

Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris

M Syukron Nawawi¹, Falentino Sembiring², Adhitha Erfina³
Universitas Nusa Putra¹, Universitas Nusa Putra², Universitas Nusa Putra³
e-mail: m.syukron_si17@nusaputra.ac.id

Abstract : During the COVID-19 pandemic, one of them resulted in a community culture for online transactions. The Lide Fashion store, which is the official agent of the Nibras company, gets a fairly high sales demand with various types of products, but the shop owner finds it difficult to determine the best-selling products, in order to optimize inventory and increase sales promotions. This study uses the k-mean clustering method and program implementation using orange data mining software. So as to facilitate and speed up the clustering process in determining which products are the best-selling. The existence of this research can provide solutions to optimize inventory and increase sales promotions. The results of this study there are 3 clusters of 282 products. There are 228 products that are not selling well (c1), 52 products that are selling well (c2) and 2 products that are categorized as selling very well (c3). Thus, these results will assist the company in increasing revenue from sales promotions and goods management that is optimal and measurable.

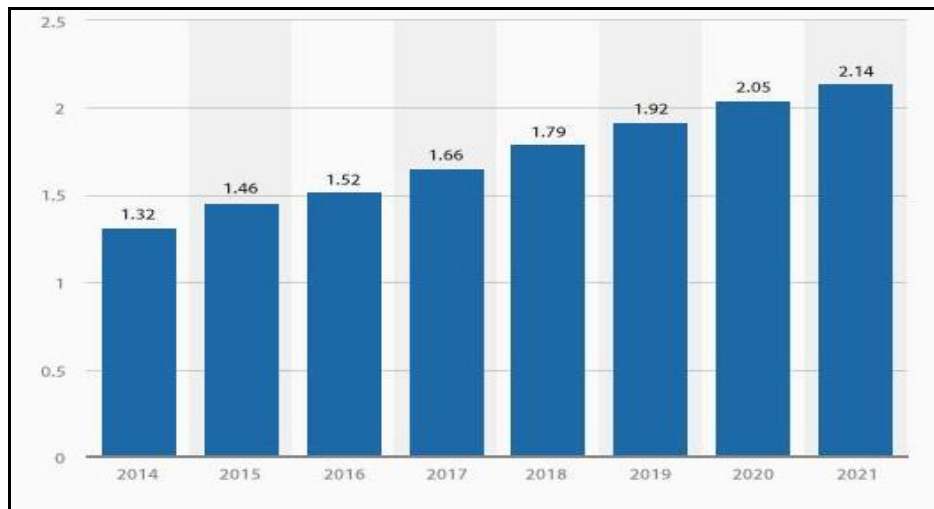
Keywords: K-Means Algorithm Clustering, Orange, Best Selling Products

Abstrak: Saat pandemi covid 19 ini, salah satunya menghasilkan budaya masyarakat untuk transaksi secara digital. Toko Lide Fashion yaitu agen resmi dari perusahaan Nibras mendapatkan permintaan penjualan yang cukup tinggi dengan berbagai jenis produk, namun pemilik toko kesulitan untuk menentukan produk laris, guna mengoptimalkan persediaan barang dan meningkatkan promosi penjualan. Penelitian ini menggunakan metode *k-mean clustering* dan implementasi program menggunakan *software* orange data mining. Sehingga memudahkan dan mempercepat proses *clustering* dalam menentukan produk mana yang paling laris. Adanya penelitian ini bisa memberikan solusi untuk mengoptimalkan persediaan barang dan meningkatkan promosi penjualannya. Hasil dari penelitian ini terdapat 3 *cluster* yaitu, kelompok pertama dengan 228 produk dikategorikan kurang laris (C1), kelompok ke dua dengan 52 produk dikategorikan cukup laris (C2) dan kelompok ke tiga dengan 2 produk dikategorikan sangat laris (C3). Dengan demikian, hasil ini akan membantu perusahaan dalam meningkatkan penghasilan dari promosi penjualan dan pengelolaan barang yang optimal dan terukur.

