

Studi Perbandingan Algoritma Klustering Dalam Pengelompokan Persediaan Produk (Studi Kasus : Subdirektorat Perencanaan Sarana Prasarana Dan Logistik PTN X)

Choirun Nisa¹, Wiyli Yustanti²

^{1,2}Jurusan Teknik Informatika/Sistem Informasi, Universitas Negeri Surabaya

¹choirun.17051214067@mhs.unesa.ac.id

²wilyliyustanti@unesa.ac.id

Abstrak— *Clustering* merupakan metode penganalisaan data yang bertujuan untuk membuat suatu kelompok atau cluster berdasarkan tingkat kemiripan karakter tiap objek. Hasil dari *Clustering* merupakan satu kelas dalam tingkat kemiripan yang tinggi ataupun tingkat kemiripan yang rendah pada antar kelas. Didalam penelitian ini dilakukan studi perbandingan ketiga algoritma *Clustering* yaitu *K-means Clustering*, *K-Medoids Clustering* dan *Agglomerative Hierarchical Clustering* dalam pengelompokan persediaan produk di Subdirektorat Perencanaan Sarana Prasarana dan Logistik pada PTN X yang bertujuan untuk membandingkan ketiga metode *Clustering* dalam mengelompokkan data persediaan produk dengan optimal. Dari pengujian validitas nilai hasil uji validitas metode *Silhouette Coefficient* yang optimal untuk membentuk *cluster* dari data persediaan produk ialah algoritma *K-means Clustering* nilai *index* yang didapatkan sebesar 0,52 sesuai dengan nilai *Silhouette Coefficient* diantara ketiga metode hanya *K-means* yang nilai *index*nya mendekati angka 1 yang dapat diinterpretasikan bahwa nilai *Silhouette Coefficient* untuk Algoritma *K-means* dalam kategori baik. Dari hasil uji yang telah dilakukan mendapati hasil bahwa algoritma *Clustering* yang terbaik ialah *K-means Clustering* maka dari itu akan dilakukannya analisa lanjutan untuk mengevaluasi hasil pengelompokan data menggunakan algoritma *K-means Clustering*. Dari analisa lanjutan yang telah dilakukan metode *K-means Clustering* berhasil membentuk 2 *cluster* yaitu *cluster* produk dalam stok rendah dan produk dalam stok tinggi.

Kata Kunci— Studi Perbandingan, *Clustering*, *Silhouette Coefficient*

I. PENDAHULUAN

Definisi persediaan atau inventory adalah stok produk yang digunakan untuk memenuhi kebutuhan permintaan *customer*.

Persediaan produk pada unit logistik di Perguruan Tinggi Negeri X bertujuan untuk memenuhi kebutuhan terhadap suatu produk atau material yang digunakan untuk menunjang kegiatan atau aktivitas para civitas akademiknya, jika persediaan tidak dikendalikan dengan baik maka akan menghambat seluruh kegiatan maupun aktivitas civitas akademika maka dari itu suatu persediaan produk sangat penting untuk memenuhi permintaan produk serta menunjang kegiatan civitas akademika di Perguruan Tinggi Negeri X.

Pengelompokan atau *Clustering* merupakan metode dalam penganalisaan data yang mana bertujuan untuk membuat kelompok data berdasarkan karakter yang sama ke dalam kelompok atau *cluster*, hasil dari *clustering* merupakan satu kelas dalam tingkat kemiripan yang tinggi ataupun tingkat kemiripan yang rendah pada antar kelas.

