Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

Proyeksi Tren Kategori Pakaian Mendatang Menggunakan Random Forest pada Data Transaksi Pelanggan

Nur' Aini Umar *, Andi Ircham Hidayat, Eka Wijaya Paula
Fakultas Teknologi Industri, Sistem & Teknologi Informasi, Institut Teknologi dan Bisnis Nobel Indonesia, Makassar, Indonesia
Email: [1*] nurainnii272@gmail.com, [2] hidayat.ircham@gmail.com, [3] wijayapaula25@gmail.com
Email Penulis Korespondensi: nurainnii272@gmail.com

Abstrak− Industri fesyen yang dinamis membutuhkan proyeksi tren yang akurat untuk pemasaran, dan pengembangan produk. Penelitian ini bertujuan memproyeksikan kategori pakaian tren mendatang menggunakan Random Forest sebagai model utama dan XGBoost sebagai model kedua. Dataset utama berisi 3.900 transaksi dengan informasi demografis, riwayat pembelian, musiman, dan kategori produk. Untuk konteks lokal, diintegrasikan data stok toko fesyen "Coffer Ruh" sebagai studi kasus pendamping. Metodologi, pra-pemrosesan, penanganan ketidakseimbangan dengan SMOTE, pembagian stratified (80:20), pelatihan Random Forest dan XGBoost, serta evaluasi menggunakan akurasi, precision, recall, F1-score, dan confusion matrix. Hasil evaluasi (Outerwear, Footwear, Bottoms, Tops, Accessories) menunjukkan Random Forest mencapai akurasi 67,56%, presisi tertimbang 68,04%, recall 67,56%, dan F1-score 67,79%, sementara XGBoost menunjukkan performa serupa dengan akurasi sekitar 68%. Model Random Forest memproyeksikan Jaket (17%), Mantel (12%), dan Sepatu (10%) sebagai tiga kategori tren global teratas. Analisis data toko menunjukkan stok tertinggi, child mask (35 pcs), cornerstick merah (33 pcs, mantel), dan drams (31 pcs, jaket). Terdapat kesesuaian, dua dari tiga produk dengan stok tertinggi termasuk outerwear sesuai tren global, namun masker bukan kategori pakaian yang diprediksi. Karena keterbatasan data toko (sampel kecil, tanpa dimensi waktu/transaksi), pendekatan transfer learning atau hybrid dataset belum dapat diterapkan, yang disebut sebagai keterbatasan dan arah riset masa depan.

Kata Kunci: Random Forest, XGBoost, prediksi tren, fesyen, data transaksi

Abstract− The dynamic fashion industry requires accurate trend projections for marketing and product development. This study aims to project future clothing trends using Random Forest as the primary model and XGBoost as the secondary model. The main dataset contains 3,900 transactions with demographic information, purchase history, seasonal data, and product categories. For local context, inventory data from the "Coffer Ruh" fashion store was integrated as a companion case study. The methodology included preprocessing, handling class imbalance with SMOTE, stratified splitting (80:20), training Random Forest and XGBoost, and evaluation using accuracy, precision, recall, F1-score, and a confusion matrix. Evaluation results (Outerwear, Footwear, Bottoms, Tops, Accessories) show that Random Forest achieved an accuracy of 67.56%, weighted precision of 68.04%, recall of 67.56%, and an F1-score of 67.79%, while XGBoost demonstrated similar performance with an accuracy of approximately 68%. The Random Forest model projected Jackets (17%), Coats (12%), and Shoes (10%) as the top three global trend categories. Store data analysis revealed the highest stock levels for children's masks (35 pcs), red cornersticks (33 pcs, coats), and drams (31 pcs, jackets). There is some alignment: two of the three products with the highest inventory are outerwear items that align with global trends; however, masks are not a predicted apparel category. Due to limitations in the store data (small sample size, lack of time/transaction dimensions), transfer learning or hybrid dataset approaches cannot yet be applied, which is identified as a limitation and a direction for future research.

Keywords: Random Forest, XGBoost, trend forecasting, fashion, transaction data,

1. PENDAHULUAN

Industri fesyen merupakan salah satu sektor ekonomi paling dinamis yang perubahannya berlangsung dengan cepat, mengikuti tren serta selera pasar yang terus berganti [1]. Perkembangan sektor ini sangat dipengaruhi oleh berbagai faktor kompleks seperti perubahan perilaku konsumen, tren sosial budaya, kondisi ekonomi global, dan inovasi teknologi [2] Kemampuan untuk mengantisipasi dan memprediksi tren masa depan menjadi kunci kesuksesan bagi pelaku industri dalam menyusun strategi pemasaran yang efektif, mengelola inventaris secara efisien, dan mengembangkan produk yang sesuai dengan kebutuhan pasar. Ketidakmampuan dalam membaca arah tren dapat mengakibatkan kerugian finansial yang signifikan akibat kelebihan stok produk yang tidak diminati atau kehilangan momentum pasar. Dalam era digital saat ini, ketersediaan data historis transaksi pelanggan membuka peluang besar untuk membangun model prediktif yang dapat mengidentifikasi pola pembelian dan mengantisipasi kategori produk yang akan populer di masa mendatang [3]. Penelitian [4] secara khusus menggunakan pendekatan metode yang memiliki karakteristik serupa dengan penelitian yang di teliti, yaitu berfokus pada deteksi dini untuk meningkatkan keberhasilan transaksi secara real-time melalui analisis data dan klasifikasi prediktif.

