

## Penerapan Algoritma K-Means Clustering untuk Menganalisis Penjualan pada Toko Ayu Collection Barbasis Web

Wahyu Tities Pambudi<sup>1</sup>, Arita Witanti<sup>2</sup>

<sup>1,2</sup> Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta (UMBY), Kampus II UMBY, Jl. Jembatan Merah 84C Gejayan, Yogyakarta  
e-mail: <sup>1</sup>wahyutities100@gmail.com, <sup>2</sup>arita@mercubuana-yogya.ac.id

Submitted Date: August 01<sup>st</sup>, 2021

Reviewed Date: August 15<sup>th</sup>, 2021

Revised Date: August 22<sup>nd</sup>, 2021

Accepted Date: October 12<sup>th</sup>, 2021

### Abstract

The main activity in business is to determine the amount of stock that must be maintained to analyze the profit of each item sold. Therefore, groups such as high and low sales categories are needed to consider the stock of goods in the sales process. Ayu Collection store is a store that sells various types of clothing and accessories that have not implemented the grouping of goods in its sales information system to provide the maximum and the minimum number of stock items to be sold at the store. The process of grouping goods is still done manually, which is based on observations from shop owners. Therefore, to maintain the stock of goods so that no items are empty, this study aims to support the process of determining the stock of goods by building a model that can group items into high and low categories in sales using k-means clustering. The group with the highest centroid will be the group with the highest selling rate, while the lowest centroid will be the group with the least demand in sales. The data used in this study was taken from sales data in 2017 and 2018. The clustering scenario uses the variable name of goods, data of incoming goods, data of goods out, and stock of goods. The results of this study are showing the value of system performance in grouping goods by 83.33%.

**Keywords:** Sales Analysis; K-Means Clustering; Website; Data Mining

### Abstrak

Kegiatan utama dalam berbisnis adalah menentukan jumlah stok yang harus dipertahankan untuk menganalisis keuntungan dari setiap item yang terjual. Oleh karena itu, dibutuhkan kelompok seperti kategori tinggi dan rendah penjualan guna mempertimbangkan stok barang dalam proses penjualan. Toko Ayu Collection adalah toko yang menjual berbagai jenis busana dan aksesoris yang belum menerapkan pengelompokan barang dalam sistem informasi penjualannya untuk menyediakan jumlah stok maksimal dan minimum barang yang akan di jual di toko tersebut. Proses dalam mengelompokkan barang masih dilakukan secara manual yaitu berdasarkan pengamatan dari pemilik toko. Oleh karena itu, untuk menjaga persediaan stok barang agar tidak ada barang yang kosong penelitian ini bertujuan untuk mendukung proses dalam menentukan stok barang dengan membangun model yang dapat mengelompokkan item ke dalam kategori tinggi dan rendah dalam penjualan menggunakan k-means clustering. Kelompok centroid yang paling tinggi akan menjadi kelompok yang tingkat penjualannya paling laris, sedangkan centroid yang rendah adalah kelompok yang kurang laris dalam penjualan. Data yang digunakan dalam penelitian ini diambil dari data penjualan tahun 2017 dan 2018. Skenario clustering menggunakan variabel nama barang, data barang masuk, data barang keluar, dan stok barang. Hasil dari penelitian ini adalah menunjukkan nilai unjuk kerja sistem dalam melakukan pengelompokan barang sebesar 83.33%.