Kata kunci : Algoritma K-Means Clustering, Orange , Produk Terlaris

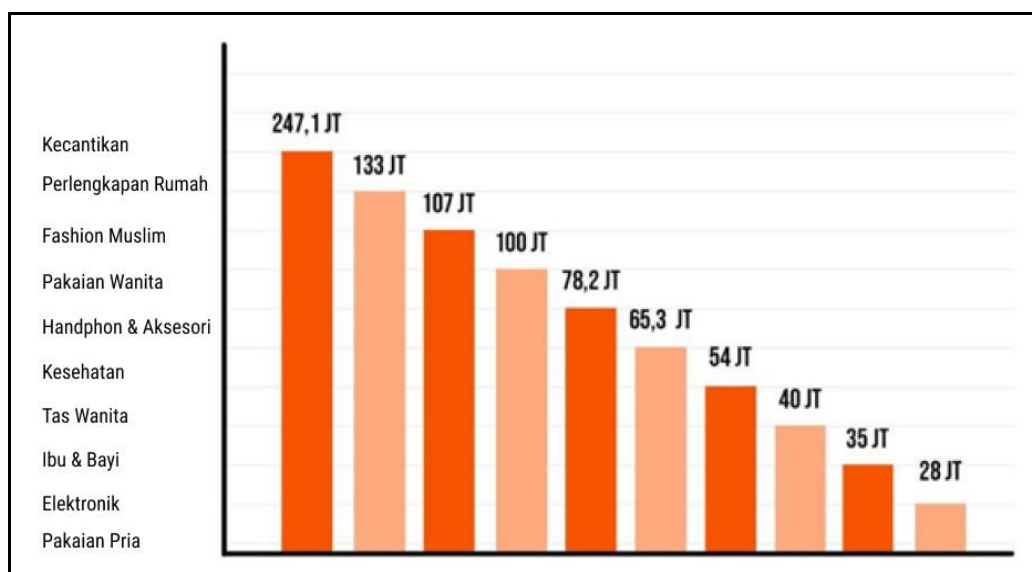
Pendahuluan

Di zaman digital saat ini berbagai masalah kebutuhan masyarakat mudah untuk didapatkan solusinya . Transaksi berbasis digital kini sudah menjadi hal yang sudah biasa di masyarakat . Terlihat dari grafik di bawah jumlah transaksi secara digital seluruh dunia mencapai angka estimasi 2,05 miliar di tahun 2020 dan di prediksi akan melesat diatas 2,14 miliar pada tahun 2021 (Fawkes, 2021).



Gambar 1. Pembelian secara digital seluruh dunia 2021

Tim Asosiasi Digital Marketing Indonesia dalam sebuah website resminya telah melakukan riset pada 2,3 juta produk yang di jual di salahsatu *e-commerce* yaitu shopee . Terdapat 10 kategori dengan penjualan terbanyak, dan kategori fashion muslim menduduki peringkat tiga dengan total 133 Juta penjualan setelah kategori perlengkapan rumah di posisi kedua (Asosiasi Digital Marketing Indonesia, 2020).



Gambar 2. 10 Kategori Produk dengan Penjualan Terbanyak

Dengan kondisi seperti di atas, salah satu agen resmi dari produk busana muslim Nibras, yaitu toko lide fashion mendapatkan peningkatan penjualan yang cukup tinggi. Namun hal tersebut cukup menyulitkan pada saat pemilik toko ingin mengetahui produk mana yang dikategorikan laris dan tidak laris demi mengoptimalkan stok barang dan meningkatkan promosi penjualan di bulan selanjutnya. Dengan demikian diperlukan suatu sistem yang dapat mendukung pemilik toko untuk pengambilan keputusan secara cepat dan tepat .

Algoritma *K-Means* adalah salah satu metode *cluster analysis* non hirarki yang mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya. Sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain (Agus Nur Khormarudin, 2016). Algoritma *k-means* juga sudah banyak digunakan untuk sebuah metode penelitian di berbagai bidang. Seperti bidang marketing (Muslih, 2015), bidang pendidikan (Yunita, 2018), bidang perbankan (Iin Parlina, Agus Perdana Windarto, Anjar Wanto, 2018), dan bidang-bidang lainnya.