Berbagai teknik data mining dan pembelajaran mesin telah dieksplorasi untuk mengekstraksi pengetahuan dari data transaksi. Salah satu metode ensemble learning yang terbukti efektif dalam menangani data dengan banyak fitur adalah Random Forest [5]. Metode ini menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi yang akurat dan robust, memiliki keunggulan dalam mengatasi overfitting, menangani data berdimensi tinggi, serta memberikan informasi mengenai kepentingan fitur (feature importance) yang dapat diinterpretasikan [6]. Penelitian [7] mengoptimalkan pemeringkatan kredit menggunakan ensemble feature selection dengan Random Forest, menunjukkan keunggulannya pada data dengan banyak fitur. [8] Memberikan tinjauan komprehensif mengenai algoritma Random Forest sebagai metode ensemble yang powerful. Lebih relevan lagi, [9] secara khusus menerapkan metode prediksi untuk produk baru di ritel fesyen dengan data terbatas, sebuah tantangan yang serupa dengan prediksi tren jangka panjang. Penelitian [10] membuktikan bahwa penerapan random forest pada data transaksi pelanggan mampu mengidentifikasi pola musiman dan

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

preferensi. Metode Random Forest lebih unggul dibandingkan decision tree tunggal dalam memprediksi tren fesyen karena kemampuannya mengurangi overfitting melalui agregasi banyak pohon keputusan [11].Penelitian lain juga menyoroti pentingnya aspek kontekstual [12]. Menekankan peran pengelolaan hubungan pelanggan (Customer Relationship Management) dan pemanfaatan data transaksi untuk memahami perilaku konsumen. [13] Memberikan gambaran umum tentang teknik data mining untuk analisis tren. [13] Menemukan bahwa faktor musiman memiliki kontribusi signifikan dalam menentukan preferensi konsumen pakaian. Sementara itu, [14] menganalisis perilaku konsumen Generasi Z dalam belanja online, memberikan wawasan tentang karakteristik demografis yang memengaruhi keputusan pembelian.

Meskipun berbagai penelitian telah menunjukkan efektivitas Random Forest dalam prediksi pada berbagai domain, masih terdapat celah penelitian (research gap) dalam penerapannya untuk prediksi tren pakaian jangka panjang, berdasarkan data transaksi pelanggan. Kebaruan yang ditawarkan dalam penelitian ini adalah penggunaan dataset tren belanja pelanggan yang tersedia publik [15] yang dipadukan dengan pendekatan Random Forest untuk memproyeksikan kategori pakaian dominan di masa depan, serta membandingkan performanya dengan algoritma XGBoost sebagai pembanding untuk memastikan robustitas hasil. Oleh karena itu, penelitian ini bertujuan untuk memprediksi tren kategori pakaian yang akan dominan pada tahun mendatang dengan memanfaatkan dataset historis transaksi pelanggan menggunakan metode Random Forest sebagai model utama. Dengan mempertimbangkan fitur-fitur seperti demografi pelanggan, musim, riwayat pembelian, dan variabel transaksional lainnya, model Random Forest diharapkan mampu memprediksi probabilitas tren pada setiap item fesyen pada tahun mendatang. Hasil prediksi ini selanjutnya dibandingkan dengan data stok dari toko fesyen lokal sebagai studi kasus pendamping untuk menguji keselarasannya dengan realitas ritel.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan research and development (R&D) dengan menerapkan metode machine learning khususnya algoritma Random Forest sebagai model utama dan XGBoost sebagai pembanding untuk memprediksi tren pakaian tahun mendatang. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, pemilihan fitur, penanganan ketidakseimbangan kelas, pembagian data, pelatihan model, evaluasi, dan prediksi. Seluruh proses dilakukan menggunakan bahasa pemrograman Python dengan pustaka scikit-learn, imbalanced-learn, xgboost, serta pustaka pendukung lainnya.

Gambar 1. Diagram alur penelitian

2.1 Sumber dan Deskripsi Data

Data yang digunakan dalam penelitian ini bersumber dari platform Kaggle dengan judul "Customer Shopping Trends Dataset". Dataset ini terdiri dari 3.900 catatan transaksi pelanggan dengan 19 atribut, meliputi informasi demografis (usia, jenis kelamin, lokasi), detail transaksi (item yang dibeli, kategori produk, jumlah pembelian, metode pembayaran), serta perilaku pelanggan (frekuensi pembelian, ulasan, status langganan). Tabel 1 menyajikan deskripsi lengkap variabel dalam dataset.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

Tabel 1. Deskripsi Variabel Dataset

No.	Nama Kolom	Tipe Data	Deskripsi
1	Customer ID	Numerik	Identitas unik pelanggan
2	Age	Numerik	Usia pelanggan
3	Gender	Kategorikal	Jenis kelamin (Male/Female)
4	Item Purchased	Kategorikal	Nama item spesifik yang dibeli
5	Category	Kategorikal	Kategori produk (Clothing, Accessories, dll.)
6	Purchase Amount (USD)	Numerik	Jumlah pembelian dalam dolar
7	Location	Kategorikal	Lokasi atau negara bagian pelanggan
8	Size	Kategorikal	Ukuran pakaian (S, M, L, XL)
9	Color	Kategorikal	Warna produk
10	Season	Kategorikal	Musim saat pembelian (Spring, Summer, dll.)
11	Review Rating	Numerik	Rating ulasan (skala 1-5)
12	Subscription Status	Kategorikal	Status langganan (Yes/No)
13	Payment Method	Kategorikal	Metode pembayaran (Cash, Credit Card, dll.)
14	Shipping Type	Kategorikal	Jenis pengiriman (Express, Standard, dll.)
15	Discount Applied	Kategorikal	Apakah diskon diterapkan (Yes/No)
16	Promo Code Used	Kategorikal	Apakah kode promo digunakan (Yes/No)
17	Previous Purchases	Numerik	Jumlah pembelian sebelumnya
18	Preferred Payment Method	Kategorikal	Metode pembayaran preferensi pelanggan
19	Frequency of Purchases	Kategorikal	Frekuensi pembelian (Weekly, Monthly, dll.)

