

Analisis Data Penjualan Pasca COVID-19 Menggunakan Algoritma K-Means

Yustia Hapsari^{1*)}, Muhammad Fikri Hidayattullah², M. Shidiq Erdi W.³, Dega Surono Wibowo⁴, M. Nishom⁵

^{1,3}Prodi Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Panca Sakti, Tegal

^{2,4,5}Prodi Sarjana Terapan Teknik Informatika, Politeknik Harapan Bersama, Tegal

^{1,3}Jl. Halmahera No.KM. 01, Mintaragen, Kec. Tegal Timur, Kota Tegal, Jawa Tengah 52121

^{2,4,5}Jl. Mataram No. 09, Pesurungan Lor, Margadana, Kota Tegal, Jawa Tengah 52147

email: ¹yustia.hapsari@upstegal.ac.id, ²fikri@poltektegal.ac.id, ³msiddik_erdwicaksono@upstegal.ac.id,

⁴dega.wibowo@poltektegal.ac.id, ⁵nishom@poltektegal.ac.id

Abstract – The Covid-19 pandemic has had a negative impact on various sectors. The economic sector is no exception. Even developed countries have been seriously affected in the economic sector from this pandemic. Business actors in various lines carried out massive transformations to rise from this adversity. One of the strategic steps taken was to change the marketing system and sales transactions from previously offline to online. This study will conduct an analysis of post-Covid-19 sales using a dataset from Narasio Data. The products to be analyzed are in the form of sales of sports and apparel products. The results of the analysis will show which products are the fastest selling, the effect of product ratings and discounts on the level of sales, as well as the relationship between the many variations of the brand and the level of sales. The analysis process uses a product segmentation approach. The algorithm used to perform segmentation is K-Means. The algorithm is very reliable for product clustering cases.

Abstrak – Pandemi Covid-19 memberikan dampak buruk ke berbagai sektor. Tak terkecuali sektor ekonomi. Bahkan negara-negara maju sekali pun ikut terkena dampak yang sangat serius di sektor ekonomi dari adanya pandemi ini. Para pelaku usaha dalam berbagai lini melakukan transformasi yang massif untuk bangkit dari keterpurukan tersebut. Salah satu langkah strategis yang dilakukan adalah dengan mengubah sistem marketing dan transaksi penjualan dari yang sebelumnya *offline* ke *online*. Penelitian ini akan melakukan analisis penjualan pasca Covid-19 dengan menggunakan *dataset* dari Narasio Data. Produk yang akan dianalisis berupa penjualan produk *sport* dan *apparel*. Hasil analisis akan menampilkan produk apa saja yang paling cepat terjual, pengaruh rating produk dan pemberian diskon dengan tingkat penjualan, serta keterkaitan antara banyaknya variasi *brand* dengan tingkat penjualan. Proses analisis menggunakan pendekatan segmentasi produk. Algoritma yang digunakan untuk melakukan segmentasi adalah K-Means. Algoritma tersebut sangat handal untuk kasus *clustering* produk.

Kata Kunci – Covid-19, analisis penjualan, K-Means, *clustering*.

I. PENDAHULUAN

Pandemi Covid-19 yang telah berlangsung selama hampir tiga tahun memberikan dampak kerugian besar ke berbagai sektor. Salah satu sektor yang terkena dampak adalah sektor ekonomi [1][2]. Tidak hanya negara Indonesia saja yang

terkena dampak ekonomi, bahkan negara-negara maju seperti Jepang, Inggris, Singapura, Selandia Baru, Korea Selatan hingga Amerika Serikat ikut terdampak secara ekonomi [3]. Fenomena tersebut membuat beberapa lini usaha merubah strategi marketing dan penjualannya untuk meningkatkan penjualan. Hal yang paling umum adalah dengan berpindah dari sistem penjualan *offline* ke *online*[4][5].

Selama masa pandemi tingkat penjualan dengan sistem dan mekanisme *online* mengalami peningkatan yang cukup tinggi. Pada tahun 2020 terdapat peningkatan sebesar 38% [6]. Kenaikan tersebut disebabkan munculnya banyak lini usaha yang membuka sistem transaksi secara *online* sekaligus karena adanya kebijakan *stay at home*.

