

# DARI DATA KE KEPUASAN:

MEMPERKIRAKAN LAMA WAKTU  
DELIVERY DENGAN CERDAS

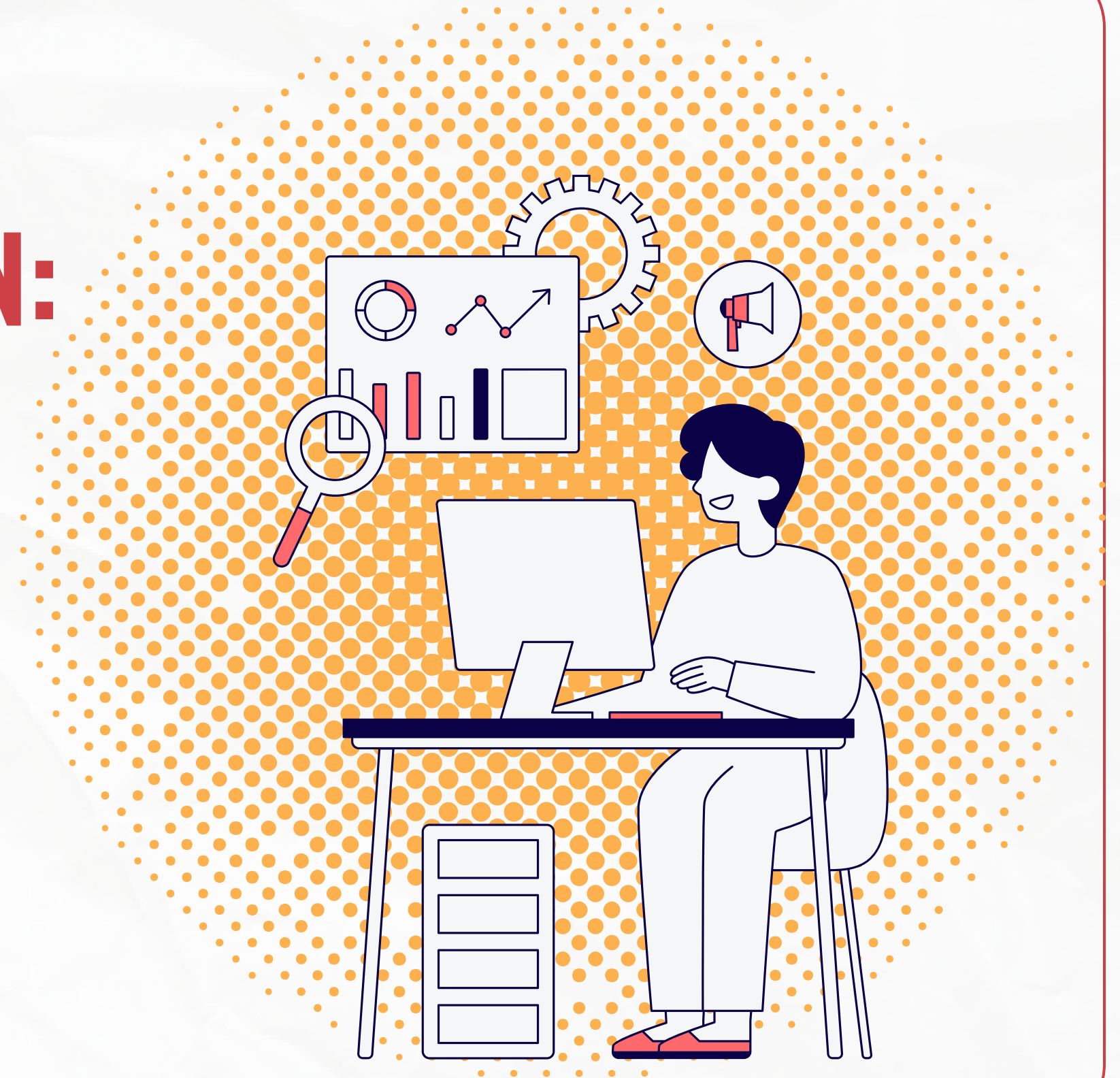
Luthfi Ridhoul Aziz



: <https://www.linkedin.com/in/luthfiridhoul/>



: <https://github.com/luthfiridhoul>



# Profil

Saya **Luthfi Ridhoul Aziz**, lulusan **S1 Akuntansi** dari **Universitas Brawijaya** dengan IPK 3,58. Selama masa kuliah, saya aktif mengikuti berbagai kegiatan akademik maupun non-akademik yang mengasah kemampuan analisis, komunikasi, dan kerja sama tim.

