

Desain Sistem Big Data untuk Deteksi Pola Perilaku Konsumen dengan Menggunakan Algoritma Machine Learning

Mariska Putri Pratiwi¹, Erni Yanti Natalia², Rahmat Fauzi³

^{1,3}Program Studi Teknik Informatika, Universitas Putera Batam, Indonesia

²Program Studi Akutansi, Universitas Putera Batam, Indonesia

Informasi Artikel

Terbit: Januari 2026

Kata Kunci:

Big data
Perilaku
Machine learning
e-commerce
Arsitektur

ABSTRAK

Penelitian ini mengembangkan sistem berbasis big data untuk mendeteksi pola perilaku konsumen menggunakan algoritma machine learning. Sistem ini memproses data besar yang diperoleh dari berbagai sumber seperti transaksi e-commerce, aplikasi mobile, dan media sosial. Fokus utama penelitian adalah menggabungkan teknik *data integration*, *preprocessing*, serta penerapan model *clustering* dan *classification* untuk memahami pola konsumen. Hasil analisis kemudian divisualisasikan melalui dashboard interaktif untuk membantu pengambilan keputusan bisnis berbasis data. Eksperimen dilakukan menggunakan dataset transaksi dan perilaku pengguna, dengan tujuan meningkatkan efisiensi dan akurasi dalam merancang strategi pemasaran yang tepat sasaran.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Mariska Putri Pratiwi,
Email: mrskaptrw@gmail.com

1. PENDAHULUAN

Era digital menghasilkan pertumbuhan data yang sangat cepat melalui interaksi pelanggan sistem digital seperti e-commerce, media sosial, dan aplikasi mobile. Data tersebut mencakup catatan transaksi, perilaku klik, penilaian produk, serta ulasan konsumen. Volume, variasi, dan kecepatan data yang tinggi menantang sistem tradisional karena tidak mampu merangkul pola yang tersembunyi dalam kumpulan data besar ini[1]. *Machine learning* kini menjadi solusi yang memungkinkan identifikasi pola dalam data besar secara otomatis dan akurat. Sistem yang dirancang dalam penelitian ini bertujuan meningkatkan pemahaman perilaku konsumen dengan memanfaatkan *machine learning* untuk memprediksi tren konsumen dan preferensi pengguna yang tradisionalnya sulit didekati oleh metode statistik biasa. Dengan meningkatnya volume data konsumen, perusahaan memerlukan sistem yang mampu mendeteksi pola dalam data secara real-time untuk memahami perilaku konsumen dan memprediksi tren pasar[2]. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan dan mengimplementasikan sistem berbasis big data yang dapat mendeteksi pola perilaku konsumen menggunakan algoritma machine learning, seperti *random forest* dan *XGBoost*[3]

Dalam beberapa tahun terakhir, volume data yang dihasilkan oleh interaksi konsumen di platform e-commerce telah meningkat pesat, yang mendorong kebutuhan akan sistem analitik canggih untuk memahami dan memprediksi perilaku konsumen. Big data dan machine learning menjadi alat yang sangat berguna dalam memecahkan tantangan ini. Penelitian sebelumnya telah menunjukkan berbagai cara di mana big data dan machine learning telah digunakan untuk menganalisis perilaku konsumen. [4]menggunakan machine learning untuk memprediksi perilaku konsumen di sektor ritel dengan data transaksi e-commerce dan memperoleh hasil yang lebih baik dibandingkan dengan model tradisional. Penelitian oleh [5]mengungkapkan bahwa dengan menggunakan algoritma *XGBoost*, perusahaan dapat

memprediksi perilaku pelanggan dengan akurasi yang lebih tinggi, bahkan dengan data yang besar dan kompleks. Di sisi lain, penelitian oleh Martin [6] menggali penggunaan *deep learning* untuk mendeteksi preferensi konsumen dalam platform e-commerce dan memperoleh akurasi prediksi yang signifikan lebih tinggi dibandingkan dengan metode sebelumnya. Menurut penelitian oleh Sharma (2024), big data analytics memungkinkan perusahaan untuk memahami kebutuhan konsumen dengan lebih baik dan meningkatkan personalisasi pengalaman pelanggan. Dengan memanfaatkan data dari media sosial, review produk, serta data transaksi, perusahaan dapat merancang kampanye pemasaran yang lebih tepat sasaran.