Pasca Covid-19 beberapa lini usaha yang telah membuka toko *online* harus melakukan analisis penjualan agar tingkat penjualannya bisa stabil dan berkelanjutan. Hasil analisis penjualan tersebut dapat dijadikan sebagai acuan untuk menentukan kebijakan pemberian diskon, penyetoran barang yang paling banyak terjual dan pengaturan tata letak posisi barang yang mudah dilihat oleh calon pembeli. Permana *et al.* [6] melakukan analisis aktivitas transaksi belanja online di beberapa *marketplace* besar seperti Shopee, Tokopedia, Bukalapak dan sebagainya. Penelitian tersebut hanya berfokus ke analisis perubahan perilaku masyarakat dalam transaksi jual beli *online*. Atmaja dan Novitaningtyas [7] melakukan analisis aspek pemasaran di media online yang dilakukan oleh para pelaku usaha UMKM. Penelitian tersebut berupaya mencari tahu seberapa tinggi tingkat pemasaran online yang telah dilakukan oleh para pelaku usaha UMKM. Sedangkan Purnama dan Putri [8] menganalisis alasan masyarakat dalam memilih berbelanja secara *online*. Adapun penelitian yang berfokus untuk melakukan analisis produk untuk mengetahui produk mana saja yang paling cepat terjual dan yang tidak belum dilakukan pada penelitian sebelumnya.

Pada penelitian ini akan melakukan analisis penjualan produk *sports* dan *apparel*. *Dataset* yang akan digunakan diperoleh dari Narasio Data. Penelitian ini bertujuan mengetahui *brand* apa saja yang paling cepat terjual, seberapa efektif pemberian diskon untuk menunjang tingkat penjualan, apakah variasi produk tiap *brand* berpengaruh terhadap angka penjualan dan pengaruh *review* dari *customer* dalam menunjang penjualan. Algoritma yang digunakan untuk melakukan segmentasi adalah K-Means. Algoritma K-

^{*)} Penulis Korespondensi : Yustia Hapsari
email: yustia@hapsari@gmail.com

Means merupakan algoritma yang unggul untuk melakukan klusterisasi atau segmentasi [9].

II. PENELITIAN YANG TERKAIT

Beberapa penelitian yang berfokus pada analisis data penjualan di masa pandemi Covid-19 telah dilakukan oleh beberapa peneliti. Berdasarkan analisis dan pembahasan yang dilakukan Permana *et al.* [6], ditarik kesimpulan bahwa pandemi Covid-19 di Indonesia berpengaruh cukup signifikan terhadap pola penggunaan *e-commerce* oleh masyarakat sebagai sarana pemenuhan kebutuhan. Peningkatan ini terlihat pada kuartal pertama tahun 2020, yang bertepatan dengan awal kemunculan kasus pertama covid-19 di Indonesia, dan terus meningkat pada kuartal kedua sebanyak 38% hingga saat ini.

Atmaja dan Novitaningtyas [7] melakukan penelitian yang bertujuan untuk mengetahui tingkat penggunaan *e-commerce* di kalangan para pelaku usaha UMKM. Hasil penelitian menunjukkan bahwa pemasaran produk UMKM melalui *e-commerce*, promosi produk dengan memanfaatkan media digital, dan menjalin relasi yang baik dengan pelanggan melalui CRM (*Customer Relationship Marketing*) dapat diimplementasikan oleh UMKM untuk meningkatkan minat beli konsumen dan volume penjualan. Dengan demikian, UMKM dapat terus menjalankan bisnisnya selama masa pandemi Covid-19.

Purnama dan Putri [8] juga membuat sebuah penelitian tentang analisis data penjualan di masa pandemi. Penelitian ini bertujuan untuk menganalisis alasan masyarakat dalam bertransaksi ke salah satu *marketplace*. Berdasarkan hasil analisis dapat disimpulkan bahwa masyarakat banyak menggunakan *e-commerce* untuk berbelanja keperluan sehari-hari dan keperluan lainnya dengan alasan harga yang lebih murah. Pada penelitian ini juga disebutkan bahwa Shopee merupakan *e-commerce* yang paling banyak dikunjungi selama masa pandemi. *Shopee* merupakan *e-commerce* yang selalu memberikan promo-promo yang dapat terus diingat oleh masyarakat. Selain itu, kepercayaan dan kualitas *web* menjadi faktor utama konsumen melakukan keputusan pembelian di *Shopee*.