**Kata Kunci :** Analisis Penjualan; K-Means Clustering; Website; Data Mining

## 1. Pendahuluan

Teknik Informatika merupakan ilmu yang memperelajari Teknologi Informasi dimana pengimplementasiannya mengarah kepada perkembangan teknologi pada saat ini. Pesatnya kemajuan teknologi memberikan kemudahan salah satunya adalah dalam mendapatkan informasi yang dibutuhkan untuk membantu dalam proses pengambilan suatu keputusan tertentu (Nawang et al., 2017). Informasi didapatkan dari adanya komunikasi antar dua individu atau lebih, sehingga dapat diinterpretasikan dengan adanya komunikasi, informasi akan mudah untuk diperoleh (Hastanti, 2011). Informasi dibutuhkan demi kelangsungan suatu produksi dari perusahaan atau organisasi tertentu (Taufani, Riyadi, & Dewantara, 2016). Dari suatu data maka akan diperoleh suatu informasi yang akan dibutuhkan untuk membantu dalam proses penjualan ataupun promosi pada suatu perusahaan atau toko untuk mencapai suatu tujuan (Sutejo, 2006). Penjualan merupakan kegiatan utama perusahaan dalam menghasilkan pendapatan dan keuntungan. Salah satu cara penjualan dan pemasaran adalah dengan menggunakan sistem berbasis web untuk meningkatkan jumlah keuntungan dimana kondisi tersebut membuat produk menjadi lebih kompleks sehingga memberikan fungsi yang besar untuk membuat pasar yang lebih baik (Hernandhi, Astuti, & Priambada, 2018).

Dalam pemasaran berbasis web, menciptakan interaktif jenis penjualan yang berbeda dan signifikan dari cara pemasaran tradisional. Penjualan dengan berbasis web membentuk persepsi pembeli tentang produk yang akan dibeli, dan bagaimana cara menganalisis barang dan produk mana saja yang paling laris dipasarkan sehingga toko bisa menyiapkan stok lebih banyak barang yang dibutuhkan

Dalam mengimplementasikan proses oprasional terdapat beberapa permasalahan yang sering dihadapi oleh sebuah perusahaan yaitu ketersediaan barang, sama halnya dengan Toko Ayu Collection yang menjual berbagai produk busana dan aksesoris untuk semua kalangan di mana toko ini setiap harinya harus memenuhi kebutuhan konsumen dengan stok yang berbeda-beda. Untuk melakukan hal tersebut, maka dibutuhkan sebuah sistem untuk menganalisis kebutuhan konsumen yang paling banyak dicari. Banyak cara yang dapat digunakan untuk menganalisis hal tersebut yaitu salah satunya dengan Metode K-Means. Metode ini dipilih karena memiliki kelebihan di antaranya mudah

diimplementasikan karena metode yang fleksibel sehingga dapat menentukan jumlah cluster yang akan dibuat dengan relatif cepat (Siregar, 2018)

## 2. Metode Penelitian

Metode yang digunakan adalah dengan menggunakan metode K-Means di mana data minat pelanggan terhadap produk Ayu Collection bisa dianalisis untuk membantu manajemen stok produk dari Ayu Collection.

Produk terbagi menjadi 2 yaitu baju dan aksesoris, karena jenis produk yang beraneka macam maka objek yang telah tergabung dalam kelompok adalah ukuran kedekatan dalam membedakan dengan kelompok lainnya. Pada proses pemilihan produk dilakukan dengan cara menghitung jarak dan jumlah kuadrat perbedaan didalam nilai untuk setiap variabel yang telah dibuat.

### 2.1 Pengumpulan Data

Pada penelitian ini dilakukan beberapa tahap pengumpulan data yaitu dengan wawancara atau dengan cara tanya jawab secara langsung dengan pemilik toko (Dr. Sandu Siyoto, SKM, M.Kes, M. Ali Sodik, 2015). Observasi dilakukan dengan mengumpulkan data secara langsung ke objek yang diteliti (Hasanah, 2017). Selanjutnya untuk mendapatkan data teoritis dilakukan dengan studi pustaka dimana penulis mengumpulkan data dari berbagai referensi tertulis seperti buku, journal dan referensi yang terakait dengan penelitian (Mirzaqon & Purwoko, 2017).