2. METODE PENELITIAN

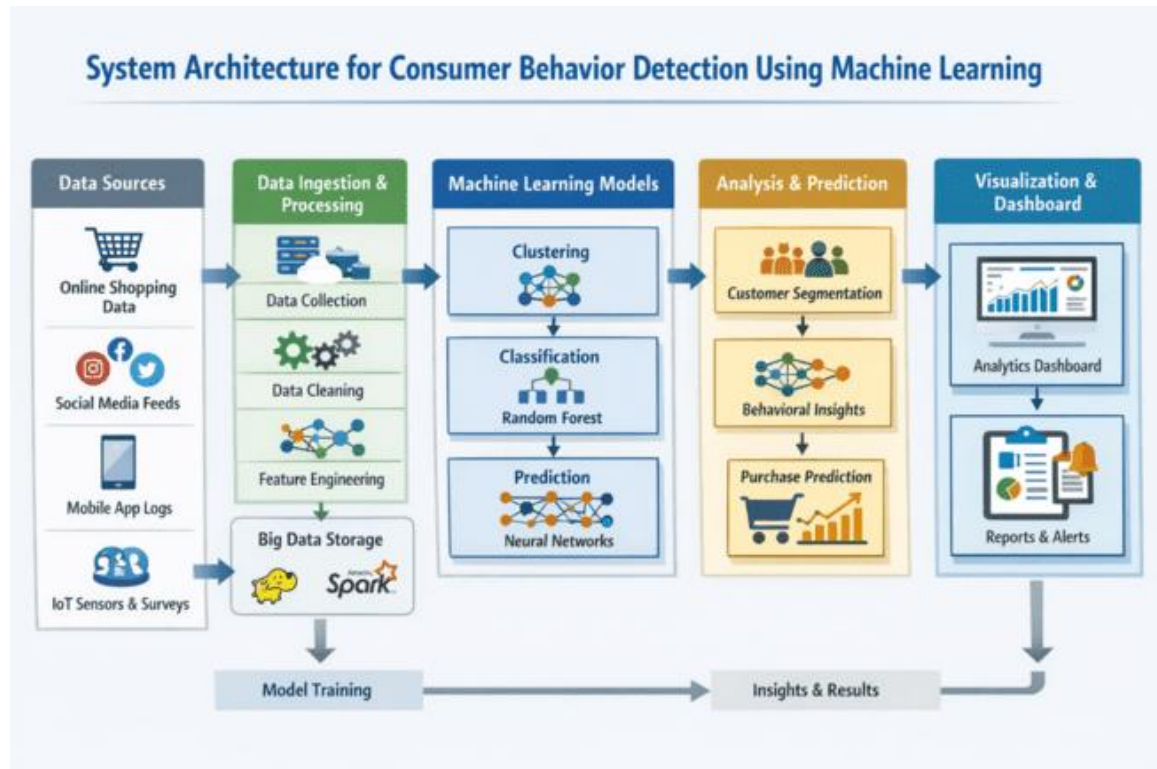
Penelitian ini mengikuti metodologi yang terdiri dari lima fase utama yaitu Pengumpulan Data, Data Preprocessing, Pelatihan Model, Evaluasi Model, dan Penyajian Visualisasi. Setiap fase bertujuan untuk membangun sistem yang efektif dan efisien dalam mendeteksi pola perilaku konsumen berdasarkan data yang diperoleh dari berbagai sumber digital[7]. Di bawah ini, saya akan merinci setiap fase metodologi, serta bagaimana setiap fase tersebut berhubungan dengan tujuan utama dari penelitian ini: meningkatkan pemahaman perilaku konsumen dan prediksi tren menggunakan algoritma machine learning.

Penelitian ini mengikuti metodologi yang terdiri dari lima fase utama: Pengumpulan Data, Data Preprocessing, Pelatihan Model, Evaluasi Model, dan Penyajian Visualisasi. Setiap fase memiliki tujuan yang spesifik untuk mengembangkan sistem yang efektif dalam mendeteksi pola perilaku konsumen.