Beberapa penelitian yang ada lebih berfokus melakukan analisis penjualan selama masa pandemi. Sedangkan saat ini pandemi Covid-19 sudah mulai mereda. Sehingga dibutuhkan sebuah penelitian yang berfokus pada analisis data penjualan pasca Covid-19.

III. METODE PENELITIAN

Metode penelitian yang digunakan menggunakan pendekatan sebagaimana tahapan dalam algoritma K-Means [10].



Gambar 1. Metode Penelitian

1. Import Libraries

Tahap ini melakukan *import* seluruh *Python library* yang dibutuhkan untuk penelitian [11].

2. Reading Dataset

Tahap ini merupakan tahap pembacaan *dataset* penjualan pasca Covid-19 yang diperoleh dari Narasio Data.

3. Data Visualization

Dataset yang telah ditampilkan dalam bentuk tabel akan divisualisasikan ke dalam bentuk grafik dengan berbagai bentuk. Tujuannya untuk mengetahui sebaran data produk yang nantinya akan dianalisis.

4. Find K Value

Algoritma K-Means bertumpu kepada jumlah kluster (K) [12]. Secara teori jumlah K bisa ditentukan secara acak. Namun akan berpengaruh ke dalam efektifitas hasil *clustering*. Oleh karena itu digunakan metode Elbow [13] dipadukan dengan Silhouette [14] untuk mencari nilai K yang optimum.

5. Clustering

Tahap klusterisasi atau segmentasi produk.

6. Model Interpretation

Hasil klusterisasi akan diinterpretasi untuk diketahui makna atau potensi di setiap *cluster*-nya [15].

7. Result

Hasil interpretasi dapat digunakan untuk penentuan kebijakan penyetokan barang, pemberian diskon dan pemberian pelayanan terbaik ke para *customer* agar *rating*-nya selalu tinggi.

IV. HASIL DAN PEMBAHASAN

1. Dataset

Dataset berisi produk *sport* dan *apparel* yang terjual pasca pandemi Covid-19 dengan total data sebanyak 3268 *data record*.

	Product Name	Product ID	Listing Price	Sale Price	Discount	Brand	Description	Rating	Reviews	Last Visited
0	Women's adidas Originals NMD_Racer Primeknit S	AN2430	14999	7899	50	Adidas ORIGINALS	Channeling the streamlined look of an 80s sneaker.	4.5	41	2020-04-13T15:06:14Z
1	Women's adidas Originals Sleek Shoes	Q27341	7999	3799	50	Adidas ORIGINALS	A modern take on adidas sport heritage, value.	3.3	24	2020-04-13T15:06:19Z
2	Women's adidas Suede Puka Slippers	C88001	999	599	40	Adidas CORE / NEO	These adidas Puka slippers for women's come w/	2.6	37	2020-04-13T15:06:19Z
3	Women's adidas Sport Inspired Quarter Ride Shoes	B44E32	6999	3499	50	Adidas CORE / NEO	Inspired by modern tech runners, these women's	4.1	35	2020-04-13T15:06:19Z
4	Women's adidas Originals Taskwonder Shoes	C86205	7999	3999	50	Adidas ORIGINALS	This design is inspired by vintage Taskwonder s	3.5	72	2020-04-13T15:06:19Z
3263	Air Jordan 8 Retro	C1236-100	15995	12797	0	Nike	The Air Jordan 8 Retro recaptures the moment	5.0	1	2020-04-13T15:41:57Z
3264	Nike Phantom Venom Club IC	AO5575-717	4995	3457	0	Nike	The Nike Phantom Venom Club IC is engineered f	0.0	0	2020-04-13T15:41:57Z
3265	Nike Mercurial Superfly 7 Academy TF	AT7975-414	8495	5947	0	Nike	The soft upper of the Nike Mercurial Superfly	5.0	1	2020-04-13T15:41:57Z
3266	Nike Air Max 95	AH6795-350	0	10395	0	Nike	The Nike Air Max 95 features the OG design in	4.0	4	2020-04-13T15:41:59Z
3267	Nike P-4000 SE	CJ8965-600	8995	6297	0	Nike	A mash-up of Pegasus' past, the Nike P-4000 SE	0.0	0	2020-04-13T15:42:57Z

Gambar 2. *Dataset* Penjualan Produk *Sport* dan *Apparel*

Ada beberapa fitur yang tidak digunakan dalam proses analisis data dalam penelitian ini, di antaranya *Product Name*, *Product Id*, *Last Visited*, dan *Description*.

