

Implementasi Algoritma *K-Means* untuk Menentukan Persediaan Barang pada *Poultry Shop*

Firman Nurdiyansyah¹, Ismail Akbar²

¹ Universitas Widyagama Malang, Indonesia

² UIN Maulana Malik Ibrahim Malang, Indonesia

Article Info

Article History

Received: 09-09-2021

Revised : 03-11-2021

Accepted: 25-11-2021

Keywords

Data Mining

Clustering

K-Means

Sale

Goods

✉ Corresponding Author

Firman Nurdiyansyah,

Universitas Widyagama Malang,

Tel. +62 822-3344-8777

firmannurdiyansyah7@gmail.com

ABSTRACT

Maintaining inventory so that the goods do not get empty is one of the ways to maintain customer satisfaction. To do this, company management must be able to analyze which items are selling well and which items are not selling well, especially in the sales department. It is not easy to CV. Muria PS because it has a large number of items, so it takes a little computational technique to simplify the problem. The K-Means clustering algorithm was chosen to solve this problem because it can group the products sold and still available into several clusters. Of the three clusters formed, cluster 1 consists of two items, cluster 2 consists of 9 items, and the remaining 25 items are included in cluster 3. From these results, CV management can take advantage of this. Muria PS to increase inventory stock and sales strategy.

PENDAHULUAN

CV. Muria PS merupakan usaha dagang dalam bidang penjualan pakan unggas. Dalam menentukan strategi penjualan, pemilik CV. Muria PS harus mengambil suatu keputusan yang tepat. Agar usaha dagangnya mampu bersaing, maka harus meningkatkan kualitas dari sebuah produk dagang. Yang sering terjadi saat ini suatu perusahaan dalam bidang jasa penjualan memerlukan persediaan dalam mengantisipasi permintaan konsumen. Tanpa adanya persediaan, perusahaan akan dihadapkan pada suatu resiko yang tidak dapat memenuhi permintaan para pelanggannya. Untuk menjaga kepuasan pelanggan CV Muria PS berusaha untuk selalu menyediakan produk yang dibutuhkan pelanggannya. Tetapi, pengelolaan data stok barang masih dilakukan secara manual dan belum ada analisa terhadap data penjualan, sehingga ada barang yang menumpuk karena kurang laku dan ada barang yang kosong.

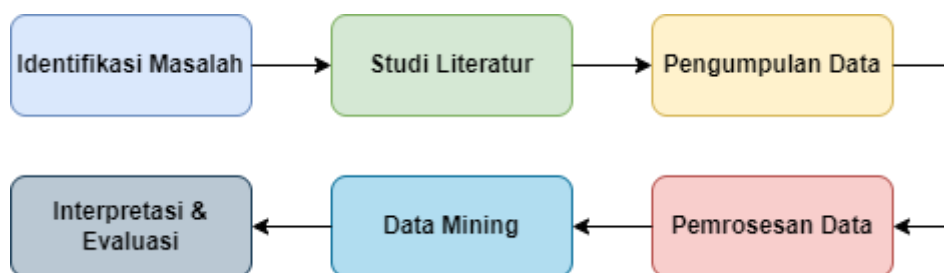
Agar tidak terjadi kesalahan dalam perkiraan penjualan yang berdampak pada hilangnya sejumlah pelanggan, untuk itu diperlukan suatu peramalan yang mendukung pengambilan keputusan. Dalam menangani masalah tersebut maka dibutuhkan suatu manajemen persediaan barang yang baik, yaitu dengan teknik *data mining*. *Data mining* merupakan penggalian data yang tersembunyi dari *database*. *Data mining* merupakan suatu proses menemukan korelasi yang berarti, pola serta kecenderungan dengan memeriksa pada sekumpulan data besar yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika[1]. Salah satu metode dalam teknik data mining untuk mengurangi

permasalahan yang digunakan pada kasus tersebut adalah *clustering*. Metode *clustering* bermanfaat untuk menemukan jumlah *cluster* yang tepat pada data penjualan[2].