2.2 Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk membersihkan dan menyiapkan data agar dapat digunakan oleh model [16] Langkah-langkahnya adalah sebagai berikut:Penghapusan Kolom Tidak Relevan, Kolom Customer ID dihapus karena tidak memiliki pengaruh terhadap pola pembelian dan bersifat unik sehingga dapat menyebabkan overfitting. [17] Mengatasi data yang kosong adalah langkah untuk memastikan hasil analisis tetap akurat dan tidak menimbulkan kesalahan, Berdasarkan eksplorasi awal, tidak ditemukan nilai hilang (missing values) pada seluruh kolom, sehingga tidak diperlukan imputasi. Encoding Variabel Kategori, Seluruh variabel kategori diubah menjadi bentuk numerik menggunakan teknik one-hot encoding dengan fungsi pd.get_dummies() dari pustaka pandas python. Metode ini dipilih karena tidak mengasumsikan adanya urutan antar kategori. Setelah encoding, jumlah fitur menjadi 147 kolom. Pemisahan Fitur dan Target, Target penelitian adalah item spesifik yang dibeli, yaitu kolom Item Purchased. Fitur (X) adalah seluruh kolom selain target. Dengan demikian, model akan mempelajari pola hubungan antara karakteristik pelanggan dan transaksi dengan item yang dipilih. Penghapusan Fitur Penyebab Kebocoran Data, Untuk menghindari data leakage, seluruh kolom hasil one-hot encoding yang mengandung frasa "Item Purchased_" dihapus dari fitur. Hal ini dikarenakan informasi tentang item yang dibeli tidak boleh diketahui sebelum prediksi. Setelah penghapusan, jumlah fitur berkurang menjadi 124.

2.3 Penanganan Ketidakseimbangan Kelas

Distribusi item dalam dataset tidak merata, dengan beberapa item memiliki frekuensi jauh lebih rendah dibanding lainnya. Ketidakseimbangan kelas dapat menyebabkan model bias terhadap kelas mayoritas. Untuk mengatasi hal ini, diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) pada data latih. SMOTE menghasilkan sampel sintetis untuk kelas minoritas dengan melakukan interpolasi antar sampel yang ada. Penerapan SMOTE dilakukan setelah pembagian data latih dan uji untuk menghindari data leakage dari data uji [18]

2.3 Pembagian Data

Dataset dibagi menjadi dua bagian: data latih (training set) dan data uji (testing set) dengan proporsi 80:20 [19]. Pembagian dilakukan secara stratified berdasarkan target Item Purchased untuk memastikan proporsi setiap kelas tetap terjaga baik di data latih maupun data uji. Parameter random state ditetapkan sebesar 42 untuk memastikan reproduksibilitas hasil.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

2.4 Pemodelan dengan Random Forest

Algoritma utama yang digunakan adalah Random Forest Classifier dari pustaka scikit-learna [20] Random Forest merupakan metode ensemble yang menggabungkan banyak pohon keputusan melalui teknik bootstrap aggregating (bagging) dan pemilihan fitur acak. Parameter yang digunakan dalam model ini adalah:

- n_estimators = 100 (jumlah pohon)
- random_state = 42
- n_jobs = -1 (menggunakan seluruh prosesor)
- class_weight = 'balanced' (memberi bobot lebih pada kelas minoritas)

Model dilatih menggunakan data yang telah di-resampling dengan SMOTE. Setelah pelatihan, model dievaluasi pada data uji.

Meskipun belum dilakukan optimasi parameter secara ekstensif, pemilihan parameter dasar (n_estimators=100, class_weight='balanced') didasarkan pada praktik standar untuk menangani ketidakseimbangan kelas dan menjaga keseimbangan antara kompleksitas model dengan waktu komputasi. Penggunaan parameter ini telah sesuai dengan rekomendasi dari studi sebelumnya [20] untuk permasalahan klasifikasi dengan jumlah kelas yang besar.

Untuk memproyeksikan tren di masa mendatang, penelitian ini menggunakan pendekatan average customer profile. Langkah pertama adalah membangun vektor fitur rata-rata ($\bar{x}$) dari seluruh data latih. Vektor ini merepresentasikan karakteristik seorang "pelanggan tipikal". Perhitungannya menggunakan rumus rata-rata sederhana:

Vektor Fitur Pelanggan Rata-rata

$$\bar{x} = 1/N \sum\_(i = 1)\^N x\_i$$

Keterangan:

$\bar{x}$ = vektor fitur rata-rata (profil pelanggan)

N = jumlah total sampel dalam data latih (3.120 sampel)

x_i = vektor fitur dari sampel pelanggan ke-i

Vektor $\bar{x}$ ini kemudian dimasukkan ke dalam model Random Forest menggunakan Rumus

$$P(y = c \mid x) = 1/T \sum\_(t = 1)\^T P\_t(y = c \mid x)$$

untuk menghitung probabilitas setiap item pakaian. Hasil probabilitas ini kemudian diurutkan untuk melihat item mana yang memiliki kecenderungan tertinggi untuk menjadi tren.

