



PREDIKSI KETERLAMBATAN PENGIRIMAN PAKET

Kelompok 14 IF-ISSI-C

ANGGOTA TIM



220660121115

— **KEMAL HAPIDZ
PRASTIAWAN**



220660121179

— **DEDE YAYAN
SUCIYANA**



220660121086

— **AGUNG FEBRIAN**

1. Ringkasan Kasus

Dalam era digital, layanan ekspedisi memegang peran penting dalam mendukung e-commerce. Namun, keterlambatan pengiriman masih sering terjadi karena berbagai faktor seperti berat paket, diskon, dan prioritas produk

Penjelasan berikuit ini bertujuan untuk memprediksi risiko keterlambatan pengiriman dengan menggunakan data historis, agar perusahaan dapat meningkatkan efisiensi dan kepuasan pelanggan

2. METODE ANALISIS

a. Data Cleaning & Preprocessing

- Menghapus data duplikat
- Meng-encode fitur kategorial:
`Warehouse_block`, `Mode_of_Shipment`, `Product_importance`
- Melakukan standarisasi fitur numerik (menggunakan `StandardScaler`)
- Menetapkan target prediksi:
`Reached.on.Time_Y.N` → (0: Terlambat, 1: Tepat Waktu)

b. Exploratory Data Analysis (EDA)

- Visualisasi distribusi keterlambatan pengiriman
- Analisis korelasi antara fitur-fitur (diskon, berat, rating)
- Boxplot menunjukkan bahwa paket dengan berat tertentu lebih rentan terlambat

c. Segmentasi Pelanggan (Clustering)

- PCA digunakan untuk reduksi dimensi agar clustering dapat divisualisasikan
- Clustering dilakukan menggunakan K-Means dengan 3 klaster
- Visualisasi menunjukkan pola yang dapat dimanfaatkan untuk segmentasi operasional

d. Model Prediksi (Supervised Learning)

- Data dibagi menjadi 80% data latih dan 20% data uji
- Model Random Forest digunakan karena kemampuannya menangani fitur kategorikal dan numerik
- Model dilatih untuk memprediksi apakah paket akan terlambat atau tidak

e. Evaluasi Model

- Model menghasilkan metrik:
 1. Accuracy: (0.66)
 2. Precision, Recall, F1-Score ditampilkan menggunakan `classification_report`
- Confusion matrix divisualisasikan untuk memahami kesalahan prediksi

f. Visualisasi & Interpretasi

1. Diskon Besar & Risiko Keterlambatan

Insight: Semakin tinggi diskon yang ditawarkan, semakin besar kemungkinan pengiriman terlambat

Strategi: Batasi diskon besar pada produk berat atau untuk pengiriman ke area rawan keterlambatan

2. Berat Paket & Ketepatan Waktu

Insight: Paket yang lebih berat (`Weight_in_gms`) cenderung lebih sering terlambat

Strategi: Prioritaskan paket berat dengan moda pengiriman yang lebih andal dan siapkan lebih awal

3. Dominasi & Risiko Moda "Ship"

Insight: Moda "Ship" paling banyak digunakan, namun memiliki tingkat keterlambatan yang perlu diwaspadai

Strategi: Lakukan evaluasi performa dan renegotiasi SLA (Service Level Agreement) untuk moda "Ship"

4. Peluang Segmentasi Pelanggan

Insight: Data menunjukkan adanya beberapa kelompok (klaster) transaksi dengan pola risiko yang berbeda

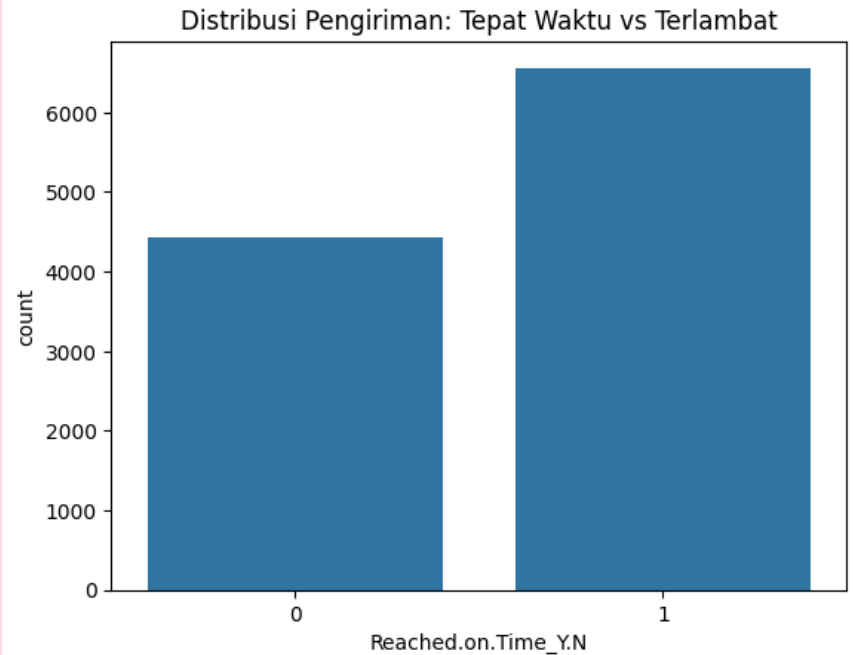
Strategi: Buat kebijakan logistik khusus untuk tiap segmen, terutama penanganan khusus untuk klaster berisiko tinggi