Penelitian ini akan menggunakan perbandingan algoritma *K-means Clustering*, *K-Medoids* dan *Agglomerative Hierarchical Clustering* data akan diolah menggunakan Bahasa pemrograman R menggunakan software RStudio. Algoritma *K-Means Clustering* merupakan Metode data *clustering* non-hirarki yang dapat mengelompokkan data ke dalam satu *cluster* atau lebih. Data atau objek akan dikelompokkan berdasarkan karakteristik yang identik dan akan dikelompokkan kedalam satu kelompok, sedangkan data atau objek yang memiliki sifat atau karakteristik berbeda juga akan dikelompokkan kedalam *cluster* lain sehingga objek yang berada dalam satu kelompok akan memiliki tingkat variasi data yang kecil [1], sedangkan Algoritma *K-Medoids* merupakan metode cluster bersifat non hirarki yang memiliki kemiripan dengan algoritma *K-means*. Sementara *K-means* mencoba meminimalkan jumlah kuadrat dalam kluster, *K-Medoids* mencoba meminimalkan jumlah jarak antara setiap titik dan medoid klusternya. Medoid adalah titik data (tidak seperti centroid) yang memiliki jarak total terkecil ke anggota lain dari clusternya. Penggunaan titik data untuk mewakili setiap pusat cluster memungkinkan penggunaan metrik jarak apa pun untuk melakukan pengelompokan data [2]. *Agglomerative Hierarchical Clustering* (AHC) adalah metode hirarki clustering merupakan pengelompokan aglomeratif yang bekerja dengan cara "bottom-up". Artinya, setiap objek pada awalnya dianggap sebagai gugus elemen tunggal. Pada setiap langkah algoritma, dua cluster yang paling mirip digabungkan menjadi cluster baru yang lebih besar. Prosedur ini diulang sampai semua titik menjadi satu kluster tunggal. Setelah mendapatkan hasil *clustering* dari ketiga metode diatas akan dilakukan uji validitas hasil *clustering*, dilakukan perhitungan menggunakan metode *Silhouette Coefficient* untuk mengukur seberapa baik objek atau

data dalam suatu *cluster* [3]. Jika nilai dari perhitungan Silhouette Coefficient berskala -1 hingga 1 [4]. Hasil dari nilai rata-rata Silhouette Index untuk keseluruhan data pada cluster mengartikan ketepatan untuk melakukan pengelompokan data. Jika nilai Index mendekati nilai 1, struktur dari clustering semakin tepat, tetapi jika nilai index -1 maka struktur dari clustering yang dihasilkan overlapping.

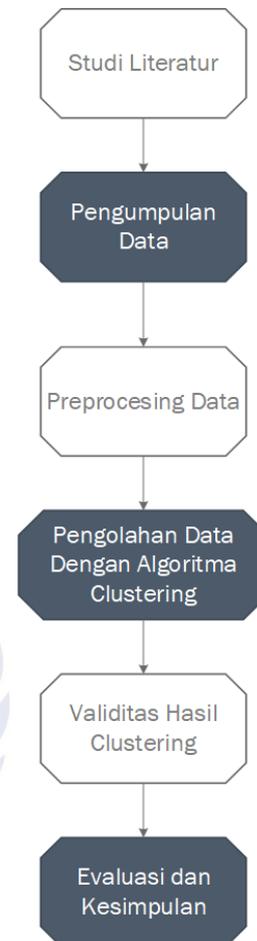
Hasil terbaik dari pengujian ketiga metode ini akan menjadi usulan atau masukkan untuk melakukan pengelompokan persediaan barang ATK dengan baik. Dataset yang digunakan merupakan data stock opname persediaan barang dalam kurun waktu 1 tahun. Atribut yang digunakan untuk pengelompokan data ini adalah kode produk, nama produk, harga produk, stok produk pada bulan 1 sampai bulan 12. Tujuan dari penelitian ini adalah studi perbandingan algoritma klustering *K-means Clustering*, *K-Medoids* dan *Agglomerative Hierarchical Clustering* dan Membentuk kelompok atau *cluster* dari data persediaan produk menggunakan algoritma *K-means Clustering*, *K-Medoids* dan *Agglomerative Hierarchical Clustering* sehingga dapat dijadikan pertimbangan PTN X untuk melakukan pengelolaan persediaan produk yang optimal menggunakan salah satu diantara ketiga metode yang akan dibandingkan.

Beberapa penelitian dengan metode serupa pernah dilakukan sebelumnya diantaranya yaitu Pemanfaatan Metode *K-means* Dalam Penentuan Persediaan Barang [4], *Clustering* Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode *K-means* Fintri Indriyani [5], Klasterisasi Pengendalian Persediaan Aki Menggunakan Metode *K-means* [6], Perbandingan Metode *Clustering* dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap [7], dari penelitian serupa diatas permasalahan pengelompokan pada persediaan barang sudah banyak dilakukan dengan beberapa algoritma yang paling banyak digunakan ialah algoritma *K-means Clustering* maka dari itu peneliti ingin membandingkan ke akurasian metode *K-means* jika dibandingkan dengan metode lain salah satunya menggunakan algoritma *K-Medoids* dan *Agglomerative Hierarchical Clustering*.