Jumlah total produk dari masing-masing *brand* dapat dilihat pada Tabel 1.

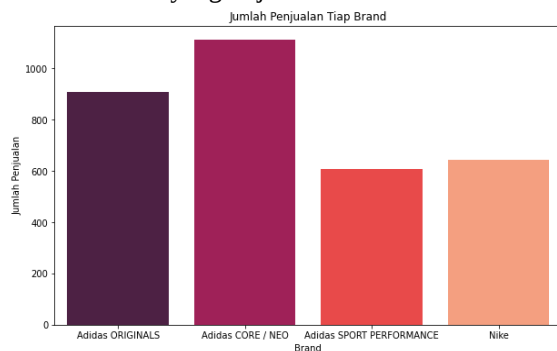
Tabel 1. Jumlah Produk di tiap *Brand*

Brand	Total Produk
Adidas SPORT PERFORMANCE	348
Adidas ORIGINALS	350
Nike	393
Adidas CORE / NEO	440

2. Visualisasi Data dan Analisis Penjualan

Tahap visualisasi data akan menampilkan sebaran informasi tentang produk dalam bentuk grafik multi dimensi. Visualisasi data memainkan peran yang vital dalam melakukan analisis penjualan.

a. Analisis *brand* yang terjual

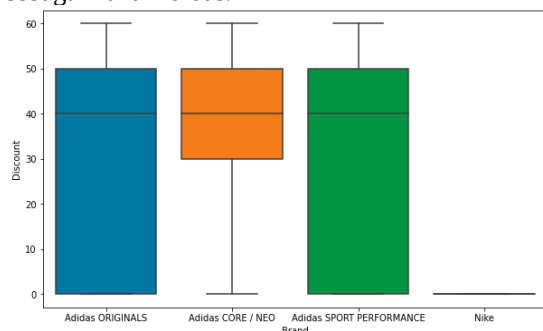


Gambar 3. Jumlah Penjualan tiap *Brand*

Berdasarkan visualisasi data pada Gambar 3. terdapat dua *brand* yang dijual di toko yaitu Adidas dan Nike. Total penjualan *brand* Adidas lebih tinggi dibandingkan dengan Nike. Penjualan paling banyak terdapat pada Adidas Core/Neo.

b. Analisis pemberian diskon

Proporsi penjualan *brand* Adidas jauh melampaui penjualan Nike. Salah satu faktor yang mempengaruhi adalah adanya pemberian diskon. Hal tersebut dapat dibuktikan melalui grafik *box plot* pada Gambar 4. *Brand* Adidas memberikan tawaran diskon pada semua kategori *brand* yang dijual. Sedangkan Nike, hampir tidak memberikan diskon sebagaimana Adidas.



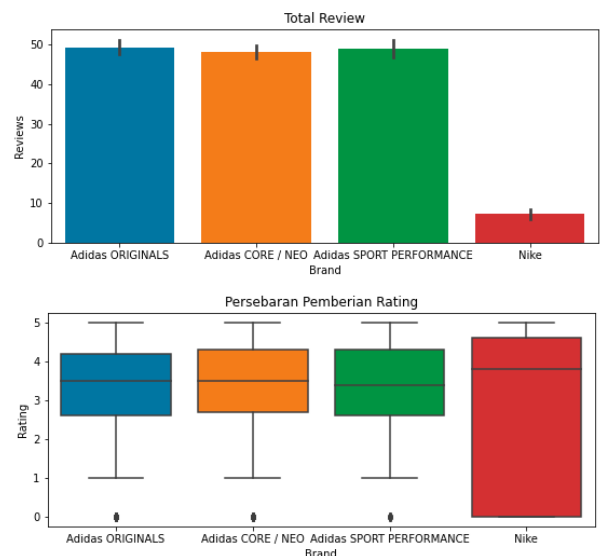
Gambar 4. Pengaruh Pemberian Diskon dalam Penjualan

c. Analisis *review* dari *customer* dan rating produk

Gambar 5. menunjukkan bahwa *brand* Adidas mendapatkan jumlah *review* yang banyak pada semua kategori. Berbeda dengan Nike yang hanya mendapatkan *review* sedikit dari *customer*. Begitu

juga pada nilai rating, rata-rata produk Adidas mendapatkan rating di atas 2. Hal tersebut mengindikasikan bahwa tingkat kepuasan *customer* cukup tinggi terhadap *brand* Adidas. Sedangkan untuk produk Nike mendapatkan rating yang bervariasi mulai dari 0 sampai 4,5. Hal ini menginformasikan bahwa ada *customer* yang merasa 'tidak puas' hingga 'sangat puas' terhadap produk Nike.

