**Prediksi Keterlambatan Pengiriman Paket**

Disusun untuk memenuhi tugas besar pada mata kuliah: Data Science

Dosen Pengampu: Yopi Hidayatul Akbar, S.Kom., M.T



Disusun oleh:

Agung Febrian (220660121086)

Kemal Hapidz Prastiawan (220660121115)

Dede Yayan Suciyana (220660121179)

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNOLOGI INFORMASI**

**UNIVERSITAS SEBELAS APRIL**

**SUMEDANG**

**2025**

# **Ringkasan Kasus**

Dalam era digital yang serba cepat, layanan ekspedisi memiliki peran vital dalam mendukung aktivitas e-commerce dan distribusi logistik. Ketepatan waktu pengiriman menjadi indikator utama dalam menilai kepuasan pelanggan terhadap layanan yang diberikan. Namun, dalam praktiknya, keterlambatan pengiriman masih kerap terjadi karena berbagai faktor, seperti kondisi operasional, karakteristik barang, hingga strategi diskon yang diterapkan.

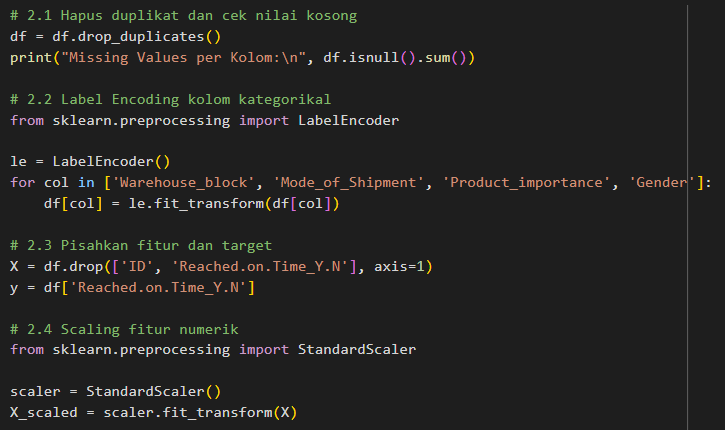
Untuk menjawab tantangan tersebut, analisis ini diarahkan untuk mengidentifikasi faktor-faktor signifikan yang memengaruhi keterlambatan pengiriman serta membangun model prediktif yang mampu memperkirakan apakah suatu pengiriman akan terlambat atau tidak. Dengan memanfaatkan data historis pengiriman, diharapkan solusi ini dapat meningkatkan efisiensi operasional serta membantu pengambilan keputusan yang lebih tepat dalam manajemen logistik.

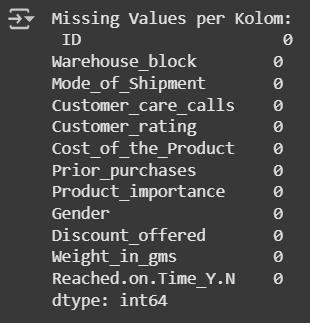
# **Metode Analisis**

Langkah-langkah analisis yang dilakukan:

## ***Data Cleaning* & *Preprocessing***

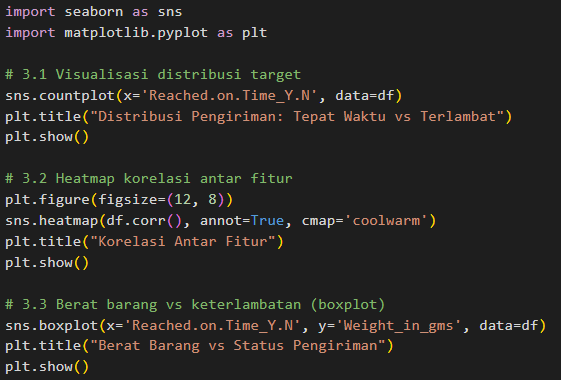
* Duplikat data dihapus
* Kolom kategorial seperti: Warehouse\_block, Mode\_of\_Shipment, dan Product\_importance diencoding menggunakan LabelEncoder
* Standarisasi dilakukan menggunakan StandardScaler
* Target prediksi adalah Reached.on.Time\_Y.N (0: terlambat, 1: tepat waktu)

