JUSIFOR: Jurnal Sistem Informasi dan Informatika
Vol. 4, No. 1, Juni 2025, hlmn. 85-91
E-ISSN: 2830-3393 | P-ISSN: 2830-2443
DOI: https://doi.org/10.70609/jusifor.v4i1.6962

Prediksi Volume Penjualan Gadget Berdasarkan Promo dan Channel Penjualan Menggunakan Random Forest

Alvi Agustina[1*], Tukino Tukino[2], Baenil Huda[3], Elfina Novalia[4]

[1,2,3,4] Sistem Infomasi, Fakultas Ilmu Komputer, Universitas Buana Perjuangan, Karawang, Jawa Barat, Indonesia

Info Artikel

Riwayat Artikel:
Diterima : **30 April 2025**
Direvisi : **18 Juni 2025**
Diterbitkan : **25 Juni 2025**

Kata Kunci:
Prediksi Volume Penjualan,
Random Forest,
Machine Learning,
Channel Penjualan

ABSTRAK

Volume penjualan gadget dipengaruhi oleh berbagai faktor seperti harga, rating pengguna, keberadaan promo, serta *channel* distribusi yang digunakan. Pemahaman terhadap pengaruh faktor-faktor tersebut sangat penting untuk merumuskan strategi pemasaran yang efektif. Penelitian ini bertujuan untuk memprediksi volume penjualan gadget menggunakan algoritma *Random Forest* berdasarkan fitur promo dan *channel* penjualan. Data yang digunakan merupakan catatan transaksi aktual dari sebuah toko gadget yang melayani penjualan secara online dan offline di wilayah Jabodetabek dalam periode tertentu. Tahapan penelitian meliputi *data preprocessing*, pembentukan target klasifikasi volume penjualan menjadi tiga kategori (rendah, sedang, dan tinggi), pelatihan model *Random Forest*, serta evaluasi performa model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Pembagian data latih dan data uji dilakukan dengan teknik *stratified sampling* untuk menjaga keseimbangan distribusi kelas. Hasil evaluasi menunjukkan bahwa model Random Forest mencapai akurasi sebesar 49,5%, dengan performa terbaik dalam mengklasifikasikan kategori volume penjualan "sedang". Angka ini mengindikasikan bahwa akurasi prediksi model masih terbatas, dengan hanya sekitar separuh dari keseluruhan data uji yang dapat diprediksi dengan benar. Meskipun demikian, temuan ini menunjukkan bahwa penerapan algoritma machine learning, khususnya Random Forest, berpotensi mendukung pengambilan keputusan berbasis data dalam meningkatkan efektivitas penjualan gadget.

***Keywords*:**
Sales Volume Prediction,
Random Forest,
Machine Learning,
Sales Channel

ABSTRACT

Gadget sales volume is influenced by various factors such as price, user ratings, promotional offers, and the distribution channels used. Understanding the impact of these factors is crucial for formulating effective marketing strategies. This study aims to predict gadget sales volume using the Random Forest algorithm based on promotional features and sales channels. The dataset consists of actual transaction records from a gadget store operating both online and offline in the Greater Jakarta area over a specified period. The research stages include data preprocessing, categorization of sales volume into three classes (low, medium, and high), Random Forest model training, and model performance evaluation using accuracy, precision, recall, and F1-score metrics. The training and testing datasets are split using a stratified sampling technique to maintain class distribution balance. The evaluation results show that the Random Forest model achieves an accuracy of 49.5%, with the best performance in classifying the "medium" sales volume category. This figure indicates that the model's prediction accuracy is still limited, with only about half of the total test data predicted correctly. Nevertheless, these findings indicate that implementing machine learning algorithms, particularly Random Forest, has the potential to support data-driven decision-making in improving gadget sales effectiveness, although further model development is required to enhance prediction accuracy for other categories due to the high potential for prediction errors that could significantly impact expected outcomes.

