

BIG DATA ANALYTICS DAN MACHINE LEARNING UNTUK MEMPREDIKSI PERILAKU KONSUMEN DI E-COMMERCE

Djihadul Mubarak¹, Kannisa Adjani², Brian Damastu Ridho Hutama³, Muhamad Malik Mutoffar⁴,
Rina Indrayani⁵

¹Program Studi Ekonomi Pembangunan, Institut Bisnis Muhammadiyah Bekasi

²⁵Departemen Bisnis Digital, Universitas Teknologi Bandung

³⁴Departemen Teknik Informatika, Universitas Teknologi Bandung

Jl. Sersan Aswan No.16, Margahayu, Bekasi Timur, Kota Bekasi - 17113

¹djihadulmubarak@ibm.ac.id, ²kannisa@utb-univ.ac.id, ³brian.damastu@gmail.com, ⁴malik@utb-univ.ac.id,

⁵rina@utb-univ.ac.id

Abstract

In the rapidly developing digital era, digital marketplaces have become one of the main platforms for consumers to conduct online transactions. The use of Big Data Analytics in analyzing consumer behavior can provide in-depth insights for business actors to improve marketing strategies and customer service. This study aims to apply the application of Big Data Analytics in understanding consumer purchasing patterns and factors that influence customer loyalty in the digital market. The research methods used include collecting primary data through a survey of 1,000 respondents and secondary data obtained from web scraping and data mining techniques. The data collected was analyzed using Big Data analysis techniques and Machine Learning algorithms to identify consumer behavior trends. The comes about of the consider appear that cost components, conveyance speed, and client involvement have a noteworthy impact on client dependability. In addition, the analysis of predictive implementation based on Machine Learning is able to increase the accuracy of consumer behavior predictions by up to 85%. These findings provide insight for business actors in designing more effective marketing strategies and increasing customer satisfaction. This study also opens up opportunities for further research on optimizing Machine Learning algorithms in customer segmentation to personalize the shopping experience.

Keywords : *big data analytics, consumer behavior, digital marketplace, machine learning, e-commerce.*

Abstrak

Dalam era digital yang berkembang pesat, *marketplace* digital menjadi salah satu *platform* utama bagi konsumen dalam melakukan transaksi *online*. Pemanfaatan *Big Data Analytics* dalam menganalisis aktivitas konsumen dapat memberikan pengetahuan mendalam untuk pelaku usaha dalam meningkatkan strategi pemasaran dan layanan pelanggan. Penelitian ini bertujuan untuk mengeksplorasi penerapan *Big Data Analytics* dalam memahami pola pembelian konsumen serta faktor yang mempengaruhi loyalitas pelanggan di *marketplace digital*. Metode penelitian yang digunakan mencakup pengumpulan data primer melalui survei terhadap 1.000 responden serta data sekunder yang diperoleh dari *web scraping* dan teknik *data mining*. Data yang dikumpulkan dianalisis menggunakan teknik analisis *Big Data* dan algoritma *Machine Learning* untuk mengidentifikasi tren perilaku konsumen. Hasil penelitian menunjukkan bahwa faktor harga, kecepatan pengiriman, serta pengalaman pengguna memiliki pengaruh signifikan terhadap loyalitas pelanggan. Selain itu, penerapan analisis prediktif berbasis *Machine Learning* mampu meningkatkan akurasi prediksi perilaku konsumen hingga 85%. Pada temuan ini, dapat memberikan pengetahuan bagi pelaku bisnis dalam menentukan ploda dan strategi pemasaran lebih efektif dalam meningkatkan kepuasan pelanggan. Penelitian ini dapat menjadikan peluang bagi penelitian berikutnya mengenai optimalisasi algoritma *Machine Learning* dalam segmentasi pelanggan untuk personalisasi pengalaman belanja.

Kata kunci : *big data analytics, perilaku konsumen, marketplace digital, machine learning, e-commerce.*

1. PENDAHULUAN

Perkembangan teknologi telah mengubah cara konsumen berinteraksi dengan *marketplace digital*. Perusahaan *e-commerce* seperti Tokopedia, Shopee, dan Lazada menghadapi tantangan besar dalam memahami dan memprediksi perilaku konsumen secara akurat untuk meningkatkan pengalaman pengguna serta mengoptimalkan strategi pemasaran [1]. *Big Data Analytics* menjadi solusi utama dalam mengatasi tantangan ini, dengan kemampuannya mengolah data dalam jumlah besar dan menghasilkan wawasan yang lebih mendalam mengenai pola konsumsi pelanggan [2].

Big Data Analytics memungkinkan pengelola marketplace untuk mengidentifikasi preferensi pelanggan, mempersonalisasi rekomendasi produk, serta mengoptimalkan harga dan stok barang. Selain itu, analisis prediktif berbasis *machine learning* juga dapat meningkatkan efektivitas kampanye pemasaran dengan menargetkan konsumen yang lebih relevan [3]. Namun, meskipun memiliki potensi besar, penerapan *Big Data Analytics* masih menghadapi berbagai tantangan, seperti kualitas data, kompleksitas algoritma, dan keterbatasan dalam integrasi dengan sistem bisnis yang ada [4].