2.5 Pemodelan dengan XGBoost (Pembanding)

Sebagai pembanding, digunakan algoritma XGBoost (Extreme Gradient Boosting) yang merupakan implementasi gradient boosting yang efisien dan banyak digunakan dalam kompetisi machine learning [21] Karena XGBoost memerlukan target dalam bentuk numerik, dilakukan label encoding pada target menggunakan LabelEncoder dari scikit-learn. Parameter yang digunakan adalah:

- n_estimators = 100
- random_state = 42
- n_jobs = -1
- use_label_encoder = False
- eval_metric = 'mlogloss'

Sama seperti Random Forest, model XGBoost dilatih pada data yang telah di-resampling SMOTE. Hasil prediksi kemudian dikonversi kembali ke label asli menggunakan inverse transform dari label encoder. Sebagai pembanding, digunakan algoritma XGBoost (Extreme Gradient Boosting) yang merupakan implementasi gradient boosting yang efisien dan banyak digunakan dalam kompetisi machine learning [22].

XGBoost merupakan metode ensemble berbasis boosting yang membangun model secara bertahap dengan meminimalkan fungsi objektif. Berbeda dengan Random Forest yang menggunakan pendekatan bagging (paralel), XGBoost membangun pohon secara berurutan di mana setiap pohon baru dikoreksi untuk mengurangi kesalahan dari pohon sebelumnya [22]. Secara matematis, XGBoost memodelkan probabilitas untuk setiap kelas c menggunakan fungsi softmax:

$$P(y = c \mid x) = exp(f\_c(x)) / \sum\_(j = 1)\^K exp(f\_j(x))$$

di mana $P(y=c|x)$ adalah probabilitas bahwa sampel x termasuk ke dalam kelas c, $fc(x)$ adalah output dari model boosting untuk kelas c, dan K adalah jumlah total. Fungsi $fc(x)$ sendiri merupakan penjumlahan dari M pohon keputusan yang dibangun secara bertahap :

$$f\_c(x) = \sum\_(m = 1)\^M h\_m\^((c))(x)$$

di mana M adalah jumlah pohon ($n\_estimators$ = 100) dan $hm(c)(x)$ adalah prediksi dari pohon ke-m untuk kelas c. Setiap pohon baru hm dibangun untuk meminimalkan fungsi objektif yang terdiri dari dua komponen: fungsi *loss* (kesalahan) dan fungsi regularisasi (mencegah *overfitting*):

$$L = \sum\_(i = 1)\^N l(y\_i,\hat(y)\_i) + \sum\_(m = 1)\^M \Omega(h\_m)$$

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

di mana $l(yi,y\hat{}i)$ adalah fungsi *loss* yang mengukur kesalahan prediksi $y\hat{}i$ terhadap nilai aktual *yi* (menggunakan *mlogloss* untuk klasifikasi multi-kelas), dan $\Omega(hm)$ adalah fungsi regularisasi yang mengontrol kompleksitas pohon, dirumuskan sebagai:

$$\backslash Omega(h) = \backslash gamma\, T + 1/2\, \backslash lambda\, \backslash sum\_(j = 1)\hat{}T\, w\_j\hat{}2$$

di mana *T* adalah jumlah daun pada pohon, *wj* adalah bobot pada daun ke-*j*, serta γ dan λ adalah parameter regularisasi yang mengendalikan penalti

Karena XGBoost memerlukan target dalam bentuk numerik, dilakukan *label encoding* pada target menggunakan Label encoder dari scikit-learn. Parameter yang digunakan adalah:

- *n_estimators* = 100
- *random_state* = 42
- *n_jobs* = -1
- *use_label_encoder* = False
- *eval_metric* = 'mlogloss'

Sama seperti Random Forest, model XGBoost dilatih pada data yang telah di-resampling SMOTE. Hasil prediksi kemudian dikonversi kembali ke label asli menggunakan inverse transform dari label encoder.

2.6 Evaluasi Model

Kinerja kedua model dievaluasi menggunakan metrik-metrik berikut :

- Akurasi: Proporsi prediksi benar terhadap total data uji.
- Precision, Recall, dan F1-Score: Diukur untuk setiap kelas guna melihat performa model pada masing-masing item.
- Confusion Matrix: Menyajikan tabulasi prediksi vs aktual untuk setiap kelas.
- Feature Importance: Mengukur kontribusi setiap fitur dalam prediksi, diambil dari atribut feature_importances_ model.

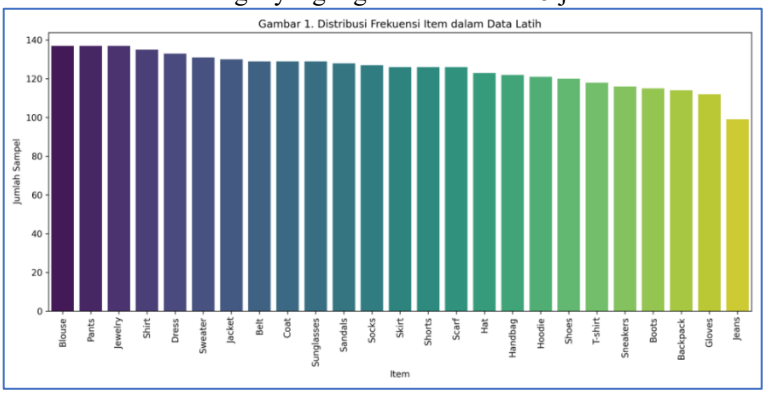
Untuk memastikan stabilitas dan generalisasi model, dilakukan Stratified 5-Fold Cross-Validation pada data latih sebelum evaluasi akhir. Metode ini dipilih karena mampu menjaga proporsi setiap kelas pada setiap lipatan (fold), sehingga mengurangi bias akibat distribusi kelas yang tidak seimbang [18] Model dilatih pada 4 lipatan dan divalidasi pada 1 lipatan secara bergantian sebanyak 5 kali, sehingga setiap sampel pada data latih berkesempatan menjadi data validasi. Nilai rata-rata akurasi dari 5 lipatan digunakan sebagai indikator awal stabilitas model.