Selain memiliki minat yang besar di bidang **finance**, saat ini saya juga mendalami dunia data, khususnya **Data Science** dan **Data Analysis**, untuk mengembangkan potensi saya di dunia kerja.

Untuk mendukung hal tersebut, saya mengikuti Bootcamp Data Science di Dibimbing.id selama kurang lebih 6 bulan, mempelajari berbagai keterampilan seperti pengolahan data, visualisasi, serta pengembangan model Machine Learning untuk pemecahan masalah bisnis.



PHONE NUMBER

**+62-811-3776-663**

EMAIL ADDRESS

**luthfiridhoulz@gmail.com**



# Goals?

Prediksi tepat 🕒 → pelanggan puas → operasional lebih lancar 🚴♂️



## Pertanyaan Utama

- Bagaimana memprediksi waktu pengantaran makanan dengan akurasi tinggi?
- Faktor apa saja yang paling berpengaruh terhadap lama pengantaran?

## Tujuan Proyek

- Membangun model prediksi waktu pengantaran yang akurat & andal.
- Memberikan estimasi waktu yang lebih realistis kepada pelanggan.
- Menyediakan insight untuk meningkatkan efisiensi operasional kurir & restoran.

# Dataset

Dataset yang digunakan berisi informasi pesanan makanan beserta faktor-faktor yang memengaruhi lama waktu pengantaran. Data ini menjadi landasan dalam pembangunan model prediksi.

## Gambaran Umum

- Jumlah data: 455 observasi
- Jumlah variabel: 9 variabel
- Variabel target: `Delivery_Time_min` (lama pengantaran dalam menit)
- Sumber dataset: **Kaggle – Food Delivery Time Prediction**

## Variabel yang Digunakan

- `Distance_km` → jarak tempuh kurir
- `Weather` → kondisi cuaca saat pengantaran
- `Traffic_Level` → tingkat kepadatan lalu lintas
- `Time_of_Day` → waktu pengantaran (pagi, siang, sore, malam)
- `Vehicle_Type` → jenis kendaraan yang digunakan kurir
- `Preparation_Time_min` → lama waktu restoran menyiapkan pesanan
- `Courier_Experience_yrs` → pengalaman kurir dalam tahun
- `Delivery_Time_min` → variabel target yang diprediksi



# Data Cleaning & Manipulation



Data perlu diproses terlebih dahulu agar bersih dan siap digunakan dalam pembangunan model prediksi. Proses preprocessing dibagi menjadi dua tahap utama: **Data Cleaning dan Data Manipulation**.

