

ANALISIS CLUSTER K-MEANS DALAM PENGELOMPOKAN KEMAMPUAN MAHASISWA

B. Poerwanto dan R.Y. Fa'rifah

Fakultas Teknik Universitas Cokroaminoto Palopo
Email: bobbypoerwanto@uncp.ac.id, 2riskayanu@uncp.ac.id

Abstract. Cluster Analysis, K-Means Algorithm, Student Classification. This study aims to classify students based on learning outcomes for subject the basic of statistics (DDS), which is measured based on attendance, task, midterm (UTS), and final exams (UAS) to further used to evaluate learning for subjects that require analysis of quantitative . This study uses k-means cluster analysis to classify the students into three groups based on learning outcomes. After grouped, there are 3 people in the low category, 27 in the medium category and over 70% in the high category.

Abstrak. Analisis Cluster K-Means dalam Pengelompokan Kemampuan Mahasiswa. Penelitian ini bertujuan untuk mengelompokkan mahasiswa berdasarkan hasil belajar mata kuliah dasar-dasar statistika (DDS) yang diukur berdasarkan variabel nilai kehadiran, tugas, ujian tengah semester (UTS), dan ujian akhir semester (UAS) untuk selanjutnya digunakan untuk mengevaluasi pembelajaran untuk mata kuliah yang membutuhkan kemampuan analisis kuantitatif yang baik. Penelitian ini menggunakan analisis cluster k-means dalam mengelompokkan mahasiswa ke dalam tiga kelompok berdasarkan hasil belajarnya. Setelah dikelompokkan, terdapat 3 orang yang masuk pada kategori rendah, 27 orang pada kategori sedang dan lebih dari 70% pada kategori tinggi.

Kata Kunci: Cluster Analysis, K-Means Algoritma, Klasifikasi Mahasiswa, Universitas Cokroaminoto Palopo

Matematika adalah ilmu yang memicu berkembangnya ilmu lain sehingga matematika disebut juga queen of science. Dalam perkembangannya, bidang teknik, ekonomi, sosial dan sains tidak akan berkembang seperti sekarang ini tanpa bantuan dari perkembangan ilmu matematika, begitupun dengan bidang computer, informatika dan komunikasi yang saat ini berkembang sangat pesat. Jika sebelumnya melakukan percakapan antar Negara adalah hal yang sulit dan mahal, saat ini sudah berubah. Percakapan bahkan bisa dilakukan oleh multi-person dalam waktu yang sama dengan biaya murah serta dapat melihat lawan bicara. Hal ini tentu saja dilatar belakangi oleh kemajuan matematika dan ilmu terapannya.

Matematika dan computer adalah dua bidang ilmu yang tidak dapat dipisahkan. Mempelajari komputer saja akan menyebabkan orang terjebak dalam penguasaan kulit tanpa mengerti akan isinya (Siang, 2009).

Salah satu cabang ilmu matematika yang juga tak kalah penting penerapannya dalam bidang computer adalah statistika. Matematika dan statistika menuntut kemampuan kuantitatif yang baik serta kemampuan berpikir terstruktur sehingga jika telah terbiasa akan memudahkan dalam pembuatan algoritma pada ilmu computer yang juga menuntut kemampuan bekerja secara terstruktur dengan sangat baik.

Mata kuliah statistika merupakan mata kuliah wajib program studi teknik informatika Universitas Cokroaminoto Palopo (UNCP) yang diprogramkan pada semester kedua. Pembelajaran statistika mahasiswa angkatan 2015 masih tergolong rendah, hal ini dibuktikan dari perolehan nilai rata-rata dari 114 mahasiswa yaitu 69,87. Melihat pembelajaran statistika yang belum maksimal, peneliti merasa perlu untuk mengelompokkan kemampuan mahasiswa berdasarkan hal-hal yang menjadi penilaian untuk selanjutnya dilakukan tindakan berdasarkan kelompoknya.