Terdapat penelitian sebelumnya yang sejalan dengan penelitian ini yaitu Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Indriyani & Irfiani, 2019). Hasil yang diperoleh didapatkan tiga cluster dimana terdapat 2 jenis barang paling laris, 8 jenis barang yang cukup laris dan 18 jenis barang yang kurang laris. Hasil ini bisa dimanfaatkan oleh manajemen Genta Corp untuk peningkatan manajemen stok dan strategi penjualannya.

Pada penelitian selanjutnya berkaitan dengan *output* pada penelitian ini yaitu penelitian dengan judul penerapan metode *first expired first out* (FEFO) pada sistem informasi gudang (Sembiring et al., 2019), dengan hasil penelitian yaitu mengantisipasi stok obat yang kosong karena terdapat batas minimal obat tersedia.

Saut Parsaoran, Felix Toknady Kesuma dan peryanto menyusun penelitian dengan judul penerapan data mining untuk menentukan penjualan sparepart toyota dengan metode *k-means clustering* (., Kesuma, & Tamba, 2020). Hasil yang diperoleh yaitu barang paling laris sebanyak 15, barang yang laris sebanyak 45 dan kurang laris sebanyak 13. Dengan adanya pengolahan data yang dilakukan dapat memberikan solusi kepada pihak perusahaan agar dapat mengetahui mana barang yang paling laris, laris dan mana barang yang tidak laris.

Perbedaan dari penelitian-penelitian yang telah diuraikan diatas yaitu implementasi algoritma *k-means* menggunakan *visual programming* yaitu orange data mining untuk *clustering* produk laris dengan data penjualan dan stok awal pada 6 bulan di tahun 2021. Penelitian ini bertujuan untuk mengimplementasikan algoritma *k-means clustering* terhadap data penjualan busana muslim selama 6 bulan. Hasil akhir pada penelitian ini akan digunakan pemilik toko dalam menentukan produk mana yang laris terjual dan yang kurang laris terjual, sehingga akan membantu pemilik toko dalam proses penentuan keputusan penyediaan barang dan peningkatan promosi penjualan.

Kajian Teori

Algoritma *K-Means Clustering*

Algoritma *K-means* adalah salah satu metode clustering yang paling populer. dimana algoritmanya menunjukkan prosedur pengelompokan *K-means*. Ide dasarnya adalah diberikan pengelompokan awal tetapi tidak optimal, pindahkan setiap titik ke pusat terdekat yang baru, perbarui pusat pengelompokan dengan menghitung rata-rata titik anggota, kemudian mengulangi proses relokasi dan pemutakhiran hingga kriteria konvergensi (seperti jumlah iterasi yang telah ditentukan, perbedaan pada nilai fungsi distorsi) terpenuhi (Jin & Han, 2010).

Algoritma *K-means clustering* juga salah satu metode cluster analysis non hirarki yang akan mempartisi objek yang ada kedalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan kedalam *cluster* yang lain (Wakhidah, 2010) .

Orange Data Mining

Orange data mining adalah paket perangkat lunak pemrograman visual berbasis komponen untuk visualisasi data *machine learning*, *data mining*, dan analisis data. Pemrograman visual diimplementasikan melalui antarmuka di mana alur kerja dibuat dengan menghubungkan widget yang telah ditentukan sebelumnya atau yang dirancang pengguna (wikipedia, n.d.). Orange menyediakan lingkungan serbaguna untuk pengembang, peneliti, dan data praktisi pertambangan. Berkat Python, bahasa skrip generasi baru dan lingkungan pemrograman, skrip penambangan data yang mungkin sederhana tetapi hasilnya sangat berkualitas (Demšar, Zupan, Leban, & Curk, 2004).

Metode

Metodologi Tinjauan Pustaka

Metode tinjauan pustaka yang digunakan pada penelitian ini menggunakan *systematic literature review* (Kitchenham, 2004), dengan menggunakan alur proses penentuan PICOC (*population, intervention, comparison, outcomes*), dan *context* beserta (RQ) *research question* mengikuti *style* (Wahono, 2015).