## Data Cleaning

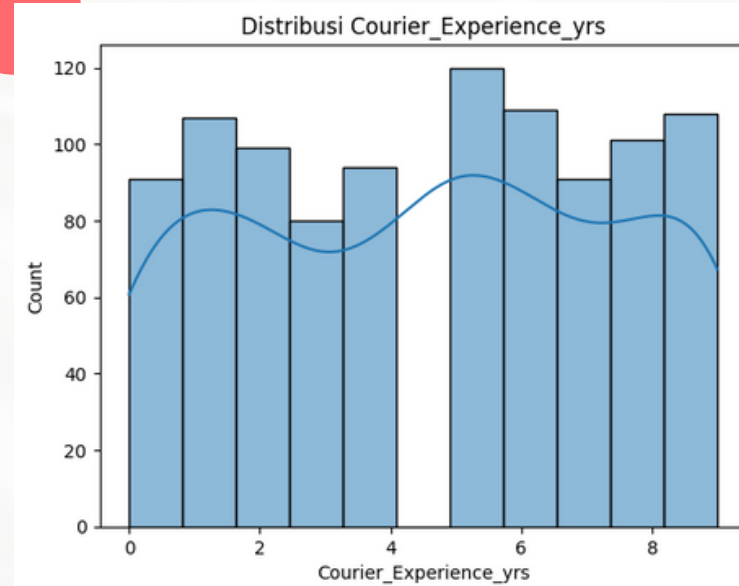
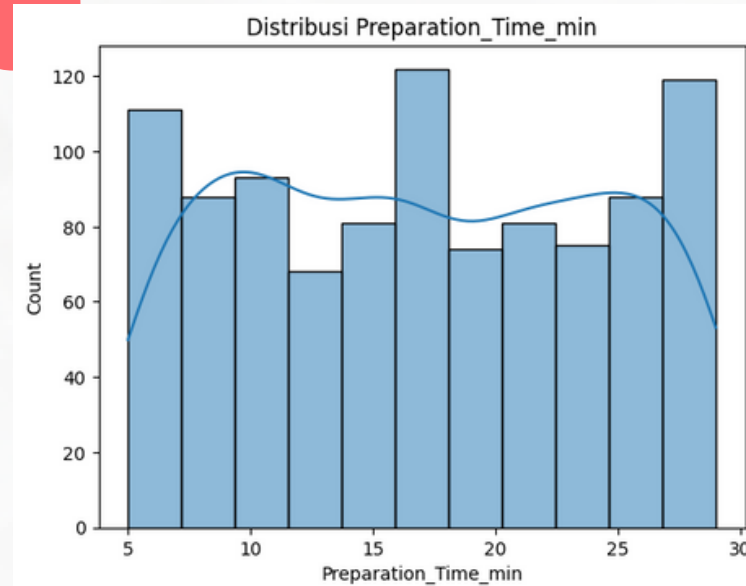
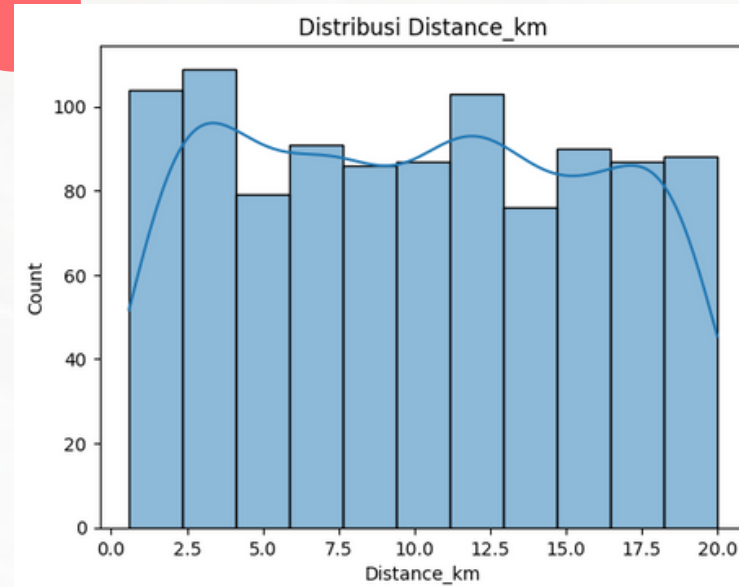
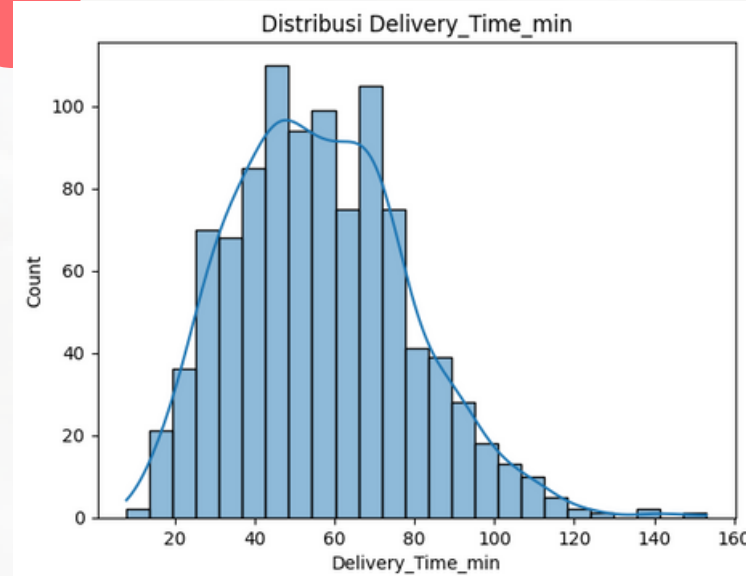
- Penanganan Missing Values
  - Variabel kategorikal (Weather, Traffic\_Level, Time\_of\_Day, Vehicle\_Type) diisi dengan modus.
  - Variabel numerik (Courier\_Experience\_yrs) diisi dengan median.
- Pemeriksaan Duplikasi
  - Baris duplikat dihapus agar data tidak bias.