Penulis Korespondensi:
Alvi Agustina,
Sistem Informasi,
Universitas Buana Perjuangan, Karawang, Jawa Barat, Indonesia
Email: si22.alviagustina@mhs.ubpkarawang.ac.id

1. PENDAHULUAN

Perkembangan teknologi informasi dalam satu dekade terakhir telah menciptakan transformasi besar dalam perilaku konsumen, model bisnis, dan strategi pemasaran di berbagai sektor, termasuk industri gadget. Gadget, sebagai produk teknologi dengan tingkat inovasi tinggi, sangat dipengaruhi oleh tren pasar, preferensi

konsumen yang cepat berubah, dan kompetisi harga yang ketat. Menurut laporan IDC (2024) [1], pertumbuhan volume penjualan *smartphone* global dipicu oleh perluasan *e-commerce channel*, digitalisasi pemasaran, serta peningkatan personalisasi penawaran produk kepada konsumen.

Dalam kondisi pasar yang semakin kompetitif, penting bagi perusahaan untuk mengadopsi pendekatan berbasis data dalam pengambilan keputusan bisnis. Salah satu kebutuhan utama adalah kemampuan untuk memprediksi volume penjualan secara akurat guna mengoptimalkan alokasi sumber daya dan strategi distribusi. Implementasi *machine learning* menjadi sangat relevan dalam memenuhi kebutuhan ini, dengan *Random Forest* sebagai salah satu algoritma yang banyak digunakan karena kemampuannya dalam menangani dataset besar, mengelola variabel kompleks, serta meminimalkan risiko *overfitting* [2][3].

Penelitian oleh Breiman [4] memperkenalkan *Random Forest* sebagai metode *ensemble* berbasis *decision tree* yang menawarkan akurasi tinggi dalam berbagai tugas prediksi. Seiring perkembangan, *Random Forest* telah diterapkan dalam berbagai domain termasuk prediksi volume penjualan retail [5], klasifikasi pelanggan *e-commerce* [6], dan analisis *churn* pelanggan [7]. Chen dan Guestrin [8] juga menekankan bahwa metode berbasis pohon keputusan seperti *Random Forest* dan *XGBoost* efektif dalam prediksi dengan data heterogen.

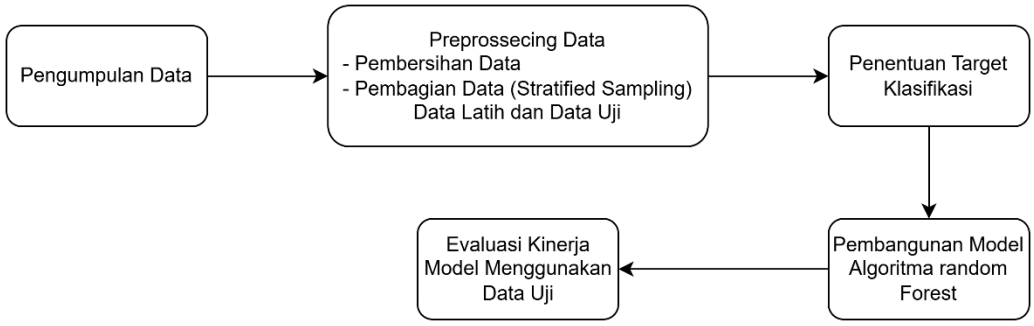
Dalam sektor pemasaran digital, penggunaan *Random Forest* untuk mengoptimalkan strategi promosi juga semakin luas. Penelitian terbaru oleh Hayat, Putra, dan Harahap [3], [9] menunjukkan bahwa integrasi *machine learning* dalam pengelolaan kampanye iklan digital mampu meningkatkan efektivitas *targeting* dan konversi penjualan hingga 30%. Di sisi lain, adaptasi *Random Forest* terhadap data *imbalance* menjadi penting, mengingat volume penjualan di beberapa *channel* cenderung tidak seimbang. Teknik *Weighted Random Forest* yang dikembangkan dalam penelitian Budianti dan Suliadi [10] terbukti meningkatkan akurasi prediksi dalam domain kesehatan dan dapat diadaptasi untuk konteks penjualan. Optimalisasi fitur menjadi faktor kritis dalam meningkatkan kinerja prediksi. Priantama dan Siswa [11] menunjukkan bahwa penerapan metode *Correlation-Based Feature Selection* (CFS) dapat meningkatkan akurasi prediksi akademik mahasiswa secara signifikan.