3. Hasil Analisis dan Model

a. Eksplorasi Data

1. Distribusi Keterlambatan

- Grafik countplot memperlihatkan distribusi antara pengiriman yang tepat waktu (1) dan terlambat (0)
- Terlihat bahwa sebagian besar pengiriman berada di kategori tepat waktu, namun keterlambatan tetap signifikan secara jumlah

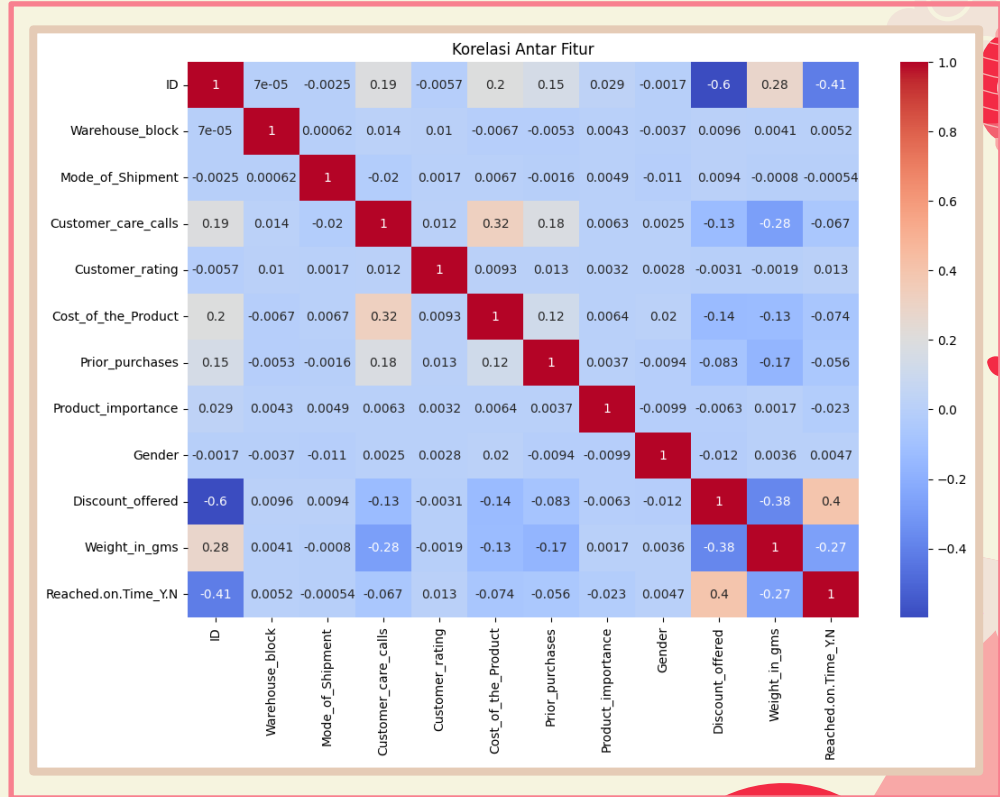


2. Korelasi Antar Fitur

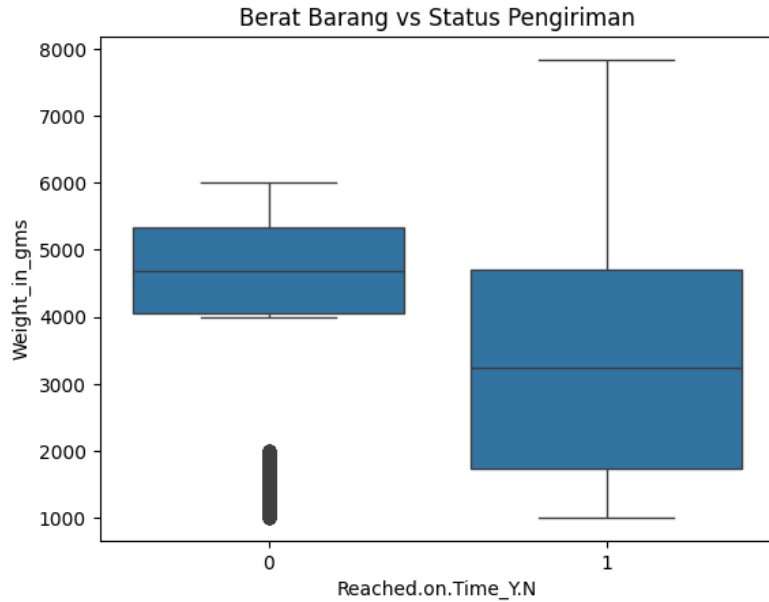
- Heatmap menunjukkan fitur yang paling berkorelasi dengan keterlambatan