K-means merupakan salah satu algoritma yang sering digunakan dalam proses *clustering* yaitu *K-means clustering*. "K" dalam namanya mengacu pada fakta bahwa algoritma mencari sejumlah *cluster* tetap yang ditentukan dalam istilah kedekatan titik data satu sama lain[3]. Algoritma *K-means* telah banyak di terapkan dalam berbagai kasus selain kasus penjualan, semisal permasalahan pada Toko Rizki Barokah sering terjadi penumpukan stok produk yang mengakibatkan produk tersebut sudah kadaluwarsa. Hal ini disebabkan adanya kesalahan dalam pengambilan keputusan tentang stok produk. Dengan banyaknya data penjualan yang tersimpan di *database*, pihak toko belum melakukan *data mining* dan pengelompokan untuk mengetahui potensi produk. Untuk mengatasi kasus tersebut, digunakan teknik *data mining* dengan metode *clustering* menggunakan algoritma *K-means*. Berdasarkan hasil perhitungan dapat diketahui bahwa 6 produk dalam permintaan dan 4 produk kurang diminati. Sehingga metode *clustering* dengan algoritma *K-means* dapat digunakan untuk pengelompokan produk di Toko Rizki Barokah, dalam penentuan stok produk[4]. Untuk itu penelitian ini bertujuan untuk menghasilkan kelompok-kelompok persediaan barang yang ada di CV. Muria PS dengan tiga kategori, yaitu laris, cukup laris dan kurang laris dari setiap *item* barang yang ada.

METODE

Dalam hal ini, metode penelitian menggambarkan langkah-langkah dan prosedur yang harus diikuti untuk memecahkan masalah dalam penelitian. Langkah-langkah penelitian selanjutnya dapat dilihat pada Gambar 1 di bawah ini.



Gambar 1. Tahapan Penelitian

1. Identifikasi Masalah

Pada tahap ini melakukan identifikasi masalah pada *poultry shop* serta merencanakan solusi untuk menyelesaikan permasalahan tersebut. Permasalahan yang ada yaitu pengendalian persediaan barang masih belum terorganisir dengan baik oleh karena itu masalah tersebut akan digunakan sebagai sumber permasalahan pada penelitian ini.

2. Studi Literatur

Langkah ini menemukan referensi teoritis yang relevan dalam hal manajemen persediaan barang dalam jurnal, artikel dan buku. Tujuan dari referensi tersebut akan digunakan sebagai dasar pemikiran pemecahan masalah dalam penelitian ini.

3. Pengumpulan Data

Pengumpulan data merupakan tahapan untuk mendapatkan informasi pada perusahaan sebagai bahan analisis penelitian ini. Data yang diperoleh merupakan data primer yaitu data yang dikumpulkan dan diolah sendiri dari *poultry shop*. Data yang digunakan penelitian ini adalah data laporan penjualan dan data barang.

4. Pemrosesan Data

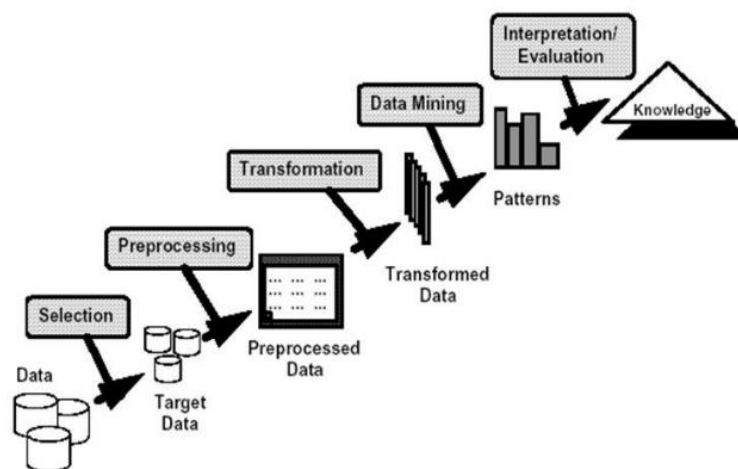
Pada tahap ini, data mengalami perubahan karena datanya masih belum sempurna, seperti data yang tidak lengkap dan tidak konsisten. Tujuan utama dari fase

preprocessing data adalah untuk membersihkan, menambah, mengurangi, dan mengurutkan data sehingga terstruktur sesuai dengan kebutuhan proses *data mining*.