Berdasarkan latar belakang, maka topik penelitian ini berjudul “**Studi Perbandingan Algoritma Klustering Dalam Pengelompokan Persediaan Produk (Studi Kasus: Subdirektorat Perencanaan Sarana Prasarana Dan Logistik PTN X)**”. Pengolahan data dengan menggunakan metode *K-means Clustering*, *K-Medoids* dan *Agglomerative Hierarchical Clustering* pada penelitian ini nantinya akan membentuk perbandingan klaster yang dapat menggambarkan pengelompokan data persediaan produk dengan lebih akurat dan optimal sehingga tujuan pada penelitian ini dapat tercapai.

II. METODELOGI PENELITIAN

Studi Perbandingan Algoritma Klustering Dalam Pengelompokan Persediaan Produk dengan menggunakan algoritma *K-means Clustering*, *K-Medoids*, dan *Agglomerative Hierarchical Clustering* memiliki beberapa tahapan dalam penelitian ini yang dapat dilihat pada gambar 3 dibawah ini.



Gbr. 2 Tahapan Penelitian

A. Studi Literatur

Sebelum melakukan pengumpulan data peneliti akan melakukan studi literatur terlebih dahulu yang akan digunakan peneliti untuk menjadi referensi penelitian yang akan dilakukan khususnya tentang Algoritma *Clustering* sumber yang menjadi referensi berasal dari buku, skripsi atau tugas akhir serta jurnal penelitian terdahulu.

B. Pengumpulan Data

Pada penelitian ini peneliti akan mengambil data persediaan produk pada Subdirektorat Perencanaan Sarana Prasarana Dan Logistik di PTN X dalam jangka waktu 1 tahun terakhir.

C. Preprocessing Data

Data yang sudah terkumpul akan diperiksa kembali sebelum dilanjutkan ke proses data mining. Data yang tidak dibutuhkan akan diseleksi kembali untuk menentukan atribut-atribut data

yang terpilih untuk dilakukannya pengolahan data menggunakan algoritma *clustering*.

D. Pengolahan Data dengan Algoritma Clustering

Pada tahap pengolahan data akan dilakukan proses pencarian pola atau sebuah informasi pada data yang telah terpilih dengan menggunakan tiga metode atau algoritma *clustering* yaitu *K-means Clustering*, *K-Medoid* dan *Agglomerative Hierarchical Clustering* dengan menggunakan *Clustering* yang akan dibuat dalam bahasa R menggunakan perangkat lunak RStudio.

E. Validitas Hasil Clustering

Setelah didapati hasil dari *clustering* dari ketiga algoritma *clustering* langkah selanjutnya dilakukan validasi dari tiap-tiap klaster pada setiap metode atau algoritma saat validasi klaster dilakukan akan terlihat besar nilai *clustering* dari masing-masing klaster yang berfungsi untuk membandingkan hasil diantara ketiga metode *clustering*.

F. Evaluasi dan Kesimpulan

Pada tahap terakhir setelah data selesai melalui tahapan validasi hasil peneliti akan mengevaluasi dan menyimpulkan hasil dari klastering dari ketiga metode *clustering* ini dan akan terbentuk sebuah kesimpulan sesuai dengan tujuan peneliti pada penelitian Studi Perbandingan Algoritma Klastering Dalam Pengelompokan Persediaan Produk.

