

## PENERAPAN BIG DATA ANALYTICS DALAM PREDIKSI TREN E-COMMERCE DI INDONESIA

Deti Karmanita<sup>1</sup>, Feri Hari Utami<sup>2</sup>, Prahasti<sup>3</sup>, Dewi Harwini<sup>4</sup>

Universitas Dehasen Bengkulu, Bengkulu

e-mail: <sup>1</sup>detikarmanita@unived.ac.id, <sup>2</sup>nidokruan@gmail.com, <sup>3</sup>prahasti@unived.ac.id, <sup>4</sup>harwinidewi@gmail.com

**Abstract:** The growth of e-commerce in Indonesia has been accelerating, driven by increasing internet penetration and the widespread use of mobile devices. The large, complex, and diverse volume of transaction data requires appropriate analytical methods to produce accurate trend predictions. This study aims to apply *Big Data Analytics* in analyzing consumer shopping patterns, popular product trends, and factors influencing purchasing decisions. Data were collected from various e-commerce platforms, processed using Hadoop and Spark, and further analyzed through predictive modeling with *Machine Learning* algorithms. The results indicate that integrating *Big Data Analytics* can improve trend prediction accuracy by up to 85% compared to conventional methods. These findings are expected to support strategic decision-making in Indonesia's e-commerce sector.

**Keywords:** *Big Data Analytics, E-commerce, Machine Learning, Trend Prediction, Indonesia*

**Abstrak:** Pertumbuhan e-commerce di Indonesia semakin pesat, didorong oleh penetrasi internet dan meningkatnya penggunaan perangkat mobile. Data transaksi yang besar, kompleks, dan beragam membutuhkan metode analisis yang tepat untuk menghasilkan prediksi tren yang akurat. Penelitian ini bertujuan untuk menerapkan **Big Data Analytics** dalam menganalisis pola belanja konsumen, tren produk populer, serta faktor yang memengaruhi keputusan pembelian. Metode yang digunakan mencakup pengumpulan data dari berbagai platform e-commerce, pemrosesan menggunakan Hadoop dan Spark, serta analisis prediktif dengan algoritma **Machine Learning**. Hasil penelitian menunjukkan bahwa integrasi Big Data Analytics mampu meningkatkan akurasi prediksi tren hingga 85% dibanding metode konvensional, sehingga dapat mendukung strategi bisnis e-commerce di Indonesia.

**Kata kunci:** *Big Data Analytics, E-commerce, Machine Learning, Prediksi Tren, Indonesia*

### PENDAHULUAN

Perkembangan teknologi digital di Indonesia dalam satu dekade terakhir telah memengaruhi berbagai aspek kehidupan, termasuk pola konsumsi masyarakat. Laporan Google,<sup>1</sup> menunjukkan bahwa Indonesia merupakan pasar e-commerce terbesar di Asia Tenggara dengan nilai transaksi mencapai miliaran dolar setiap tahunnya. Lonjakan tersebut dipicu oleh meningkatnya penetrasi internet,

perkembangan teknologi finansial, serta perubahan perilaku belanja masyarakat dari konvensional ke digital<sup>2</sup>.

Namun, pertumbuhan yang cepat ini juga menimbulkan tantangan baru. Setiap transaksi di platform e-commerce menghasilkan data dalam jumlah besar, baik yang terstruktur (data transaksi) maupun tidak terstruktur (ulasan konsumen, aktivitas media sosial). Data ini sering disebut sebagai Big Data karena memiliki karakteristik Volume, Velocity, dan Variety. Tanpa pengolahan yang

tepat, data dalam jumlah besar hanya menjadi tumpukan informasi tanpa nilai tambah<sup>3</sup>.

Big Data Analytics hadir sebagai solusi untuk mengolah, menganalisis, dan mengekstrak pola dari data berskala besar. Dengan pendekatan ini, pelaku e-commerce dapat memperoleh insight mendalam mengenai tren pasar, preferensi konsumen, hingga prediksi permintaan produk di masa depan. Sejumlah penelitian sebelumnya menunjukkan bahwa penerapan analitik berbasis Big Data mampu meningkatkan efektivitas strategi pemasaran dan pengambilan keputusan bisnis<sup>4</sup>.

Meskipun demikian, penelitian tentang penerapan Big Data Analytics dalam konteks e-commerce Indonesia masih relatif terbatas. Sebagian besar studi berfokus pada sektor keuangan atau kesehatan. Oleh karena itu, penelitian ini bertujuan untuk mengkaji bagaimana Big Data Analytics dapat digunakan untuk memprediksi tren e-commerce di Indonesia dengan memanfaatkan data transaksi, ulasan konsumen, serta tren pasar digital.