5. Data Mining

Data mining mengolah data mentah yang tersimpan di dalam basis data sehingga menghasilkan informasi yang berguna. Data mentah merupakan data yang disimpan hanya sebagai dokumentasi. Pengolahan data mentah dapat dimanfaatkan untuk pembuatan laporan, mengontrol kebutuhan, menemukan informasi dan pengetahuan baru. *Data mining* menelusuri data pada *database* untuk membangun model dan menggunakannya untuk mengenali pola data lain yang tidak tersimpan dalam basis data[5]. Tujuan utama *data mining* adalah memanfaatkan data suatu basis data dengan mengolahnya sehingga mendapatkan informasi baru yang berguna[6].

Secara sederhana *data mining* biasa dikatakan sebagai proses penyaring atau "menambang" pengetahuan dari sejumlah data yang besar. Istilah lain untuk *data mining* adalah *Knowledge Discovery in Database* (KDD). Walaupun *data mining* sendiri adalah bagian dari tahapan proses KDD seperti yang terlihat pada Gambar 2[2].



Gambar 2. Proses *Knowledge Discovery in Databases* (KDD)

Knowledge Discovery in Databases (KDD) terdiri dari beberapa proses seperti[4]:

- Pahami domain aplikasi Anda untuk mengetahui dan menggali pengetahuan baru dan apa yang menjadi tujuan pengguna Anda.
- Buat kumpulan data target yang mencakup pengambilan data dan fokus pada sub-kumpulan data.
- Pembersihan dan transformasi data meliputi eliminasi *noise*, *outlier*, nilai yang hilang, dan pengurangan dimensi serta pemilihan fitur.
- Penggunaan algoritma *data mining* yang terdiri dari asosiasi, urutan, klasifikasi, dll.
- Interpretasi, evaluasi, dan visualisasi pola untuk melihat apakah ada sesuatu yang baru dan menarik dan dilakukan jika diperlukan.

6. Interpretasi dan Evaluasi

Langkah terakhir adalah memvisualisasikan hasil proses *data mining* dalam bentuk gambar atau tabel dan mengevaluasi apakah hasil *data mining* tersebut dapat digunakan untuk menyelesaikan masalah penelitian ini.

Pada penelitian ini bersifat kuantitatif, yang berarti pada penelitian ini akan dilakukan proses perhitungan dengan menggunakan angka-angka dalam menemukan solusi dari permasalahan yang ada. Prosedur yang digunakan dalam penelitian ini adalah *data mining* dengan menggunakan algoritma *K-means*. *K-Means* merupakan salah satu algoritma dengan

partitional, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. Algoritma *K-Means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan jumlah *cluster* akhir sebagai *output*. Jika algoritma diperlukan untuk menghasilkan *cluster* K maka akan ada K awal dan K akhir[7].

Selain itu juga Algoritma *K-Means* merupakan salah satu metode klasterisasi yang berbasis partisi yaitu metode yang digunakan untuk memisah data menjadi kelompok. Kelompok yang memiliki kemiripan yang sama di masukan dalam kelompok yang sama dan kelompok yang berbeda di masukan dalam kelompok yang berbeda. Untuk memisahkan kelompok tersebut metode *K-Means* membagi data dengan mengelompokkan objek berdasarkan nilai k yang telah ditentukan. Tujuan membagi data adalah untuk mengetahui kelompok produk yang diminati dan kurang diminati[8].

Metode *K-Means* akan memilih pola k sebagai titik awal *centroid* secara acak. Jumlah iterasi untuk mencapai *cluster centroid* akan dipengaruhi oleh calon *cluster centroid* awal secara *random* dimana jika posisi *centroid* baru tidak berubah. Nilai K yang dipilih menjadi pusat awal, akan dihitung dengan menggunakan rumus *Euclidean Distance* yaitu mencari jarak terdekat antara titik *centroid* dengan data/objek. Data yang memiliki jarak pendek atau terdekat dengan *centroid* akan membentuk sebuah *cluster*[9]. Dengan demikian sesuai dengan penelitian yang dilakukan oleh S, Kapil, dkk[8], bisa diujikan pada penelitian ini untuk mengelompokkan data penjualan laris, cukup laris dan kurang laris. Dengan langkah-langkah Algoritma *K-Means* sebagai berikut[10]:

1. Menentukan nilai k sebagai jumlah *cluster* yang ingin di bentuk.
2. Menentukan nilai acak atau *random* untuk pusat *cluster* awal *centroid* sebanyak k, untuk menghitung jarak setiap data *input* terhadap masing-masing *centroid* dengan menggunakan rumus jarak *Euclidean Distance* yaitu:

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2}$$

3. Mengelompokkan setiap data berdasarkan kedekatannya dengan *centroid* atau mencari jarak terkecil.
4. Memperbaharui nilai *centroid* baru, nilai *centroid* baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus yaitu:

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j$$

5. Apabila data setiap *cluster* belum berhenti, lakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

HASIL DAN PEMBAHASAN

Langkah awal pada perhitungan Algoritma *K-means* dilakukan penentuan banyaknya *cluster* dan nilai *centroid* awal, jumlah *cluster* pada penelitian ini ditentukan tiga *cluster* berdasarkan *variable* pengelompokkan data penjualan laris, cukup laris dan kurang laris.

Tabel 1. Nilai Awal *Centroid*

Cluster	Centroid Awal	
	Jumlah Stock	Terjual
1	211,188	209,806
2	40,442	39,329
3	435	430

Dari tiga nilai *centroid* diatas, lalu kita hitung jarak data dengan *centroid* dengan rumus *Euclidean Distance* dan langsung kita kelompokkan data berdasarkan kedekatannya pada Tabel 2.

Tabel 2. Jarak *Centroid* pada Iterasi 1

Nama Barang	JC1	JC2	JC3	C1	C2	C3
124 P	15102	235514	291268	*		
324 K	295544	54267	1535			*
124	0	241281	297078	*		
511 B	128938	112345	168143		*	
144	250090	8826	46988		*	
BP 104 B	229390	11982	67727		*	
K 36	174747	66535	122331		*	
K 202	286704	45423	10401			*
FB 504	297013	55736	68			*
B 201	297495	56218	417			*
AS 100 B	296614	55337	466			*
BSG 101	288383	47104	8698			*
324	297675	56398	598			*
...
524	285909	44631	11169			*
BURAS (B214)	297367	56090	289			*
611B	292161	50884	4917			*
PRORAS	297350	56073	272			*

Pada proses iterasi ke-2, menghitung rata-rata *centroid*nya sehingga mendapat nilai *centroid* baru yaitu C1: 213874.00, 202749.00; C2: 69597.90, 65585.10; C3: 4769.96, 4547.22. kemudian kita ulangi prosesnya mulai dari langkah kedua dengan hasil di Tabel 3.

Tabel 3. Jarak *Centroid* pada Iterasi 2

Nama Barang	JC1	JC2	JC3	C1	C2	C3
124 P	7551	196280	285292	*		
324 K	292556	93485	4444			*
124	7551	202107	291101	*		
511 B	126020	73178	162166		*	
144	247117	48062	41011			*
BP 104 B	226366	27294	61748		*	
K 36	171811	27447	116354		*	
K 202	283702	84630	4440			*
FB 504	294026	94954	5914			*
B 201	294507	95436	6396			*
AS 100 B	293626	94555	5514			*
BSG 101	285392	86321	2720			*
BR I P	275126	76058	12996			*
...
593	293801	94730	5690			*
324	294688	95617	6576			*
524	282922	83852	5192			*
BURAS (B214)	294379	95308	6268			*

611B	289176	90105	1071	*
PRORAS	294363	95292	6251	*

Setelah dihasilkan iterasi kedua maka dihitung kembali rata-rata *centroid* sehingga diperoleh nilai *centroid* baru C1: 213874.00, 202749.00; C2: 73574.44, 69149.44; C3: 5979.92, 5753.83. Dan dikarenakan pada iterasi ke-2 ada anggota *cluster* yang berubah maka dilanjutkan ke iterasi tiga seperti Tabel 4.

