

## **Data Scrapping untuk Optimalisasi Penjualan Produk Skincare di Marketplace Aryan Eka Prastya Nugraha<sup>1</sup>, Riyanto<sup>2</sup>, Novika Wahyuhastuti<sup>3</sup>, Inayah Adi Sari<sup>4</sup>**

<sup>123</sup>Program Studi Pendidikan Ekonomi, Universitas PGRI Semarang

Email: [aryaneka@upgris.ac.id](mailto:aryaneka@upgris.ac.id)

### **ABSTRACT**

The research aims to understand the use and benefits of marketplace data to determine consumer behavior on Skincare Brands in Indonesia, aiming to create a picture for consumers to choose the right skincare brand from producers and sellers. The method used in this study is to take advantage of server log files that have been scraped using the Octoparse tool. General data from sellers will be processed to be analyzed using Xlstat with the exploratory factor analysis (or EFA) technique, which is a method that reveals the possible existence of underlying factors that provide an overview of the information contained in a large number of measured variables. The findings in this study are It can be concluded that the rating and the number of items sold are interrelated and indirectly able to attract consumer attention in making choices on products.

**Keywords:** *data scrapping, marketplace, optimization*

### **ABSTRAK**

Penelitian bertujuan untuk memahami penggunaan dan manfaat data marketplace untuk menentukan perilaku konsumen pada Brand Skincare di Indonesia, dengan tujuan sebagai seller mampu membuat gambaran bagi konsumen dalam kepastian untuk memilih brand skincare yang tepat dari produsen maupun seller. Metode yang digunakan dalam penelitian ini yaitu dengan memanfaatkan file log server yang sudah di scraping menggunakan tool Octoparse serta data umum dari seller akan diproses untuk dianalisis menggunakan Xlstat dengan teknik Analisis faktor eksploratori (atau EFA) adalah metode yang mengungkapkan kemungkinan adanya faktor-faktor yang mendasari yang memberikan gambaran umum tentang informasi yang terkandung dalam sejumlah besar variabel yang diukur. Temuan dalam penelitian ini yaitu Hal ini dapat disimpulkan bahwa rating dan banyaknya barang terjual saling terkait dan secara tidak langsung mampu menarik perhatian konsumen dalam melakukan pemilihan alternatif pada produk.

**Kata Kunci:** *data scrapping, marketplace, optimalisasi*

## PENDAHULUAN

Meningkatnya penggunaan media sosial dalam kehidupan sehari-hari untuk berbagai keperluan, seperti berkomunikasi, mencari informasi, hingga membeli produk. Hal ini menghasilkan jumlah data yang sangat besar dan beragam, yang dikenal sebagai big data pada sosial media. Menurut sebuah laporan dari Hootsuite dan We Are Social, pada Januari 2021, ada lebih dari 4,2 miliar pengguna media sosial di seluruh dunia. Besarnya data yang dihasilkan dan dibagikan oleh bisnis, administrasi publik, berbagai sektor industri dan nirlaba, serta penelitian ilmiah, telah meningkat secara tak terukur. Data ini mencakup konten tekstual (yaitu terstruktur, semi-terstruktur, dan tidak terstruktur), hingga konten multimedia (misalnya video, gambar, audio) di berbagai platform (misalnya telekomunikasi mesin-ke-mesin, situs media sosial, jaringan sensor, sistem siber-fisik, dan *Internet of Things* (IoT) (Sivarajah et al., 2017).

Era big data e-commerce bukanlah pemasaran massal. Sebaliknya, ini menyediakan pemasaran yang dipersonalisasi dari big data (Tang, 2021). Dalam bidang bisnis, big data di sosial media dapat digunakan untuk meningkatkan strategi pemasaran dan memahami perilaku konsumen. Dalam bidang politik, big data sosial media dapat digunakan untuk memantau opini publik dan memprediksi hasil pemilihan. Namun, penggunaan big data sosial media juga memunculkan beberapa masalah, seperti masalah privasi dan keamanan data, serta penyebaran informasi palsu dan hoaks. Perilaku konsumen menghasilkan sejumlah besar jenis data tertentu, salah satunya adalah data aliran klik, yang didefinisikan sebagai catatan elektronik penggunaan Internet yang dikumpulkan dari log server situs web atau layanan pihak ketiga. Data ini biasanya berisi data mengenai klik mouse navigasi individu dan klik pada elemen lain dari halaman web yang dilakukan pengunjung saat mengunjungi situs web tertentu (Vanhala et al., 2020). Masih banyak penelitian yang harus dilakukan mengenai pemanfaatan Internet karena perilaku memilih di Internet dalam banyak hal sangat berbeda