Metodologi Penelitian

Metode pada penelitian ini menggunakan algoritma *K-Means*, dimana algoritma *K-Means* melakukan clusterisasi pada data yang di uji. Tahap dalam pengelompokan menggunakan algoritma *K-Means* menurut Nur wakhidah (Wakhidah, 2010) adalah sebagai berikut :

1. Tentukan jumlah *k* (*cluster*) dan *centroid* awal secara acak
2. Penentuan jarak data dan pusat *cluster* dihitung menggunakan *Euclidean Distance* yang di rumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2}$$

Dimana :

$D(i,j)$ = Jarak data ke *i* ke pusat *cluster j*

Xki = Data ke *i* pada atribut data ke *k*

Xkj = Titik pusat ke *j* pada atribut ke *k*

3. Kelompoka data kedalam *cluster* yang dengan jarak paling dekat
4. Hitung Pusat *cluster* yang baru dengan mencari nilai rata-rata pada kelompok data
5. Pusat *cluster* ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat
6. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* di ulangi sampai nilai *centroid* tidak berubah lagi

Hasil

Hasil penelitian yang di dapatkan sebagai berikut :

1. Kelompok pertama, terdiri atas produk kurang laris berjumlah 228 produk, produk cukup laris berjumlah 0 produk dan produk yang dikategorikan sangat laris yang berjumlah 0 produk. Kelompok tersebut adalah produk yang dikategorikan " KURANG LARIS " sebab nilai terbesar berada pada kelompok sangat laris. Produk-produknya adalah produk dengan nomor ,1-109,115,118-132,135,138-143,145-151,154-160,162-170,186,189,198,202-203,dan 215-282.

2. Kelompok kedua, terdiri atas produk kurang laris berjumlah 0 produk, produk cukup laris berjumlah 52 produk dan produk yang dikategorikan sangat laris yang berjumlah 0 produk. Kelompok tersebut adalah produk yang dikategorikan " CUKUP LARIS " sebab nilai

terbesar berada pada kelompok cukup laris. Produk-produknya adalah produk dengan nomor 110-114, 116,117,133 -137,144,152,153,161, 171-178,180-185,187,188,190-197,199-201 dan 205- 214.

3. Kelompok ketiga, terdiri atas produk kurang laris berjumlah 0 produk, produk cukup laris berjumlah 0 produk dan produk yang dikategorikan sangat laris yang berjumlah 2 produk. Kelompok tersebut adalah produk yang dikategorikan " SANGAT LARIS " sebab nilai terbesar berada pada kelompok cukup laris. Produk-produknya adalah produk dengan nomor 179 dan 204.

Pembahasan

Pada penelitian ini data yang di proses dalam pengolahan data menggunakan *software* orange data mining meliputi tahapan – tahapan yang diuraikan sebagai berikut :

Tahap Pengumpulan Data

Dalam penelitian ini data yang diolah yaitu data penjualan dari data inventori toko Lide Fashion selama 6 bulan.

Cleaning

Dari data inventori penjualan selama 6 bulan hanya di ambil data kategori, nama barang, warna, stok barang dan jumlah penjualan. Parameter tersebut yang akan digunakan sebagai meta data dan target. Berikut data hasil proses *cleaning* :