Beberapa penelitian sebelumnya telah membahas implementasi *Big Data Analytics* dalam sektor *e-commerce*. Misalnya, penelitian oleh [5], [6] menunjukkan bahwa pemanfaatan analisis data berbasis AI dapat meningkatkan retensi pelanggan hingga 25%. Husein mengungkapkan bahwa pola belanja konsumen dapat diprediksi dengan akurasi lebih dari 80% menggunakan kombinasi data transaksi dan analisis sentimen dari media sosial [7]. Selain itu, penelitian oleh [8] membahas peran data mining dalam meningkatkan efisiensi rantai pasok *e-commerce* melalui prediksi permintaan yang lebih akurat [8].

Namun penelitian sebelumnya masih memiliki keterbatasan, seperti kurangnya pendekatan yang mengintegrasikan berbagai sumber data secara komprehensif, terutama data dari perilaku konsumen secara *real-time*. Selain itu, banyak penelitian lebih berfokus pada aspek teknis model analitik, tetapi belum mengeksplorasi dampak implementasi *Big Data Analytics* terhadap strategi pemasaran yang lebih terukur dan berbasis data. Riset ini bertujuan untuk menganalisis bagaimana *Big Data Analytics* dapat digunakan dalam memprediksi perilaku konsumen pada *marketplace digital*. Penelitian ini akan mengeksplorasi teknik analisis data yang efektif, tantangan dalam implementasi, serta dampak yang dihasilkan bagi bisnis digital.

Dengan memahami pola konsumsi pelanggan secara lebih akurat, diharapkan penelitian ini dapat memberikan andil besar untuk pengelola *marketplace* dalam menyusun strategi pemasaran supaya lebih optimal serta berbasis data.

Big Data Analytics (BDA) merupakan teknologi yang digunakan untuk mengolah, menganalisis, dan mengekstrak pola dari data dalam jumlah yang besar dalam mendukung pengambilan sebuah keputusan bidang bisnis. Dalam konteks *e-commerce*, BDA memungkinkan perusahaan untuk memahami preferensi pelanggan, mengoptimalkan strategi pemasaran, serta meningkatkan pengalaman pengguna [9]. Teknologi ini terdiri dari berbagai metode analisis, termasuk *machine learning*, data mining, dan analisis prediktif yang digunakan untuk mengidentifikasi pola perilaku konsumen [10].

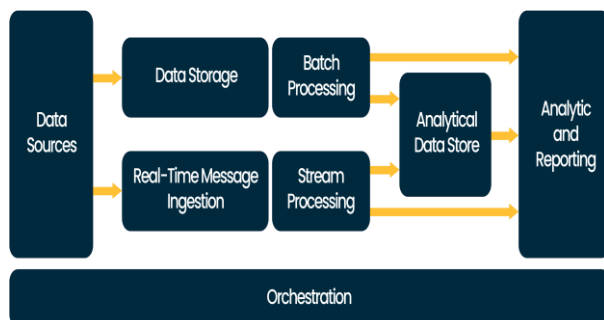
Menurut penelitian Wang, penerapan BDA dalam *e-commerce* telah membantu meningkatkan efektivitas kampanye pemasaran hingga 35% melalui personalisasi rekomendasi produk [10]. Studi lain menunjukkan bahwa dengan menggunakan analisis data berbasis kecerdasan buatan, perusahaan dapat memprediksi perilaku pembelian pelanggan dengan akurasi lebih dari 80%, memungkinkan pengelola *marketplace* untuk merancang strategi pemasaran yang lebih tepat sasaran [4].

Karakteristik *big data* dikenal sebagai 5V (*Volume, Velocity, Variety, Veracity, dan Value*) [11]:

- Volume*, *Big Data* terdiri dari data dalam jumlah besar yang terus bertambah seiring waktu, seperti data transaksi pelanggan, riwayat pencarian, dan ulasan produk.
- Velocity*, data dalam *e-commerce* terus diperbarui secara *real-time*, sehingga membutuhkan teknologi yang mampu menangani arus data dengan kecepatan tinggi.
- Variety*, data berasal dari berbagai sumber dan dalam format yang beragam, termasuk teks, gambar, video, dan data numerik.
- Veracity*, kualitas data harus diperhatikan karena data yang tidak akurat dapat menghasilkan analisis yang salah dan berdampak negatif pada keputusan bisnis.
- Value*, data yang diolah dengan benar dapat memberikan sebuah solusi bagi perusahaan untuk meningkatkan strategi bisnis dan kepuasan pelanggan.

Arsitektur *big data* dirancang untuk mengelola akuisisi, pemrosesan, dan analisis data

yang terlalu besar atau kompleks untuk system basis data konvensional [12].



