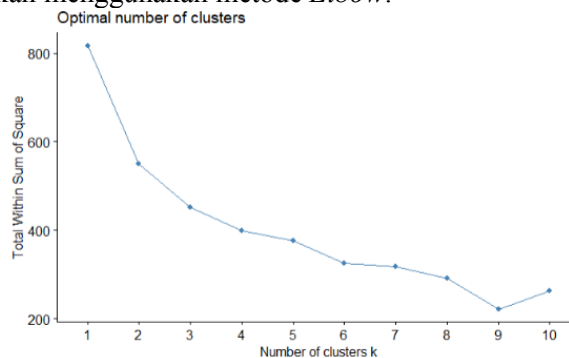


Gambar. 2. Grafik Hasil Uji Multikolinieritas

Gambar 2 menunjukkan hasil grafik pengujian multikolinieritas dimana antar atribut didapati nilai $VIF < 10$, sehingga dapat disimpulkan bahwa dataset dinyatakan non multikolinieritas.

D. Penentuan Cluster Optimal

Dalam menentukan jumlah cluster optimal, pada penelitian ini digunakan metode *Elbow* dimana metode ini yang paling umum digunakan dengan cara kerja melakukan perbandingan dari nilai antar jumlah *cluster* dalam bentuk grafik yang menunjukkan sudut grafik penurunan. Penurunan yang paling signifikanlah yang akan dijadikan sebagai jumlah *cluster* optimal [17]. Melalui dataset yang ada dicari seberapa nilai K optimalnya. Sebelum itu proses awal komputasi yang dilakukan pada RStudio adalah dengan dilakukan terlebih dahulu proses standarisasi data untuk menyamakan satuan data yang tidak sama. Selanjutnya akan masuk ke tahap mencari berapa cluster optimal dari dataset yang digunakan menggunakan metode *Elbow*.



Gambar. 3. Hasil Nilai K-Optimal dengan Model *Elbow*

Gambar 3 menunjukkan bahwa melalui model *Elbow* yang ada pada RStudio bahwa nilai K optimal ada pada cluster ke 2, hal ini ditandai dengan curamnya sudut yang diciptakan dan dengan penurunan yang lebih stabil pada nilai *Sum of Square* pada nilai K ke 2. Namun pada penelitian ini penulis menggunakan 3 *cluster* dengan tujuan untuk mempermudah melakukan analisa terkait hasil *cluster* yang dihasilkan.

III. HASIL DAN PEMBAHASAN

Setelah mendapatkan jumlah *cluster* optimal sebanyak 3 *cluster*, selanjutnya diterapkan menggunakan algoritma K-Means dengan memanfaatkan *library* yang sudah ada pada RStudio dengan menggunakan enam variabel sebagai acuan dalam pembentukan *cluster*.

```

K-means clustering with 3 clusters of sizes 68, 41, 28

Cluster means:
      IPK      rx1      rx2      rx3      rx4      ry
1  0.04359803 -0.1389142 -0.1482809  0.1336863 -0.2475787  0.1360706
2  0.05909966  0.9359578  1.0431686  0.7942361  0.9305440  0.6733269
3 -0.19241973 -1.0331467 -1.1673862 -1.4876554 -0.7613198 -1.3164001

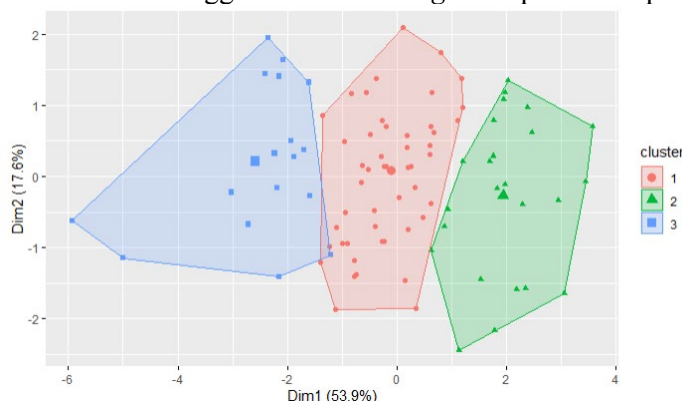
Clustering vector:
 [1] 1 1 2 1 3 2 2 1 3 1 3 2 1 2 1 1 1 1 1 2 3 2 1 3 3 3 1 2 1 2 1 3 2 3 3 1 2 2 2 1 1 2 2 2 1 2 3 1 2 1 1 3
 [53] 1 3 1 1 1 1 2 2 2 1 3 1 1 2 2 2 1 1 1 3 1 1 1 1 1 1 1 1 1 3 1 1 1 1 3 1 1 1 1 1 1 1 1 1 1 1 2 1 1 2 1 3 1 3 2 1 2
 [105] 1 1 1 1 1 2 3 2 1 3 3 3 1 2 1 2 1 3 2 3 3 1 2 2 2 1 1 2 2 2 1 2 3

Within cluster sum of squares by cluster:
 [1] 190.7223 125.5454 133.7238
 (between_SS / total_SS = 44.9 %)

Available components:
 [1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"
 [8] "iter"         "ifault"
    
```