Dalam implementasinya di bisnis skala kecil dan menengah (UMKM), *Random Forest* berhasil mengklasifikasikan kategori penjualan berdasarkan platform distribusi, sebagaimana ditunjukkan oleh Rindiyani et al. [12]. Penelitian di bidang retail oleh Verdiyanto et al. [13] mengembangkan aplikasi *Point of Sales* yang mampu memprediksi penjualan harian dengan *error* rendah menggunakan *Random Forest Regression*. Penggunaan *Random Forest* dalam analisis prediktif di sektor pendidikan untuk memprediksi risiko *drop-out* mahasiswa, seperti yang ditunjukkan Marzuqi et al. [14], semakin memperlihatkan fleksibilitas dan keunggulan algoritma ini.

Berdasarkan kajian literatur tersebut, penelitian ini bertujuan untuk membangun model prediksi volume penjualan gadget berdasarkan dua variabel utama, yaitu intensitas promosi dan *channel* distribusi. Model ini menggabungkan pemanfaatan algoritma *Random Forest* dengan analisis dua faktor kunci bisnis, sehingga diharapkan dapat memberikan kontribusi nyata dalam mengoptimalkan strategi pemasaran berbasis data, meningkatkan efektivitas promosi, serta mendorong peningkatan volume penjualan gadget secara berkelanjutan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Diagram Alur Penelitian

Penelitian ini dilaksanakan melalui tahapan sistematis yang meliputi pengumpulan data, *preprocessing* data, pembentukan target klasifikasi, pembangunan model Random Forest, dan evaluasi kinerja model. Setiap tahap dirancang untuk mengikuti standar penelitian *machine learning* modern, di mana data yang telah dikumpulkan diproses, dibersihkan, dan dibagi menjadi data latih dan data uji menggunakan teknik *stratified sampling* untuk menjaga proporsi distribusi kelas [2]. Diagram alur keseluruhan proses penelitian ini ditampilkan pada Gambar 1.

2.2 Dataset

Dataset yang digunakan berasal dari catatan transaksi penjualan sebuah toko gadget yang beroperasi di wilayah Jabodetabek melalui kanal *online* dan *offline*. Dataset mencakup 1.000 baris data transaksi dengan atribut: tanggal transaksi, merek gadget, model, harga, rating pengguna, *channel* penjualan (*online/offline*), status promo (ya/tidak), dan jumlah unit terjual. Penggunaan data riil ini selaras dengan praktik umum penelitian *machine learning* berbasis data aktual sebagai sumber utama [6]. Cuplikan contoh data transaksi disajikan pada Tabel 1.

Tabel 1. Cuplikan Dataset

Tanggal	Merek	Model	Harga (Rp)	Jumlah Terjual	Rating	Channel Penjualan	Promo
2023-01-01	Xiaomi	Note 11 Pro 5G	4.099.000	15	4,1	Offline	Ya
2023-01-02	Oppo	Find N2 Flip	14.999.000	8	4,9	Online	Ya
2023-01-03	Vivo	V23 5G	3.150.000	9	4,6	Online	Ya
2023-01-04	Samsung	A73 5G	7.799.000	10	4,4	Offline	Tidak
2023-01-05	Realme	GT Master Edition	5.299.000	13	3,7	Offline	Tidak
2023-01-06	Xiaomi	10 5G	2.599.000	10	3,7	Online	Ya
2023-01-07	Vivo	Y27 5G	2.500.000	11	3,6	Online	Ya
2023-01-08	Oppo	Reno 10 Pro+ 5G	10.999.000	8	4,8	Online	Ya

2.3 *Preprocessing* Data

Tahap *preprocessing* data merupakan tahap penting untuk mempersiapkan dataset sebelum model dibangun. Proses pertama adalah melakukan *encoding* pada atribut kategorikal seperti merek dan *channel* penjualan menggunakan *LabelEncoder* [15]. Selanjutnya, pembentukan target dilakukan dengan mengklasifikasikan jumlah unit terjual ke dalam tiga kategori: Rendah (≤7 unit), Sedang (8–12 unit), dan Tinggi (>12 unit), untuk memudahkan proses klasifikasi. Data kemudian dibagi menjadi dua bagian, yakni 80% sebagai data latih dan 20% sebagai data uji, menggunakan teknik *stratified sampling* agar proporsi antar kategori tetap seimbang [3].