## Data Manipulation

- Identifikasi Variabel
  - Kategorikal: Weather, Traffic\_Level, Time\_of\_Day, Vehicle\_Type.
  - Numerik: Distance\_km, Preparation\_Time\_min, Courier\_Experience\_yrs, Delivery\_Time\_min (target).
- Transformasi Data
  - Variabel kategorikal dengan One-Hot Encoding. Variabel numerik dengan StandardScaler (standarisasi skala data).
- Pembagian Dataset
  - Training set: 80% & Testing set: 20%



# Exploratory Data Analysis (Univariate)



Analisis univariat dilakukan untuk melihat **distribusi** masing-masing **variabel numerik**, termasuk variabel target.

## Hasil Analisis

- Delivery\_Time\_min: distribusi agak condong ke kanan (right-skewed), sebagian besar pengantaran memakan waktu 40–70 menit.
- Distance\_km: distribusi merata, jarak pesanan bervariasi hingga  $\pm 20$  km.
- Courier\_Experience\_yrs: pengalaman kurir relatif seimbang, mayoritas 0–9 tahun.
- Preparation\_Time\_min: umumnya 5–30 menit, dengan penyebaran cukup merata.

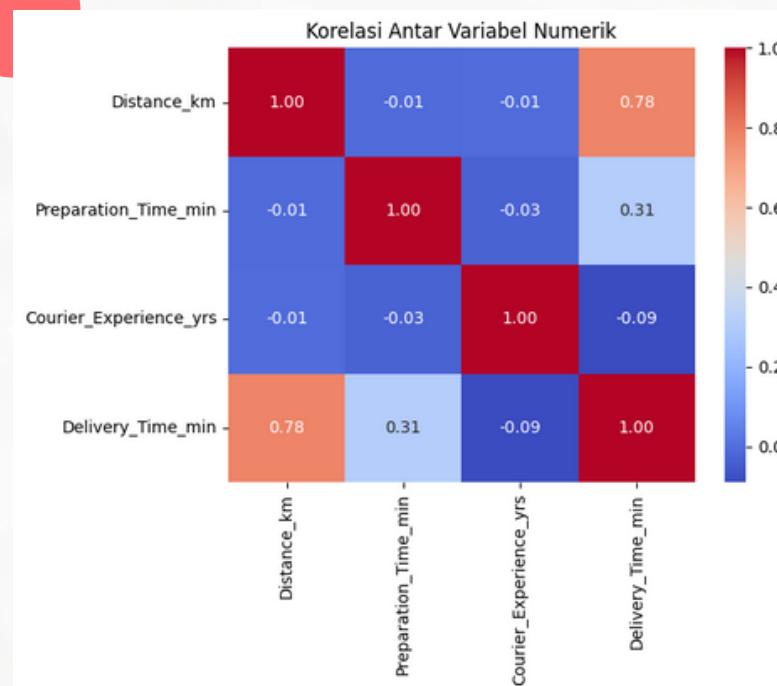
# Exploratory Data Analysis (Bivariate)



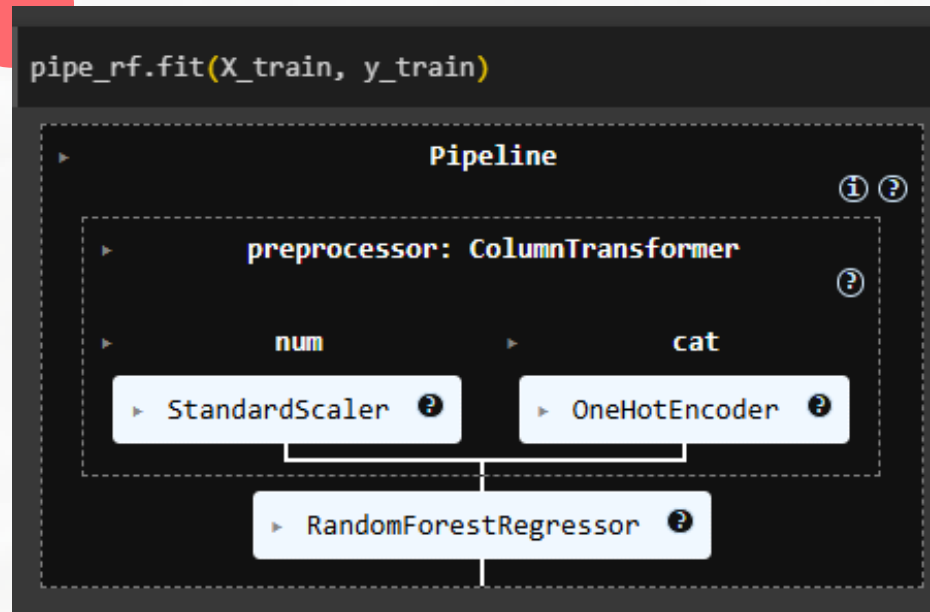
Analisis bivariat dilakukan untuk melihat **hubungan** antara variabel-variabel **independen** dengan variabel **target** (Delivery\_Time\_min).