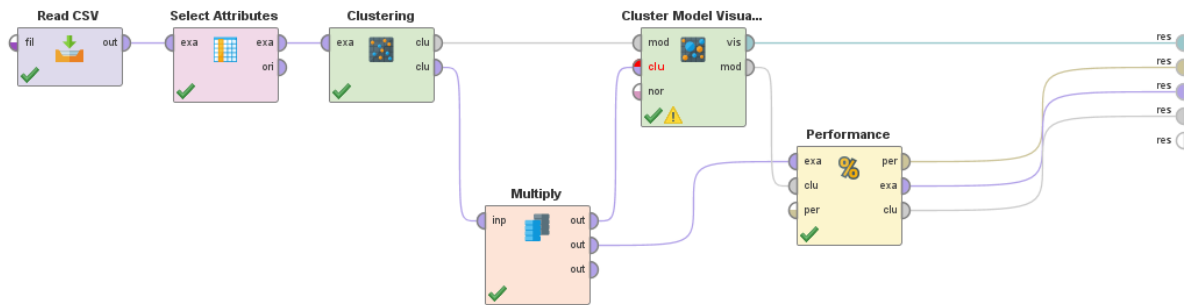
Tabel 4. Jarak *Centroid* pada Iterasi 3

Nama Barang	JC1	JC2	JC3	C1	C2	C3
124 P	7551	190939	283585	*		
324 K	292556	98824	6153			*
124	7551	196778	289392	*		
511 B	126020	67855	160457		*	
144	247117	53402	39302			*
BP 104 B	226366	32632	60040		*	
K 36	171811	22169	114645		*	
K 202	283702	89968	2758			*
FB 504	294026	100293	7622			*
B 201	294507	100775	8104			*
AS 100 B	293626	99893	7223			*
BSG 101	285392	91660	1016			*
BR I P	275126	81397	11287			*
KG P	291463	97729	5073			*
KL 36	181953	11940	104466		*	
BR I SUPER	238296	44576	48114		*	
112	288461	94730	2060			*
122	279396	85662	7049			*
CP 11	287585	93853	1184			*
521	292746	99014	6342			*
594	230647	36969	55788		*	
591	225189	31531	61249		*	
B101	294544	100811	8140			*
K 102	294090	100358	7687			*
511	271754	78022	14650			*
5104 P	287492	93760	1097			*
KL 36 PLUS	165252	28619	121185		*	
324 SCK	179017	23046	109641		*	
PL 1	272776	79049	13638			*
592	292368	98636	5964			*
593	293801	100069	7398			*
324	294688	100955	8284			*
524	282922	89191	3484			*
BURAS (B214)	294379	100647	7976			*
611B	289176	95444	2773			*
PRORAS	294363	100630	7959			*

Pada hasil iterasi tiga, nilai anggota *cluster* sudah tidak ada yang berubah sehingga proses iterasi berakhir pada iterasi tiga dari *clustering* didapatkan hasil sebagai berikut :

1. Kluster pertama memiliki nilai *centroid* (213874.00; 202749.00), yang dapat diartikan sebagai kelompok produk atau barang paling laris (diminati), sehingga jumlah stok barang yang ada harus banyak. Barang yang masuk kluster pertama ini adalah 124 P dan 124.
2. Kluster kedua memiliki *centroid* (73574.44; 69149.44) yang dapat diartikan sebagai kelompok produk yang cukup laris (tingkatan sedang). Barang yang masuk kategori ini sebagai berikut 511 B, BP 104 B, K 36, KL 36, BR I SUPER, 594, 591, KL 36 PLUS, 324 SCK.
3. Kluster tiga memiliki *centroid* (5979.92; 5753.83) yang dapat diartikan sebagai kelompok produk yang kurang diminati (kurang laris). Sehingga toko tidak menyetok barang yang termasuk kategori ini terlalu banyak. Barang yang masuk kategori ini adalah 324 K, 144, K 202, FB 504, B 201, AS 100 B, BSG 101, BR 1 P, KG P, 112, 122, CP 11, 521, B101, K102, 511, 5104 P, PL1.

Setelah melakukan proses hitung manual pada algoritma *K-means*, kemudian dibandingkan dengan menggunakan *tools* Rapidminer.



Gambar 3. Desain Algoritma *K-Means* pada Rapidminer

Dari hasil pengujian dengan Rapidminer menghasilkan rata-rata dari jarak *centroid* sebesar 416152856.758, yang artinya jarak antar tiap *cluster* pada algoritma *K-means* mempunyai nilai tetap tersebut. Sehingga proses penentuan titik tengah dan penempatan data dalam *cluster* diulangi sampai nilai titik tengah dari semua *cluster* yang terbentuk tidak berubah lagi atau sudah mempunyai nilai tetap. Dari hasil tersebut dapat dilihat melalui sebaran nilai *cluster* dalam bentuk diagram *scatter* pada gambar 4.