Untuk memproyeksikan tren pakaian pada tahun mendatang, digunakan pendekatan average customer profile. Vektor fitur rata-rata dihitung dari seluruh data latih (X_train.mean()) yang merepresentasikan karakteristik pelanggan tipikal. Vektor ini kemudian dimasukkan ke dalam model terlatih (Random Forest sebagai model utama dan XGBoost sebagai kedua) untuk mendapatkan probabilitas setiap item. Probabilitas yang dihasilkan menunjukkan seberapa besar kemungkinan suatu item dipilih oleh pelanggan rata-rata di masa depan. Hasil prediksi disusun dalam bentuk tabel dan diurutkan berdasarkan probabilitas tertinggi.

3. HASIL DAN PEMBAHASAN

3.1 Eksplorasi dan Distribusi Data

Dataset awal terdiri dari 3.900 transaksi dengan 19 atribut. Setelah penghapusan kolom Customer ID dan one-hot encoding pada variabel kategorikal, diperoleh 147 fitur. Fitur yang berasal dari Item Purchased dihapus untuk mencegah kebocoran data, sehingga tersisa 124 fitur. Target yang digunakan adalah 25 jenis item.



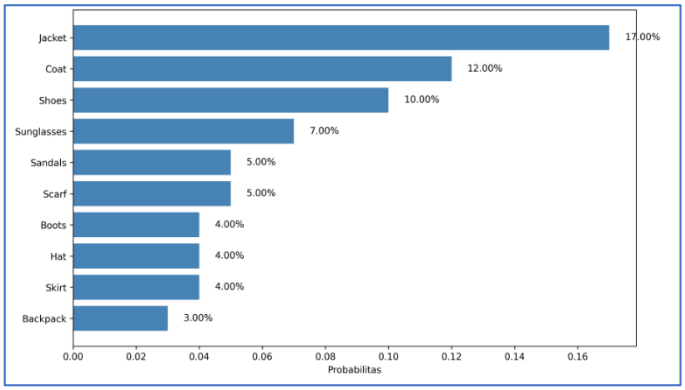
Gambar 2. Distribusi frekuensi item dalam data latih

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

Berdasarkan Gambar 2, terlihat bahwa distribusi item dalam dataset tidak merata. Beberapa item seperti Jacket, Coat, dan Shoes memiliki frekuensi yang relatif lebih tinggi dibandingkan item lainnya seperti Socks atau Jeans. Ketidakseimbangan ini menjadi pertimbangan utama dalam penerapan teknik SMOTE pada tahap pelatihan model.

3.2 Prediksi Tren Menggunakan Random Forest

Proyeksi tren pakaian dilakukan menggunakan pendekatan average customer profile, yaitu dengan menghitung vektor fitur rata-rata dari seluruh data latih yang merepresentasikan karakteristik pelanggan. Vektor ini kemudian dimasukkan ke dalam model Random Forest yang telah dilatih untuk mendapatkan probabilitas setiap item. Gambar 3 menyajikan 10 item dengan probabilitas tertinggi menurut model Random Forest.

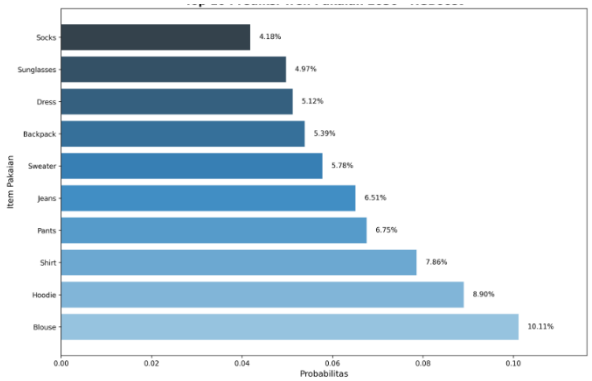


Gambar 3. Item teratas prediksi mendatang (Random Forest)

Berdasarkan hasil prediksi model Random Forest, Jacket menempati peringkat pertama dengan probabilitas 17,00% , diikuti oleh Coat (12,00%) dan Shoes (10,00%). Dominasi ketiga item ini mengindikasikan bahwa outerwear dan alas kaki merupakan kebutuhan pokok konsumen dengan pola pembelian yang stabil dan mudah dikenali oleh model.

3.3 Prediksi Tren Menggunakan XGBoost dan Perbandingan

Proyeksi serupa dilakukan menggunakan model XGBoost sebagai pembanding. Gambar 4 menunjukkan 10 item dengan probabilitas tertinggi menurut model XGBoost.