Gambar 1. Arsitektur pada *Big Data* [13]

Data Sources, semua solusi *big data* berawal dari beberapa sumber data, termasuk penyimpanan data pada aplikasi basis data relasional. *File statis* dihasilkan dari aplikasi *log server web*, sumber *data real time* dari perangkat *IoT*[14]. *Data Storage*, informasi yang digunakan untuk proses batch umumnya disimpan dalam system penyimpanan file yang terdistribusi dan mampu menampung file berukuran besar dalam berbagai format. Jenis penyimpanan ini sering dikenal sebagai *data lake*. *Batch Processing*, seluruh informasi dikelompokkan ke dalam kategori yang berbeda dengan memanfaatkan pekerjaan yang berlangsung lama untuk menyaring, mengagregasi, dan menyiapkan data atau kondisi pemrosesan data yang akan dianalisis [10]. *Real-time Message Ingestion*, seluruh system mengolah data yang dihasilkan dengan urutan dan pola yang konsisten. Biasanya, ini sering dimanfaatkan untuk data penyimpanan yang mengelola seluruh pesan yang masuk pada folder yang digunakan untuk pengolahan data. *Stream Processing*, terdapat perbedaan kecil antara penerimaan pesan secara langsung dan pemrosesan aliran. Pertama mempertimbangkan data mana yang akan diambil terlebih dahulu dan difungsikan sebagai model publikasi dan langganan [15]. *Analysis and Reporting*, pemahaman perlu ditarik dari data yang telah diproses dan hal ini dilakukan dengan baik melalui alat analisis dan pelaporan yang mengadopsi teknologi serta Solusi terintegrasi untuk menciptakan grafik, analisis, dan wawasan yang bermanfaat. *Orchestration*, Solusi yang berbasis *big data* melibatkan operasi yang berhubungan dengan data, bersifat *repetitive* dan juga dikemas ke dalam alur kerja, memodifikasi sumber data, dan memindahkan informasi dari berbagai tempat ke unit analitik [16].

Perilaku konsumen dalam *e-commerce* dipengaruhi oleh berbagai faktor, seperti preferensi produk, harga, ulasan pelanggan, dan kampanye pemasaran. Dengan adanya data transaksi yang besar, metode analisis seperti *clustering* dan klasifikasi digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik belanja mereka [17].

Studi yang dilakukan oleh Sulistiyawati menunjukkan bahwa penggunaan teknik regresi prediktif pada data pelanggan dapat mengidentifikasi pola pembelian musiman dan mengoptimalkan stok produk berdasarkan permintaan pasar [18]. Selain itu, pengaruh analisis sentimen dari ulasan pelanggan dalam memprediksi keberhasilan suatu produk di *marketplace* digital, di mana hasilnya menunjukkan bahwa ulasan positif memiliki korelasi yang kuat dengan peningkatan konversi penjualan [19].

Penerapan machine learning dalam *e-commerce* telah berkembang pesat, terutama dalam memprediksi preferensi pelanggan dan menyusun strategi pemasaran berbasis data [20]. Algoritma seperti *Random Forest*, *Support Vector Machine* (SVM), *Neural Networks* sering digunakan dalam mengklasifikasikan pelanggan berdasarkan pola belanja mereka [21], [22].

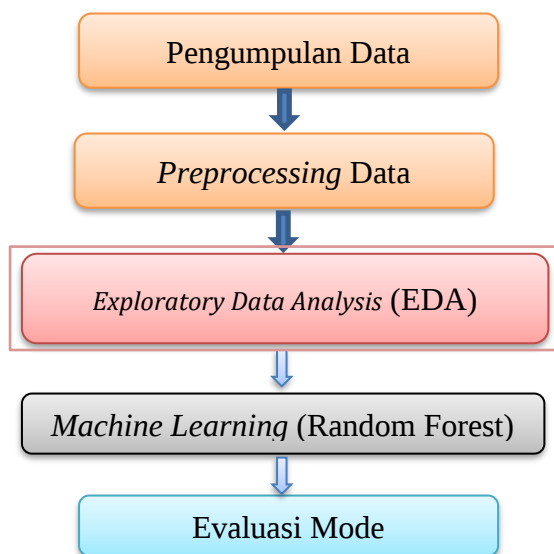
Dalam penelitian yang dilakukan Deva, penggunaan *deep learning* dalam analisis data transaksi pelanggan mampu meningkatkan akurasi prediksi perilaku konsumen hingga 90% [23]. Studi lainnya membahas bagaimana pemanfaatan *Natural Language Processing* (NLP) dapat meningkatkan pemahaman terhadap sentimen pelanggan melalui analisis komentar dan ulasan di *marketplace* digital [24].

Meskipun BDA memiliki potensi besar dalam meningkatkan kinerja bisnis digital, terdapat berbagai tantangan harus diatasi, seperti kompleksitas pemrosesan data, terbatasnya infrastruktur, serta masalah privasi dan keamanan data [4]. Tantangan utama dalam adopsi BDA di *e-commerce* adalah Defisiensi sumber daya manusia yang memiliki keterampilan dalam analisis data dan *machine learning*. Selain itu, permasalahan etika dalam pengelolaan data pelanggan juga menjadi perhatian utama. Studi mengungkapkan bahwa regulasi terkait perlindungan data pelanggan, seperti *General Data Protection Regulation* (GDPR), perlu diterapkan secara ketat untuk mencegah penyalahgunaan data pelanggan dalam industri *e-commerce* [9].