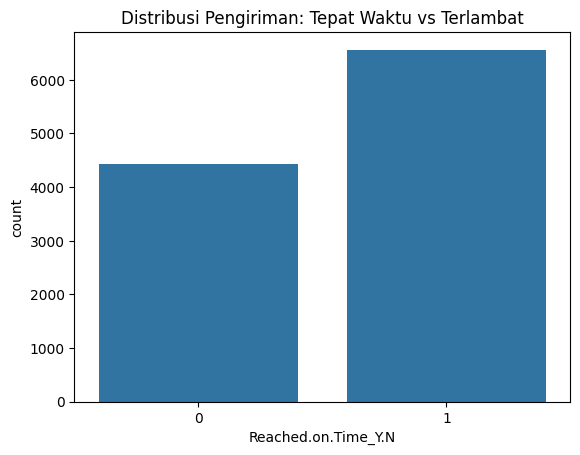


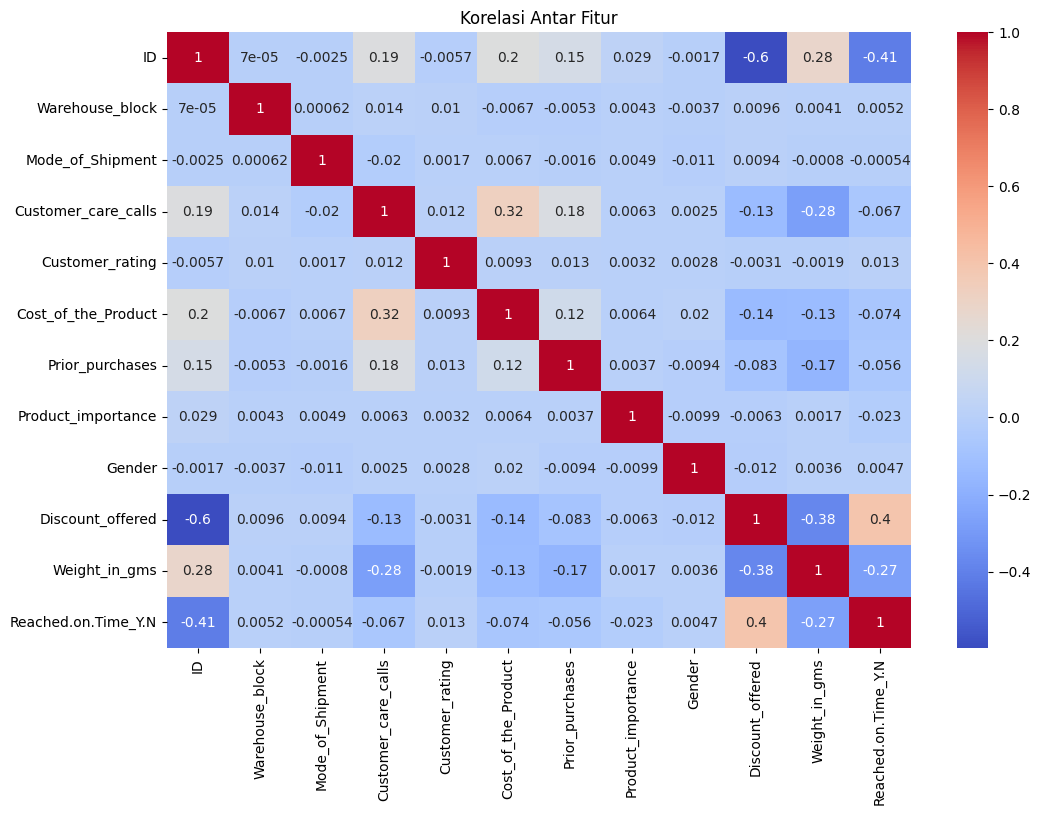


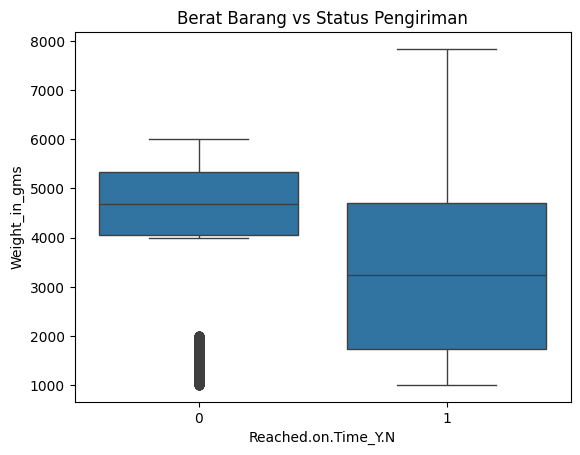
## ***Exploratory* *Data* *Analysis* (EDA)**