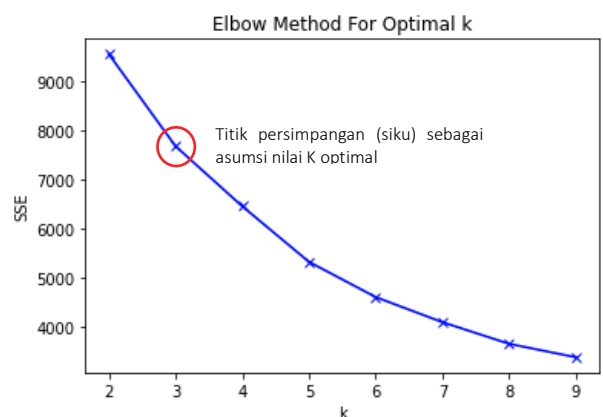
Dari hasil analisis terhadap *review* dari *customer* dan rating produk dapat diketahui bahwa tingginya penjualan *brand* Adidas bukan hanya karena banyaknya penawaran diskon. Namun, juga karena faktor kualitas produk dan strategi *marketing* yang diaplikasikan.



Gambar 5. *Customer Reviews* dan *Product Rating*

3. Nilai K optimal

Jumlah *cluster* ditentukan berdasarkan nilai K pada algoritma K-means. Nilai K optimal dapat diketahui dengan metode Elbow. Titik persimpangan yang dihasilkan dari metode Elbow diasumsikan sebagai nilai K optimal. Nilai K yang diperoleh dari metode Elbow selanjutnya akan divalidasi menggunakan *silhouette score*.



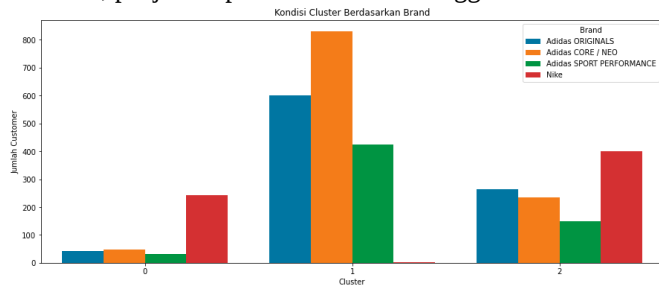
Gambar 6. Penentuan Nilai K Optimal menggunakan Metode Elbow

Berdasarkan grafik pada Gambar 6. dapat diketahui bahwa nilai K optimal adalah 3. Hasil validasi *silhouette*

score juga menyatakan bahwa nilai K optimal = 3. Jumlah *cluster* yang terbentuk sebanyak 3 sesuai dengan jumlah nilai K.

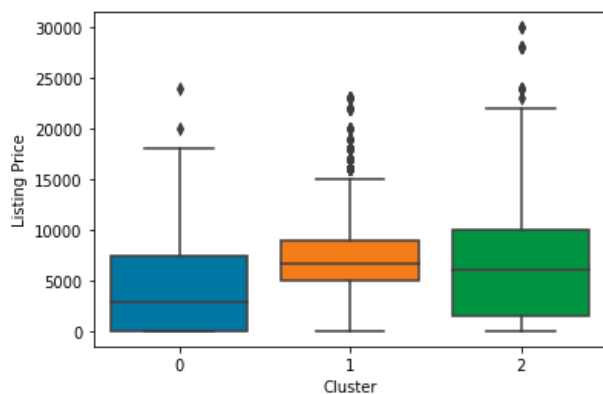
4. Klasterisasi Produk

Persebaran *brand* pada tiap *cluster* dapat dilihat di Gambar 7. Terlihat pada *cluster* 1 jumlah produk Nike yang terjual sangat sedikit. Sedangkan pada *cluster* 0 dan 2, penjualan produk Nike lebih tinggi dari Adidas.