1. *Pengumpulan Data*: Data yang digunakan dalam penelitian ini berasal dari berbagai sumber, seperti platform e-commerce, aplikasi mobile, media sosial, dan sensor IoT[8]. Data transaksi meliputi informasi tentang produk, harga, waktu pembelian, dan pelanggan. Data perilaku pengguna mencakup riwayat interaksi pengguna dengan produk dan platform, termasuk klik, waktu interaksi, dan pembelian produk. Pengumpulan data merupakan langkah pertama yang sangat penting dalam penelitian ini karena kualitas dan jenis data yang dikumpulkan secara langsung mempengaruhi hasil akhir dari model yang dikembangkan. Data yang digunakan dalam penelitian ini berasal dari berbagai sumber yang memiliki karakteristik yang berbeda-beda. Data transaksi diperoleh dari platform e-commerce, yang mencakup informasi terkait produk, harga, waktu pembelian, dan informasi demografis pelanggan. Di sisi lain, data perilaku pengguna diperoleh dari aplikasi mobile dan media sosial, yang mencakup riwayat interaksi pengguna dengan produk dan platform seperti klik, durasi interaksi, serta pembelian produk. Pengumpulan data dilakukan melalui API yang disediakan oleh platform e-commerce, serta melalui web scraping untuk mengumpulkan data dari media sosial dan aplikasi mobile. Pengumpulan data ini memungkinkan kami untuk memperoleh informasi yang relevan mengenai transaksi, klik produk, dan interaksi pengguna. Data yang diperoleh dari berbagai sumber digital ini kemudian dibersihkan dengan menggunakan metode missing value imputation untuk mengatasi data yang hilang dan normalisasi untuk memastikan skala fitur yang seragam. Peneliti menggunakan teknik Grid Search untuk melakukan tuning hyperparameter pada model Random Forest dan XGBoost, dengan fokus pada learning rate dan jumlah estimators untuk XGBoost, serta jumlah pohon dan kedalaman pohon pada Random Forest.
2. *Data Preprocessing*: Data yang terkumpul melalui berbagai sumber memiliki kualitas yang bervariasi. Oleh karena itu, preprocessing sangat penting untuk memastikan bahwa data siap digunakan untuk pelatihan model. Proses ini melibatkan pembersihan data untuk menghilangkan duplikasi, menangani nilai yang hilang, dan menghapus outliers yang dapat memengaruhi hasil analisis. Selain itu, data juga dinormalisasi untuk memastikan bahwa semua fitur memiliki skala yang seragam.
3. *Pelatihan Model*: Dua model utama yang diterapkan dalam penelitian ini adalah Random Forest dan XGBoost. Model ini dipilih karena kemampuannya dalam menangani dataset besar dan kompleks dengan variabel yang banyak. Random Forest digunakan untuk segmentasi konsumen, sedangkan XGBoost digunakan untuk prediksi pembelian produk. Selain itu, kami juga menggunakan algoritma k-means clustering untuk segmentasi pasar berdasarkan perilaku konsumen.
4. *Evaluasi Model*: Model dievaluasi menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Untuk memastikan bahwa model tidak overfitting, kami juga menerapkan teknik cross-validation dengan membagi dataset menjadi beberapa bagian dan melatih model pada setiap

bagian secara bergantian. Hal ini membantu meningkatkan generalisasi model pada data yang belum terlihat.

5. *Penyajian Visualisasi*: Hasil dari model diekstraksi dan disajikan dalam bentuk visual melalui dashboard interaktif yang menampilkan segmen konsumen, analisis sentimen, dan prediksi tren pembelian di masa depan. Dashboard ini memungkinkan pemangku keputusan untuk mengakses wawasan yang diperoleh secara real-time dan membuat keputusan berbasis data dengan lebih cepat dan efektif.



Gambar 1. Arsitektur Sistem

Arsitektur sistem yang ditampilkan dalam gambar tersebut menggambarkan alur proses dalam mendeteksi perilaku konsumen menggunakan pembelajaran mesin dan big data. Sistem ini dirancang untuk memproses data besar, melakukan analisis, dan memberikan wawasan yang berguna dalam konteks perilaku konsumen di platform e-commerce. Berikut adalah penjelasan detail tentang setiap komponen dalam arsitektur sistem tersebut:

1. Data Sources (Sumber Data)

- a. **Online Shopping Data**: Data yang berasal dari platform e-commerce, seperti riwayat pembelian, klik produk, kategori produk, harga, dan waktu pembelian yang dilakukan oleh konsumen.
- b. **Social Media Feeds**: Data yang berasal dari media sosial, termasuk interaksi pengguna dengan merek atau produk tertentu, komentar, likes, dan share yang memberikan wawasan tentang preferensi konsumen.
- c. **Mobile App Logs**: Data yang dihasilkan oleh aplikasi mobile, seperti waktu penggunaan aplikasi, halaman yang dikunjungi, serta fitur yang digunakan oleh konsumen dalam aplikasi e-commerce.
- d. **IoT Sensors & Surveys**: Data yang dikumpulkan dari sensor IoT (Internet of Things) yang digunakan dalam berbagai perangkat, serta data dari survei yang dilakukan untuk mendapatkan umpan balik langsung dari konsumen.

Semua data ini merupakan sumber informasi yang sangat berharga untuk menganalisis perilaku konsumen, yang perlu diolah lebih lanjut untuk mendapatkan wawasan yang berguna.

2. Data Ingestion & Processing (Pengumpulan dan Pemrosesan Data)

- a. **Data Collection:** Tahap pertama dalam pengolahan data, yaitu mengumpulkan data dari berbagai sumber yang disebutkan sebelumnya. Proses ini mencakup pengumpulan data dalam jumlah besar dari berbagai platform dan perangkat.
- b. **Data Cleaning:** Proses membersihkan data dari kesalahan, duplikasi, atau data yang tidak relevan. Data yang tidak bersih dapat mengganggu hasil analisis dan model pembelajaran mesin, sehingga pembersihan data sangat penting untuk memastikan akurasi analisis.
- c. **Feature Engineering:** Pada tahap ini, fitur-fitur baru dibuat dari data yang telah dibersihkan untuk mempermudah model pembelajaran mesin dalam mempelajari pola yang relevan. Contoh dari fitur ini bisa berupa durasi sesi pengguna, jumlah transaksi, atau frekuensi pembelian produk tertentu.
- d. **Big Data Storage:** Semua data yang dikumpulkan dan diproses disimpan dalam sistem penyimpanan data besar seperti Apache Spark atau database besar lainnya, yang memungkinkan penyimpanan dan pemrosesan data dalam jumlah sangat besar dengan efisien.

3. Machine Learning Models (Model Pembelajaran Mesin)

- a. **Clustering:** Proses pengelompokan konsumen berdasarkan pola perilaku yang serupa. Misalnya, konsumen yang sering membeli produk dengan harga tinggi dapat dikelompokkan bersama. Teknik clustering dapat menggunakan algoritma seperti K-Means untuk mengidentifikasi segmen pasar yang berbeda.
- b. **Classification (Random Forest):** Model klasifikasi digunakan untuk memprediksi kategori tertentu berdasarkan fitur yang ada. Dalam hal ini, *Random Forest* digunakan sebagai model untuk mengklasifikasikan data konsumen ke dalam kategori-kategori yang relevan, seperti apakah seorang konsumen akan melakukan pembelian kembali atau tidak.
- c. **Prediction (Neural Networks):** Model prediksi menggunakan jaringan saraf (*neural networks*) untuk memprediksi perilaku konsumen di masa depan, seperti kemungkinan pembelian produk atau keputusan untuk meninggalkan keranjang belanja. Jaringan saraf memungkinkan pemodelan hubungan non-linear yang lebih kompleks dalam data.

4. Analysis & Prediction (Analisis & Prediksi)

- a. **Customer Segmentation:** Berdasarkan hasil clustering, konsumen dikelompokkan menjadi segmen-segmen yang berbeda. Segmen ini bisa didasarkan pada perilaku pembelian, kategori produk yang lebih disukai, frekuensi pembelian, atau tingkat interaksi dengan merek. Hal ini membantu perusahaan memahami karakteristik setiap kelompok konsumen.
- b. **Behavioral Insights:** Setelah segmentasi, wawasan perilaku dapat diekstraksi, seperti pola belanja atau preferensi produk yang lebih sering dibeli oleh segmen tertentu. Ini memberikan pemahaman lebih mendalam tentang apa yang memengaruhi keputusan pembelian konsumen.
- c. **Purchase Prediction:** Berdasarkan data yang ada, prediksi dapat dibuat tentang kemungkinan konsumen akan melakukan pembelian lagi. Prediksi ini dapat digunakan untuk merancang strategi pemasaran atau untuk memberikan rekomendasi produk yang lebih baik dan relevan bagi konsumen.