dengan perilaku yang telah dieksplorasi secara menyeluruh di lingkungan ritel-toko tradisional (Van Den Poel & Buckinx, 2005). Dengan perkembangan teknologi analisis big data dalam ekonomi jaringan modern, banyak platform jaringan atau perusahaan yang berpartisipasi dalam e-commerce akan mengumpulkan informasi perilaku yang dipersonalisasi dari konsumen, dan menggunakan analisis data dan teknologi pemrosesan untuk mengintegrasikan dan mengekstrak informasi yang efektif, dan membuat rekomendasi yang ditargetkan kepada konsumen secara akurat (Zhang & Tan, 2020). Data berubah menjadi big data ketika volume, kecepatan, atau keragamannya melampaui kemampuan sistem operasional TI untuk mengumpulkan, menyimpan, menganalisis, dan memprosesnya (Khade, 2016). Ketika data menjadi lebih besar, lebih kompleks, dan lebih sulit dijelaskan, kapasitas mental manusia yang terbatas menimbulkan kesulitan dalam menguraikan dan menafsirkan lingkungan yang tidak diketahui (Erevelles et al., 2016).

Dengan pertumbuhan digitalisasi, triliunan data yang terkait dengan perilaku pembelian konsumen dihasilkan setiap sepersekian detik. Selain itu, karena Konsumen memiliki sifat yang bervariasi, perilaku pembelian satu konsumen mungkin sangat berbeda dari yang lain. Oleh karena itu, kompleksitas data yang meningkat berlipat ganda membuat data ini sangat sulit untuk dianalisis dengan metode analisis data konvensional. Dalam skenario ini, menjadi tak terelakkan untuk mengikuti pendekatan baru untuk mendapatkan wawasan yang lebih bermakna dari data ini. Teknologi Big Data yang memiliki kemampuan untuk mengekstrak, memproses, dan menganalisis set data yang sangat besar dan kompleks menjadi jelas (Budakoti, 2022). Oleh karena itu, penelitian diperlukan untuk memahami penggunaan dan manfaat data marketplace untuk menentukan perilaku konsumen pada Brand Skincare di Indonesia, dengan tujuan sebagai seller mampu membuat gambaran bagi konsumen dalam kepastian untuk memilih brand skincare yang tepat dari produsen maupun seller.

## **METODE PENELITIAN**

Kami menggunakan data dari seller di tokopedia yang menjual produk skincare berbagai merek di pasar Indonesia. Semua file log server yang sudah di scraping menggunakan tool Octoparse serta data umum dari seller akan diproses untuk dianalisis menggunakan Xlstat dengan teknik *Analisis faktor eksploratori (atau EFA)* adalah metode yang mengungkapkan kemungkinan adanya faktor-faktor yang mendasari yang memberikan gambaran umum tentang informasi yang terkandung dalam sejumlah besar variabel yang diukur. Struktur yang menghubungkan faktor dengan variabel pada awalnya tidak diketahui dan hanya jumlah faktor yang dapat diasumsikan.

## **HASIL DAN PEMBAHASAN**

Temuan penelitian ini menunjukkan manfaat dari pemanfaatan big data yang ada pada beberapa sosial media dan marketplace yang digunakan untuk memprediksi perilaku konsumen yang menggunakan brand skincare di Indonesia. Dengan menganalisis data media sosial tentu produsen dapat memperoleh wawasan real-time mengenai opini, preferensi, dan sentimen konsumen. Informasi ini memungkinkan produsen maupun seller skincare untuk membuat keputusan yang gesit dan terinformasi dalam menanggapi perubahan tren konsumen. Beberapa Teknik digunakan dalam penelitian dan berikut masing – masing dari analisis data beserta penjelasannya, yang pertama menggunakan teknik data scraping dengan tools octoparse. Teknik ini merupakan proses otomatisasi yang digunakan untuk mengumpulkan informasi atau data dari berbagai situs web dan menyimpannya dalam format yang dapat diakses dan dianalisis. Teknik ini memungkinkan pengguna untuk mengambil informasi seperti teks, gambar, harga produk, ulasan pelanggan, atau data lainnya dari halaman web secara otomatis.