2.4 Pemodelan Random Forest

Random Forest dipilih sebagai algoritma utama karena keandalannya dalam mengatasi masalah klasifikasi dengan data kompleks dan ketahanannya terhadap *overfitting* [4]. *Random Forest* bekerja dengan membangun banyak pohon keputusan (*decision trees*) dan menggunakan prinsip *majority voting* untuk menentukan hasil akhir. Prediksi model *Random Forest* dapat dinyatakan secara matematis dengan rumus berikut:

$$\hat{y} = mode\{h_1(x), h_2(x), \ldots, h_N(x)\} \quad (1)$$

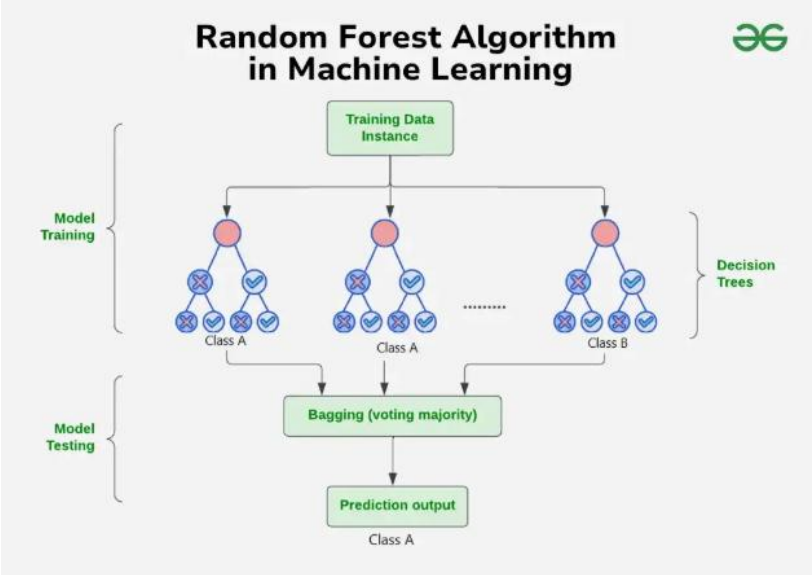
Keterangan:

- $\hat{y}$ = Prediksi akhir atau hasil klasifikasi gabungan dari semua model dalam *ensemble*.
- $mode\{.\}$ = Fungsi modus – yaitu nilai yang paling sering muncul di antara semua hasil prediksi dari model-model $h_i(x)$. Ini adalah inti dari *majority voting*.

- $h_i(x)$ = Prediksi dari model ke-i terhadap input x.
  Di sini, setiap h_i adalah satu model individu dalam *ensemble* (misalnya satu *decision tree* dalam *random forest*).
- N = Jumlah total model dalam *ensemble*.
- x = Input data (fitur yang ingin diklasifikasikan

Di mana $h_i(x)$ adalah hasil prediksi dari pohon ke-i terhadap input x, dan N adalah jumlah pohon dalam ensemble.

Dalam penelitian ini, model *Random Forest* dibangun dengan parameter 100 pohon keputusan (*n_estimators = 100*) dan *random_state* disetel ke 42 untuk memastikan reproduktibilitas hasil. *Random Forest* dipilih berdasarkan kemampuannya menangani data dengan *outlier* maupun *noise*, sesuai dengan pengembangan metode *Noise-Robust Random Forest* [7]. Struktur umum algoritma *Random Forest* diperlihatkan pada Gambar 2.



Gambar 2. Struktur Umum Algoritma Random Forest, sumber: geeksforgeeks.org

2.5 Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur performa klasifikasi. Penggunaan beberapa metrik ini bertujuan untuk memberikan gambaran yang lebih komprehensif terhadap kinerja model dalam mengklasifikasikan kategori volume penjualan yang berbeda [8].

3. HASIL DAN ANALISIS

1.1. Hasil Pelatihan dan Evaluasi Model

Model prediksi volume penjualan gadget dikembangkan menggunakan algoritma *Random Forest* dengan dataset berisi 1.000 data transaksi. Fitur-fitur utama yang digunakan dalam model ini meliputi promo (0 untuk tidak promo, 1 untuk promo), harga, *rating* pengguna, dan *channel* penjualan (online/offline). Dataset dibagi menjadi data latih sebanyak 80% dan data uji sebanyak 20%, menggunakan teknik *stratified sampling* untuk menjaga distribusi kelas yang proporsional.