(Reached.on.Time_Y.N) adalah:

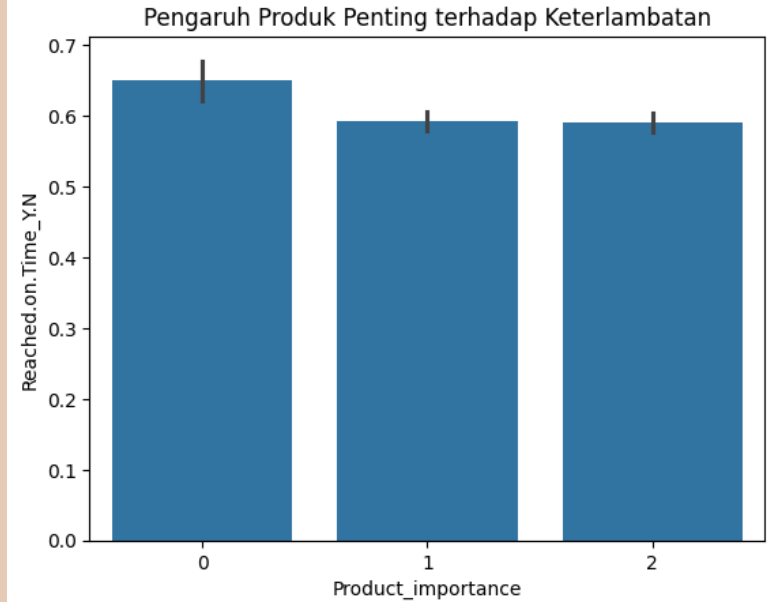
- Discount_offered** (positif)
- Weight_in_gms** (positif)
- Ini menunjukkan bahwa diskon besar dan berat paket tinggi meningkatkan risiko keterlambatan



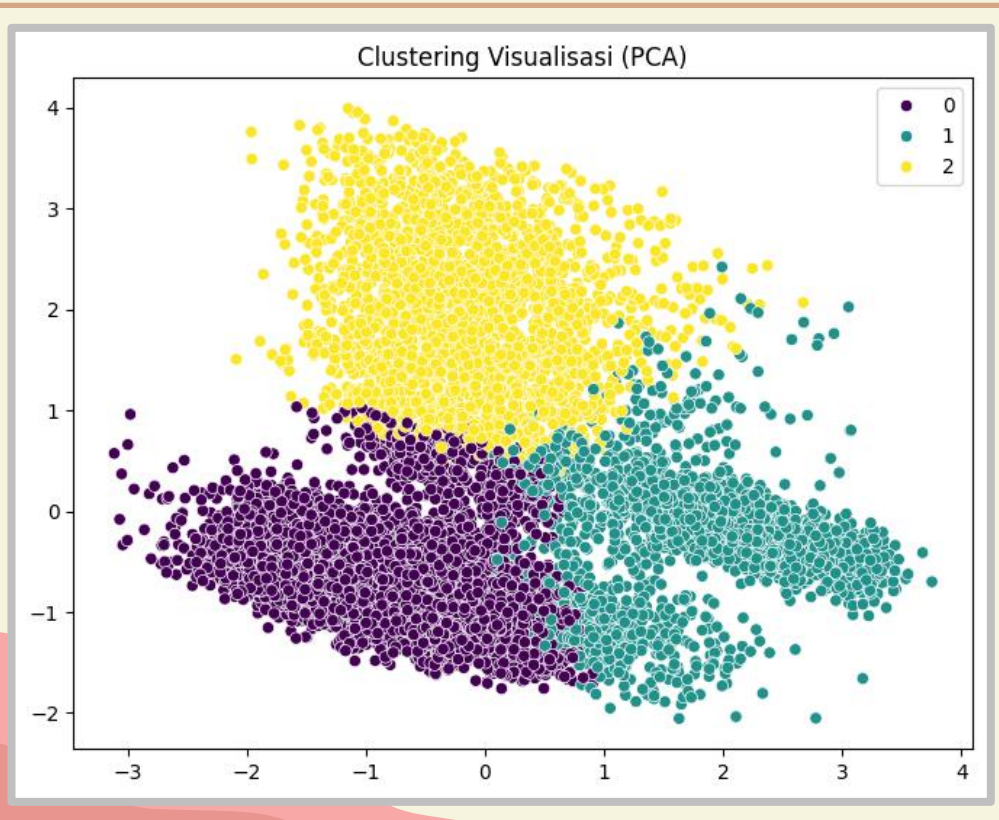
3. Berat Paket & Status Pengiriman



4. Importance Produk vs Keterlambatan



b. Clustering (Segmentasi Transaksi)