Gambar 4. Hasil data cluster

Gambar 2 menunjukkan hasil pengolahan komputasi yang telah dilakukan dimana data telah dibagi menjadi tiga *cluster* dengan komposisi *cluster* pertama sebanyak 68 anggota (49,63%), *cluster* kedua sebanyak 41 anggota (29,92%), *cluster* ke 3 sebanyak 28 anggota (20,45%). Dari proses tersebut didapati juga tingkat keberhasilan *cluster* sebesar 44.9%. Hasil visualisasi sebaran anggota dari *clustering* ini dapat dilihat pada gambar 3.



Gambar 3. Visualisasi hasil data clustering

Dari hasil pengelompokan data yang telah dilakukan menggunakan algoritma K-Means didapati bahwa *cluster* pertama dengan jumlah 68 anggota (49.63%) dari total responden memiliki karakteristik anggota yang memiliki nilai yang tinggi pada tiga unsur capaian pembelajaran lulusan dari program studi Teknik Informatika UMM yang diwakili atribut Rx1, Rx2, Rx3 (sikap, keterampilan umum, dan pengetahuan), sementara untuk atribut Rx4 (keterampilan khusus) anggota dari *cluster* ini sebagian mendapatkan nilai yang cukup. Selain itu anggota dari *cluster* ini melalui nilai Ry sebagai atribut pengukuran kesesuaian pekerjaan mendapat nilai yang tinggi dimana hal ini dapat dikatakan bahwa *cluster* pertama ini merupakan *cluster* lulusan yang mendapat pekerjaan yang relevan dengan bidang Teknik Informatika walaupun penguasaan pada keterampilan khususnya kurang.

Sementara itu *cluster* kedua dengan jumlah anggota sebanyak 41 (29.92%) dari total responden memiliki karakteristik dimana mereka memenuhi ke empat unsur capaian pembelajaran lulusan yang ditetapkan oleh program studi Teknik Informatika UMM dan mendapatkan pekerjaan yang relevan dengan apa yang mereka pelajari selama kuliah.

Sebaliknya untuk *cluster* ketiga dengan anggota sebanyak 28 (20.45%) dari total responden ini memiliki karakteristik anggota yang memenuhi capaian pembelajaran lulusan untuk ke 3 unsur (Rx1, Rx2, Rx3) namun unsur keterampilan khususnya (Ry) rendah dan mereka mendapatkan pekerjaan yang kurang relevan dengan bidang teknik Informatika.