Gambar 4. Item teratas prediksi mendatang (XGBoost)

Berbeda dengan Random Forest, XGBoost memproyeksikan Blouse (10,11%) , Hoodie (8,90%) , dan Shirt (7,86%) sebagai tiga item teratas. Tingginya probabilitas Blouse dan Hoodie menunjukkan bahwa XGBoost lebih sensitif terhadap preferensi kasual dan busana santai. Item-item seperti Jeans (6,51%), Pants (6,75%), dan Sweater (5,78%) juga masuk dalam 10 besar, menunjukkan bahwa pakaian dasar sehari-hari tetap memiliki probabilitas signifikan. Sementara itu, item seperti T-shirt (2,09%) dan Jacket (1,24%) justru berada di peringkat bawah, yang mengindikasikan bahwa XGBoost lebih menekankan pada item dengan pola pembelian yang lebih spesifik dan kasual dibandingkan item fungsional atau universal.

Tabel 2. Perbandingan Top 10 Prediksi Random Forest dan XGBoost

Peringkat	Random Forest	Probabilitas	XGBoost	Probabilitas
1	Jacket	17,00%	Blouse	10,11%
2	Coat	12,00%	Hoodie	8,90%
3	Shoes	10,00%	Shirt	7,86%

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

4	Sunglasses	7,00%	Pants	6,75%
5	Sandals	5,00%	Jeans	6,51%
6	Scarf	5,00%	Sweater	5,78%
7	Boots	4,00%	Backpack	5,39%
8	Hat	4,00%	Dress	5,12%
9	Skirt	4,00%	Sunglasses	4,97%
10	Backpack	3,00%	Socks	4,18%

Random Forest memproyeksikan Jacket, Coat, dan Shoes sebagai tiga besar, yang mengindikasikan dominasi outerwear dan alas kaki sebagai kebutuhan pokok konsumen. Sebaliknya, XGBoost menghasilkan Blouse, Hoodie, dan Shirt sebagai tiga besar, yang mencerminkan preferensi terhadap busana kasual dan santai. Perbedaan ini menunjukkan bahwa Random Forest dengan pendekatan bagging menghasilkan prediksi yang lebih stabil dan konservatif, sedangkan XGBoost dengan pendekatan boosting lebih sensitif terhadap variasi pola lokal.

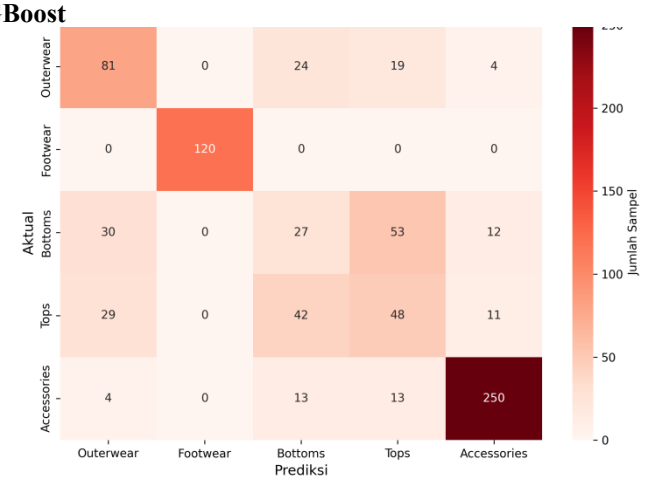
3.4 Evaluasi Performa Model

3.4.1 Confusion Matrix Random Forest

Gambar 5. Confusion Matrix – Random Forest

Berdasarkan Gambar 5, model Random Forest mencapai akurasi sebesar 67,56% , presisi tertimbang 68,04%, recall 67,56% , serta F1-score 67,79%. Kategori Footwear mencapai performa sempurna (100% di seluruh metrik), diikuti oleh Accessories dengan F1-score 89%, yang mengindikasikan bahwa kedua kategori ini memiliki karakteristik pembelian yang sangat khas dan mudah dibedakan oleh model. Kategori Outerwear yang mencakup Jacket dan Coat—dua item dengan probabilitas tren tertinggi—mencapai F1-score 60%. Kategori Bottoms (30%) dan Tops (35%) masih menjadi tantangan karena kemungkinan adanya tumpang tindih karakteristik antar item di dalamnya.

3.4.2 Confusion Matrix XGBoost



Gambar 6. Confusion Matrix – XGBoost

Model XGBoost menunjukkan performa yang hampir identik dengan Random Forest, dengan akurasi keseluruhan mencapai sekitar 68%. Kategori Footwear kembali menunjukkan performa sempurna (100%), menegaskan bahwa pola

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

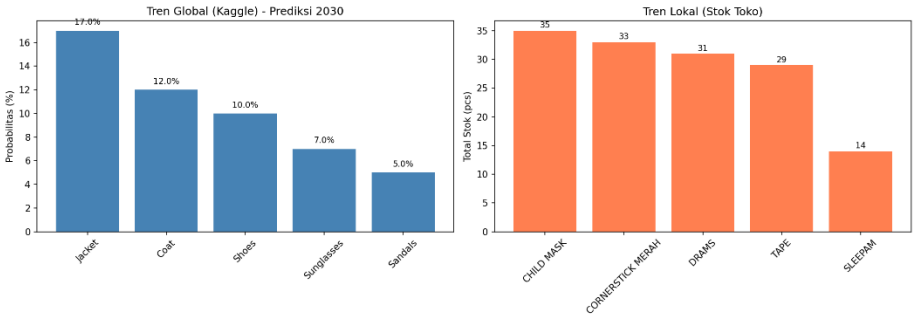
pembelian alas kaki sangat khas dan konsisten dikenali oleh kedua model. Kategori Accessories menunjukkan F1-score di atas 90%, mengindikasikan bahwa aksesoris memiliki karakteristik pembelian yang unik dan mudah dibedakan. Kategori Outerwear mencapai F1-score sekitar 60%, sedangkan Bottoms dan Tops masih menjadi tantangan dengan F1-score di bawah 40%, yang mengindikasikan perlunya penambahan fitur yang lebih diskriminatif seperti preferensi merek, harga, atau data perilaku konsumen yang lebih kaya.