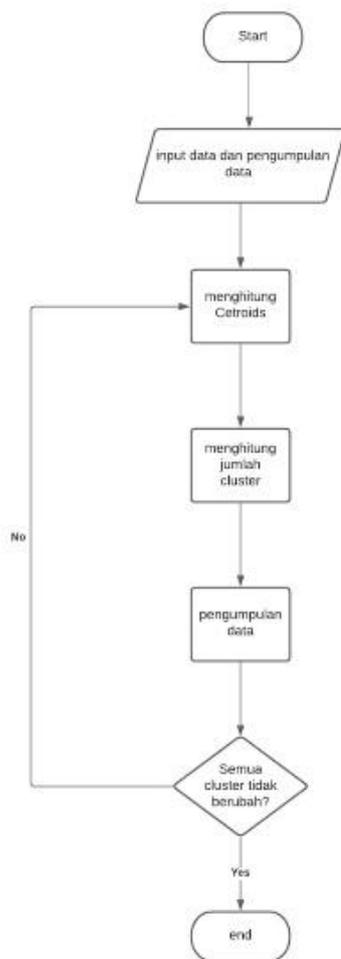
### 2.2 Pre-Processing Data

Setelah semua data terkumpul, lalu dilakukan preprocessing agar data bisa diolah untuk dilanjutkan ke proses selanjutnya pada sistem yang akan dibuat. Pada kasus ini data yang terkumpul pada toko Ayu Collection yang memiliki masing-masing data yang berbeda pada setiap item barang. Integrasi

Data yang diperoleh adalah data dari tahun 2017 sampai dengan 2018 yang digabungkan sehingga menjadi satu data yang sudah terintegrasi selanjutnya dilakukan Cleaning data untuk pembersihan data-data yang kosong dan memastikan data tersebut relevan satu dengan yang lainnya untuk menghasilkan informasi yang diinginkan.

### 2.3. Membuat rancangan sistem

Dalam membuat rancangan sistem bisa dilihat pada gambar 1 di mana terdapat tahapan-tahapan yang dimulai dari pengumpulan data barang yang tinggi penjualannya dan yang tidak, setelah itu akan dihitung rata-ratanya dan masing-masing akan dimasukkan ke dalam label, kemudian data akan dialokasikan kedalam centroid dan dicari dengan nilai rata-rata terdekat.



Gambar 1. Flowchart sistem

## 2.4 K-Means

K-Means adalah salah satu teknik dalam mengelompokan data berdasarkan partitioned clustering(Oktarina, Notodiputro, & Indahwati, 2020). Prinsip dalam pengelompokan hierarchical clustering dilakukan dengan beberapa tahapan dan pada setiap iterasi pengelompokan hierarchical clustering memiliki satu penggabungan dalam suatu item terhadap item lainnya. K-Means dikenal karena kemudahan dan kemampuannya untuk mengelompokan data besar dan outlier dengan

sangat cepat(Devi, Soleman, Pramaita, & Sudarma, 2020)

Algoritma K-Means adalah metode non-hierarki yaitu mengambil sebagian komponen populasi pada tahapan awal untuk dijadikan pusat cluster pertama(Supardi & Kanedi, 2020). Pada tahap ini pusat cluster dipilih secara acak dari sekumpulan populasi data yang ada dan masing-masing dari komponen populasi akan ditandai pada pusat cluster. Posisi pusat cluster akan di hitung kembali sampai semua komponen data digolongkan kedalam tiap-tiap pusat cluster dan terakhir akan terbentuk posisi pusat cluster yang baru.

Algoritma K-Means juga mudah untuk diimplementasikan dan memiliki kompleksitas waktu dan ruang yang relatif kecil. Algoritma ini juga merupakan algoritma yang cukup efisien dalam komputasinya dan memberikan hasil yang cukup baik dan memuaskan jika clusternya compact, hyperspherical dalam shape dan mampu memisahkan fitur-fitur ruangnya dengan baik.

Berdasarkan(Dwitri, Tampubolon, Prayoga, Zer, & Hartama, 2020) Langkah-langkah dasar untuk algoritma K-Means adalah:

1. Menentukan nilai K cluster sesuai dengan yang diinginkan.
2. Pilih titik-titik atau sampel yang menjadi anggota cluster secara acak.
3. Tentukan nilai centroid atau titik tengah dari cluster tersebut dengan rumus:

$$M_k = \left( \frac{1}{n_k} \right) \sum_{i=1}^{n_k} x$$

4. Di mana  $x_{ik}$  = sampel yang ada yang ada dalam cluster  $c_k$
5. Hitunglah *square error* untuk tiap cluster  $c_k$  yang merupakan jumlah kuadrat dari jarak Euclidean antara tiap sampel dalam dan titik tengah centroidnya. Error ini dikenal juga dengan nama within cluster variation (WCV), dengan rumus