Pelatihan model dilakukan dengan parameter *default*, menggunakan 100 pohon keputusan (*n_estimators = 100*) dan *random_state* diset ke 42 untuk menjaga reprodusibilitas hasil. Hasil evaluasi model pada data uji menunjukkan akurasi sebesar 49,5%. Perhitungan akurasi menggunakan formula:

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Total\ Prediksi} \times 100\% \quad (2)$$

Angka ini menunjukkan bahwa akurasi prediksi model masih terbatas, dengan hanya sekitar separuh dari keseluruhan data uji yang dapat diprediksi dengan benar. Dalam konteks aplikasi bisnis, tingkat akurasi seperti ini dinilai belum memadai untuk dijadikan dasar pengambilan keputusan strategis yang akurat dan dapat diandalkan, karena potensi kesalahan prediksi yang tinggi dapat berdampak signifikan pada hasil yang diharapkan.

Untuk mendapatkan pemahaman lebih mendalam mengenai kinerja model, evaluasi dilakukan menggunakan metrik *precision*, *recall*, dan *F1-score* untuk masing-masing kelas volume penjualan. Hasil evaluasi disajikan dalam Tabel 2.

Tabel 2. Hasil Evaluasi *Random Forest* per Kelas

Kelas	Precision	Recall	F1-Score
Rendah	0.12	0.07	0.08
Sedang	0.59	0.81	0.68
Tinggi	0.18	0.08	0.11

Analisis hasil evaluasi menunjukkan bahwa:

1. Kelas Sedang memiliki performa terbaik dengan *F1-score* sebesar 0.68, menandakan model relatif mampu membedakan volume penjualan dalam kategori ini.
2. Kelas Rendah dan Tinggi menunjukkan *F1-score* yang sangat rendah (0.08 dan 0.11), mengindikasikan ketidakmampuan model membedakan data kelas minoritas secara efektif.

Evaluasi *confusion matrix* memberikan informasi lebih rinci terkait distribusi prediksi yang salah dan benar.

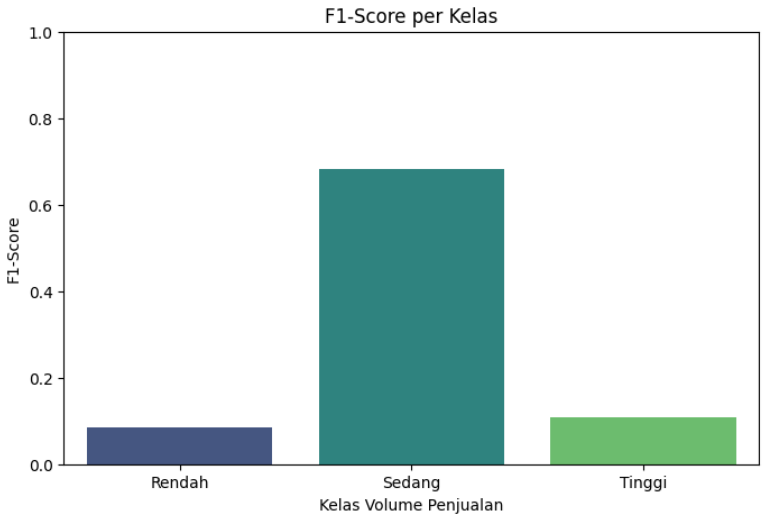
Tabel 3. Confusion Matrix Model Random Forest

	Pred: Rendah	Pred: Sedang	Pred: Tinggi
Actual Rendah	3	38	5
Actual Sedang	13	93	9
Actual Tinggi	9	27	3

Dari confusion matrix terlihat bahwa sebagian besar data kelas "Rendah" dan "Tinggi" salah diprediksi menjadi kelas "Sedang". Sebagai contoh:

1. Dari 46 data kelas "Rendah", hanya 3 yang diprediksi dengan benar sebagai "Rendah".
2. Dari 39 data kelas "Tinggi", hanya 3 yang diklasifikasikan dengan benar.
3. Sementara, 93 dari 115 data kelas "Sedang" berhasil diprediksi dengan benar, menunjukkan bias model terhadap kelas mayoritas.