3.5 Studi Kasus Toko Lokal dan Perbandingan Tren

Selain menggunakan dataset utama [15], penelitian ini mengintegrasikan data penjualan dari toko fesyen lokal "Coffer Ruh" sebagai studi kasus pendamping untuk mengkaji keselarasan antara proyeksi global dan realitas ritel lokal. Data toko diperoleh dari catatan stok fisik yang terdiri dari 45 rekaman dengan 15 produk unik.

Tabel 3. Sepuluh Produk dengan Total Penjualan Tertinggi (Data Toko Lokal)

Peringkat	Nama Produk	Total (pcs)	Kategori yang Dipetakan
1	Child mask	35	Aksesoris (Masker)
2	Cornerstick merah	33	Outerwear (Mantel)
3	Drams	31	Outerwear (Jaket)
4	Tape	29	Alas Kaki (Sepatu)
5	Alaba po	12	Blus
6	Sleepam	14	Pakaian Umum
7	Acaba pee	10	Sweter
8	Dave for your self	12	Jeans
9	Cornerstick hitay	11	Hoodie
10	1312 mask	8	Aksesoris



Gambar 7. menyajikan perbandingan antara proyeksi global (Random Forest) dan indikasi lokal

Hasil pada Gambar 7 menunjukkan keselarasan parsial antara proyeksi random forest dan indikasi lokal. Dua dari tiga produk dengan stok tertinggi di toko Mantel dan Jaket sesuai dengan tren global yang diproyeksikan oleh model Random Forest. Namun, produk dengan stok tertinggi, yaitu child mask (masker), tidak termasuk dalam kategori pakaian yang diprediksi oleh model, sehingga mengurangi tingkat keselarasan. Meskipun demikian, kesesuaian pada kategori outerwear mengindikasikan bahwa tren global yang dihasilkan dari data transaksi Kaggle memiliki relevansi dengan persediaan toko lokal, khususnya pada kategori pakaian luar dan alas kaki, yang memperkuat validitas hasil proyeksi model pada level kategori.

4. KESIMPULAN

Berdasarkan hasil proyeksi menggunakan Random Forest sebagai model utama, tiga item pakaian yang memiliki probabilitas tertinggi untuk tren mendatang adalah Jacket (17,00%) , Coat (12,00%) , dan Shoes (10,00%). Dominasi outerwear dan alas kaki ini menunjukkan bahwa kedua kategori tersebut memiliki pola pembelian yang relatif stabil dan konsisten di berbagai segmen pelanggan. Perbandingan dengan XGBoost menunjukkan perbedaan karakteristik yang cukup menarik. Random Forest dengan pendekatan bagging cenderung menghasilkan proyeksi yang lebih stabil dengan dominasi outerwear, sementara XGBoost yang menggunakan teknik boosting lebih sensitif terhadap preferensi kasual seperti blouse, hoodie, dan kemeja. Perbedaan ini memperkuat alasan pemilihan Random Forest sebagai model yang lebih

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

cocok untuk proyeksi jangka panjang. Hasil ini setidaknya bisa menjadi bahan pertimbangan awal bagi pelaku industri fesyen, terutama dalam memprioritaskan pengembangan produk jacket, coat, dan shoes. Aksesori pelengkap yang masuk dalam 10 besar prediksi juga tetap patut diperhatikan. Sementara itu, dari sisi stok toko lokal, terlihat kesesuaian parsial di mana dua produk lokal (cornerstick dan drams) termasuk dalam kategori outerwear yang sejalan dengan tren global, meskipun produk dengan stok tertinggi (masker) berada di luar kategori pakaian yang diprediksi.

Perlu dicatat bahwa hasil ini masih bersifat proyeksi berdasarkan data historis dan tentu saja belum teruji dengan data aktual di masa mendatang. Ke depannya, akan lebih baik jika model diuji pada data periode mendatang serta dilengkapi dengan fitur-fitur tambahan seperti preferensi merek, harga, atau tren media sosial untuk meningkatkan akurasi prediksi di level item yang lebih spesifik

REFERENCES

[1] M. Fernanda, R. I. S. Setiawati, and M. Wahed, "Penerapan Analisis Keranjang Belanja Pasar untuk Manajemen Ketersediaan Stok dalam Ekonomi Industri: Mengantisipasi Perubahan Tren," *JURNAL RUMPUN MANAJEMEN DAN EKONOMI*, vol. 1, no. 1, pp. 184–190, Mar. 2024, doi: 10.61722/jrme.v1i1.1153.

[2] S. Roslow, T. Li, and J. A. F. Nicholls, "Impact of situational variables and demographic attributes in two seasons on purchase behaviour," *European Journal of Marketing*, vol. 34, no. 9–10, pp. 1167–1180, Oct. 2000, doi: 10.1108/03090560010342548.

[3] A. Rahman, "URGENSI PEMBUATAN MODEL PREDIKTIF DALAM TATA KELOLA BISNIS," *Jurnal Ilmiah Multidisiplin Ilmu*, vol. 2, no. 2, pp. 78–87, Apr. 2025, doi: 10.69714/jvtm6q52.