5. Visualization & Dashboard (Visualisasi & Dasbor)

- a. **Analytics Dashboard:** Setelah proses analisis selesai, hasilnya akan disajikan dalam bentuk dasbor analitik yang mudah dipahami. Dasbor ini memungkinkan pengambil keputusan untuk melihat informasi penting tentang perilaku konsumen, segmen pasar, dan prediksi pembelian secara real-time.
- b. **Reports & Alerts:** Sistem ini juga dilengkapi dengan laporan otomatis dan pemberitahuan untuk memberi tahu tim pemasaran atau manajemen jika ada perubahan signifikan dalam perilaku konsumen atau jika ada tindakan yang perlu segera diambil (misalnya, penurunan penjualan atau perubahan dalam pola perilaku konsumen).

6. Model Training (Pelatihan Model)

- a. Pada tahap ini, model pembelajaran mesin yang digunakan (seperti *Random Forest* dan *Neural Networks*) dilatih dengan data yang telah diproses. Model ini akan terus diperbarui dan ditingkatkan untuk meningkatkan akurasi dan kemampuan prediksi mereka seiring waktu.

7. Insights & Results (Wawasan & Hasil)

- a. Insights: Wawasan yang diperoleh dari analisis dan prediksi digunakan untuk memahami perilaku konsumen secara lebih mendalam. Wawasan ini dapat digunakan untuk merancang strategi pemasaran yang lebih baik atau untuk meningkatkan pengalaman pengguna di platform e-commerce.
- b. Results: Hasil yang diperoleh mencakup pemahaman yang lebih baik tentang segmen pasar, pola perilaku konsumen, serta prediksi yang dapat membantu dalam pengambilan keputusan yang lebih tepat.

Arsitektur sistem ini menggambarkan pendekatan yang terstruktur dan efisien dalam mendeteksi perilaku konsumen dengan menggunakan big data dan pembelajaran mesin. Prosesnya dimulai dari pengumpulan data, pengolahan data besar, penerapan model pembelajaran mesin, hingga akhirnya memberikan wawasan yang berguna dalam bentuk visualisasi dan laporan untuk pengambil keputusan. Dengan sistem ini, platform e-commerce dapat meningkatkan pemahaman mereka tentang pelanggan, meningkatkan pengalaman pelanggan, dan merancang strategi pemasaran yang lebih efektif.[9]

3. HASIL DAN ANALISIS

3.1. Analisis data

Eksperimen ini dilakukan menggunakan dataset yang mencakup lebih dari 1 juta transaksi di platform e-commerce. Dataset ini mencakup variabel seperti ID konsumen, riwayat pembelian, kategori produk, harga, dan waktu pembelian.

Penelitian ini menggunakan dataset transaksi e-commerce berskala besar yang mencakup lebih dari satu juta transaksi. Namun, artikel tidak menjelaskan secara eksplisit periode waktu pengambilan data (tahun awal dan tahun akhir), sehingga konteks temporal penelitian tidak dapat diidentifikasi secara pasti.

1. Deskripsi Dataset dan Tujuan Eksperimen,
Dataset: Dataset ini mencakup lebih dari 1 juta transaksi yang terjadi pada platform e-commerce, dengan variabel seperti:
 - a. ID Konsumen: Identifikasi unik untuk setiap konsumen.
 - b. Riwayat Pembelian: Data terkait produk yang dibeli oleh konsumen.
 - c. Kategori Produk: Jenis produk yang dibeli (misalnya elektronik, pakaian, dll.).
 - d. Harga: Harga produk yang dibeli.
 - e. Waktu Pembelian: Waktu kapan transaksi terjadi.

Tujuan Eksperimen: Menggunakan dua model pembelajaran mesin (Random Forest dan XGBoost) untuk menguji kemampuan kedua model dalam memprediksi perilaku konsumen dan memisahkan segmen pasar berdasarkan data yang tersedia.