Ada beberapa metode yang dapat digunakan dalam mengelompokkan kemampuan mahasiswa, salah satu yang sering digunakan adalah analisis cluster k-means. Metode cluster k-means mengidentifikasi objek yang memiliki kesamaan karakteristik tertentu, dan kemudian menggunakan karakteristik tersebut sebagai centroid (Nasari, 2015). Metode ini telah banyak digunakan dalam berbagai bidang, seperti yang pernah dilakukan oleh Edmira Rivani (2010) yang mengelompokkan provinsi produksi padi, jagung, kedelai, dan kacang hijau. J. Ong (2013) menggunakan analisis cluster untuk menentukan strategi marketing President University. Selain itu ada juga E. Muningsih (2015) yang menggunakan k-means untuk clustering produk online shop dalam penentuan stok barang. Di luar negeri metode ini pun juga sering digunakan, contohnya Niszczota (2015) yang menerapkan analisis cluster pada kasus turis di Polandia Timur

METODE

Penelitian ini menggunakan data primer yang berasal dari rekam nilai mahasiswa semester dua mata kuliah dasar-dasar statistika, program studi Teknik Informatika FTKOM UNCP tahun akademik 2015/2016. Jumlah mahasiswa yang menjadi obyek penelitian adalah sebanyak 114 mahasiswa. Beberapa variabel independen yang digunakan dalam penelitian pada Tabel 1.

Tabel 1. Variabel Penelitian

| Variabel | Nama Variabel | Tipe Data |
|----------|-----------------------|-----------|
| X_1 | Nilai Kehadiran | Kontinu |
| X_2 | Nilai Rata-rata Tugas | Kontinu |
| X_3 | Nilai UTS | Kontinu |
| X_4 | Nilai UAS | Kontinu |

Penelitian ini akan mengelompokkan mahasiswa berdasarkan nilai mahasiswa pada mata kuliah dasar-dasar statistika dengan menggunakan algoritma K-Means. K-Means adalah salah satu algoritma non hirarki dimana proses clustering-nya berdasarkan jarak terdekat dengan titik pusat yang ditentukan. Salah satu jarak yang sering digunakan adalah Euclidean, yang dapat diperoleh dengan persamaan:

$$d(x_i, x_j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

x_i, x_j adalah dua data yang dihitung jaraknya dan p adalah dimensi dari data yang digunakan. Penentuan titik pusat cluster dapat dilihat dari persamaan di bawah ini :

$$C_{m(q)} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_{i(q)}$$

dimana,

$C_{m(q)}$: pusat kelompok ke-m variabel p

m : 1, 2, ..., k

n_m : Jumlah objek pada kelompok ke-m

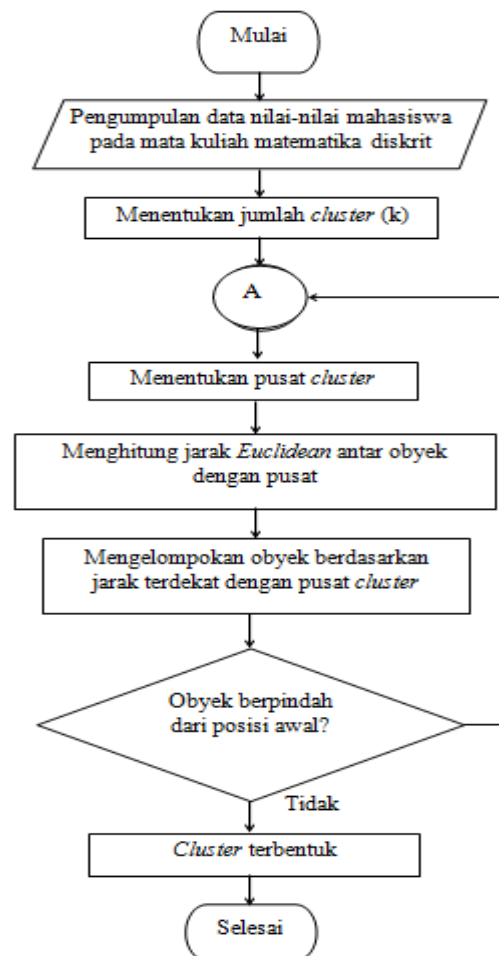
k : Jumlah cluster

q : 1,2, ..., p

x_i : Nilai pengamatan objek ke-i variabel ke-q

i : 1, 2, ..., n_m

Langkah analisis cluster dengan algoritma k-means dalam penelitian ini adalah sebagai berikut :