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



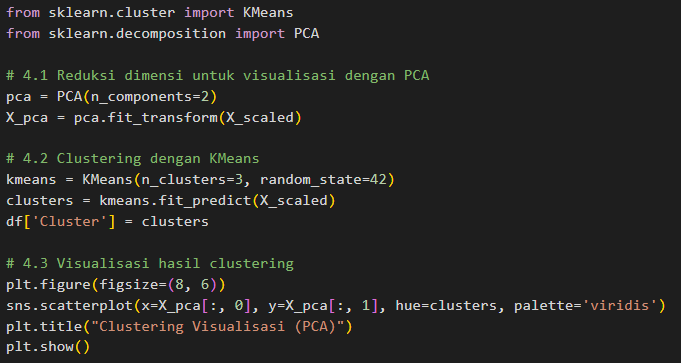


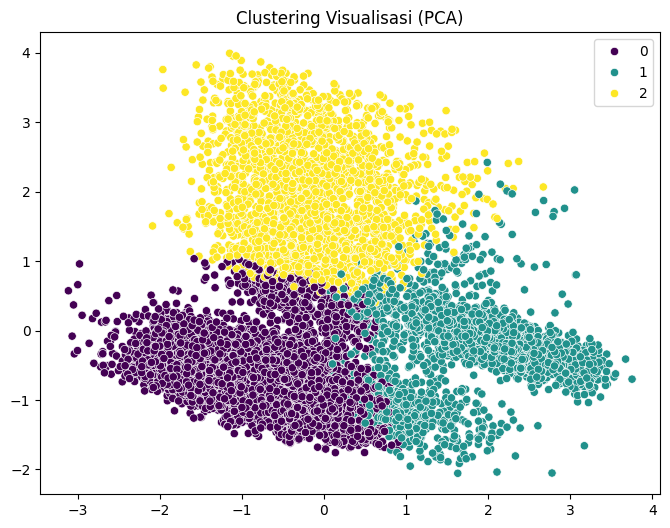




## **Segmentasi Pelanggan (*Clustering*)**

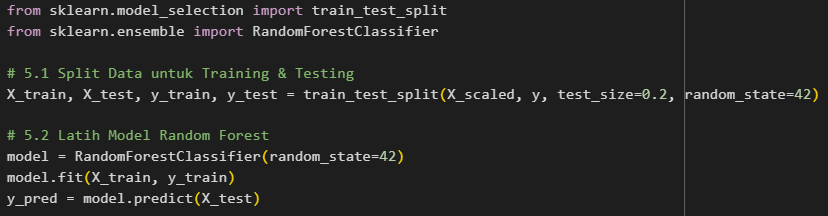
* PCA digunakan untuk reduksi dimensi agar *clustering* dapat divisualisasikan
* *Clustering* dilakukan menggunakan K-Means dengan 3 klaster
* Visualisasi menunjukkan pola yang dapat dimanfaatkan untuk segmentasi operasional





## **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

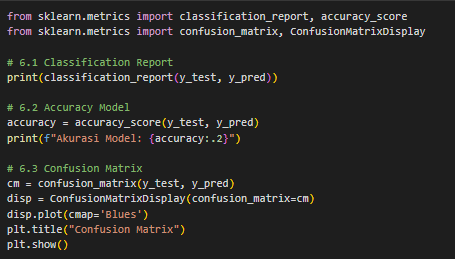


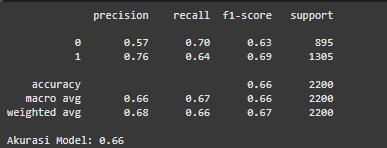
## **Evaluasi Model**

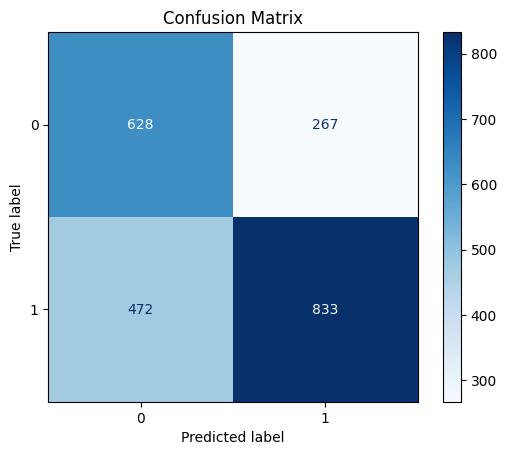
* Model menghasilkan metrik:

1. *Accuracy*: (0.66)
2. *Precision*, *Recall*, F1-Score ditampilkan menggunakan classification\_report

* Confusion matrix divisualisasikan untuk memahami kesalahan prediksi







## **Visualisasi & Interpretasi**

Berdasarkan hasil eksplorasi dan visualisasi data (heatmap korelasi, distribusi keterlambatan, dan proporsi moda pengiriman), diperoleh beberapa insight penting untuk mendukung strategi bisnis logistik, antara lain:

1. **Diskon Besar Meningkatkan Risiko Keterlambatan**

Ditemukan korelasi positif antara `Discount\_offered` dengan keterlambatan. Artinya, semakin besar diskon yang diberikan, semakin tinggi kemungkinan pengiriman mengalami keterlambatan.

Strategi: Diskon besar sebaiknya tidak diberikan pada produk berat atau daerah rawan keterlambatan. Perlu kebijakan promosi berbasis kapasitas logistik.

1. **Berat Paket Mempengaruhi Ketepatan Waktu**

Paket dengan `Weight\_in\_gms` yang tinggi lebih sering terlambat dibandingkan paket ringan.

Strategi: Segmentasikan pengiriman berdasarkan berat. Produk berat dapat diprioritaskan menggunakan moda pengiriman yang lebih andal atau disiapkan lebih awal.

1. **Moda Pengiriman Dominan dan Risiko**

Moda `Ship` merupakan yang paling banyak digunakan, namun jika dikaitkan dengan tingkat keterlambatan, perlu perhatian lebih.

Strategi: Evaluasi performa mitra logistik untuk moda pengiriman dominan dan lakukan renegosiasi SLA (*Service Level Agreement*) jika diperlukan.

1. **Peluang Segmentasi Pelanggan dan Produk**

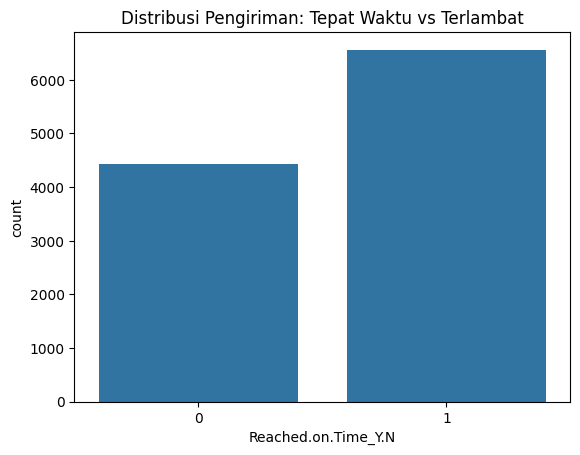
Hasil clustering menunjukkan pola berbeda antar kelompok transaksi.

Strategi: Terapkan kebijakan logistik atau layanan yang berbeda untuk tiap segmen, misalnya penanganan khusus untuk klaster berisiko tinggi keterlambatan.