## Hasil Analisis

- Jarak (Distance\_km) berhubungan paling kuat dengan waktu pengantaran. Scatterplot memperlihatkan pola linear: semakin jauh, semakin lama waktu antar.
- Preparation\_Time\_min juga berpengaruh, meski tidak sekuat jarak.
- Courier\_Experience\_yrs hampir tidak berkorelasi dengan waktu antar.
- Heatmap menunjukkan korelasi paling tinggi adalah antara Distance dan Delivery Time ( $r = 0.78$ ).



# Data Modeling and Machine Learning Tools (Random Forest)



```
best_rf = rf_randcv.best_estimator_
y_pred_best = best_rf.predict(X_test)
mae_b, rmse_b, r2_b = mean_absolute_error(
    y_test, y_pred_best), rmse_func(y_test, y_pred_best),
    r2_score(y_test, y_pred_best)
print("\n=== Random Forest (Tuned) - Test ===")
print(f"MAE : {mae_b:,.3f} | RMSE: {rmse_b:,.3f} | R²: {r2_b:,.3f}")

=== Random Forest (Tuned) - Test ===
MAE : 6.608 | RMSE: 9.519 | R²: 0.798
```

Model pertama yang digunakan adalah Random Forest, yaitu algoritma ensemble berbasis decision tree yang mampu menangani data kompleks dan variabel campuran.

## Hasil Evaluasi (Tuned Model)

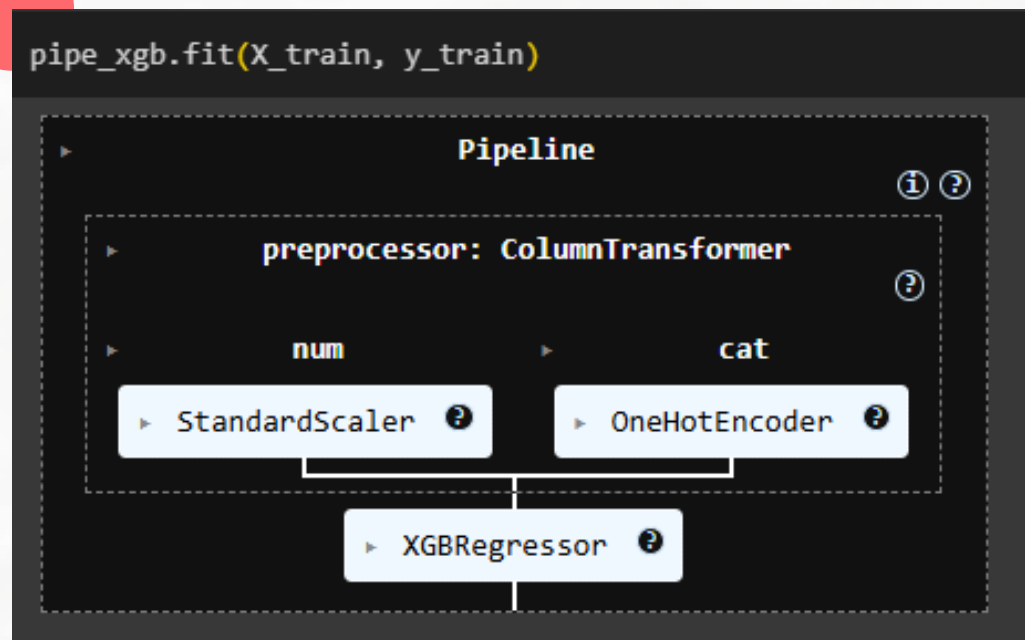
- MAE : 6.608 menit
- RMSE : 9.519 menit
- $R^2$  : 0.798
- Cross Validation (5-Fold)
  - Mean RMSE : 11.802
  - Std Dev : 0.895

## Interpretasi

- Model cukup akurat, dengan rata-rata kesalahan prediksi sekitar 6–7 menit.
- Menjelaskan sekitar 80% variasi waktu pengantaran.
- Performa stabil berdasarkan hasil cross-validation.