HASIL DAN PEMBAHASAN

Deskriptif

Mahasiswa TI FTKOM UNCP yang mengambil mata kuliah dasar-dasar statistik (DDS) ini dijadikan sebagai obyek penelitian yang terdiri dari 3 kelas, yaitu kelas 2A, 2C, dan 2D. statistika deskriptif digunakan untuk mengetahui bagaimana gambaran umum tentang penyebaran nilai DDS mahasiswa pada masing-masing. Selain itu juga dapat diketahui gambaran tentang kemampuan mahasiswa dalam menguasai materi yang diperoleh. Statistika Deskriptif nilai DDS untuk masing-masing kelas dapat dilihat pada tabel 1, 2, dan 3. Sedangkan tabel 4 adalah statistika deskriptif untuk melihat secara keseluruhan dari nilai mahasiswa.

Tabel 1. Statistika Deskriptif Nilai DDS Kelas 2A

| | Kehadiran | Tugas | UTS | UAS |
|-----------|-----------|--------|--------|--------|
| N | 35 | 35 | 35 | 35 |
| N missing | 0 | 0 | 0 | 0 |
| Rata-rata | 84.64 | 66.25 | 75.51 | 63.86 |
| Modus | 93.75 | 68.33 | 77.25 | 60 |
| Median | 93.75 | 68.33 | 77.25 | 60 |
| Varians | 416.57 | 163.33 | 194.88 | 567.39 |
| Range | 100 | 82.67 | 90 | 100 |

Berdasarkan tabel 1, rata-rata nilai DDS kelas 2A adalah sebesar 84.64 untuk kehadiran, 66.25 untuk tugas, UTS sebesar 75.51, dan 63.86 untuk UAS. Nilai yang paling banyak diperoleh di kelas 2A untuk kehadiran, tugas, UTS, dan UAS adalah 93.75, 68.33, 77.25, dan 60. Nilai tengah dari kelas adalah 93.75 pada kehadiran, 68.33 untuk tugas, 77.25 untuk UTS, dan UAS adalah 60. Nilai UAS dan kehadiran di kelas 2A lebih beragam dari tugas dan UTS, karena varians yang dihasilkan sangat tinggi, yaitu 567.39 untuk UAS dan kehadiran sebesar 416.57. Nilai tugas lebih homogen bila dibandingkan dari nilai UAS, kehadiran maupun UTS, karena varians yang dihasilkan dari nilai tugas adalah yang paling kecil, yaitu 163.33.

Tabel 2. Statistika Deskriptif Nilai DDS Kelas 2C

| | Kehadiran | Tugas | UTS | UAS |
|-----------|-----------|--------|--------|--------|
| N | 40 | 40 | 40 | 40 |
| N missing | 0 | 0 | 0 | 0 |
| Rata-rata | 91.25 | 67.48 | 76.41 | 65.93 |
| Modus | 100 | 68.67 | 78.75 | 60 |
| Median | 93.75 | 68.33 | 78.5 | 60 |
| Varians | 251.86 | 112.15 | 161.04 | 262.12 |
| Range | 87.5 | 76.04 | 86.75 | 100 |

Tabel 2 menunjukkan bahwa rata-rata tertinggi dari 40 mahasiswa di kelas 2C adalah untuk nilai kehadiran, yaitu 91.25, dilanjut dengan nilai UTS sebesar 76.41, tugas sebesar 68.67, dan terakhir adalah UAS sebesar 65.93. Nilai yang paling banyak diperoleh di kelas 2C adalah 100 untuk kehadiran, tugas adalah 68.67, UTS adalah 78.75, dan UAS 60. Nilai tengah di kelas 2C untuk kehadiran, tugas, UTS, dan UAS masing-masing adalah 93.75, 68.33, 78.50, dan 60. Nilai UAS di kelas 2C lebih beragam dari pada nilai-nilai yang lain, karena varians untuk UAS adalah yang tertinggi, yaitu sebesar 262.12. Selanjutnya adalah kehadiran dengan varians sebesar 251.86, UTS sebesar 161.04, dan nilai mahasiswa untuk tugas adalah yang lebih homogen, karena memiliki varians yang paling kecil, yaitu 112.15.