III. HASIL DAN PEMBAHASAN

A. Preprocessing Data

Untuk mendapatkan hasil *clustering* sebelum *processing* data telah disiapkan data stock opname persediaan barang dalam kurun waktu 1 tahun. Atribut pada data ini sebanyak 15 atribut data dan 519 baris data, meliputi kode produk, nama produk, kategori harga, stok produk pada bulan 1 sampai bulan 12, pada tahap *processing* yang digunakan adalah data yang bertipe numerik sebelum dilakukan klasterisasi atau *processing*, data yang sudah ada harus dinormalisasi terlebih dahulu guna normalisasi disini ialah untuk membuat variable-variabel pada data dalam skala yang sama agar memperkecil perbedaan tiap variabelnya, setelah dilakukan proses normalisasi dilakukannya penentuan optimal *cluster* untuk menentukan pusat *cluster* atau jumlah *cluster* yang akan dibentuk. Pusat *cluster* yang dipilih adalah membentuk sebanyak 2 *cluster*. Langkah selanjutnya adalah menentukan jarak dari setiap algoritma yang akan digunakan untuk proses *clustering* setelah dilakukan perhitungan jarak antara data dengan pusat *cluster* maka data-data itu akan dikelompokkan ke dalam *cluster*.

B. Hasil Clustering

Pada tahapan ini pengelompokkan atau *clustering* dilakukan menggunakan 3 algoritma *clustering*. Berikut hasil *clustering* dari setiap algoritma.

TABEL I
 JUMLAH DATA TIAP METODE

Cluster	Algoritma		
	K-means	K-Medoids	AHC
1	413	275	518
2	106	244	1

Pada tabel 1 terlihat hasil dari *clustering* dari tiap-tiap metode berbeda tergantung dari proses *cluster* masing-masing algoritma *clustering*.

C. Hasil Validitas Clustering

Validitas *clustering* diperlukan untuk mengetahui tingkat validasi atau ketepatan dari setiap hasil *clustering* pada setiap algoritma yang telah diperoleh agar diketahui hasil yang terbaik dari masing-masing algoritma. Dibawah ini merupakan hasil dari validitas *clustering* menggunakan metode *Silhouette Coefficient* :

TABEL II
 SILHOUTTE COEFFICIENT INDEX 2 CLUSTER

No	Algoritma	Index Silhoutte
1	K-means	0,52
2	K-Medoids	0,40
3	AHC	0,41

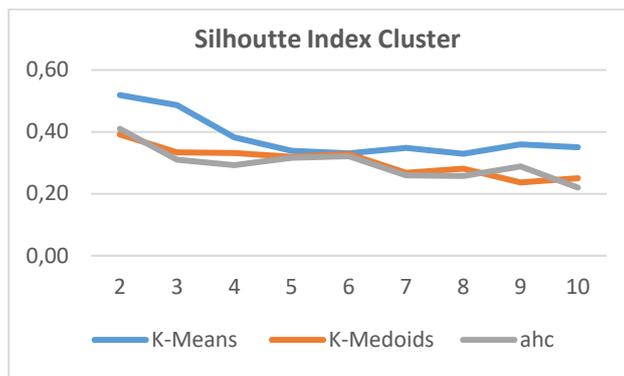
Pada tabel diatas terlihat hasil dari nilai index dari setiap algoritma *Clustering* dari ketiga metode terlihat algoritma *K-means* lebih baik karena memiliki nilai sebesar 0,52 atau mendekati nilai 1 maka dari itu hasil *Clustering* yang terbaik untuk melakukan pengelompokkan pada persediaan produk ialah algoritma *K-means*.

Untuk membuktikan kembali bahwa *K-means* merupakan metode yang baik untuk melakukan pengelompokkan persediaan produk maka dilakukan pengujian indeks *Silhouette* dari ketiga metode dengan membentuk 2-10 *cluster*.

TABEL III
 SILHOUTTE INDEX CLUSTER

Jumlah Klaster	K-means	K-Medoids	AHC
2	0,52	0,40	0,41
3	0,49	0,33	0,31
4	0,38	0,33	0,29
5	0,34	0,32	0,32
6	0,33	0,33	0,32
7	0,35	0,27	0,26
8	0,33	0,28	0,26
9	0,36	0,24	0,29
10	0,35	0,25	0,22

Pada tabel diatas terlihat hasil dari perhitungan nilai *Silhouette Index* dalam setiap pembentukan *cluster* dari setiap metode, dari membentuk 2 sampai 10 *cluster*.