Visualisasi distribusi performa model antar kelas ditunjukkan dalam Gambar 3. Gambar 3 menunjukkan visualisasi distribusi performa model antar kelas berdasarkan F1-score. Terlihat jelas bahwa kelas "Sedang" memiliki F1-score yang jauh lebih tinggi dibandingkan dengan kelas "Rendah" dan "Tinggi". Disparitas ini mengindikasikan adanya perbedaan signifikan dalam kemampuan model untuk memprediksi masing-masing kelas.



Gambar 3. Grafik F1-Score Volume Penjualan per Kelas

1.2. Analisis Penyebab Rendahnya Akurasi

Berdasarkan visualisasi pada Gambar 3 dan hasil evaluasi sebelumnya, kinerja *Random Forest* dalam penelitian ini memperlihatkan kecenderungan bias terhadap kelas dengan jumlah data terbanyak, yaitu "Sedang". Fenomena ini dikenal sebagai *class imbalance problem*. Ketika distribusi data antar kelas tidak seimbang, model cenderung lebih 'menghafal' pola dari kelas dominan, menyebabkan *recall* dan *precision* untuk kelas minoritas menjadi rendah.

Hasil ini menunjukkan bahwa model *Random Forest* tanpa teknik *balancing* data akan cenderung gagal mendeteksi kelas minoritas dengan baik. Untuk mengatasi masalah ini, teknik seperti SMOTE (*Synthetic Minority Over-sampling Technique*) dapat digunakan untuk mensintesis data baru pada kelas minoritas.

1.3. Pengaruh Keterbatasan Fitur

Selain ketidakseimbangan data, keterbatasan fitur juga berkontribusi terhadap rendahnya akurasi. Fitur "Promo" yang hanya berupa biner (Ya/Tidak) tidak mengandung informasi tentang besar diskon, jenis promo, atau durasi promo, sehingga kurang informatif. Padahal, intensitas diskon memiliki pengaruh signifikan terhadap peningkatan volume penjualan di platform *e-commerce*.

Fitur "Harga" dan "Rating" antar kelas juga menunjukkan distribusi yang saling tumpang tindih, sehingga tidak memberikan pemisahan data yang tajam di dalam *decision tree*. Studi Kumar et al. (2020) [5] menegaskan bahwa *overlapping features* menurunkan kemampuan model *Random Forest* dalam membentuk split yang efektif, menyebabkan prediksi menjadi tidak akurat.

1.4. Diskusi Perbandingan dengan Studi Sebelumnya

Dibandingkan penelitian lain, performa akurasi sebesar 49,5% tergolong rendah. Sejati et al. (2022) [16] menunjukkan bahwa *Random Forest* dapat mencapai akurasi 73,61% dalam prediksi pendaftaran mahasiswa. Pada aplikasi penjualan retail, Verdiyanto et al. (2025) [13] bahkan berhasil mencapai tingkat kesalahan prediksi hanya 2,85% setelah melakukan *hyperparameter tuning* dan *feature engineering* intensif.

Kondisi ini menunjukkan bahwa untuk mencapai kinerja lebih tinggi, perlu dilakukan pengembangan lanjutan pada:

1. *Balancing* data (menggunakan SMOTE, ADASYN, atau teknik *reweighting*).
2. *Enrichment* fitur (menambahkan data tentang tipe promosi, rating ulasan, lokasi geografis, dsb).
3. *Tuning hyperparameter* (mengatur kedalaman pohon, jumlah minimum sampel per daun, jumlah pohon).
4. Eksperimen model lain seperti *XGBoost* atau *LightGBM* yang lebih sensitif terhadap ketidakseimbangan data [8].