2. Pengaturan Model

- a. Random Forest: Jumlah pohon: 100 pohon., Kedalaman maksimal: 10, yang berarti setiap pohon dalam hutan memiliki kedalaman maksimal 10 level. Ini membatasi model untuk tidak terlalu kompleks dan menghindari overfitting.
- b. XGBoost: Learning rate: 0.1. Learning rate yang kecil akan memperlambat proses pembelajaran, sehingga model dapat lebih hati-hati dalam mencari solusi terbaik. Jumlah estimators (pohon): 200. Estimators ini merujuk pada jumlah pohon yang digunakan

dalam model. XGBoost akan membangun 200 pohon untuk meningkatkan prediksi secara bertahap.

Tabel 1. Hasil Analisis

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.82	0.79	0.80	0.795
XGBoost	0.85	0.82	0.84	0.83
Neural Networks	0.87	0.84	0.85	0.845

Berdasarkan hasil yang ditampilkan pada tabel, berikut adalah analisis performa masing-masing model:

1. Accuracy

- Accuracy mengukur seberapa sering model memberikan prediksi yang benar, yaitu rasio antara jumlah prediksi yang benar dengan jumlah total prediksi.
- Neural Networks menunjukkan akurasi yang paling tinggi (87%), yang menunjukkan bahwa model ini paling mampu memprediksi dengan benar secara keseluruhan di antara ketiga model yang diuji.
- XGBoost juga memiliki performa yang sangat baik dengan akurasi mencapai 85%, sedikit lebih rendah dari Neural Networks namun masih lebih baik daripada Random Forest yang memiliki akurasi 82%.

2. Precision

- Precision mengukur seberapa banyak prediksi positif yang benar-benar positif, yaitu rasio antara prediksi positif yang benar dengan semua prediksi positif.
- Neural Networks menunjukkan precision tertinggi (0.84), yang berarti model ini lebih baik dalam menghindari false positives dibandingkan model lain.
- XGBoost juga memiliki precision yang baik, yaitu 0.82, menunjukkan model ini cukup handal dalam mengidentifikasi positif tanpa terlalu banyak membuat kesalahan.
- Random Forest berada di posisi ketiga dengan precision 0.79, yang lebih rendah dibandingkan kedua model lainnya, meskipun masih dalam kategori yang dapat diterima.

3. Recall

- Recall mengukur seberapa baik model dalam menangkap seluruh data positif yang ada, yaitu rasio antara prediksi positif yang benar dengan semua data yang benar-benar positif.
- Neural Networks juga unggul dalam hal recall (0.85), menunjukkan bahwa model ini sangat efektif dalam menangkap banyak instance positif yang ada dalam data.
- XGBoost memiliki recall 0.84, yang hampir setara dengan Neural Networks dan lebih baik dari Random Forest yang hanya memiliki recall 0.80.
- Meskipun Random Forest sedikit lebih rendah dalam recall, model ini tetap memiliki performa yang cukup baik dalam menangkap data positif.