$$e_k^2 = \sum_{i=1}^{n_k} (x_{ik} - m_k)$$

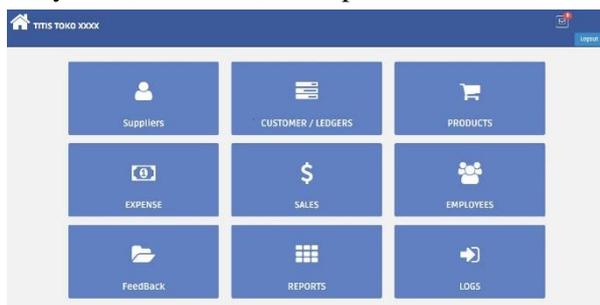
6. Selanjutnya jumlah dari keseluruhan error dari k-cluster juga dihitung dengan rumus

$$E_K^2 = \sum_{k=1}^k e$$

7. Mengelompokkan kembali semua sampel berdasarkan jarak minimum dari masing-masing pusat  $M_1, M_2, \dots, M_k$  Sehingga diperoleh distribusi baru dari sampel sesuai dengan clusternya. Untuk memperoleh distribusi sampel baru tersebut dapat dilakukan dengan menghitung jarak di  $(M_k, x_k)$ . Perhitungan jarak dari masing-masing titik tersebut dapat menggunakan beberapa metode.
8. Tuliskan hasil anggota cluster baru sesuai dengan hasil yang diperoleh pada langkah ke-5. Ulangi langkah ke-3 sampai beberapa iterasi sehingga nantinya ditemukan nilai total square error turun secara signifikan.

### 3. Hasil Dan Pembahasan

Pengujian dilakukan menggunakan beberapa cluster yang berbeda untuk mendapatkan jumlah cluster yang paling optimal dalam penjualan di toko Ayu Collections yang akan di implementasikan pada proses selanjutnya dengan jumlah data sebanyak 2350 yang mana data tersebut merupakan data penjualan produk pada satu periode penjualan yaitu bulan Agustus 2017 sampai dengan Mei 2018. Di mana sistem akan mengelompokkan produk yang tinggi penjualan dengan yang tidak. Pada gambar 2. terdapat tampilan dashboard yang terdiri dari beberapa menu yaitu supplier, customer, produk, biaya, sales, karyawan, komentar, dan laporan



Gambar 2. Halaman Dashboard

Pada gambar 3. merupakan hasil data barang yang tinggi penjualannya untuk satuan minggu



Gambar 3. Tampilan Data Barang laris

Id	kode_barang	tanggal	harga	stock	jumlah	
0	34838	GANG	2017-09	5000	45	1
1	34839	GIS	2017-09	175000	47	1
2	34842	MKCR	2017-09	20000	48	1
3	34845	GNG	2017-09	10000	40	1
4	34849	CUK	2017-09	6000	35	1
5	34851	LSTIK	2017-09	20000	50	1
6	34855	BSS	2017-09	3500	60	1
7	34857	KCRMAN	2017-09	4000	52	1
8	34862	LEKAW	2017-09	2500	47	1
9	34863	KS	2017-09	135000	62	1
10	34864	PDG	2017-09	8750	55	1
11	34865	LEKAW	2017-09	10000	60	1
12	34867	ESHACOM	2017-09	15000	37	1
13	34870	GIS	2017-09	175000	48	1
14	34879	KUNG	2017-09	22000	70	1
15	34880	GANG	2017-09	5000	68	1
16	34881	PFUM	2017-09	10000	45	1
17	34883	BBEL	2017-09	5000	45	1
18	34884	CCIN	2017-09	10000	72	1
19	34885	AWNG	2017-09	8000	50	1
20	34887	ST	2017-09	50000	95	1