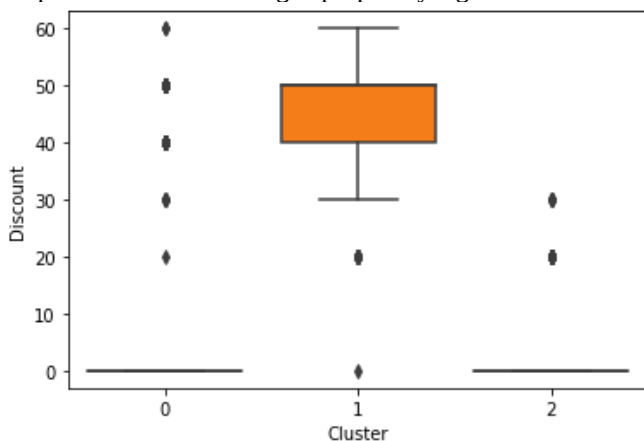
Gambar 7. Persebaran *Brand* di tiap *Cluster*

Jika dilihat dari harga yang ditawarkan, *range* harga tertinggi ada pada *cluster* 2, disusul dengan *cluster* 1 kemudian *cluster* 0.



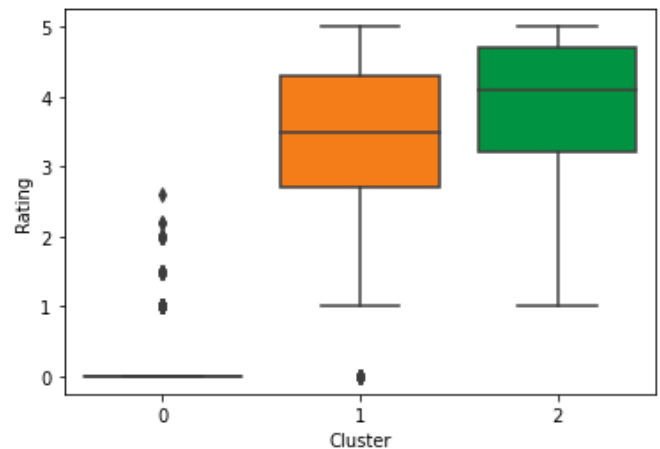
Gambar 8. Pengelompokan *Cluster* berdasarkan Harga

Analisis pengaruh pemberian diskon pada tiap *cluster* dapat dilihat pada Gambar 9. Terlihat *cluster* 1 banyak memberikan penawaran diskon. Sedangkan, di *cluster* 0 terdapat beberapa produk yang diberi diskon dengan nilai proporsi yang tinggi. *Cluster* 2 menunjukkan produk yang diberi penawaran diskon sangat minim jumlahnya, namun ada juga beberapa produk yang diberi penawaran diskon dengan proporsi yang rendah.



Gambar 9. Pengelompokan *Cluster* berdasarkan Pemberian Diskon

Selain pemberian diskon, faktor lain yang memiliki pengaruh terhadap tingkat penjualan produk adalah rating. *Cluster* 0 mengindikasikan banyak berisi *customer* yang merasa tidak puas dengan produk yang dibelinya. Sedangkan pada *cluster* 1 dan 2, rata-rata *customer* merasa puas.



Gambar 10. Kondisi *Cluster* berdasarkan Rating

5. Interpretasi Hasil

Setelah melakukan analisis dari masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 1 akan dikelompokkan ke dalam **Produk dalam Diskon**. Produk ini terdiri dari produk dengan keluaran lama namun masih banyak peminat. Pemberian diskon memberikan pengaruh yang positif dari segi penjualan. *Cluster* 0 dikelompokkan ke dalam **Produk Terbaru**. Produk dalam kategori ini adalah produk yang kurang banyak diminati dan memiliki reputasi yang kurang bagus. Produk yang ditawarkan pada kategori ini memiliki harga yang tinggi dan tidak dibantu dengan teknik dan metode penjualan opsional lainnya. Walaupun produk pada *cluster* ini merupakan keluaran yang cukup baru namun tingkat penjualan rendah. *Cluster* 2 dikelompokkan ke dalam **Produk Favorit**. Produk Favorit merupakan kategori produk unggulan, walaupun produk dibanderol dengan harga yang tinggi dan tanpa penawaran dari segi pemberian diskon, namun respon *customer* masih sangat bagus, dan *record* penjualan sangat memuaskan.