Tabel 3. Statistika Deskriptif Nilai DDS Kelas 2D

| | Kehadiran | Tugas | UTS | UAS |
|-----------|-----------|--------|--------|--------|
| N | 39 | 39 | 39 | 39 |
| N missing | 0 | 0 | 0 | 0 |
| Rata-rata | 91.99 | 76.84 | 49.62 | 48.72 |
| Modus | 100 | 90 | 50 | 40 |
| Median | 100 | 90 | 50 | 50 |
| Varians | 287.64 | 491.51 | 326.89 | 804.86 |
| Range | 100 | 75 | 80 | 100 |

Rata-rata nilai dari 39 mahasiswa kelas 2D yang tercantum pada tabel 3 untuk kehadiran adalah sebesar 91.99, 76.84 untuk tugas, UTS sebesar 49.62, dan UAS adalah sebesar 48.72. Berdasarkan tabel 3, nilai kehadiran yang paling banyak diperoleh adalah 100, tugas adalah 90, UTS 50, dan UAS 40. Nilai tengah yang diperoleh mahasiswa kelas 2D untuk kehadiran adalah 100, nilai tengah pada tugas adalah 90, UTS dan UAS adalah 50. Keragaman nilai diukur oleh besarnya varians, UAS di kelas 2D adalah yang paling beragam dr kehadiran, tugas, dan UTS, yaitu sebesar 804.86. Sedangkan yang paling homogen adalah nilai kehadiran, yaitu dengan varians sebesar 287.67. Varians untuk tugas adalah sebesar 491.51, dan UTS adalah 326.89.

Tabel 4. Statistika Deskriptif Nilai DDS Mahasiswa TI UNCP

| | Kehadiran | Tugas | UTS | UAS |
|-----------|-----------|-------|-------|-------|
| N | 114 | 114 | 114 | 114 |
| N missing | 0 | 0 | 0 | 0 |
| Rata-rata | 89.47 | 70.30 | 66.97 | 59.41 |
| Modus | 100 | 90 | 75 | 60 |
| Median | 93.75 | 68.33 | 77 | 60 |

| | | | | |
|---------|--------|--------|--------|--------|
| Varians | 319.69 | 275.89 | 382.20 | 592.43 |
| Range | 100 | 75 | 80 | 100 |

Tabel 4 memperlihatkan statistika deskriptif untuk semua kelas, yaitu kelas 2A, 2C, dan 2D. Rata-rata nilai DDS untuk 114 mahasiswa adalah 89.47 untuk kehadiran, 70.30 untuk tugas, UTS sebesar 66.97, dan UAS sebesar 59.41. Nilai yang paling banyak diperoleh untuk seluruh mahasiswa di kelas 2A, 2C, dan 2D yang mengambil mata kuliah DDS adalah 100 untuk kehadiran, 90, untuk nilai tugas, UTS 75, dan UAS adalah 60. Nilai tengah dari keseluruhan yang diperoleh mahasiswa adalah 93.75 untuk kehadiran, 68.33 untuk tugas, UTS sebesar 77, dan UAS adalah 60. Keragaman nilai untuk mata kuliah DDS di kelas 2A, 2C, dan 2D yang paling tinggi adalah nilai UAS, yaitu 592.43, berikutnya nilai UTS yaitu 382.20, kehadiran adalah 319.69, dan tugas sebesar 275.89.

Cluster Nilai Kelas 2A, 2C, dan 2D

Cluster analysis nilai DDS pada mahasiswa Teknik Informatika (TI) UNCP semester II menggunakan algoritma K-means dengan jumlah *cluster* adalah 3. *Cluster 1* adalah untuk kelompok nilai rendah (di bawah rata-rata), *cluster 2* untuk nilai sekitar rata-rata berdasarkan jarak yang digunakan dari *centroid*-nya, dan *cluster 3* adalah untuk kelompok nilai tinggi (di atas rata-rata). Hasil pembentukan kelompok berdasarkan nilai DDS mahasiswa TI kelas 2A, 2C, dan 2D dapat dilihat pada tabel 5, dan 6.