Berikutnya data mentah tersebut terkumpul sejumlah 198 data yang berasal dari marketplace. Selanjutnya data tersebut dilakukan proses data cleaning atau juga dikenal sebagai data cleansing atau data scrubbing, adalah proses mengidentifikasi, memperbaiki, dan menghilangkan kesalahan atau inkonsistensi dalam data agar data tersebut menjadi lebih akurat, konsisten, dan berguna. Terkumpul sejumlah 148 data yang layak untuk diolah dengan beberapa komponen yaitu harga, diskon, lokasi, store, rating dan banyaknya terjual. Berdasarkan data mentah yang sudah diolah dan dianalisis menggunakan Faktor analisis ditunjukkan pada tabel berikut:

Tabel. 1 Matrik Korelasi

Variables	Harga	Diskon	Lokasi	Store	Rating	terjual
Harga	<b>1</b>	0.439	0.543	0.654	0.543	0.231
Diskon	0.439	<b>1</b>	0.423	0.654	0.543	0.454
Lokasi	0.543	0.423	<b>1</b>	0.543	0.054	0.754
Store	0.654	0.654	0.543	<b>1</b>	0.702	0.883
Rating	0.543	0.543	0.054	0.702	<b>1</b>	0.838
Terjual	0.231	0.454	0.754	0.883	0.838	<b>1</b>

Pada tabel tersebut diketahui bahwa adanya korelasi yang cukup tinggi yaitu pada variable penjualan atau terjual dan rating yaitu (0.883 dan 0.838). Sedangkan Cronbach's alpha 0.823. Hasil berikutnya yang ingin kita lihat adalah muatan faktor setelah rotasi varimax. Hasil ini digunakan untuk menginterpretasikan makna dari faktor (yang dirotasi).

Tabel. 2 Korelasi antara variabel dan faktor setelah rotasi Varimax:

	D1	D2
Harga	0.432	0.541
Diskon	0.224	0.321
Lokasi	0.043	0.543
Store	0.643	0.051

Rating 0.864 0.742

Terjual 0.875 0.531

---

Dapat diketahui bahwa faktor pertama ada yang berhubungan positif dan memiliki nilai cukup tinggi diantaranya rating dan banyaknya barang yang terjual. Hal ini dapat disimpulkan bahwa rating dan banyaknya barang terjual saling terkait dan secara tidak langsung mampu menarik perhatian konsumen dalam melakukan pemilihan alternatif pada produk.

## SIMPULAN DAN SARAN

Berdasarkan pembahasan tersebut maka dapat disimpulkan bahwa rating dan banyaknya barang terjual saling terkait dan secara tidak langsung mampu menarik perhatian konsumen dalam melakukan pemilihan alternatif pada produk. Hal tersebut dapat memberikan manfaat bagi para seller untuk tetap menjaga rating dan meningkatkan penjualan dengan strategi yang masing-masing sesuai dengan karakteristik produk yang ditawarkan.

## DAFTAR PUSTAKA

- Budakoti, S. (2022). Understanding The Role of Big Data Technology in Analysing Consumer Behavior. *International Journal of Engineering Research & Technology (IJERT)*, 10(1). [www.ijert.org](http://www.ijert.org)
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904. <https://doi.org/10.1016/j.jbusres.2015.07.001>
- Khade, A. A. (2016). Performing Customer Behavior Analysis using Big Data Analytics. *Procedia Computer Science*, 79, 986–992. <https://doi.org/10.1016/j.procs.2016.03.125>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Tang, R. (2021). *The Influence of Big Data on Purchasing Behavior*. <https://www.wjx.cn/report/4566279.aspx>.
- Van Den Poel, D., & Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research*, 166(2), 557–575. <https://doi.org/10.1016/j.ejor.2004.04.022>

- Vanhala, M., Lu, C., Peltonen, J., Sundqvist, S., Nummenmaa, J., & Järvelin, K. (2020). The usage of large data sets in online consumer behaviour: A bibliometric and computational text-mining–driven analysis of previous research. *Journal of Business Research*, *106*, 46–59. <https://doi.org/10.1016/j.jbusres.2019.09.009>
- Zhang, C., & Tan, T. (2020). The Impact of Big Data Analysis on Consumer Behavior. *Journal of Physics: Conference Series*, *1544*(1). <https://doi.org/10.1088/1742-6596/1544/1/012165>