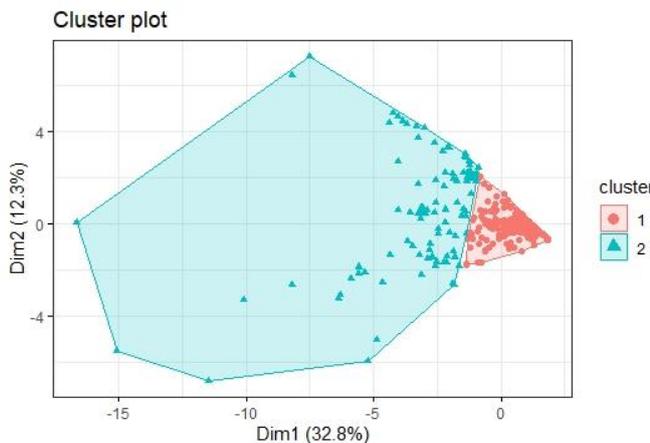
Gbr. 3 Grafik Silhouette Index Cluster

Pada gambar 4 terlihat grafik garis dari hasil silhouette index cluster dari ketiga metode, dilihat dari grafik arti sumbu x merupakan jumlah klaster dan sumbu y adalah nilai silhouette indeks cluster. Dari grafik tersebut metode *K-means* Clustering memiliki nilai lebih tinggi dibandingkan dengan *K-Medoid* Clustering dan *Agglomerative Hierarchical Clustering*. Sehingga hal itu membuktikan bahwa *K-means* Clustering adalah metode terbaik untuk menghasilkan pengelompokan persediaan produk.

D. Evaluasi Hasil K-means Clustering

Dari hasil uji yang telah dilakukan mendapati hasil bahwa algoritma *clustering* yang terbaik ialah *K-means Clustering* maka dari itu akan dilakukannya analisa lanjutan untuk mengevaluasi hasil pengelompokan data menggunakan algoritma *K-means Clustering*.

Pengelompokan persediaan produk menggunakan algoritma *K-means* menghasilkan cluster plot yang dapat dilihat pada gambar 5.



Gbr. 4 Hasil Cluster Plot K-means

Dari gambar 5 dapat dilihat bahwa hasil *cluster plot* membentuk 2 *Cluster* er yang dimana *cluster 1* memiliki anggota sebanyak 413 sedangkan *cluster 2* memiliki anggota sebanyak 106.

Dari hasil *clustering* yang telah dilakukan didapatkan nilai *centroid* pada masing-masing *cluster* sebagai berikut :

TABEL IV
NILAI CENTROID

Cluster	Nilai Centroid
1	6,48
2	62,17

Dari tabel dapat diketahui bahwa *cluster 1* memiliki nilai centroid 6.48 dan *cluster 2* memiliki nilai centroid 62.17.

Dari nilai centroid yang telah diketahui maka dapat dihitung jarak tiap barang terhadap centroid *cluster 1* dan centroid *cluster 2*. Berikut adalah tabel perhitungan jarak centroid.

TABEL V
JARAK CENTROID

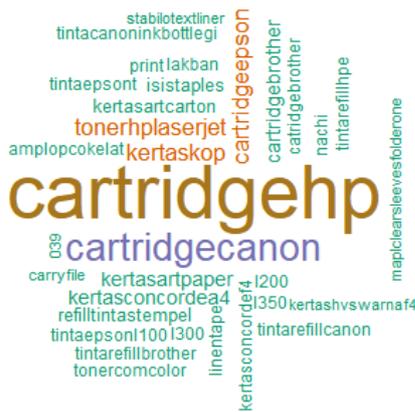
No.	Nama Barang	Jarak Centroid C1	Jarak Centroid C2	Cluster
1	Acrylic Brosur Display A5	2,98	58,67	1
2	Album CD Hitam Medium Veditra	1,85	53,84	1
3	Alunet Plate Cleaner	0,85	54,84	1
4	Amplop Cokelat A3	5,44	50,25	1
5	Amplop Cokelat A5	11,94	43,75	1
6	Amplop Cokelat D J-Plus	32,69	23,00	2
7	Amplop Cokelat Double Folio	28,94	26,75	2
8	Amplop Cokelat Double Folio	23,52	32,17	1
9	Amplop Cokelat Samson Deluxe	30,52	25,17	2
10	Amplop Coklat Samson Folio	18,85	36,84	1
....
510	Toner ComColor X1 Yellow	0,98	56,67	1
511	Toner HP Laserjet 119 A Black	9,27	46,42	1
512	Toner HP Laserjet 119 A Cyan	8,52	47,17	1
513	Toner HP Laserjet 119 A Magenta	8,52	47,17	1
514	Toner HP Laserjet 119 A Yellow	8,52	47,17	1
515	Toner HP Laserjet 307 A Black	0,98	56,67	1