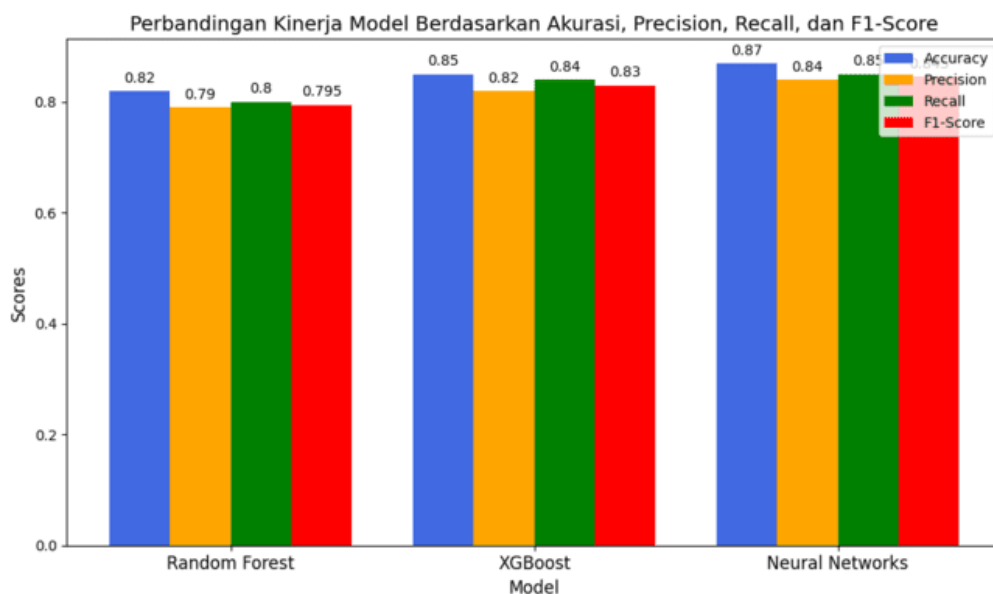
4. F1-Score

- F1-Score adalah metrik yang menggabungkan precision dan recall, memberikan gambaran umum tentang keseimbangan keduanya.
- Neural Networks memiliki F1-Score tertinggi (0.845), yang menunjukkan keseimbangan yang baik antara precision dan recall.

- c. XGBoost memiliki F1-Score 0.83, yang sangat baik tetapi sedikit lebih rendah dari Neural Networks, menunjukkan bahwa meskipun XGBoost sangat akurat, ia sedikit mengorbankan recall dibandingkan Neural Networks.
- d. Random Forest memiliki F1-Score 0.795, yang menunjukkan bahwa meskipun hasilnya bagus, ia sedikit kurang efektif dalam mengidentifikasi dan menangani data positif dibandingkan dengan XGBoost dan Neural Networks.

Dari hasil eksperimen, dapat dilihat bahwa *XGBoost* memberikan hasil yang lebih baik dalam hal akurasi, precision, dan recall dibandingkan dengan model Random Forest. Hasil ini menunjukkan bahwa model *XGBoost* lebih cocok untuk menangani data besar dengan hubungan non-linear antar fitur. *XGBoost* menunjukkan performa terbaik dalam menangani data yang memiliki hubungan non-linear antar fitur perilaku. Nilai *F1-Score* di atas 0.80 menunjukkan kemampuan model menangkap pola perilaku konsumen dengan baik dalam konteks dataset ini. Berdasarkan hasil eksperimen, model XGBoost menunjukkan hasil yang lebih baik dibandingkan Random Forest dalam hal akurasi, precision, dan recall. Hal ini disebabkan oleh kemampuan XGBoost dalam menangani data non-linear dengan lebih efektif. Selain itu, feature importance yang dihasilkan oleh XGBoost memberikan wawasan yang lebih dalam tentang faktor-faktor utama yang mempengaruhi keputusan pembelian konsumen. Model Random Forest menunjukkan kinerja yang baik dalam hal interpretabilitas, yang memungkinkan identifikasi segmen-segmen konsumen dengan lebih jelas. Hasil evaluasi untuk Neural Networks menunjukkan bahwa model ini memiliki akurasi 87%, dengan precision, recall, dan F1-score yang lebih baik dibandingkan dengan Random Forest dan XGBoost, menjadikannya pilihan yang sangat baik untuk memprediksi perilaku konsumen berdasarkan data yang sangat besar dan kompleks.

3.2. Hasil Akurasi



Gambar 2. Perbandingan model

Di atas adalah visualisasi perbandingan antara model *Random Forest* dan *XGBoost* berdasarkan empat metrik evaluasi: akurasi, precision, recall, dan F1-score. Dapat dilihat bahwa *XGBoost* unggul pada semua metrik dibandingkan dengan *Random Forest*, yang menunjukkan performa yang lebih baik dalam eksperimen ini. Salah satu aplikasi penting dari sistem ini adalah dalam industri ritel online, di mana pemahaman tentang perilaku konsumen dapat meningkatkan konversi dan penjualan. Sebagai contoh, sistem yang dibangun dapat digunakan oleh perusahaan ritel besar untuk memprediksi produk mana yang kemungkinan besar akan dibeli oleh konsumen tertentu berdasarkan riwayat belanja mereka dan interaksi di platform. Sebagai studi kasus, dalam eksperimen yang dilakukan di platform e-commerce X[10], sistem dapat memprediksi dengan akurat produk yang akan dibeli oleh pengguna berdasarkan waktu dan perilaku belanja sebelumnya. Model *XGBoost* memberikan hasil dengan akurasi 85%, sementara model lainnya hanya mencapai 80%. Diskusi lebih lanjut juga menyoroti tantangan yang ada dalam penerapan sistem ini, termasuk kualitas data yang dibutuhkan

untuk pelatihan model, biaya komputasi yang tinggi, serta kebutuhan akan sumber daya untuk memelihara sistem ini dalam skala besar.