V. KESIMPULAN

Terdapat tiga kelompok *cluster* yang dihasilkan dalam penelitian ini, yaitu Produk dalam Diskon, Produk Terbaru dan Produk Favorit. Produk dalam Diskon berisi produk-produk keluaran lama yang masih memiliki banyak peminat. Produk Terbaru merupakan produk keluaran terbaru dengan tingkat penjualan yang masih rendah dikarenakan *customer* belum mengetahui kualitas dari produk tersebut. Sedangkan Produk Favorit adalah produk unggulan yang sangat diminati oleh *customer* serta tingkat penjualannya tidak begitu terpengaruh oleh harga dan pemberian diskon. Selain itu juga, berdasarkan analisis yang dilakukan dapat diketahui bahwa pemberian diskon serta kualitas produk yang unggul merupakan faktor utama yang sangat berpengaruh dalam penjualan produk *sport* dan *apparel*.

DAFTAR PUSTAKA

- [1] S. Hanoatubun, "Dampak Covid – 19 terhadap Perekonomian Indonesia," *J. Educ. Psycology Couns.*, vol. 2, no. 2, 2020.
- [2] F. R. Yamali and R. N. Putri, "Dampak Covid-19 Terhadap Ekonomi Indonesia," *Ekon. J. Econ. Bus.*, vol. 4, no. 2, p. 384, 2020.
- [3] D. Junaedi and F. Salistia, "Dampak Pandemi COVID-19 Terhadap Pertumbuhan Ekonomi Negara-Negara Terdampak," *Simposium Nasional Keuangan Negara*, pp. 995–1115, 2020.
- [4] F. Nidaul Khasanah *et al.*, "Pemanfaatan Media Sosial dan Ecommerce Sebagai Media Pemasaran Dalam Mendukung Peluang Usaha Mandiri Pada Masa Pandemi Covid 19," *J. Sains Teknol. dalam Pemberdaya. Masy.*, vol. 1, no. 1, pp. 51–62, 2020.
- [5] E. Maria, S. Suharyadi, and R. K. Hudiono, "Implementasi pemasaran digital berbasis website sebagai strategi kenormalan baru Dusun Srumbung Gunung pasca Covid-19," *Riau J. Empower.*, vol. 4, no. 1, pp. 1–10, 2021.
- [6] A. E. Permana, A. M. Reyhan, H. Rafli, and N. A. Rakhmawati, "Analisa Transaksi Belanja Online Pada Masa Pandemi Covid-19," *J. Teknoinfo*, vol. 15, no. 1, p. 32, 2021.
- [7] I. Atmaja, Hanung Eka; Novitaningtyas, "Analisis Aspek Pemasaran UMKM di Masa Resesi Global Dampak dari Pandemi Covid-19," *J. Ilm. Poli Bisnis*, vol. 13, no. 1, pp. 60–69, 2021.
- [8] N. I. Purnama and L. P. Putri, "Analisis Penggunaan E-Commerce Di Masa Pandemi," *Semin. Nas. Teknol. Edukasi Sos. dan Hum.*, vol. 1, no. 1, pp. 556–561, 2021.
- [9] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.
- [10] E. Liberty, R. Sriharsha, and M. Sviridenko, "An algorithm for online k-means clustering," *Proc. Work. Algorithm Eng. Exp.*, vol. 2016-January, pp. 81–89, 2016.
- [11] P. Lemenkova, "Python libraries matplotlib, seaborn and pandas for visualization geospatial datasets generated by QGIS," *Analele Stiint. ale Univ. "Alexandru Ioan Cuza" din Iasi - Ser. Geogr.*, vol. 1, no. 1, pp. 13–32, 2020.
- [12] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J. Multidiscip. Sci. J.*, vol. 2, no. 2, pp. 226–235, 2019.
- [13] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means menggunakan Metode Elbow pada Data Pengguna Narkoba Dengan Pemrograman Python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [14] F. Wang, H. H. Franco-Penya, J. D. Kelleher, J. Pugh, and R. Ross, "An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10358 LNAI, pp. 291–305, 2017.
- [15] D. Hallac, S. Vare, S. Boyd, and J. Leskovec, "Toeplitz inverse covariance-based clustering of multivariate time series data," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F129685, pp. 215–223, 2017.