516	Toner HP Laserjet 307 A Cyan	0,98	56,67	1
517	Toner HP Laserjet 307 A Magenta	0,98	56,67	1
518	Toner HP Laserjet 307 A Yellow	0,98	56,67	1
519	Underlay Blanket	5,31	61,00	1

Stok Bulan 10	6,90	247,21	0	120
Stok Bulan 11	2,12	166,61	0	100
Stok Bulan 12	14,60	476,56	0	150

Dari tabel dapat diketahui bahwa kolom jarak centroid C1 adalah jarak barang terhadap nilai centroid *cluster 1* sedangkan kolom jarak centroid C2 adalah jarak barang terhadap nilai centroid *cluster 2*. Pada baris pertama, Acrylic Brosur Display A5 memiliki jarak terhadap centroid C1 sebesar 4.87 dan memiliki jarak terhadap centroid C2 sebesar 58.75583. hal ini menunjukkan bahwa jarak terdekat Acrylic Brosur Display A5 adalah dengan centroid C1 yang berarti bahwa barang tersebut termasuk ke dalam *cluster 1* dan begitu juga seterusnya untuk barang-barang selanjutnya hingga barang pada data ke-519.

Dari uji yang dilakukan dan didapati anggota setiap *cluster*nya selanjutnya dilakukan *word cloud* berbasis nama produk dari anggota tiap-tiap *cluster* yang bertujuan untuk mencari tahu mana produk yang paling dominan pada setiap *cluster*nya.



Gbr. 5 Word Cloud Cluster 1

Dari gambar 5 diatas terbentuk *word cloud* dari *cluster 1* dapat diketahui bahwa produk yang paling dominan pada *cluster 1* adalah Cartridge HP, Cartridge Canon, Cartridge Epson, Toner HP Laserjet, dan Kertas Kop.

TABEL VI
HASIL KARAKTERISTIK CLUSTER 1

Variabel	Mean	St.dv	Min	Max
Stok Bulan 1	23,00	596,20	0	200
Stok Bulan 2	4,05	239,02	0	120
Stok Bulan 3	2,59	162,51	0	60
Stok Bulan 4	8,98	348,02	0	100
Stok Bulan 5	8,21	404,53	0	120
Stok Bulan 6	6,79	363,46	0	150
Stok Bulan 7	9,86	384,51	0	120
Stok Bulan 8	4,68	286,16	0	150
Stok Bulan 9	7,12	412,25	0	200

Pada tabel 6 diatas terlihat karakteristik pada *cluster 1* dilihat dari nilai statistik deskriptif yaitu mean, standar deviasi, nilai minimum dan nilai maximum dari setiap variable dari stok bulan 1 sampai dengan stok bulan 12, fungsi standar deviasi disini ialah untuk menunjukkan seberapa dekat nilai-nilai pada data yang menyebar atau berkumpul di dekat nilai rata-rata.

Dapat disimpulkan bahwa *cluster 1* memiliki karakteristik dengan jumlah rata-rata stok rendah hal ini ditunjukkan dari jumlah keseluruhan nilai rata-rata dari *cluster 1*. Sehingga *cluster 1* dapat dinamakan dengan *cluster* stok rendah.

Pada *cluster 1* yaitu barang stok rendah yang mendominasi adalah Cartridge HP, Cartridge Canon, Cartridge Epson, Toner HP Laserjet, dan Kertas Kop barang tersebut merupakan barang dengan permintaan rendah dan termasuk barang habis pakai dengan jangka waktu yang cukup lama.