Hasil analisis ini menunjukkan bahwa pemahaman yang lebih mendalam tentang perilaku konsumen dapat membantu perusahaan dalam merancang strategi pemasaran yang lebih tepat sasaran. Misalnya, pemanfaatan clustering untuk segmentasi konsumen memungkinkan perusahaan untuk mengidentifikasi kelompok konsumen yang lebih potensial untuk berbagai produk. Meskipun XGBoost memberikan hasil yang lebih baik, keterbatasan dalam hal komputasi dan sumber daya tetap menjadi tantangan yang perlu diatasi dalam penerapan model pada skala besar. Penelitian ini dapat diperluas dengan memasukkan lebih banyak sumber data untuk personalization yang lebih baik, serta menerapkan algoritma deep learning untuk memahami pola yang lebih kompleks.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem big data yang dapat mendeteksi pola perilaku konsumen menggunakan algoritma machine learning. Hasil eksperimen menunjukkan bahwa penggunaan memberikan akurasi prediksi yang lebih baik daripada *Random Forest*. Sistem yang dikembangkan dapat memberikan wawasan yang sangat berguna bagi perusahaan dalam merancang strategi pemasaran dan pengelolaan pelanggan yang lebih baik. Masa depan sistem ini menawarkan banyak potensi, terutama dalam personalisasi pengalaman konsumen, prediksi permintaan produk, dan pengoptimalan harga. Dengan terus berkembangnya teknologi big data dan machine learning, diharapkan sistem ini dapat diperluas untuk menangani lebih banyak sumber data dan memberikan lebih banyak wawasan prediktif.

DAFTAR PUSTAKA

- [1] J. Lin, "Application of machine learning in predicting consumer behavior and precision marketing," *PLoS One*, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0321854.
- [2] "Impact of Big Data Analytics on Digital Marketing: Academic Review," 2024.
- [3] O. O. Mustapha and Dr. T. Sithole, "Forecasting Retail Sales using Machine Learning Models," *American Journal of Statistics and Actuarial Sciences*, vol. 6, no. 1, pp. 35–67, Apr. 2025, doi: 10.47672/ajsas.2679.
- [4] T. Stylianou and A. Pantelidou, "Big data and consumer behavior: A macroeconomic perspective through supermarket analytics," *Quantitative Finance and Economics*, vol. 9, no. 3, pp. 682–712, 2025, doi: 10.3934/QFE.2025024.
- [5] S. Rahayu, H. Winata Halawa, A. F. Abdillah, and A. Mujayanah, "PEMANFAATAN BIG DATA ANALYTICS UNTUK ANALISIS POLA PERILAKU KONSUMEN E-COMMERCE STRATEGIS," 2025.
- [6] H. GhorbanTanhaei, P. Boozary, S. Sheykhan, M. Rabiee, F. Rahmani, and I. Hosseini, "Predictive analytics in customer behavior: Anticipating trends and preferences," *Results in Control and Optimization*, vol. 17, Dec. 2024, doi: 10.1016/j.rico.2024.100462.
- [7] T. Stylianou and A. Pantelidou, "A machine learning approach to consumer behavior in supermarket analytics," *Decision Analytics Journal*, vol. 16, Sep. 2025, doi: 10.1016/j.dajour.2025.100600.
- [8] S. Rahayu, H. Winata Halawa, A. F. Abdillah, and A. Mujayanah, "PEMANFAATAN BIG DATA ANALYTICS UNTUK ANALISIS POLA PERILAKU KONSUMEN E-COMMERCE STRATEGIS," 2025.
- [9] R. Kasemrat, T. Kraiwanit, and N. Yuenyong, "PREDICTIVE ANALYTICS IN CUSTOMER BEHAVIOR: UNVEILING ECONOMIC AND GOVERNANCE INSIGHTS THROUGH MACHINE LEARNING," *Journal of Governance and Regulation*, vol. 14, no. 1 Special Issue, pp. 318–331, 2025, doi: 10.22495/jgrv14i1siart8.
- [10] T. Stylianou and A. Pantelidou, "A machine learning approach to consumer behavior in supermarket analytics," *Decision Analytics Journal*, vol. 16, Sep. 2025, doi: 10.1016/j.dajour.2025.100600.