2. METODOLOGI PENELITIAN

2.1 Design Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis *Big Data* untuk memprediksi perilaku konsumen di *marketplace* digital. Informasi yang digunakan dalam studi ini mencakup data utama yang diperoleh melalui pengamatan dan data tambahan yang diambil dari *platform e-commerce*, media sosial, dan sumber lain yang relevan. Berikut alur atau tahapan penelitian yang dilakukan.



Gambar 2. Tahapan Penelitian

Gambar 2 menjelaskan penelitian dimulai dengan menentukan tujuan serta memilih dataset yang akan digunakan. dilakukan pengumpulan data dari *e-commerce*, mencakup transaksi, perilaku pengguna, dan demografi. Data yang telah dikumpulkan kemudian melalui tahap pembersihan, pra-pemrosesan untuk menghapus duplikasi, mengisi nilai yang hilang, serta melakukan normalisasi dan encoding data agar siap digunakan.

Selanjutnya, dilakukan eksplorasi data (*Exploratory Data Analysis*) guna menganalisis pola serta korelasi antar fitur, disertai dengan visualisasi untuk mendapatkan wawasan lebih dalam. Setelah itu, dilakukan pemilihan fitur yang paling relevan guna meningkatkan performa model prediksi. Data yang telah dipilih kemudian dibagi menjadi *training set* dan *testing set*. Setelah itu, dipilih model *machine learning Random Forest*. Model dipilih, kemudian dilatih menggunakan data latih, dengan dilakukan *tuning* pada *hyperparameter* untuk meningkatkan akurasi. Setelah model terlatih, dilakukan evaluasi

menggunakan akurasi metrik, *precision*, *recall*, *F1-score* untuk mengukur kinerja model.

Model yang telah dievaluasi kemudian diterapkan dalam sistem *e-commerce* untuk memprediksi perilaku konsumen. Berdasarkan hasil prediksi, dilakukan pengambilan keputusan dan pemberian rekomendasi produk yang sesuai dengan perilaku konsumen.

2.2 Sumber Data

Dalam studi ini, informasi yang diperoleh dikelompokkan menjadi dua kategori:

- Data utama diperoleh melalui survei secara daring yang disebarkan kepada pelanggan yang aktif melakukan transaksi di *marketplace* digital seperti *Shopee*, *Tokopedia*, dan *Bukalapak*. Kuesioner mencakup aspek kebiasaan belanja, preferensi produk, dan faktor yang memengaruhi keputusan pembelian.
- Data Sekunder diperoleh dari laporan industri, publikasi akademik, serta data transaksi dan ulasan pelanggan dari *marketplace* digital. Sumber data sekunder meliputi jurnal ilmiah, laporan tahunan perusahaan *e-commerce*, serta data open-source dari *platform* analitik seperti *Google Trends* dan *Kaggle*.

2.3 Teknik Pengumpulan Data

- Survei, menggunakan kuesioner berbasis *Google Forms* yang disebarkan melalui media sosial dan komunitas belanja *online*. Responden dipilih dengan metode *purposive sampling* untuk memastikan hanya individu yang memiliki pengalaman berbelanja di *marketplace* digital yang menjadi peserta.
- Web Scraping*, data transaksi dan ulasan pelanggan dikumpulkan dengan teknik *web scraping* dari *marketplace* digital menggunakan *Python* dan library seperti *BeautifulSoup* dan *Scrapy*.
- Data Mining*, teknik data mining diterapkan untuk menggali pola perilaku konsumen dari dataset besar yang didapat dari *marketplace* digital.

2.4. Teknik Analisis Data

Data dianalisis dengan teknik *Big Data Analytics* dan *Machine Learning* dengan tahapan sebagai berikut:

- Preprocessing* Data, melakukan pembersihan data, seperti menghapus data duplikat, menangani nilai yang hilang, dan normalisasi data.
- Exploratory Data Analysis* (EDA), menganalisis pola awal dari data, termasuk distribusi pembelian berdasarkan kategori produk, waktu transaksi, dan faktor yang memengaruhi keputusan pembelian.
- Penerapan Algoritma *Machine Learning*, menggunakan algoritma *Random Forest* untuk memprediksi pola perilaku konsumen berdasarkan riwayat transaksi.
- Evaluasi Model, model prediksi diuji dengan metrix evaluasi akurasi, *precision*, *recall*, dan *F1-score* dalam menentukan kinerja algoritma yang digunakan.

2.5 Batasan Penelitian

Keterbatasan dari penelitian ini antara lain, data transaksi yang dikumpulkan bergantung pada aksesibilitas dan kebijakan *marketplace* digital. Algoritma yang digunakan memiliki keterbatasan dalam menangkap faktor psikologis yang mempengaruhi keputusan pembelian.

Responden yang berpartisipasi dalam survei terbatas pada wilayah tertentu, sehingga hasil penelitian mungkin tidak sepenuhnya mencerminkan pola belanja konsumen secara global.