Dari hasil ketiga *cluster* tersebut terdapat kecenderungan para lulusan dari jurusan informatika banyak yang mendapat pekerjaan yang sesuai dengan bidangnya walaupun variabel keterampilan khusus yang harusnya dikuasai hanya mendapatkan nilai yang kurang. Disisi lain terdapat sekitar 20% lulusan yang mendapatkan pekerjaan yang kurang relevan dengan apa yang mereka pelajari di bidang informatika juga menjadi faktor yang cukup penting untuk diperhatikan oleh pengelola program studi. Hal ini tentu akan dapat menjadi bahan evaluasi bagi program

studi untuk dapat mengatur rencana pengembangan kurikulum yang menunjang tingkat tercapainya kemampuan mahasiswa pada unsur keterampilan khusus program studi yang bisa mengikuti perkembangan dan kebutuhan dunia kerja yang semakin berkembang pesat. Selain itu hasil *cluster* juga dilakukan pengujian nilai *Silhouette Coefficient* dimana nilai rata-rata yang didapat adalah sebesar 0.2287. Nilai ini menunjukkan bahwa hasil *cluster* yang berhasil dibuat masuk dalam kategori yang baik karena nilai yang dihasilkan dibawah 0.5 [18]

Hasil penelitian ini juga menunjukkan kapabilitas dari K-Means dalam menangani data yang variabelnya lebih banyak dibandingkan penelitian yang dilakukan oleh [18], [19] yang hanya menggunakan 3 variabel sebagai acuan dalam pembentukan *cluster* serta mendukung implementasi yang dilakukan oleh [20] yang menggunakan enam variabel acuan dalam pembentukan *cluster*.

IV. KESIMPULAN

Melalui penelitian ini dapat disimpulkan bahwa K-Means membagi data menjadi tiga *cluster* terhadap data relevansi pekerjaan yang digunakan. Dengan pengelompokan hasil *cluster* dapat dilihat bahwa 29.92% lulusan Teknik Informatika UMM mampu memenuhi capaian pembelajaran lulusan dan mendapatkan pekerjaan yang relevan dengan apa yang mereka pelajari, 49.63% lulusan lainnya juga mendapat pekerjaan yang relevan dengan jurusannya walaupun kurang menguasai keterampilan khusus yang diukur pada capaian pembelajaran lulusan, dan 20.45% lulusan lainnya mendapatkan pekerjaan yang kurang relevan dengan bidang teknik Informatika.

Secara kinerja, algoritma K-Means yang diterapkan pada penelitian ini dengan total 3 *cluster* yang dibentuk memberikan tingkat keberhasilan sebesar 44.9% dimana hal ini mendukung pernyataan [2] dimana pemilihan centroid secara acak merupakan salah satu kelemahan utama dari algoritma ini. Sehingga perlunya penelitian lebih lanjut dan pengembangan versi algoritma K-means.