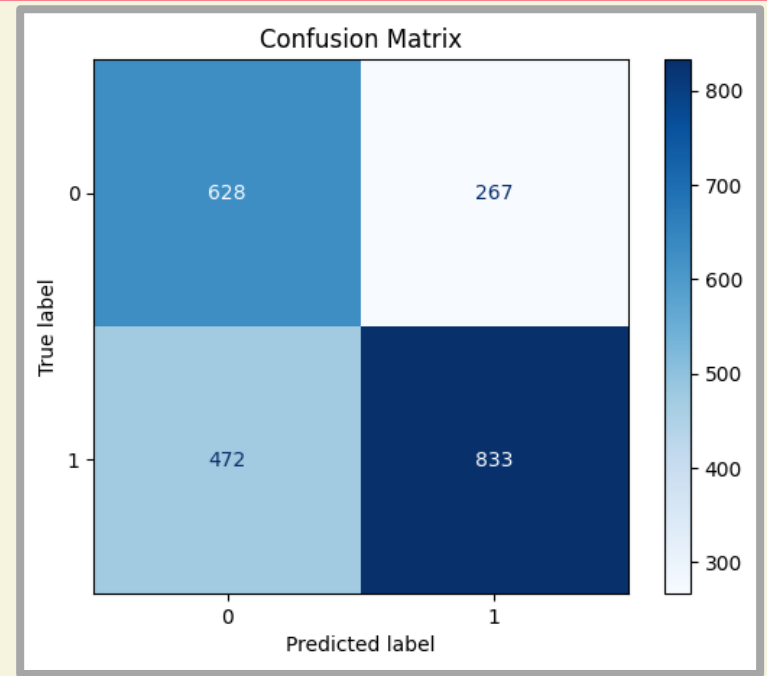
- Clustering dilakukan menggunakan K-Means ($k = 3$)
- Data direduksi menggunakan PCA agar dapat divisualisasikan dalam 2 dimensi
- Setiap kluster merepresentasikan pola pembelian/logistik yang serupa

c. Model Prediksi (Random Forest)

- Model: Random Forest Classifier
- Data dibagi menjadi 80% data latih dan 20% data uji
- Fitur penting: **Discount_offered**, **Weight_in_gms**, **Product_importance**, dll
- Hasil prediksi dievaluasi dengan Classification Report dan Confusion Matrix

	precision	recall	f1-score	support
0	0.57	0.70	0.63	895
1	0.76	0.64	0.69	1305
accuracy			0.66	2200
macro avg	0.66	0.67	0.66	2200
weighted avg	0.68	0.66	0.67	2200

Akurasi Model: 0.66



4. Kesimpulan dan Saran

a. Kesimpulan

1. Model Random Forest menghasilkan akurasi sebesar 66% atau 0.66, dengan F1-score yang cukup baik untuk kedua kelas keterlambatan dan ketepatan waktu
2. Model lebih mampu mengenali pengiriman yang terlambat daripada yang tepat waktu, ditunjukkan oleh precision dan recall yang lebih tinggi pada kelas 1
3. Fitur Discount_offered dan Weight_in_gms menunjukkan korelasi yang signifikan terhadap kemungkinan keterlambatan
4. Hasil clustering memetakan transaksi ke dalam beberapa kelompok yang menunjukkan pola risiko logistik berbeda, yang dapat dijadikan acuan segmentasi

b. Saran

1. Perlu peningkatan kualitas model, misalnya:
 - Menggunakan algoritma lain seperti XGBoost
 - Melakukan hyperparameter tuning
 - Mengatasi ketidakseimbangan data dengan teknik seperti SMOTE
2. Model ini sebaiknya digunakan sebagai alat bantu awal, bukan keputusan mutlak, karena masih terdapat cukup banyak false negative
3. Disarankan melakukan analisis lanjutan terhadap fitur waktu jika tersedia (misalnya waktu pengiriman, tanggal order) untuk meningkatkan akurasi
4. Perusahaan dapat menggunakan hasil prediksi ini untuk memfilter risiko keterlambatan lebih awal dan menyesuaikan strategi pengiriman pada produk dengan risiko tinggi

TERIMA KASIH



Kelompok 14 IF-ISSI-C