# **Hasil Analisis dan Model**

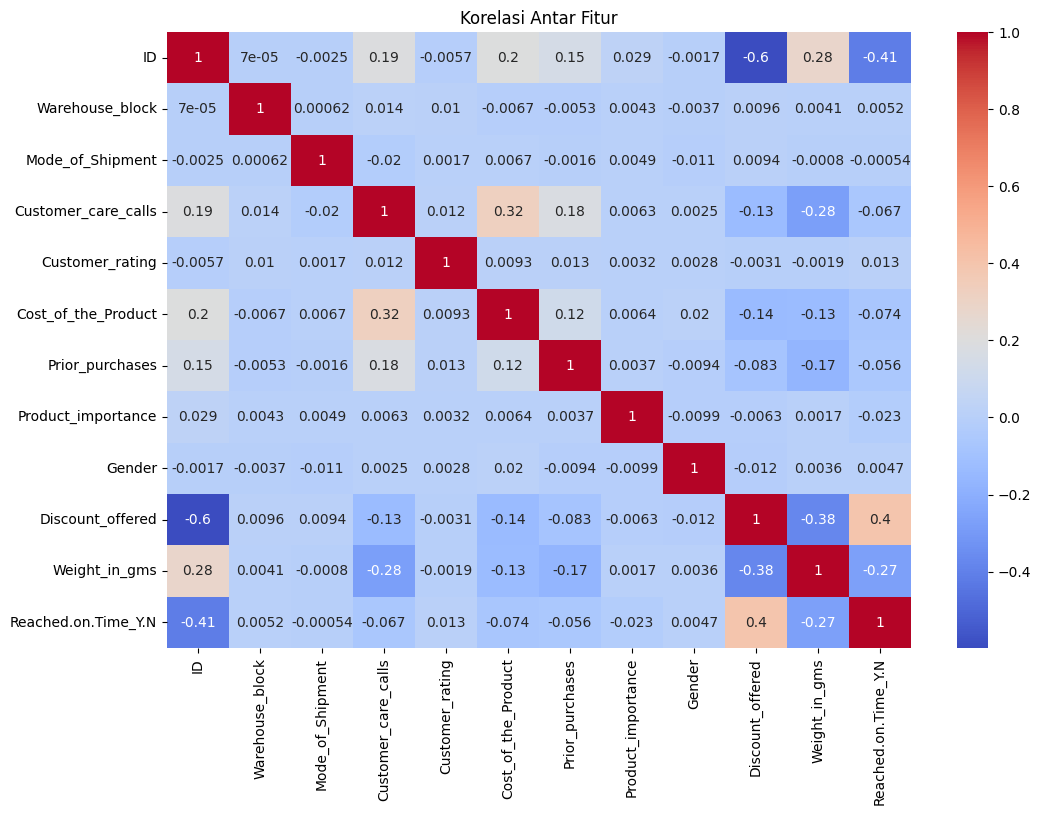
1. **Eksplorasi Data**
2. **Distribusi Keterlambatan**

* 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.

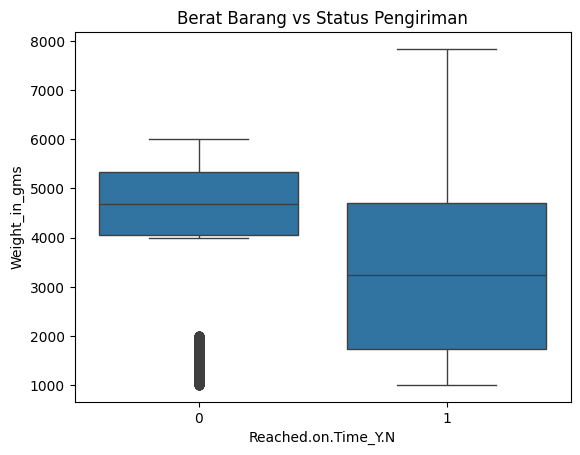


1. **Korelasi Antar Fitur**

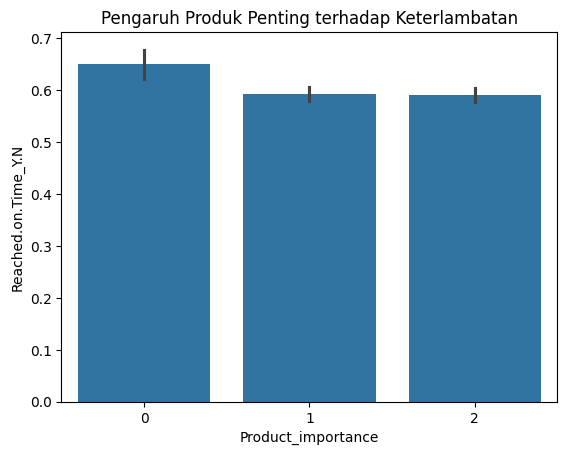
* *Heatmap* menunjukkan fitur yang paling berkorelasi dengan keterlambatan (Reached.on.Time\_Y.N) adalah:
* Discount\_offered (positif)
* Weight\_in\_gms (positif)
* Ini menunjukkan bahwa diskon besar dan berat paket tinggi meningkatkan risiko keterlambatan.