Tabel 5. *Cluster Centroids* Masing-masing Nilai

| Variabel | Cluster 1 | Cluster 2 | Cluster 3 | Grand Centroid |
|-----------|-----------|-----------|-----------|----------------|
| Kehadiran | 4.17 | 91.44 | 91.89 | 89.47 |
| Tugas | 19.13 | 68.92 | 80.29 | 70.30 |
| UTS | 0 | 44.61 | 76.55 | 66.97 |
| UAS | 0 | 31.48 | 70.5 | 59.41 |

Tabel 5 berisi tentang besarnya nilai rata-rata dari variabel yang ada pada setiap *cluster* dari masing-masing nilai, yaitu kehadiran, tugas, UTS, dan UAS. Terlihat bahwa, pada *cluster 1* memiliki nilai yang sangat rendah, karena rata-rata yang dihasilkan sangat kecil, yaitu 4.17 untuk kehadiran, 19.13 untuk tugas, dan 0 untuk UTS dan UAS. Sedangkan untuk *cluster 2* dan 3 lebih tinggi dari *cluster 1*, namun *cluster 3* yang paling tinggi. *Grand centroid* merupakan rata-rata dari masing-masing nilai, dimana 89.47

untuk kehadiran, 70.30 untuk tugas, UTS sebesar 66.97, dan UAS adalah 59.41. Berdasarkan nilai *cluster centroid* yang dihasilkan untuk setiap *cluster* dari masing-masing nilai, maka *cluster 1* merupakan kelompok nilai rendah, *cluster 2* adalah kelompok nilai yang berada di sekitar rata-rata, dan *cluster 3* adalah kelompok nilai mahasiswa yang bernilai tinggi.

Tabel 6. *Cluster* Nilai DDS

| Cluster | Jumlah Mahasiswa | Jarak Cluster dari Centroid |
|-------------------|------------------|-----------------------------|
| Rendah | 3 | 19.75 |
| Sekitar rata-rata | 27 | 28.88 |
| Tinggi | 84 | 18.41 |

Jumlah mahasiswa yang termasuk dalam *cluster 1*, *cluster 2*, dan 3 dapat dilihat pada tabel 6. Tabel 6 menunjukkan bahwa jumlah mahasiswa yang termasuk pada kelompok nilai rendah adalah sebanyak 3 orang, masing-masing 1 orang dari kelas 2A, 2C, dan 2D. Mahasiswa yang memiliki nilai disekitar rata-rata ada sebanyak 27 orang, yaitu 2 orang di kelas 2A dan 25 di kelas 2D. Sedangkan yang termasuk pada kelompok nilai tinggi ada sebanyak 84 mahasiswa, yaitu 32 kelas 2A, 39 kelas 2C, dan 13 di kelas 2D. Berdasarkan pengelompokan berdasarkan nilai kehadiran, tugas, UTS, dan UAS, terlihat bahwa kelas 2C adalah yang paling baik nilainya bila dibanding kelas 2A dan 2D.

PEMBAHASAN

Hasil analisis statistika deskriptif menunjukkan bahwa kelas 2C TI FTKOM UNCP adalah kelas yang mendapatkan rata-rata tertinggi untuk mata kuliah DDS dan kelas 2D adalah yang terendah. Nilai yang diperoleh mahasiswa kelas 2A, 2C, dan 2D sangat beragam, karena varians yang dihasilkan dari ketiga kelas untuk nilai kehadiran, tugas, UTS dan UAS tersebut tinggi. Namun, nilai yang keragamannya paling tinggi dari ketiga kelas tersebut adalah kelas 2D dan yang paling rendah keragamannya adalah kelas 2C. Hal tersebut dapat dilihat dari varians yang dihasilkan di masing-masing kelas berdasarkan tabel 1, 2, dan 3. Hasil analisis *cluster* yang terdiri dari 3 kelompok menunjukkan bahwa *cluster 1* merupakan kelompok mahasiswa dengan nilai rendah ada 3 mahasiswa yang berasal dari kelas