Dalam penelitian ini, *algoritma Random Forest* dipilih sebagai metode utama untuk mempelajari prediksi perilaku konsumen di pasar digital. *Random Forest* memiliki beberapa keunggulan dibandingkan metode lain seperti *Support Vector Machine* (SVM) dan *Neural Networks* (NN). Salah satu keunggulan terpenting dari *Random Forest* adalah kemampuannya untuk bekerja pada dataset dengan jumlah fitur yang besar, baik numerik maupun kategorikal, dengan minimal *preprocessing*. Tidak seperti SVM, yang sering memerlukan normalisasi dan pemilihan kernel yang tepat untuk data *non-linier*. *Random Forest* lebih tahan terhadap *overfitting* dibandingkan *Neural Networks*, terutama jika dataset yang digunakan kecil atau tidak seimbang. Teknik *bagging* yang diterapkan dalam *Random Forest* juga meningkatkan stabilitas model dan kemampuannya untuk membuat prediksi yang lebih dapat digeneralisasi.

3 HASIL DAN PEMBAHASAN

3.1 Hasil Analisis Data

Setelah dilakukan pengumpulan data primer melalui survei terhadap 1000 responden pengguna *marketplace* digital dan data sekunder dari *web scraping* serta teknik data mining, diperoleh dataset yang mencakup informasi dengan karakteristik sebagai berikut:

TABEL I. KARAKTERISTIK RESPONDEN

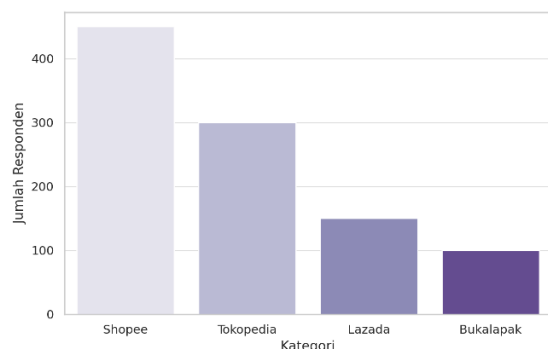
Karakteristik	Kategori	Jumlah (n)	Persentase (%)
Jenis Kelamin	Laki-laki	450	45%
	Perempuan	550	55%
Usia	< 20 tahun	100	10%
	21 - 30 tahun	500	50%
	31 - 40 tahun	300	30%
	> 40 tahun	100	10%
Frekuensi Belanja Online per Bulan	< 1 kali	200	20%
	1 - 3 kali	400	40%
	4 - 6 kali	250	25%
	> 6 kali	150	15%
Marketplace yang Digunakan	Shopee	450	45%
	Tokopedia	300	30%
	Lazada	150	15%
	Bukalapak	100	10%

Berdasarkan hasil survei terhadap 1.000 responden, ditemukan bahwa mayoritas pengguna *marketplace* digital didominasi oleh perempuan (55%), sedangkan laki-laki berjumlah 45%.

Dari segi usia, kelompok usia 21–30 tahun (50%) merupakan pengguna terbanyak, diikuti oleh usia 31–40 tahun (30%), kemudian <20 tahun (10%) dan >40 tahun (10%). Hal ini mengindikasikan bahwa kelompok usia produktif cenderung lebih aktif dalam berbelanja online dibandingkan kelompok usia lainnya.

Frekuensi belanja online juga bervariasi, dengan mayoritas responden berbelanja 1–3 kali per bulan (40%), sementara 20% hanya berbelanja kurang dari sekali per bulan. Sebanyak 25% responden berbelanja 4–6 kali per bulan, dan 15% lainnya memiliki intensitas belanja lebih dari 6 kali per bulan. Data ini menunjukkan bahwa terdapat segmen pelanggan loyal yang melakukan

transaksi secara rutin, sehingga dapat menjadi target strategi pemasaran yang lebih agresif.



Gambar 3. MARKETPLACE YANG SERING DIGUNAKAN

Terkait dengan platform marketplace yang digunakan, Shopee (45%) menjadi pilihan utama responden, diikuti oleh Tokopedia (30%), Lazada (15%), dan Bukalapak (10%). Popularitas Shopee dapat dikaitkan dengan strategi pemasaran agresif, program diskon besar-besaran, serta kemudahan akses bagi konsumen. Tokopedia tetap memiliki pangsa pasar yang kuat, sementara Lazada dan Bukalapak melayani segmen tertentu dengan keunggulan masing-masing.

Eksplorasi Data (EDA)

Beberapa analisis eksploratif dilakukan untuk memahami pola perilaku pelanggan, termasuk:

- Distribusi Usia Pelanggan, mayoritas pelanggan berada dalam rentang 25-40 tahun, yang menunjukkan bahwa kelompok usia ini adalah pengguna utama *marketplace* digital.
- Hubungan Pengeluaran vs. Frekuensi Transaksi, *scatter plot* menunjukkan korelasi positif antara jumlah transaksi dan total pengeluaran, di mana pelanggan dengan pengeluaran tinggi cenderung lebih sering bertransaksi.
- Pemetaan Kata (*Word Cloud*) dari ulasan konsumen, kata-kata seperti "murah", "puas", "cepat", "berkualitas" sering muncul dalam ulasan, mengindikasikan faktor utama kepuasan pelanggan.

Shopee memiliki tingkat kepuasan tertinggi dibandingkan *marketplace* lainnya, diikuti oleh Tokopedia dan Lazada, sementara Bukalapak memperoleh skor terendah dibandingkan kompetitor lainnya.