Gambar 4. Sebaran kluster dari Algoritma *K-means*

Dari sebaran kluster tersebut didapatkan nilai dari *cluster_0* berjumlah 26 barang, yang artinya kelompok barang yang kurang laris, pada *cluster_1* berjumlah 8 barang, yang masuk dalam kategori barang cukup laris. Sedangkan *cluster_2* merupakan kategori yang laris dengan jumlah 2 barang.

SIMPULAN DAN SARAN

Pada penelitian ini yang menerapkan Algoritma *K-means clustering* pada CV. Muria PS dapat menghasilkan kelompok persediaan barang yang laris, cukup laris dan kurang laris dari data jumlah stok barang dan jumlah barang terjual pada setiap *item* barang yang ada. Baik dari hasil perhitungan secara manual maupun menggunakan *tools* Rapidminer didapatkan jumlah barang yang laris terdiri dari 2 barang, yang dimana kategori barang ini harus mempunyai jumlah stok yang banyak. Sedangkan hasil yang sedikit berbeda pada kategori cukup laris, yang dimana hasil manual memiliki 9 barang dan hasil dari Rapidminer 8 barang dan pada kategori kurang laris, hasil manual memiliki 25 barang dan 26 barang pada Rapidminer. Jadi dari hasil tersebut perhitungan secara manual dan menggunakan Rapidminer tidak jauh berbeda. Sehingga hasil dari Algoritma *K-Means* tersebut dapat dijadikan rujukan bagi manajemen untuk mengatur stok persediaan barang agar pelanggan yang membeli produk pada CV. Muria PS tidak mengecewakan. Oleh karena ini, dari penelitian ini dapat dijadikan referensi bagi pihak lain untuk mengatur stok persediaan barang pada usaha yang dimiliki dan dapat juga dikembangkan dengan menambahkan sebuah fitur prediksi pada penjualan barang.

DAFTAR PUSTAKA

- [1] P. P. Putra and A. S. Chan, "Pengembangan Aplikasi Perhitungan Prediksi Stock Motor Menggunakan Algoritma C 4.5 Sebagai Bagian dari Sistem Pengambilan Keputusan (Studi Kasus di Saudara Motor)," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 24, 2018, doi: 10.35314/isi.v3i1.296.
- [2] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [3] K. Kusriani, "Grouping of Retail Items by Using K-Means Clustering," *Procedia Comput. Sci.*, vol. 72, pp. 495–502, 2015, doi: 10.1016/j.procs.2015.12.131.
- [4] M. Imron, U. Hasanah, and B. Humaidi, "Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping," *IJIS Int. J. Informatics Inf. Syst.*, vol. 3, no. 1, pp. 12–22, 2020, doi: 10.47738/ijiis.v3i1.3.
- [5] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innov. Res. Informatics*, vol. 1, no. 2, pp. 63–69, 2019, doi: 10.37058/innovatics.v1i2.918.
- [6] M. R. L. Iin Parlina, Agus Perdana Windarto, Anjar Wanto, "Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center," *Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assessment Cent. Untuk Clust. Progr. Sdp*, vol. 3, no. 1, pp. 87–93, 2018.
- [7] M. G. H. Omran, A. P. Engelbrecht, and A. Salman, "An overview of clustering methods," *Intell. Data Anal.*, vol. 11, no. 6, pp. 583–605, 2007, doi: 10.3233/ida-2007-11602.
- [8] S. Kapil, M. Chawla, and M. D. Ansari, "On K-means data clustering algorithm with

genetic algorithm,” 2016, doi: 10.1109/PDGC.2016.7913145.

- [9] A. Agrawal and H. Gupta, “Global K-Means (GKM) Clustering Algorithm: A Survey,” *Int. J. Comput. Appl.*, vol. 79, no. 2, pp. 20–24, 2013, doi: 10.5120/13713-1472.
- [10] mohamad jajuli nurul rohmawati, sofi defiyanti, “Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa,” *Jitter 2015*, vol. I, no. 2, pp. 62–68, 2015.