# Data Modeling and Machine Learning Tools (XGBoost)



```
best_xgb = xgb_randcv.best_estimator_
y_pred_best = best_xgb.predict(X_test)
mae_b, rmse_b, r2_b = mean_absolute_error(
    y_test, y_pred_best), rmse_func(y_test, y_pred_best),
    r2_score(y_test, y_pred_best)
print("\n=== XGBoost (Tuned) - Test ===")
print(f"MAE : {mae_b:,.3f} | RMSE: {rmse_b:,.3f} | R²: {r2_b:,.3f}")

=== XGBoost (Tuned) - Test ===
MAE : 6.596 | RMSE: 9.207 | R²: 0.811
```

Model kedua adalah XGBoost, algoritma boosting yang lebih kompleks dan sering unggul dalam prediksi dengan akurasi tinggi.

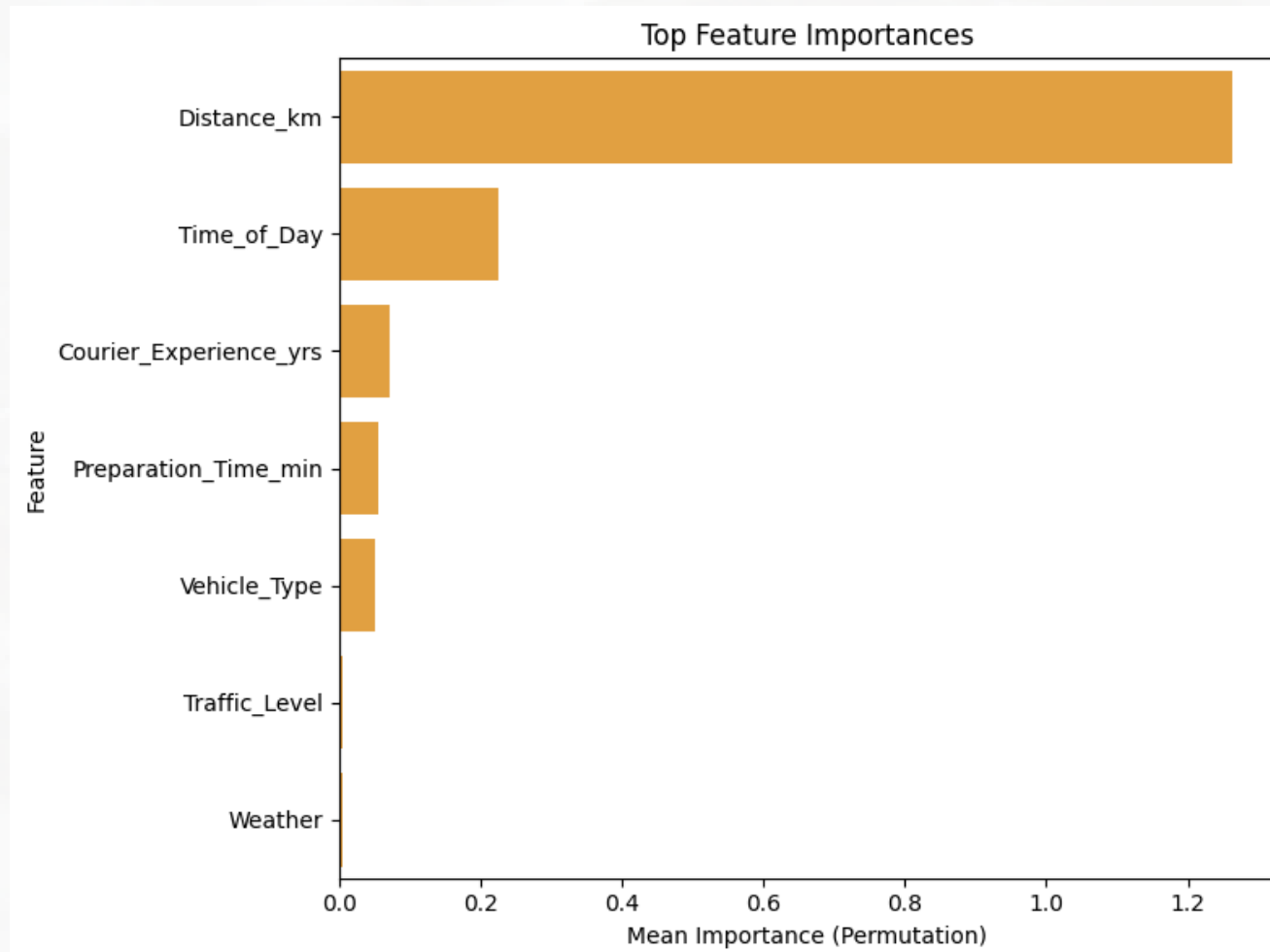
## Hasil Evaluasi (Tuned Model)