Tabel 1. Data Penjualan Selama 6 Bulan

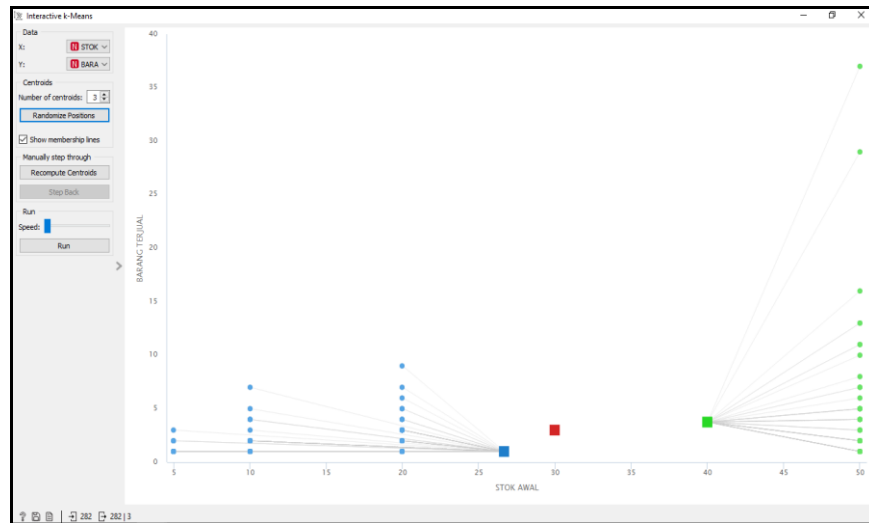
No	Kategori	Nama Barang	Warna	Stok Awal	Barang Terjual
1	Koko Dewasa	Carina	Burgundy	20	3
2	Koko Dewasa	Nk 071	Maroon	10	1
3	Koko Dewasa	Aurora	Purple	20	1
4	Koko Dewasa	Copdesign	Black	10	2
5	Gamis Dewasa	Carina	Burgundy	20	3
6	Gamis Dewasa	Gm 096	Mint	10	2
7	Gamis Dewasa	Gm 096	Plumeria	10	2
8	Gamis Dewasa	Nb A61	Brown	10	1
...
278	Celana Dewasa	Ip 004	Grey	10	2
279	Gamis Anak	Deanna	Cinnamon	20	1
280	Gamis Anak	Bianca	Yellow	20	1
281	Outer	Grasiella	Navy	5	1
282	Baju Olahraga	Asri 017	Peach	5	1

Tahap K-Means Clustering

Berdasarkan jumlah data yaitu sebanyak 282 yang sudah tercantum sebelumnya maka tahapan pada proses *K-Means Clustering* yaitu sebagai berikut :

1. Jumlah k (*cluster*) yang di tentukan secara acak yaitu 3 *cluster* dengan atribut dalam hal ini kurang laris,cukup laris dan sangat laris.

2. Penentuan centroid awal di lakukan secara acak. Data yang digunakan adalah data yang secara otomatis diacak oleh mesin pada aplikasi orange. Seperti terlihat pada gambar dibawah ini :

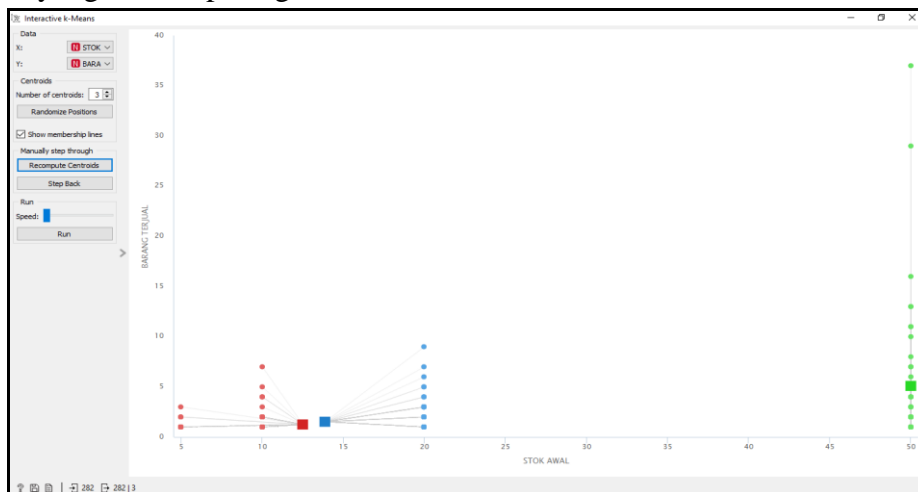


Gambar 3. Penentuan Centroid Awal

3. Penentuan jarak data dan pusat *cluster* dihitung menggunakan mesin orange dan menggunakan *Euclidean Distance* yang di rumuskan sebagai berikut :