1.5. Makna dan Implikasi Temuan

Meskipun performa awal model masih rendah, penelitian ini berhasil menunjukkan bahwa variabel promo dan *channel* penjualan memiliki pengaruh terhadap prediksi volume penjualan. Hal ini mempertegas pentingnya mengelola strategi promosi dan distribusi secara cermat, serta memperkaya data bisnis untuk mendukung penerapan *machine learning* dalam pengambilan keputusan pemasaran

4. KESIMPULAN

Penelitian ini dilakukan untuk membangun model prediksi volume penjualan gadget berdasarkan fitur promo dan *channel* penjualan menggunakan algoritma *Random Forest*, sebagaimana telah dinyatakan dalam pendahuluan. Berdasarkan hasil dan analisis, model berhasil dikembangkan dan menunjukkan kemampuan prediksi meskipun dengan akurasi yang masih terbatas, yaitu sebesar 49,5%, serta performa terbaik pada kategori volume penjualan "Sedang". Hal ini membuktikan bahwa tujuan awal penelitian telah tercapai, yakni menghasilkan model prediksi berbasis data transaksi aktual, meskipun diperlukan pengembangan lebih lanjut untuk mengatasi ketidakseimbangan kelas dan keterbatasan fitur. Prospek pengembangan ke depan meliputi perluasan data fitur promosi, penggunaan teknik balancing data seperti SMOTE, tuning *hyperparameter*, serta eksplorasi algoritma alternatif seperti *XGBoost* atau *LightGBM* untuk meningkatkan akurasi dan generalisasi model. Selain itu, penerapan studi lanjut dapat diarahkan pada analisis prediksi musiman, segmentasi perilaku pelanggan, dan integrasi data real-time guna meningkatkan keandalan sistem prediksi dalam mendukung pengambilan keputusan strategis di bidang pemasaran gadget.

REFERENSI

[1] International Data Corporation (IDC), "Worldwide Smartphone Forecast Update, 2024–2028," Framingham, MA, Jan. 2024.
[2] Brownlee J, *Ensemble Learning Algorithms With Python*. Machine Learning Mastery, 2021.
[3] M. Putra and E. Harahap, "Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest ," *Jurnal Riset Mahasiswa* , vol. 4, no. 2, 2024.
[4] L. Breiman, "Random Forests," 2001.
[5] Kumar S, Choudhury S, and Verma A, "Sales forecasting for retail chains using ensemble models," *Procedia Comput Sci*, vol. 167, pp. 2400–2409, 2020.
[6] Wu Y, Chen Q, and Zhang L, "Customer Segmentation Based on RFM and Machine Learning," *Journal of Retailing and Consumer Services*, vol. 45, pp. 162–172, 2018.
[7] Suryawanshi M, "Predicting Customer Churn Using Random Forest," *Int J Comput Appl*, 2019.
[8] Chen T and Guestrin C, "XGBoost: A scalable tree boosting system," 2016, pp. 785–894.
[9] A. Hayat, M. U. Ashraf, M. Sajawal, S. Usman, and H. S. Alshaikh, "Predictive Analysis of Retail Sales Forecasting using Machine Learning Techniques," 2023, doi: 10.54692/lgurjcsit.2022.06004399.
[10] L. Budianti and Suliadi, "Metode Weighted Random Forest dalam Klasifikasi Prediksi Kelangsungan Hidup Pasien Gagal Jantung ," *BCSS*, vol. 2, no. 2, 2022.
[11] Y. Priantama and T. A. Y. Siswa, "Optimasi Correlation-Based Feature Selection untuk Random Forest Classifier ," *JIKO*, vol. 6, no. 2, 2022.
[12] R. Rindiyani, A. Primadewi, Maimunah, and A. H. Purwantini, "Klasifikasi Penjualan Berdasarkan Platform pada UMKM Omah Branded ," *JURIKOM*, vol. 9, no. 5, 2022.
[13] R. Verdiyanto, D. Hartanti, and E. Purwanto, "Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian ," *Jurnal Ilmu dan Teknologi (JIT)* , vol. 8, no. 1, 2025.
[14] T. A. Marzuqi, E. Kristiani, and Marcel, "Prediksi Mahasiswa Drop-Out di Universitas XYZ ," *JTI&IK* , 2024.
[15] Z. Zhou and Y. Ding, "Improving Robustness of Random Forest Under Label Noise ," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa Village, Hawaii, USA: IEEE, Jan. 2019, pp. 1170–1178.
[16] P. Sejati, M. Munawar, M. Pilliang, and H. Akbar, "Studi Komparasi Naive Bayes, K-Nearest Neighbor, dan Random Forest untuk Prediksi Calon Mahasiswa ," *JTI&IK* , 2022.