- MAE : 6.596 menit
- RMSE : 9.207 menit
- $R^2$  : 0.811
- Cross Validation (5-Fold)
  - Mean RMSE : 11.949
  - Std Dev : 1.055

## Interpretasi

- Hasil lebih baik dibanding Random Forest, dengan error yang lebih rendah.
- Menjelaskan sekitar 81% variasi waktu pengantaran.
- Cocok dipilih sebagai model utama karena memberikan prediksi yang lebih konsisten dan akurat.

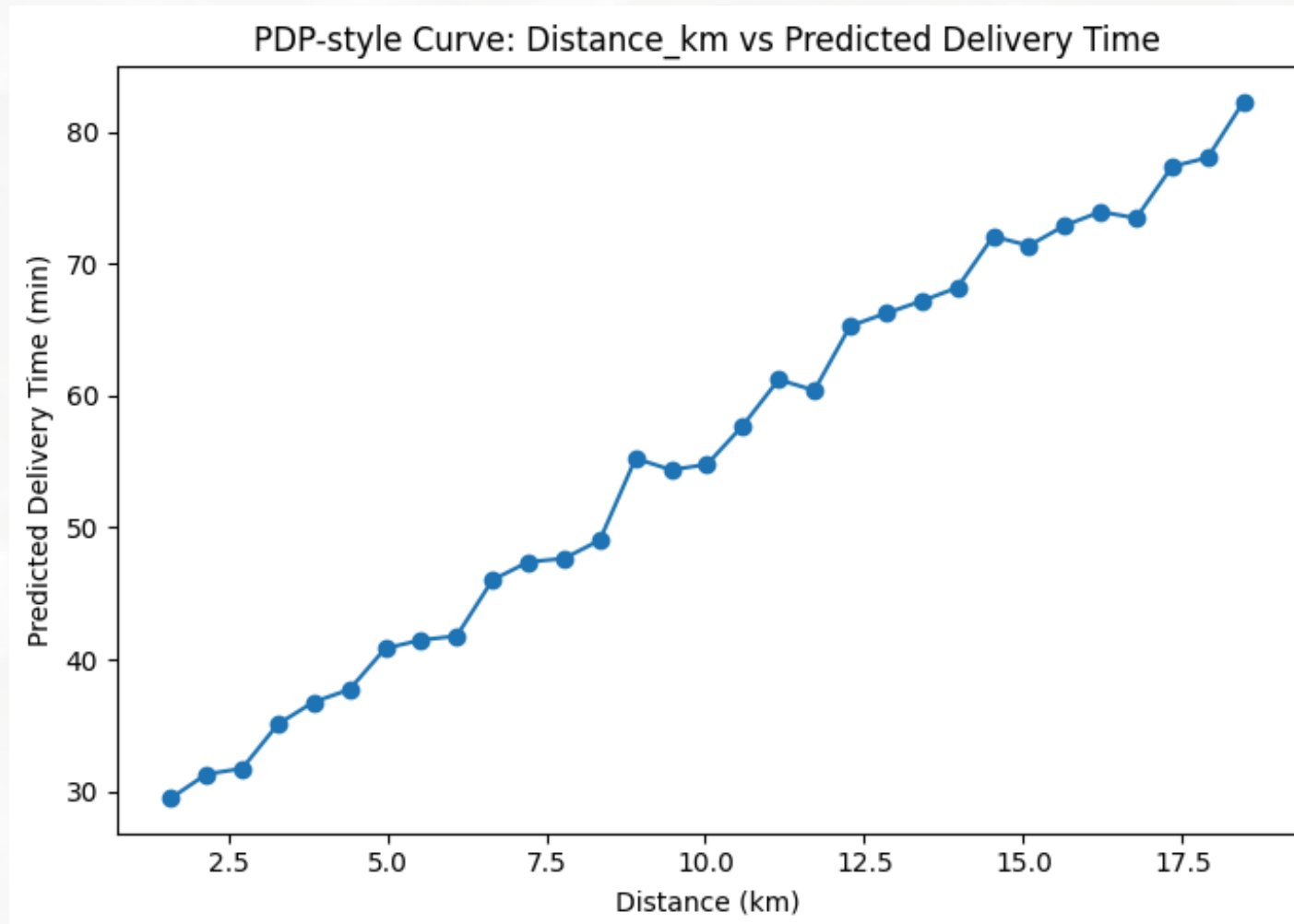
# Interpretasi Umum & Feature Importance