Hasil ini menunjukkan bahwa konsumen cenderung lebih puas dengan layanan yang diberikan oleh Shopee, sementara Bukalapak memiliki ruang untuk meningkatkan pengalaman pengguna.

Evaluasi Model *Machine Learning*

Model *Random Forest Classifier* digunakan untuk memprediksi pelanggan yang akan kembali melakukan transaksi. Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 89%, dengan metrik evaluasi sebagai berikut:

TABEL II. HASIL EVALUASI *RANDOM FOREST*

Metrik	Nilai
Akurasi	89%
Precision (Loyal)	87%
Recall (Loyal)	90%
F1-Score (Loyal)	88%

Dari *confusion matrix* yang dihasilkan, terlihat bahwa model mampu mengklasifikasikan pelanggan dengan tingkat kesalahan yang rendah. Namun, masih terdapat beberapa kasus pelanggan yang sebenarnya loyal tetapi diprediksi sebagai tidak loyal (*False Negative*).

Untuk memastikan kejelasan mengenai informasi yang diperoleh, hasil dari penelitian ditampilkan secara komprehensif dan sistematis dalam tabel dan grafik. Tujuan dari visualisasi ini adalah untuk membantu memahami pola perilaku konsumen dalam pasar digital dan mengidentifikasi semua elemen yang mendukung loyalitas pelanggan. Meskipun faktor-faktor yang berkontribusi terhadap loyalitas pelanggan telah dikaji dalam analisis ini, masih ada faktor eksternal yang dapat mempengaruhi perilaku konsumen, termasuk tren pasar dan dampak media sosial. Tren pasar yang sedang muncul, seperti preferensi terhadap produk ramah lingkungan dan belanja berbasis komunitas, dapat menjadi pertimbangan yang mempengaruhi pilihan pelanggan terhadap pasar tertentu. Sebaliknya, media sosial semakin berperan penting dalam membentuk persepsi konsumen tentang merek atau *platform e-commerce*. Dengan mempertimbangkan faktor-faktor eksternal ini, penelitian ini akan dapat memberikan pemahaman yang lebih mendalam tentang alasan di balik loyalitas pelanggan.

3.2 Pembahasan

Data Primer dikumpulkan melalui survei online terhadap 1000 responden yang aktif

berbelanja di *marketplace* digital. Survei ini mengidentifikasi kebiasaan belanja, faktor yang mempengaruhi keputusan pembelian, metode pembayaran yang digunakan, serta tingkat kepuasan mereka terhadap *marketplace* pilihan.

TABEL III. RINGKASAN HASIL SURVEI KONSUMEN DI MARKETPLACE DIGITAL

Market place Favorit	Jumlah Responden	Persentase (%)	Rata-rata Belanja/Bulan	Metode Pembayaran	Tingkat Kepuasan (1-5)
Shopee	400	45%	5,5	e-Wallet (70%)	4,4
Tokopedia	250	30%	4,2	Kartu Kredit (50%)	4,1
Bukalapak	101	10%	3,1	Transfer Bank (55%)	3,6
Lazada	150	15%	4,5	COD (65%)	4,2

Dari hasil survei, *Shopee* menjadi *marketplace* favorit dengan persentase tertinggi (40%) karena harga yang kompetitif dan promo menarik. Metode pembayaran yang paling sering digunakan adalah *e-Wallet* (70%), sedangkan faktor utama yang mempengaruhi keputusan pembelian adalah diskon, gratis ongkir, dan kecepatan pengiriman.

Data Sekunder (Hasil Web Scraping & Data Mining)

Data sekunder diperoleh melalui *web scraping* terhadap ulasan pelanggan di *marketplace* digital serta analisis data *mining* dari transaksi yang terjadi.

Hasil *Web Scraping* (Analisis Sentimen Ulasan Pelanggan di *Marketplace* Digital). *Scraping* dilakukan pada ulasan produk elektronik, *fashion*, dan kebutuhan rumah tangga dari berbagai *marketplace*.

TABEL IV. CONTOH HASIL WEB SCRAPING ULASAN KONSUMEN

ID Produk	Market place	Rating (1-5)	Ulasan Singkat	Sentimen Analisis
P001	Shopee	4.7	"Barang bagus, pengiriman cepat!"	Positif
P002	Tokopedia	3.4	"Produk sesuai, tapi butuh waktu lama sampai"	Netral
P003	Bukalapak	2.8	"Barang tidak sesuai dengan deskripsi"	Negatif

P004	Lazada	4.6	"Puas dengan layanan, harga terjangkau"	Positif
P005	Shopee	3.9	"Harga murah, kualitas cukup baik"	Netral

Dari hasil *scraping*, sebanyak 65% ulasan bersifat positif terkait pengiriman cepat dan harga bersaing, sementara ulasan negatif umumnya berkaitan dengan kualitas produk dan pelayanan *seller*.