2A, 2C, dan 2D. Berarti dari 114 mahasiswa TI FTKOM UNCP yang mengambil mata kuliah DDS terdapat 2.63% yang memperoleh nilai rendah. Terdapat 27 mahasiswa dari kelas 2A 2 orang dan 25 di kelas 2D yang berada di *cluster* 2, yaitu nilai disekitar rata-rata. Sehingga ada 23.68% mahasiswa TI yang memperoleh nilai DDS yang berada disekitar rata-rata. Sedangkan yang memperoleh nilai tinggi adalah 73.68% atau sebanyak 84 mahasiswa dari 114 mahasiswa TI FTKOM UNCP, dimana 32 kelas 2A, 39 kelas 2C, dan 13 di kelas 2D. Penelitian tentang analisis *cluster* dengan menggunakan algoritma k-means juga pernah dilakukan oleh Poerwanto dan Fa'rifah (2016), yaitu melihat nilai Matematika Diskrit yang diperoleh mahasiswa TI FTKOM UNCP. Hasil penelitian menunjukkan bahwa dari 102 mahasiswa TI FTKOM UNCP terdapat 8 mahasiswa terdiri dari 5 orang di kelas Gabungan 2 dan 3 orang di Gabungan 3 yang memperoleh nilai rendah. Yang memiliki nilai disekitar rata-rata ada 13 mahasiswa yang terdiri dari 2 orang di kelas Gabungan 1, 7 orang di Gabungan 2 dan di Gabungan 3 ada 4 orang. Sedangkan untuk yang nilai tinggi ada 81 mahasiswa terdiri dari 38 orang berada di kelas Gabungan 1, 18 orang di Gabungan 2, dan 25 orang di Gabungan 3. Dengan kata lain mahasiswa TI yang mengambil mata kuliah Matematika Diskrit dengan nilai rendah ada 17.65%, disekitar rata-rata ada 12.75%, dan nilai tinggi ada 79.41%. Jumlah mahasiswa yang mengambil mata kuliah Matematika Diskrit pada *cluster* 1 lebih besar dari DDS. Sedangkan jumlah mahasiswa jumlah mahasiswa yang mengambil mata kuliah DDS pada *cluster* 2 dan 3 lebih banyak dari mahasiswa pada mata kuliah Matematika Diskrit. Akan tetapi secara keseluruhan, pengetahuan dan pemahaman mahasiswa TI FTKOM UNCP tentang analisis kuantitatif termasuk pada kategori baik, karena mahasiswa yang masuk pada kelompok nilai tinggi adalah yang paling banyak dari nilai rendah dan disekitar rata-rata untuk masing-masing mata kuliah, yaitu 81 mahasiswa untuk Matematika Diskrit dan 84 untuk DDS.

SIMPULAN

Berdasarkan hasil analisis yang telah dibahas, diketahui bahwa keragaman nilai mata kuliah DDS mahasiswa yang paling tinggi adalah kelas 2D dan yang paling rendah adalah kelas 2C. Secara umum rata-rata nilai tertinggi yang dilihat dari kehadiran, tugas, UTS, dan UAS adalah kelas 2C, dan yang terendah adalah kelas 2D. Hasil *clustering* dengan menggunakan algoritma k-means memperlihatkan bahwa pema-haman kelas 2C terhadap mata kuliah DDS lebih tinggi jika dibandingkan dengan kelas 2A, dan 2D. Hal ini dapat dilihat bahwa dari 40 mahasiswa TI FTKOM UNCP kelas 2C yang berada di *cluster* 3 (nilai tinggi) ada 39 mahasiswa, dan yang berada di *cluster* 1 (nilai rendah) ada 1 mahasiswa

DAFTAR PUSTAKA

- Borkowska-Niszczoła, M., 2015. Tourism Clusters in Eastern Poland - Analysis of Selected Aspects of the Operation. *Procedia - Social and Behavioral Sciences*, 213, pp.957–964. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1877042815058668>.
- Muningsih, E. & Kiswati, S., 2015. Penerapan Metode K-Means untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang. *Jurnal Bianglala Informatika*, 3(1), pp.10–17.
- Nasari, F., Darma, S. & Informasi, S., 2015. Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru. *Seminar Nasional Teknologi Informasi dan Multimedia 2015*, 10(2), pp.73–78.
- Ong, J.O., 2013. Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *Jurnal Ilmiah Teknik Industri*, 12(1), pp.10–20.
- Rivani, E., 2010. Aplikasi K- Means Cluster Untuk Pengelompokan Provinsi Berdasarkan Produksi Padi , Jagung , Kedelai , Dan Kacang Hijau Tahun 2009. *Mat Stat*, 10(2), pp.122–134.
- Poerwanto, B. & Fa'rifah, R.Y., 2016. Analisis Cluster Menggunakan Algoritma K-Means. *d'ComputarE*, 6(2), pp.62–67