1. **Berat Paket & Status Pengiriman**

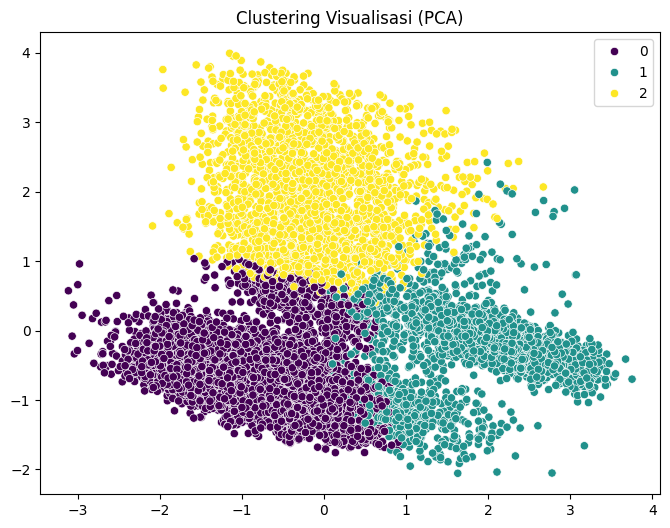


1. ***Importance* Produk vs Keterlambatan**



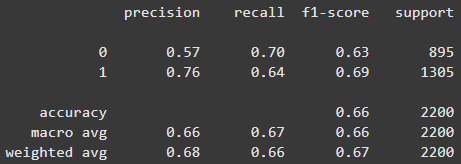
1. ***Clustering* (Segmentasi Transaksi)**

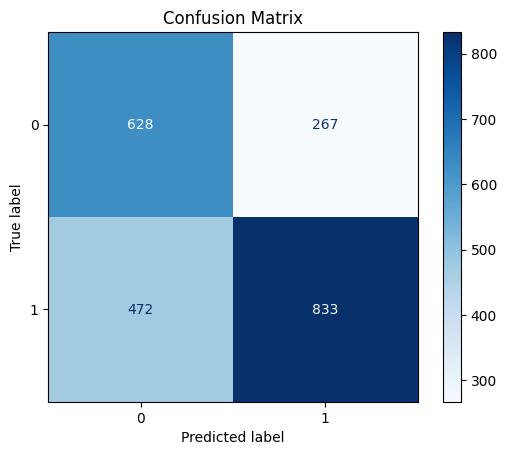
* *Clustering* dilakukan menggunakan K-Means (k = 3).
* Data direduksi menggunakan PCA agar dapat divisualisasikan dalam 2 dimensi.
* Setiap klaster merepresentasikan pola pembelian/logistik yang serupa.



1. **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, dsb.
* Hasil prediksi dievaluasi dengan Classification Report dan Confusion Matrix.





# **Kesimpulan dan Saran**

## **Kesimpulan**

* Model Random Forest menghasilkan akurasi sebesar 66% atau 0,66 dengan F1-score yang cukup baik untuk kedua kelas keterlambatan dan ketepatan waktu.
* Model lebih mampu mengenali pengiriman yang terlambat daripada yang tepat waktu, ditunjukkan oleh precision dan recall yang lebih tinggi pada kelas 1.
* Fitur Discount\_offered dan Weight\_in\_gms menunjukkan korelasi yang signifikan terhadap kemungkinan keterlambatan.
* Hasil *clustering* memetakan transaksi ke dalam beberapa kelompok yang menunjukkan pola risiko logistik berbeda, yang dapat dijadikan acuan segmentasi.

## **Saran**

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

# ***Tools* dan *Library***

1. Python (pandas, seaborn, matplotlib, scikit-learn)

2. Google Colab: Pada link berikut: [colab/kelompok-14](https://colab.research.google.com/drive/1tjsrxE7N0xd8b-327G529eFQPBc3YVZx?usp=sharing)

3. Dataset: *E*-*Commerce* *Shipping* Data ([Kaggle/@prachi13](https://www.kaggle.com/datasets/prachi13/customer-analytics))

# ***Repository***

[github.com/kelomppok-14](https://github.com/agungfbrrn/Prediksi-Keterlambatan-Pengiriman-Paket-pada-Layanan-Ekspedisi)