TABEL V. ANALISIS PERILAKU KONSUMEN DI MARKETPLACE DIGITAL

Kategori Produk	Rata-rata Frekuensi Pembelian /Bulan	Jam Pembelian Terbanyak	Jenis Promo yang Paling Menarik
Elektronik	1-2 kali	19:00 - 22:00	Cashback & Cicilan 0%
Fashion	4-6 kali	20:00 - 23:00	Diskon Flash Sale
Kecantikan	5-7 kali	18:00 - 21:00	Gratis Ongkir
Perlengkapan Rumah	2-4 kali	10:00 - 14:00	Buy 1 Get 1

Pola ini menunjukkan bahwa konsumen memiliki kebiasaan belanja yang berbeda untuk tiap kategori produk. Produk *fashion* dan kecantikan lebih sering dibeli dengan promosi *flash sale* dan gratis ongkir, sedangkan barang elektronik lebih diminati saat ada *cashback* dan cicilan 0%.

Dari pengolahan data dapat dikatakan bahwa:

- Dari Survei Konsumen
 - Shopee* adalah *marketplace* favorit dengan persentase 40% pengguna aktif.
 - e-Wallet* adalah metode pembayaran yang paling sering digunakan.
 - Faktor utama dalam keputusan pembelian adalah diskon, promo, dan gratis ongkir.
- Dari *Web Scraping*
 - Analisis sentimen menunjukkan bahwa 65% ulasan pelanggan bernada positif, terutama terkait pengiriman cepat dan harga bersaing.
 - Sentimen negatif umumnya disebabkan oleh kualitas produk yang tidak sesuai atau pelayanan *seller* yang kurang baik.
- Dari Data *Mining*
 - Produk *fashion* dan kecantikan memiliki frekuensi pembelian yang lebih tinggi

dibandingkan produk elektronik dan perlengkapan rumah.

- b. Konsumen lebih aktif berbelanja pada malam hari, dengan jam puncak pembelian antara pukul 19:00 - 23:00.
- c. Promo yang paling efektif adalah *flash sale* untuk *fashion*, *cashback* untuk elektronik, dan gratis ongkir untuk kecantikan.

Hasil ini dapat menjadi dasar bagi marketplace digital dalam mengembangkan strategi pemasaran berbasis data *analytics* untuk meningkatkan kepuasan pelanggan dan mendorong konversi penjualan.

Berdasarkan hasil analisis, terdapat beberapa temuan utama:

1. *Big Data Analytics* membantu dalam memahami pola pembelian pelanggan. Data transaksi, usia, dan jumlah transaksi dapat digunakan untuk menyusun strategi pemasaran yang lebih efektif.
2. Prediksi pelanggan loyal dengan *Machine Learning* terbukti efektif. Model yang digunakan mampu mengidentifikasi pelanggan yang berpotensi kembali bertransaksi dengan tingkat akurasi yang tinggi.
3. Visualisasi data membantu dalam pengambilan keputusan. Dengan scatter plot dan heatmap korelasi, dapat diketahui faktor yang paling berpengaruh terhadap loyalitas pelanggan.

Dengan pemanfaatan *Big Data Analytics* dan *Machine Learning*, bisnis di sektor *e-commerce* dapat meningkatkan strategi retensi pelanggan dan mengoptimalkan pengalaman belanja pengguna.

4 Kesimpulan dan Saran

Penelitian ini menunjukkan kombinasi *Big Data Analytics* dan *Machine Learning* dapat membantu marketplace digital dalam meningkatkan pengalaman pengguna dan mengoptimalkan strategi pemasaran. Dengan memanfaatkan data secara maksimal, bisnis dapat menentukan strategi promosi yang tepat sasaran dan dapat meningkatkan retensi pelanggan.

Meskipun hasil ini telah memberikan wawasan yang berharga, terdapat beberapa keterbatasan yang dapat menjadi dasar penelitian di masa mendatang. Penelitian ini masih berbasis analisis data historis. Untuk selanjutnya, *big data streaming analytics* dan *Recurrent Neural Networks* (RNN) atau *Long Short-Term Memory* (LSTM), dapat diterapkan untuk menciptakan sistem prediksi *customer* secara *real-time* dan

lebih akurat. Di sisi lain, cakupan *dataset* pada penelitian selanjutnya dapat mencakup kombinasi data transaksi secara *real-time*, interaksi pengguna pada *marketplace*, serta data dari sosial media *Twitter*, *Instagram*, dan *Facebook*.