Gbr.6 Word cloud cluster 2

Dari gambar 6 diatas terbentuk *word cloud* dari *cluster 2* dapat diketahui bahwa produk yang paling dominan pada *cluster 2* adalah Kertas HVS, Spidol, Amplop, dan Business File

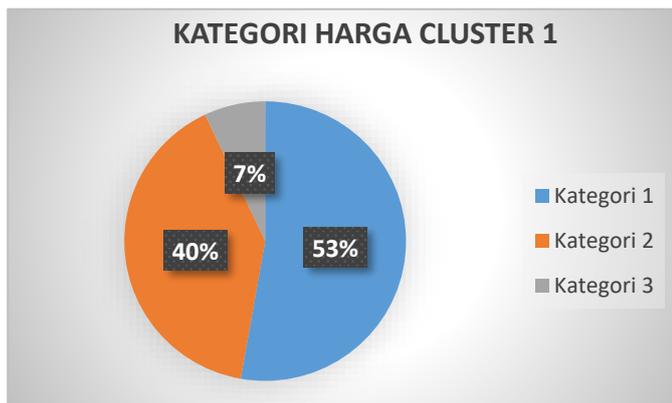
TABEL VII
HASIL KARAKTERISTIK CLUSTER 1

Variabel	Mean	St.dv	Min	Max
Stok Bulan 1	156,00	110,56	0	500
Stok Bulan 2	24,40	52,04	0	240
Stok Bulan 3	32,80	54,01	0	250
Stok Bulan 4	73,40	100,35	0	500
Stok Bulan 5	58,80	94,40	0	456
Stok Bulan 6	26,20	54,73	0	250
Stok Bulan 7	76,00	86,66	0	300
Stok Bulan 8	50,80	85,04	0	323
Stok Bulan 9	53,00	85,64	0	400
Stok Bulan 10	49,00	79,23	0	480
Stok Bulan 11	23,60	50,19	0	314
Stok Bulan 12	122,00	107,86	0	517

Pada tabel 7 diatas terlihat karakteristik pada *cluster 2* dilihat dari nilai statistik deskriptif dari setiap variable dari stok bulan 1 sampai dengan stok bulan 12.

Sehingga dapat disimpulkan bahwa *cluster 2* memiliki karakteristik dengan jumlah rata-rata stok tinggi hal ini ditunjukkan dari jumlah keseluruhan nilai rata-rata dari *cluster 2*. Sehingga *cluster 2* dapat dinamakan dengan *cluster* stok tinggi.

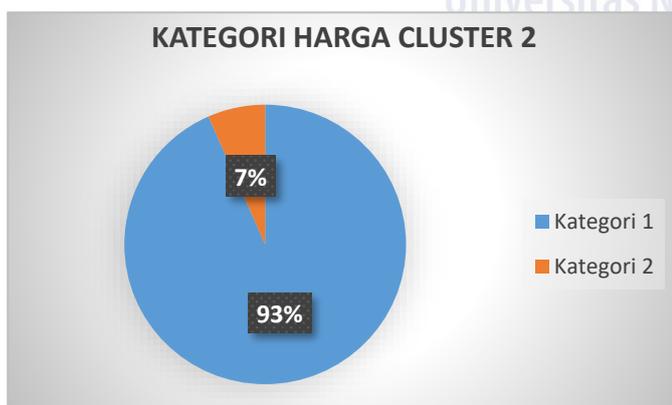
Pada *cluster 2* yaitu barang dengan *cluster* stok tinggi yang mendominasi di *cluster* ini adalah Kertas HVS, Spidol, Amplop, dan Business File barang tersebut merupakan barang dengan permintaan tinggi dan termasuk dalam barang habis pakai dengan jangka waktu yang singkat.



Gbr.7 Presentase Variabel Kategori Harga Cluster 1

Pada gambar 7 diatas terlihat presentase dari variabel data kategori harga pada *cluster 1*, sebelumnya peneliti membuat range kategori harga agar memudahkan dalam proses pengolahan data pada kategori 1 range harga produk mulai dari Rp500–Rp99.000, sedangkan untuk kategori 2 range harga produk dimulai dari Rp102.000–Rp995.000, dan pada pada kategori 3 range harga produk mulai dari Rp1.008.467–Rp3.875.000 .

Pada *cluster 1* kategori harga yang terlihat dominan ialah kategori harga 1.