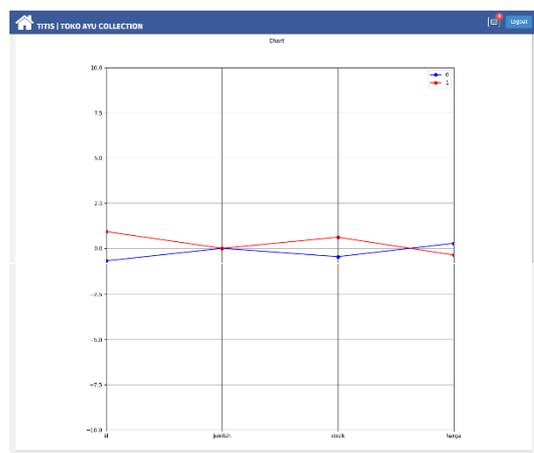
Gambar 4. Tampilan Perhitungan Cluster K-Means

Pada gambar 5. Merupakan tampilan proses data untuk menentukan proses cluster k-means.

kode_barang	1	2	3	
0	0.016	-1.51487	0.03389	0.13676
1	0.016	-1.51487	0.03389	0.13676
2	0.016	-1.51487	0.03389	0.13676
3	0.016	-1.51487	0.03389	0.13676
4	0.016	-1.51487	0.03389	0.13676
5	0.016	-1.51487	0.03389	0.13676
6	0.016	-1.51487	0.03389	0.13676
7	0.016	-1.51487	0.03389	0.13676
8	0.016	-1.51487	0.03389	0.13676
9	0.016	-1.51487	0.03389	0.13676
10	0.016	-1.51487	0.03389	0.13676
11	0.016	-1.51487	0.03389	0.13676
12	0.016	-1.51487	0.03389	0.13676
13	0.016	-1.51487	0.03389	0.13676
14	0.016	-1.51487	0.03389	0.13676
15	0.016	-1.51487	0.03389	0.13676
16	0.016	-1.51487	0.03389	0.13676
17	0.016	-1.51487	0.03389	0.13676
18	0.016	-1.51487	0.03389	0.13676
19	0.016	-1.51487	0.03389	0.13676
20	0.016	-1.51487	0.03389	0.13676

Gambar 5. Data barang yang terjual berdasarkan bulan

Dari hasil data yang telah diolah maka bisa diketahui produk yang memiliki penjualan paling tinggi dan paling rendah berdasarkan hasil presentase jumlah produk yang dapat dilihat pada gambar 6.



Gambar 6. Hasil pengolahan data  
 Tabel 1. Rata-Rata Cluster

Id	Jumlah	Stok	Harga
0.724765	0.0	-0.074560	0.174913
-0.944321	0.0	-0.046467	-0.210877
-0.273765	0.0	12.151882	-0.33166
-0.105747	0.0	-0.152635	3.780555

Tabel 2. Rata-Rata Terdekat

Id	Stok	Harga	Prediksi
0.724765	-0.074560	-0.174913	0
-0.944321	-0.046467	-0.210877	1
-0.273765	12.151882	-0.33166	2
-0.105747	-0.152635	3.780555	3

Tabel 3. Data Training dan Testing

No	Kode barang	Data latih	Data uji
1	AIN	1	1
2	AING	1	1
3	ASAN	0	0
4	ATAN	1	1
5	ATON	0	0
6	B	1	1
7	BAK	0	1
8	BANG	1	1
9	BBL	0	0
10	BDANA	1	1
11	BDHI	1	1
12	BDO	1	0
13	BDUL	1	1
14	BIK	1	1
15	BS	0	1
16	BSS	1	1
17	BU	1	1
18	CANA	1	1

Berdasarkan perbandingan Cluster data didapat 18 barang yang memiliki hasil Cluster yang berbeda, sehingga dapat dihitung unjuk kerja sistem dengan rumus sebagai berikut

$$Uk = ((TD - JDM) / TD) \times 100\%$$

Keterangan : UK (Unjuk kerja sistem)

Td : Total data

JDM : Jumlah data meleset

Hasil dari data tabel menunjukkan unjuk kerja sistem dalam melakukan pengelompokan barang yang tinggi dan rendah dalam penjualan berjalan dengan cukup baik dan diperoleh hasil dengan menggunakan persamaan berikut

$$Uk = ((18 - 3) / 18) \times 100\% = 83.33\%$$

Dari hasil perhitungan diatas, Maka unjuk kerja sistem dalam mengelompokkan data penjualan bisa diimplementasikan dengan cukup baik dengan nilai unjuk kerja sebesar 83.33%