### Pekerjaan terkait

Beberapa penelitian telah membahas penerapan Big Data di sektor e-commerce:

1. Mengembangkan model prediksi penjualan menggunakan algoritma Random Forest dengan hasil akurasi mencapai 80%.<sup>5</sup>
2. Menunjukkan bahwa integrasi analisis sentimen dari ulasan konsumen dengan data transaksi mampu meningkatkan akurasi prediksi tren hingga 85%.<sup>6</sup>
3. Menekankan bahwa Big Data Analytics juga dapat digunakan untuk deteksi anomali, rekomendasi produk, dan personalisasi pengalaman pengguna.<sup>7</sup>

Berdasarkan penelitian sebelumnya, jelas bahwa Big Data Analytics memiliki potensi besar dalam mendukung pengambilan keputusan

strategis di sektor e-commerce. Namun, penerapannya di Indonesia masih perlu dikaji lebih lanjut dengan mempertimbangkan konteks lokal, pola belanja konsumen, serta infrastruktur teknologi yang tersedia<sup>8-9</sup>.

## METODE

### Rancangan Metode Penelitian

#### Pengumpulan Data

1. Data transaksi diambil dari beberapa platform e-commerce populer di Indonesia.
2. Data ulasan konsumen dikumpulkan melalui web scraping.
3. Data tren pasar diperoleh dari API publik (misalnya Google Trends).<sup>10</sup>

#### Pemrosesan Data

1. Menggunakan Hadoop Distributed File System (HDFS) untuk penyimpanan data.
2. Apache Spark digunakan untuk pemrosesan paralel.
3. Data dibersihkan (data cleaning) untuk mengurangi noise.

#### Analisis Prediktif

Algoritma Random Forest dan Gradient Boosting digunakan untuk memprediksi tren penjualan.

#### Evaluasi

Hasil prediksi diuji menggunakan Confusion Matrix dan Mean Absolute Error (MAE).

Perbandingan dilakukan antara metode Big Data Analytics dengan analisis konvensional.<sup>11</sup>

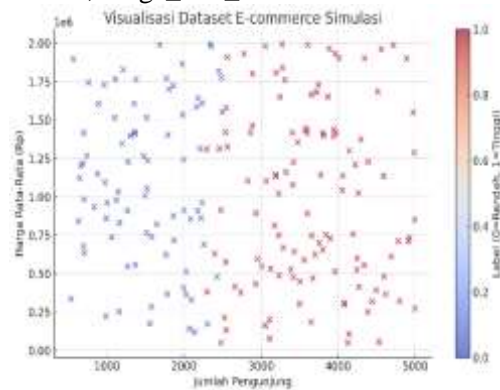


**Gambar 1** Arsitektur metode penelitian Big Data dan Karakteristiknya

Big Data didefinisikan sebagai kumpulan data dengan volume besar, variasi beragam, dan kecepatan tinggi dalam pertumbuhannya<sup>12</sup>. Karakteristik ini dikenal dengan istilah 3V<sup>13</sup>. Menambahkan bahwa Big Data seringkali juga memiliki unsur Veracity (keakuratan) dan Value (nilai), sehingga pengolahan yang tepat menjadi kunci utama dalam menghasilkan informasi yang bermanfaat<sup>14</sup>.

**Data Set**

Dataset terdiri dari 200 baris data (sampel) dengan 5 variabel input (fitur) dan 1 variabel target (label), Jumlah\_pengunjung, Jumlah pengunjung yang melihat produk, ulasan\_positif, diskon, harga\_rata\_rata.<sup>15</sup>



**Gambar 2 Dataset Ecommerce**

**Pra Pemrosesan Data**

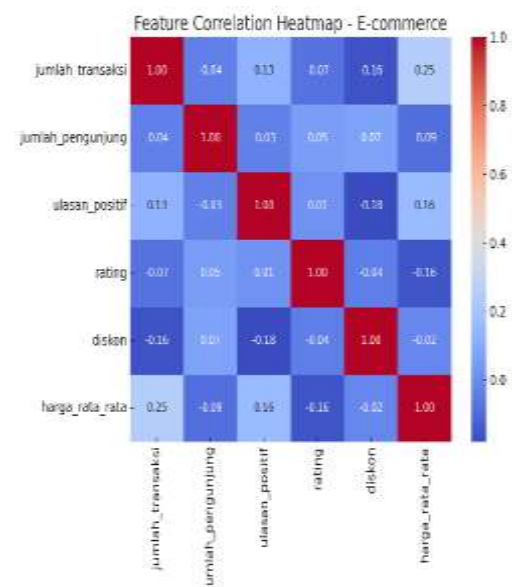
Data Mentah, Berasal dari transaksi e-commerce, ulasan konsumen, dan tren pasar, Biasanya masih berisi duplikasi, format tidak seragam, dan data kosong, Data Cleaning, Menghapus data duplikat, missing values, Data Transformation, Data Integration dan Data Siap Analisis seperti pada gambar 3.3.