[4] R. Kartika and R. D. H. U. N. R. Saputra, "DECISION TREE-BASED PREDICTIVE MODELING FOR MOBILE PAYMENT TRANSACTION SUCCESS: A CASE STUDY OF SHOPEEPAY AND DANA," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, pp. 8895–8899, Jul. 2025, doi: 10.36040/jati.v9i5.15245.

[5] G. Almuzadid and E. R. Subhiyakto, "Stroke Risk Classification Using the Ensemble Learning Method of XGBoost and Random Forest," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 828–837, Jun. 2025, doi: 10.30871/jaic.v9i3.9528.

[6] H. Haeruddin, E. Erick, and H. W. Aripradono, "Perbandingan Support Vector Machine, Random Forest Classifier, dan K-Nearest Neighbour dalam Pendeteksian Anomali pada Jaringan DDos," *Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 1, pp. 23–33, Jan. 2025, doi: 10.35746/jtim.v7i1.628.

[7] A. Fauziah, "Optimizing Credit Scoring Performance Using Ensemble Feature Selection with Random Forest," *Jurnal Matematika, Statistika dan Komputasi*, vol. 21, no. 2, pp. 560–572, Jan. 2025, doi: 10.20956/j.v21i2.42032.

[8] A. S. Sulaeman, T. Y. Agustian, and A. S. Sunge, "Komparasi Algoritma MultipleLinearRegression, Random Forest, dan XGBoostuntuk Prediksi Pengangguran Terbuka Berdasarkan Tingkat Pendidikan di Jawa Barat," vol. 7, no. 3, 2025.

[9] "Big Data in fashion: transforming the retail sector | Journal of Business Strategy | Emerald Publishing." Accessed: Mar. 04, 2026. [Online]. Available: https://www.emerald.com/jbs/article-abstract/41/4/21/196981/Big-Data-in-fashion-transforming-the-retail-sector?redirectedFrom=fulltext

[10] S. Wardani, S. S. Lubis, and R. W. Dewantoro, "Analisis Big Data untuk Prediksi Permintaan Produk dalam E-commerce," *Jurnal Penelitian Teknologi Informasi dan Sains*, vol. 3, no. 1, pp. 74–81, Apr. 2025, doi: 10.54066/jptis.v3i1.3066.

[11] A. Riansah, O. Nurdiawan, and R. Herdiana, "PENERAPAN ALGORITMA RANDOM FOREST DAN DECISION TREE UNTUK MENINGKATKAN AKURASI KLASIFIKASI PENJUALAN PADA TOKO BANGUNAN," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4242–4249, May 2025, doi: 10.36040/jati.v9i3.13622.

[12] S. Riptiono, "Pemasaran Digital dan Customer Relationship Management: Upaya Meningkatkan Kemampuan UMKM dalam Membangun Relasi Pelanggan," *JCSE: Journal of Community Service and Empowerment*, vol. 6, no. 1, pp. 65–73, Apr. 2025, doi: 10.32639/sgm19c75.

[13] J. Oh, "Classification and regression tree approach for the prediction of the seasonal apparel market: focused on weather factors," *Journal of Fashion Marketing and Management: An International Journal*, vol. 28, no. 5, pp. 893–910, Nov. 2023, doi: 10.1108/JFMM-12-2022-0266.

[14] S. M. S. Rana, S. M. F. Azim, A. R. K. Arif, M. S. I. Sohel, and F. N. Priya, "Investigating online shopping behavior of generation Z: an application of theory of consumption values," *Journal of Contemporary Marketing Science*, vol. 7, no. 1, pp. 17–37, Jan. 2024, doi: 10.1108/JCMARS-03-2023-0005.

[15] "Customer Shopping Trends Dataset." Accessed: Mar. 04, 2026. [Online]. Available: https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset

[16] A. Prakash, "Pre-processing techniques for preparing clean and high-quality data for diabetes prediction," *International Journal of Research Publication and Reviews*, vol. 5, pp. 458–465, Feb. 2024, doi: 10.55248/gengpi.5.0224.0412.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 141 - 150
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2776
https://journal.fkpt.org/index.php/MIS

[17] D. Mualfah, W. Fadila, and R. Firdaus, "Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 2, pp. 107–113, Aug. 2022, doi: 10.37859/coscitech.v3i2.3912.

[18] E. R. Susanto and E. Saputra, "Implementation of Deep Learning with Multilayer Perceptron (MLP) for Heart Disease Prediction Using the SMOTE-ENN Technique," *Journal of Applied Informatics and Computing*, vol. 9, no. 3, pp. 1034–1041, Jun. 2025, doi: 10.30871/jaic.v9i3.9337.

[19] B. N. Azmi, A. Hermawan, and D. Avianto, "Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver," *Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 281–290, Feb. 2023, doi: 10.35746/jtim.v4i4.298.

[20] E. Permana and J. Susilo, "Optimasi Prediksi Jumlah Kontainer Aktual di Kapal Menggunakan Random Forest dan XGBoost dengan Hyperparameter Tuning," *Jurnal Informatika dan Bisnis*, vol. 14, no. 2, pp. 114–132, 2025, doi: 10.46806/jib.v14i2.1921.

[21] H. Wijaya, D. P. Hostiadi, and E. Triandini, "Optimization of XGBoost Algorithm Using Parameter Tunning in Retail Sales Prediction," *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 13, no. 3, pp. 769–786, Dec. 2024, doi: 10.23887/janapati.v13i3.82214.