#### 4. Kesimpulan

Hasil dari penelitian ini, penulis mengambil beberapa kesimpulan diantaranya :

1. Implementasi K-Means Clustering berbasis web ini dapat mengelompokkan data barang yang tinggi penjualannya dengan yang tidak.
2. Dengan menggunakan metode ini bisa dilihat produk mana yang paling tinggi dan rendah penjualannya sehingga pemilik toko bisa menyiapkan stok produk yang laris dipasaran dan memasang strategi pemasaran baru untuk barang yang rendah penjualannya
3. Metode K-Means bisa mengelompokkan produk yang tinggi dan rendah penjualannya dengan tingkat unjuk kerja yang cukup tinggi yaitu sebesar 83.33% .

#### Referensi

- Devi, C., Soleman, O., Pramaita, N., & Sudarma, M. (2020). *Classification Of Loyalty Customer Using K-Means Clustering , Studi Case : PT . Sucofindo ( Persero )*. 5(2).
- Dr. Sandu Siyoto, SKM, M.Kes, M. Ali Sodik, M. (2015). *Buku Metode Penelitian Kualitatif dan Kuantitatif*.
- Dwitri, N., Tampubolon, J. A., Prayoga, S., Zer, F. I. R. ., & Hartama, D. (2020). Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia. *JurTI (Jurnal Teknologi Informasi)*, 4(1), 128–132. Retrieved from <http://jurnal.una.ac.id/index.php/jurti/article/view/1266>
- Hasanah, H. (2017). TEKNIK-TEKNIK

- OBSERVASI (Sebuah Alternatif Metode Pengumpulan Data Kualitatif Ilmu-ilmu Sosial). *At-Taqaddum*, 8(1), 21. <https://doi.org/10.21580/at.v8i1.1163>
- Hastanti, R. P. (2011). *Analisis Dan Perancangan Sistem Penjualan Berbasis Web ( E-Commerce )*. 1–8.
- Hernandhi, D. T., Astuti, E. S., & Priambada, S. (2018). Desain Sistem Informasi Pemasaran Berbasis Website Untuk Promosi ( Studi Kasus pada Kedai Ayam Geprak & Sambal Bawang Malang ). *Jurnal Administrasi Bisnis*, 55(1), 1–10.
- Mirzaqon, A. T., & Purwoko, B. (2017). Studi Kepustakaan Mengenai Landasan Teori Dan Praktik Konseling Expressive Writing Library. *Jurnal BK UNESA*, 1–8.
- Nawang, M., Kurniawati, L., Duta, D., Akuntansi, K., Informasi, S., & Akuntansi, K. (2017). *Rancang Bangun Sistem Informasi Pengolahan Data Persediaan Barang Berbasis Dekstop Dengan Model*. 13(2), 233–238.
- Oktarina, C., Notodiputro, K. A., & Indahwati, I. (2020). Comparison of K-Means Clustering Method and K-Medoids on Twitter Data. *Indonesian Journal of Statistics and Its Applications*, 4(1), 189–202.
- <https://doi.org/10.29244/ijsa.v4i1.599>
- Siregar, M. H. (2018). Data Mining Klasterisasi Penjualan Alat-Alat Bangunan Menggunakan Metode K-Means (Studi Kasus Di Toko Adi Bangunan). *Jurnal Teknologi Dan Open Source*, 1(2), 83–91. <https://doi.org/10.36378/jtos.v1i2.24>
- Supardi, R., & Kanedi, I. (2020). Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis. *JurTI (Jurnal Teknologi Informasi)*, 4(2), 270–277. Retrieved from <http://jurnal.una.ac.id/index.php/jurtti/article/view/1444>
- Sutejo, B. S. (2006). Internet Marketing : Konsep Dan. *Jurnal Manajemen*, 6(1), 41–57.
- Taufani, M., Riyadi, R., & Dewantara, R. (2016). Analisis Dan Desain Sistem Informasi Pemasaran (Studi pada Sistem Informasi Pemasaran untuk Promosi CV. Intan Catering). *Jurnal Administrasi Bisnis SI Universitas Brawijaya*, 38(2), 1–10.