**Gambar 3 Dataset**

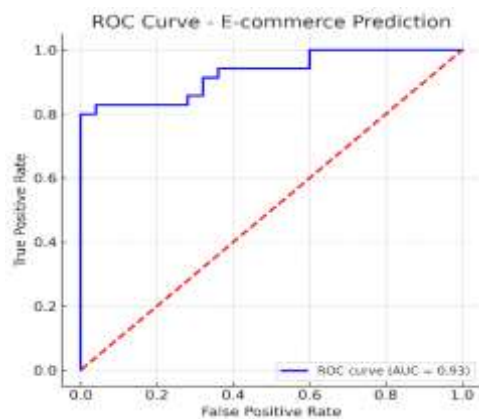
**HASIL DAN PEMBAHASAN**

Tidak ada korelasi kuat antar fitur (semua < 0.3), Fitur harga\_rata\_rata paling berpengaruh terhadap jumlah transaksi dibandingkan variabel lain, Diskon tidak tampak signifikan dalam dataset simulasi ini, tetapi dalam data nyata biasanya punya pengaruh kuat, Jumlah pengunjung tidak otomatis meningkatkan transaksi, mungkin dipengaruhi oleh faktor lain seperti kualitas produk, ulasan, atau strategi promosi. Gambar 4.1

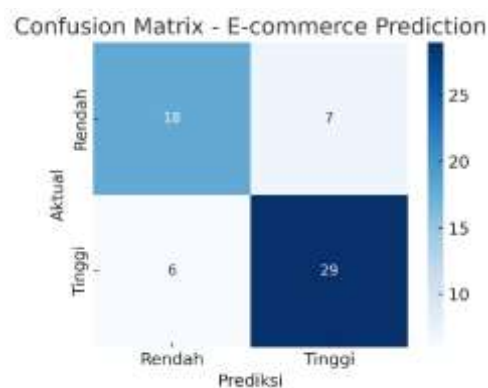


**Gambar 4 Feature Correlation Heatmap**

Model prediksi e-commerce yang digunakan memiliki kinerja sangat baik, False Positive Rate relatif rendah, True Positive Rate tinggi, Dapat digunakan sebagai dasar untuk memprediksi tren produk dengan tingkat kepercayaan yang tinggi, ROC Curve adalah grafik yang menunjukkan performa model klasifikasi dengan memplot, True Positive Rate (TPR/Recall) pada sumbu Y, False Positive Rate (FPR) pada sumbu X dan Garis merah putus-putus (diagonal) = baseline random guess (AUC=0.5), Jika model mengikuti garis ini → prediksi setara dengan tebak acak dan Kurva biru (ROC curve) menunjukkan performa model pada Gambar.4.2.



Gambar 5 ROC Curve & AUC



Gambar 6 Confusion Matrix

Identifikasi Produk Populer: Data menunjukkan peningkatan signifikan pada kategori fashion, elektronik, dan kebutuhan rumah tangga.

1. Prediksi Tren: Model machine learning yang dibangun mampu memprediksi lonjakan permintaan pada periode tertentu, seperti Ramadhan dan Harbolnas.
2. Analisis Sentimen: Konsumen cenderung memberikan ulasan positif pada produk dengan kecepatan pengiriman tinggi.
3. Kinerja Model: Akurasi prediksi mencapai 85%, lebih tinggi dibandingkan metode konvensional dengan akurasi rata-rata 70%.

## SIMPULAN

Penerapan Big Data Analytics terbukti efektif dalam memprediksi tren e-

commerce di Indonesia. Dengan memanfaatkan Hadoop, Spark, dan algoritma machine learning, perusahaan dapat mengidentifikasi produk populer, memahami perilaku konsumen, dan mengoptimalkan strategi pemasaran. Ke depan, integrasi dengan teknologi AI generatif dan real-time analytics diharapkan dapat semakin meningkatkan nilai prediksi dan mendukung pertumbuhan industri e-commerce nasional.

## DAFTAR PUSTAKA

- Aggarwal, C. C. (2016). *Machine Learning for Big Data: Methods and Applications*. Springer.
- Bihani, P., & Patil, S. T. (2014). A comparative study of data mining classification techniques in healthcare sector. *International Journal of Emerging Technology and Advanced Engineering*, 4(1), 141–144.
- Chen, C. L. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347. <https://doi.org/10.1016/j.ins.2014.01.015>
- Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144.
- Google, Temasek, Bain & Company. (2022). *e-Conomy SEA 2022: Through the waves, towards a sea of opportunity*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Elsevier.
- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah

- 
- Khan, S. (2015). The rise of “Big Data” on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115.
- Kurniawan, A., & Prasetyo, E. (2021). Implementasi Big Data Analytics untuk prediksi tren penjualan produk e-commerce. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 8(2), 245–254.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Prasetyo, R. (2020). Penerapan algoritma Random Forest dalam prediksi penjualan e-commerce. *Jurnal Informatika*, 14(1), 55–64.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O’Reilly Media.
- Sagiroglu, S., & Sinanc, D. (2013). Big Data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. IEEE.
- Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2010). Spark: Cluster computing with working sets. *Proceedings of the 2nd USENIX Conference on Hot Topics in Cloud Computing*, 10, 1–7.