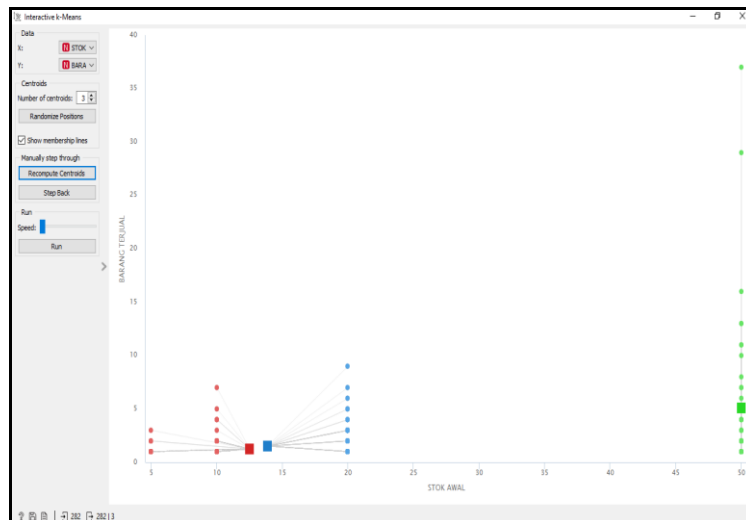
$$D(i,j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + .. + (Xki - Xkj)^2}$$

4. Kemudian dilakukan pengelompokan data kedalam *cluster* yang dengan jarak paling dekat.seperti yang terlihat pada gambar di bawah ini :



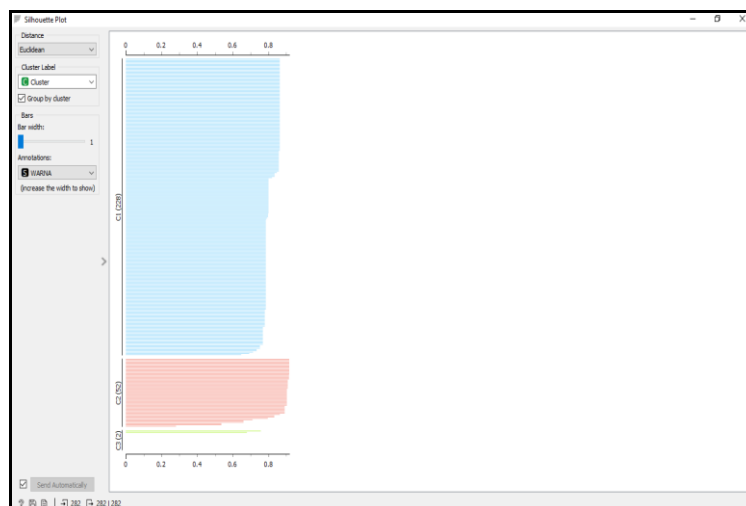
Gambar 4. Pengelompokan Data

5. Tahap selanjutnya yaitu penentuan pusat *cluster* , yang mana pada tahap ini hasil yang terjadi yaitu posisi pusat *cluster* dalam posisi yang sama, begitupuan pada saat dilakukan pengelompokannya. Seperti yang terlihat pada gambar berikut :



Gambar 5 Hasil Penentuan *Cluster* Baru Beserta Pengelompokannya

6. Dengan berakhirnya proses iterasi yang di sebabkan karna sudah tidak ada lagi nilai yang berubah, maka proses *K-Means Clustering* dinyatakan selesai dan didapatkan hasil sebagai berikut :



Gambar 6. Hasil Akhir Proses *K-Means Clustering*

Kesimpulan

Dari hasil penelitian yang di uraikan diatas maka dapat disimpulkan sebagai berikut :

1. Kelompok pertama, terdiri atas produk kurang laris berjumlah 228 produk, produk cukup laris berjumlah 0 produk dan produk yang dikategorikan sangat laris yang berjumlah 0 produk. Kelompok tersebut adalah produk yang dikategorikan “ KURANG LARIS “ sebab nilai terbesar berada pada kelompok sangat laris. Produk-produknya adalah produk dengan nomor,1-109,115,118-132,135,138-143,145-151,154-160,162-170,186,189,198,202-203,dan 215- 282.

2. Kelompok kedua, terdiri atas produk kurang laris berjumlah 0 produk, produk cukup laris berjumlah 52 produk dan produk yang dikategorikan sangat laris yang berjumlah 0 produk. Kelompok tersebut adalah produk yang dikategorikan “ CUKUP LARIS “ sebab nilai terbesar berada pada kelompok cukup laris. Produk-produknya adalah produk dengan nomor

110-114, 116,117,133-137,144,152,153,161, 171-178,180-185,187,188,190-197,199-201 dan 205-214.