DAFTAR PUSTAKA

- [1] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation," *Electronics* 2020, Vol. 9, Page 1295, vol. 9, no. 8, p. 1295, Aug. 2020, doi: 10.3390/ELECTRONICS9081295.
- [2] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, pp. 52–61, Oct. 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
- [3] A. Maghawry, Y. Omar, and A. Badr, "Self-organizing map vs initial centroid selection optimization to enhance k-means with genetic algorithm to cluster transcribed broadcast news documents," *International Arab Journal of Information Technology*, vol. 17, no. 3, pp. 316–324, May 2020, doi: 10.34028/iajit/17/3/5.
- [4] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl Inf Syst*, vol. 14, no. 1, pp. 1–37, Dec. 2008, doi: 10.1007/S10115-007-0114-2/METRICS.
- [5] N. Wang, G. Guo, B. Wang, and C. Wang, "Traffic Clustering Algorithm of Urban Data Brain Based on a Hybrid-Augmented Architecture of Quantum Annealing and Brain-Inspired Cognitive Computing," *Tsinghua Sci Technol*, vol. 25, no. 6, pp. 813–825, Dec. 2020, doi: 10.26599/TST.2020.9010007.
- [6] L. ChaoYing, W. X. Da, and Z. E. Hui, "Research on modeling of government debt risk comprehensive evaluation based on multidimensional data mining," *Soft comput*, vol. 26, no. 16, pp. 7493–7500, Aug. 2022, doi: 10.1007/S00500-021-06478-7/METRICS.
- [7] Z. Wang, "Higher Education Management and Student Achievement Assessment Method Based on Clustering Algorithm," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/4703975.
- [8] L. Sun, "Research on Mining Balanced Competition Strategy in Financial Market Based on Computer Data Mining Method," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/6202890.
- [9] M. Andryan, W. Saputra, and S. Harini, "Java Island Health Profile Clustering using K-Means Data Mining," *International Journal on Information and Communication Technology (IJoICT)*, vol. 8, no. 1, pp. 1–9, Jul. 2022, doi: 10.21108/IJOICT.V8I1.606.
- [10] D. Gultom *et al.*, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," (*JurTI*) *Jurnal Teknologi Informasi*, vol. 4, no. 1, pp. 146–151, Jun. 2020, doi: 10.36294/JURTI.V4I1.1263.
- [11] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 356–363. doi: 10.1016/j.procs.2022.12.146.
- [12] Z. Zhang and A. Kusiak, "Monitoring wind turbine vibration based on SCADA data," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 134, no. 2, May 2012, doi: 10.1115/1.4005753/455682.
- [13] D. Kumalasari, A. Bramanto, W. Putra, A. Fanany, and O. Gaffar, "Speech classification using combination virtual center of gravity and k-means clustering based on audio feature extraction," *Jurnal Informatika*, vol. 14, no. 2, pp. 85–94, May 2020, doi: 10.26555/JIFO.V14I2.A17390.
- [14] A. Triayudi, Iksal, and R. Haerani, "Data Mining K-Means Algorithm for Performance Analysis," *J Phys Conf Ser*, vol. 2394, no. 1, p. 012031, Dec. 2022, doi: 10.1088/1742-6596/2394/1/012031.
- [15] R. Asih, D. Alonzo, and T. Loughland, "The critical role of sources of efficacy information in a mandatory teacher professional development program: Evidence from Indonesia's underprivileged region," *Teach Teach Educ*, vol. 118, p. 103824, Oct. 2022, doi: 10.1016/J.TATE.2022.103824.
- [16] C. A. W. Heryanto, C. S. F. Korangbuku, M. I. A. Djeen, and A. Widayati, "Pengembangan dan Validasi Kuesioner untuk Mengukur Penggunaan Internet dan Media Sosial dalam Pelayanan Kefarmasian," *Indonesian Journal of Clinical Pharmacy*, vol. 8, no. 3, pp. 175–187, Sep. 2019, doi: 10.15416/IJCP.2019.8.3.175.
- [17] A. Basri, A. Mubarak, H. Kurniadi Sirajuddin, and S. Do. Abdullah, "Penentuan Jumlah Klaster Terbaik Pada K-Means Dalam Melihat Pola Klasifikasi Data Mahasiswa Yang Telah Lulus | Jurnal Jaringan dan Teknologi Informasi," *Jurnal Jaringan dan Teknologi Informati*. Accessed: Apr. 08, 2024. [Online]. Available: <https://e-journal.unkhair.ac.id/index.php/jati/article/view/60>
- [18] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *Journal of Artificial Intelligence and Technology Information (JAITI)*, vol. 1, no. 1, pp. 20–28, Feb. 2023, doi: 10.58602/jaiti.v1i1.22.
- [19] E. M. Fitri, R. R. Suryono, and A. Wantoro, "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada PT XYZ," *Jurnal Komputasi*, vol. 11, no. 2, pp. 157–168, Nov. 2023, doi: 10.23960/KOMPUTASI.V11I2.12582.
- [20] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *Jurnal Tekno Kompak*, vol. 15, no. 2, pp. 25–36, 2021.