Gbr.8 Presentase Variabel Kategori Harga Cluster 1

Pada gambar 8 diatas terlihat presentase dari variabel data kategori harga pada *cluster 2*, pada *cluster 2* kategori yang dominan juga berasal dari kategori 1, bila grafik sebelumnya yaitu kategori harga *cluster 1* terdapat 3 kategori harga, namun pada anggota dari *cluster 2* memang tidak mendapati kategori yang bernilai 3.

IV. KESIMPULAN

Dari pengujian yang telah dilakukan dapat disimpulkan bahwa algoritma dari pengelompokan data pada persediaan produk yang hasilnya paling baik ialah algoritma *K-means Clustering* dengan dibuktikan dari hasil uji validitas *K-means* memiliki rata-rata pada *index Silhouette Coefficient* sebesar 0,52 dibandingkan algoritma *K-Medoid* yang hanya memiliki rata-rata pada *index Silhouette Coefficient* sebesar 0,40 dan algoritma *Agglomerative Hierarchical Clustering* sebesar 0,41.

Algoritma *K-means* membentuk 2 *cluster*, *cluster* stok rendah dengan *cluster* stok tinggi setelah melakukan *word cloud* pada setiap *cluster* didapati hasil yang mendominasi pada *cluster 1* yaitu barang dengan stok rendah adalah Cartridge HP, Cartridge Canon, Cartridge Epson, Toner HP Laserjet, dan Kertas Kop dan yang mendominasi pada *cluster 2* yaitu barang dengan stok tinggi adalah Kertas HVS, Spidol, Amplop, dan Business File.

Dilihat dari perhitungan nilai statistik deskriptif pada *cluster 1* dapat disimpulkan bahwa data pada *cluster 1* memiliki variasi yang kecil antar objeknya dikarenakan nilai rata-rata stok yang rendah sebesar 8,24 sedangkan untuk nilai rata-rata pada standart deviasi sebesar 16,78 dan bisa dikatakan variasi pada data cenderung bersifat homogen, sementara untuk *cluster 2* memiliki nilai rata-rata stok tinggi sebesar 62,17 sedangkan untuk nilai rata-rata pada standart deviasi sebesar 80,44 dan variasi data dapat dikatakan cenderung bersifat heterogen.

Dari pengujian yang telah dilakukan di atas algoritma *K-means* sangat disarankan untuk mengelompokkan data pada persediaan produk di PTN X untuk mendapatkan hasil yang optimal.

V. SARAN

Saran yang akan diberikan berdasarkan hasil pada penelitian ini, diharapkan untuk peneliti selanjutnya dapat melakukan studi perbandingan algoritma *K-Means Clustering* dengan metode yang berbeda untuk mengetahui apakah algoritma *K-Means* adalah algoritma yang baik untuk membentuk sebuah kelompok.

REFERENSI

- [1] Agus Nur Khormarudin, Teknik Data Mining: Algoritma *K-means Clustering*, *Jurnal Ilmu Komputer*, 1–12. <https://ilmukomputer.org/category/datamining/>, 2016.
- [2] Pulungan, N., Suhada, S., & Suhendro, D, Penerapan Algoritma *K-Medoids* Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 3(1), 329–334. <https://doi.org/10.30865/komik.v3i1.1609>, 2019.
- [3] Anggara, et al., Pemilihan Distance Measure Pada *K-means Clustering* Untuk Pengelompokan Member Di Alvaro Fitness, *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*. Vol. 1, No. 1, 2016.
- [4] Setiawan, S, Pemanfaatan Metode *K-means* Dalam Penentuan Persediaan Barang, *PIKSEL: Penelitian Ilmu Komputer Sistem*

- Embedded and Logic, 6(1), 41–48.
<https://doi.org/10.33558/piksel.v6i1.1398>, 2018.
- [5] Indriyani, F., & Irfiani, E, *Clustering* Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode *K-means*, JUITA : Jurnal Informatika, 7(2), 109, <https://doi.org/10.30595/juita.v7i2.5529>, 2019
- [6] Andean, R., Fendy, S., & Nugroho, A, Klasterisasi Pengendalian Persediaan Aki Menggunakan Metode *K-means*, JOINTECS (Journal of Information Technology and Computer Science), 4(1), 5. <https://doi.org/10.31328/jointecs.v4i1.998>, 2019.
- [7] Sistem, R, *Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas*, 1(10), 5, 2021.