3. Kelompok ketiga, terdiri atas produk kurang laris berjumlah 0 produk, produk cukup laris berjumlah 0 produk dan produk yang dikategorikan sangat laris yang berjumlah 2 produk. Kelompok tersebut adalah produk yang dikategorikan "SANGAT LARIS" sebab nilai terbesar berada pada kelompok cukup laris. Produk-produknya adalah produk dengan nomor 179 dan 204.

Dengan demikian, hasil ini akan membantu perusahaan dalam meningkatkan penghasilan dari promosi penjualan dan pengelolaan persediaan barang yang optimal dan terukur. Berikut disajikan grafik hasil clusterisasi produk kurang laris, cukup laris dan sangat laris pada penjualan busana muslim toko Lide Fashion selama 6 bulan.

Daftar Pustaka

- . F., Kesuma, F. T., & Tamba, S. P. (2020). Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 67–72. <https://doi.org/10.34012/jusikom.v2i2.376>
- Agus Nur Khormarudin. (2016). Teknik Data Mining : Algoritma K-Means Clustering. *IlmuKomputer.Com*, 1–12. Retrieved from <https://ilmukomputer.org/category/datamining/>
- "Asosiasi Digital Marketing Indonesia". (2020). 10 Kategori Produk Terlaris Shopee | Asosiasi Digital Marketing. Retrieved June 12, 2021, from digimind.id website: <https://digimind.id/10-kategori-produk-terlaris-shopee/>
- Demšar, J., Zupan, B., Leban, G., & Curk, T. (2004). Orange: From experimental machine learning to interactive data mining. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3202(August 2014), 537–539. https://doi.org/10.1007/978-3-540-30116-5_58
- Fawkes, G. (2021). Tren Internet 2021. Statistik & Fakta di Amerika Serikat dan Seluruh Dunia. Retrieved June 12, 2020, from id.vpnmentor.com website: <https://id.vpnmentor.com/blog/trend-internet-vital/>
- "In Parlina, Agus Perdana Windarto, Anjar Wanto, M. R. L. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assessment Center. *Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assessment Center Untuk Clustering Program Sdp*, 3(1), 87–93.
- "Indriyani, F., & Irfiani, E. (2019). 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>
- "Jin, X., & Han, J. (2010). K-Means Clustering. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 563–564). https://doi.org/10.1007/978-0-387-30164-8_425
- "Kitchenham, B. (2004). Procedures for Performing Systematic Reviews Kitchenham, B., 2004. *Keele, UK, Keele University*, 33(2004), 1–26. "
- Muslih, M. (2015). Pengembangan E-Marketing Penerimaan Mahasiswa Baru (PMB) Dengan Menggunakan Metode Service Oriented Architecture (SOA) : Studi Kasus Pada STT Nusa Putra Sukabumi tahun 2015. *Jurnal Nusantara*, 1–12. Retrieved from <http://jurnal.nusaputra.ac.id/index.html>
- "Sembiring, F., Sari, D. P., Sukmawan, D., Permana, A., Studi, P., Informasi, S., ... Indonesia, U. P. (2019). Penerapan Metode First Expired First Out (Fefo) Pada Sistem Informasi Gudang. *1(2)*, 19–25. "
- "Wahono, R. S. (2015). A Systematic Literature Review of Software Defect Prediction:

Research Trends, Datasets, Methods and Frameworks. *Journal of Software Engineering*, 1(1), 1–16. ”

Wakhidah, N. (2010). Clustering Menggunakan K-Means Algorithm. *Jurnal Transformatika*, 8(1), 33. <https://doi.org/10.26623/transformatika.v8i1.45>

wikipedia. (n.d.). Orange(Software). Retrieved June 12, 2021, from wikipedia.org website: [https://en.wikipedia.org/wiki/Orange_\(software\)](https://en.wikipedia.org/wiki/Orange_(software))

Yunita. ” (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus : Universitas Islam Indragiri). *Jurnal Sistemasi*, 7(September), 238–249. ”