## Akurasi Prediksi

- Model terbaik (XGBoost Tuned) menghasilkan:
  - MAE : 6.596 menit
  - RMSE : 9.207 menit
  - $R^2$  : 0.811 → berarti 81% variasi waktu pengantaran dapat dijelaskan oleh model.
- Distance\_km → faktor dominan, semakin jauh semakin lama.
- Time\_of\_Day → jam sibuk menambah waktu antar.
- Preparation\_Time\_min → semakin lama restoran menyiapkan, semakin lama total waktu.
- Traffic\_Level → macet memperlambat waktu antar.
- Weather → hujan/cuaca buruk menambah keterlambatan.

# Feature Importance ( Distance )



**MAE: 6.425 | RMSE: 9.336 | R<sup>2</sup>: 0.806 AME**  
**(Distance\_km): ~3.34 menit per km (sekitar IQR)**

- **+1 km  $\approx$  5.65 menit**
- **+2 km  $\approx$  8.32 menit**

- **Distance\_km** adalah variabel terkuat di model ini. Artinya, variasi waktu antar paling banyak dijelaskan oleh jarak.
- Secara intuisi operasional, efek jarak terhadap waktu cenderung monoton naik: **makin jauh**  $\rightarrow$  **makin lama**. Polanya bisa mendekati linear pada jarak pendek–menengah, lalu kadang melengkung (non-linear) di jarak panjang
- Terhitung juga setiap penambahan **1 km** jarak, maka waktu tempuh pengantaran akan sebanyak **5,56 menit**. Sehingga semakin jauh maka waktu juga akan semakin bertambah.



# Perbandingan & Kesimpulan

Model	MAE ↓	RMSE ↓	R <sup>2</sup> ↑
Random Forest (Tuned)	6.608	9.519	0.798
XGBoost (Tuned)	6.596	9.207	0.811

## Kesimpulan

- Kedua model sudah cukup baik.
- XGBoost lebih unggul, karena prediksinya lebih dekat dengan kenyataan (MAE & RMSE lebih rendah, R<sup>2</sup> lebih tinggi).
- Dengan model ini, estimasi waktu antar bisa dibuat lebih akurat dan realistis.

## Perbandingan Model (Setelah Tuning)

### • Random Forest

- MAE : 6.608 menit → rata-rata meleset sekitar 6–7 menit.
- RMSE : 9.519 menit → sebagian prediksi bisa meleset hingga  $\pm 9$  menit.
- R<sup>2</sup> : 0.798 → mampu menjelaskan hampir 80% variasi waktu antar pesanan.

### • XGBoost

- MAE : 6.596 menit → rata-rata meleset sekitar 6 menit.
- RMSE : 9.207 menit → kesalahan lebih kecil dibanding Random Forest.
- R<sup>2</sup> : 0.811 → mampu menjelaskan lebih dari 80% variasi waktu antar pesanan.

# ● Rekomendasi Bisnis

## 1. Tampilkan estimasi yang lebih realistis di aplikasi.

- Berdasarkan prediksi model, rata-rata error hanya  $\pm 6$  menit.

## 2. Atur jumlah kurir sesuai kondisi.

- Hasil model menunjukkan jarak, lalu lintas, dan waktu sibuk sangat berpengaruh. Penambahan kurir di jam sibuk atau area macet akan mengurangi keterlambatan.

## 3. Bekerja sama dengan restoran.

- Karena preparation time berpengaruh, restoran perlu mempercepat proses masak agar pesanan tidak menumpuk.

## 4. Mitigasi saat cuaca buruk.

- Hujan menambah waktu antar, maka perlu strategi seperti armada khusus atau rute alternatif.



# Thank You

🌐 [Link Github](#)