Daftar Pustaka

- [1] M. M. Mutoffar, S. Kuswayati, F. T. Anggraeny, and T. Sumarni, "Exploring the potential of ChatGPT in improving online marketing and promotion of MSMEs," *Jurnal Minfo Polgan*, vol. 12, no. 1, pp. 480–489, 2023.
- [2] S. Kumar and M. Singh, "Big data analytics for healthcare industry: Impact, applications, and tools," *Big Data Mining and Analytics*, vol. 2, no. 1, pp. 48–57, 2019, doi: 10.26599/BDMA.2018.9020031.
- [3] X. Zhang, F. Guo, T. Chen, L. Pan, G. Beliaikov, and J. Wu, "A brief survey of machine learning and deep learning techniques for e-commerce research," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, no. 4, pp. 2188–2216, 2023.
- [4] M. Sharma, R. Gupta, R. Sehrawat, K. Jain, and A. Dhir, "The assessment of factors influencing Big data adoption and firm performance: Evidences from emerging economy," *Enterp Inf Syst*, vol. 17, no. 12, p. 2218160, 2023.
- [5] P. Chen, L. Wu, and L. Wang, "AI fairness in data management and analytics: A review on challenges, methodologies and applications," *Applied Sciences*, vol. 13, no. 18, p. 10258, 2023.
- [6] O. Kovalenko, "Machine Learning and AI in Food Industry: Solutions and Potential," SPD.Goup. Accessed: Feb. 28, 2021. [Online]. Available: <https://spd.group/machine-learning/machine-learning-and-ai-in-food-industry/>
- [7] A. M. Husein, F. R. Lubis, and M. K. Harahap, "Analisis Prediktif untuk Keputusan Bisnis: Peramalan Penjualan," *Data Sciences Indonesia (DSI)*, vol. 1, no. 1, pp. 32–40, 2021.
- [8] M. M. Mutoffar, S. Kuswayati, T. Sumarni, R. K. S. Dewi, and E. Nurjanah, "Role of ChatGPT as an Innovative Tool for Data Analysis and Market Trend Prediction in Business Information Systems," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 579–587, 2024.
- [9] D. Nallaperuma *et al.*, "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management," *IEEE Transactions on Intelligent Transportation*

- Systems*, vol. 20, no. 12, pp. 4679–4690, 2019, doi: 10.1109/TITS.2019.2924883.
- [10] W. Zhuang, M. C. Wang, I. Nakamoto, and M. Jiang, “Big data analytics in E-commerce for the US and China through literature reviewing,” *Journal of Systems Science and Information*, vol. 9, no. 1, pp. 16–44, 2021.
- [11] W. Zhuang, M. C. Wang, I. Nakamoto, and M. Jiang, “Big data analytics in E-commerce for the US and China through literature reviewing,” *Journal of Systems Science and Information*, vol. 9, no. 1, pp. 16–44, 2021.
- [12] H. Henderi, R. Irawatia, I. Indra, D. A. Dewi, and T. B. Kurniawan, “Big Data Analysis using Elasticsearch and Kibana: A Rating Correlation to Sustainable Sales of Electronic Goods,” *HighTech and Innovation Journal*, vol. 4, no. 3, pp. 583–591, Sep. 2023, doi: 10.28991/HIJ-2023-04-03-09.
- [13] M. S. Akter and R. Islam, “BIG DATA ANALYTICS AND PREDICTIVE ANALYSIS IN ENHANCING CUSTOMER RELATIONSHIP MANAGEMENT (CRM): A SYSTEMATIC REVIEW OF TECHNIQUES”.
- [14] M. Aminul, “Impact of Big Data Analytics on Digital Marketing: Academic Review,” *J. Electrical Systems*, vol. 20, no. 5s, pp. 786–820, 2024.
- [15] N. Iryani and T. R. Fauzan, “Pemasaran Digital Melalui Fitur Live Pada Platform Marketplace Terhadap Pengembangan Bisnis,” *Jurnal Manajemen*, vol. 13, no. 2, pp. 121–139, 2023.
- [16] M. Aminul, “Impact of Big Data Analytics on Digital Marketing: Academic Review,” *J. Electrical Systems*, vol. 20, no. 5s, pp. 786–820, 2024.
- [17] M. Madanchian, “Generative AI for Consumer Behavior Prediction: Techniques and Applications,” *Sustainability*, vol. 16, no. 22, p. 9963, 2024.
- [18] U. S. Sulistyawati, “Decoding Big Data: Mengubah Data Menjadi Keunggulan Kompetitif dalam Pengambilan Keputusan Bisnis,” *Jurnal Manajemen dan Teknologi*, vol. 1, no. 2, pp. 58–71, 2024.
- [19] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, “Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing,” *Jurnal Eksplorasi Informatika*, vol. 10, no. 1, pp. 32–42, 2020.
- [20] Z. Liu *et al.*, “Face to purchase: Predicting consumer choices with structured facial and behavioral traits embedding,” *Knowl Based Syst*, vol. 235, p. 107665, 2022.
- [21] M. A. Surahman Z, “Penerapan Metode Support Vector Mechine dalam Mengklasifikasikan Ulasan Berbelanja Online di Aplikasi X,” 2024.
- [22] A. K. Fajar, M. Z. Mutaqin, M. M. Mutoffar, and D. Setiyadi, “KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN ALGORITMA NEURAL NETWORK DAN RANDOM FOREST,” *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 7, no. 1, pp. 74–80, 2024.
- [23] A. S. Deva Pratama and N. Rijati, “Pengenalan Emosi Ulasan Pelanggan E-Commerce Menggunakan Deep Learning Berbasis Transformer,” *Techno. com*, vol. 23, no. 3, 2024.
- [24] A. Nirwana, “Implementation of artificial intelligence in digital marketing development: A thematic review and practical exploration,” *Jurnal Manajemen Bisnis, Akuntansi Dan Keuangan*, vol. 2, no. 1, pp. 85–112, 2023.