



THT DS PROJECT

FOOD DELIVERY TIME

PREDICTION

I MADE BAYU SATRIA WARDHANA



CONTENT

01

About Me

02

Background

03

Problem

04

Methodology

05

Data Cleaning

06

Exploratory
Data Analystist

07

Feature
Engineering

08

Modelling

09

Conclusion

10

Business Rec.

ABOUT ME

I Made Bayu Satrai Wadhana

Data Science Enthusiast

Computer Engineering graduate with hands-on experience in data analysis, cloud computing, and web development. Passionate about uncovering insights through data and transforming information into impactful decisions.

Education Background :

Telkom University (2020-2024)

Computer engineering graduate (GPA : 3.39/4.00)

Dbimbing Data Scientist Bootcamp (2025-Now)

Learned data science methodologies, Python, and data visualization.

Working Experience :

Web Developer Intern PT.Indonusa Multijaya (june 2023 - august 2023)

Create a website for company profiling



BACKGROUND

Ketepatan waktu dalam layanan food delivery sangat menentukan kepuasan pelanggan dan efisiensi operasional. Faktor eksternal seperti jarak, kondisi cuaca, lalu lintas, serta pengalaman kurir seringkali membuat estimasi konvensional tidak akurat. Oleh karena itu, dibutuhkan model prediktif berbasis data yang mampu memberikan estimasi waktu pengiriman lebih andal.

PROBLEM STATEMENT



Variasi Waktu Pengiriman

Waktu antar pesanan sangat bervariasi dipengaruhi jarak, traffic, cuaca, dan jam sibuk.



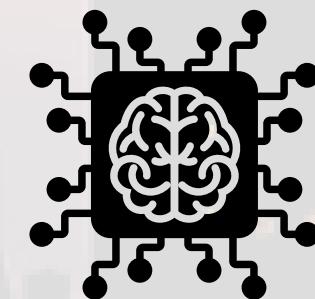
Keterbatasan Estimasi Konvensional

Estimasi ETA masih statis dan belum memperhitungkan faktor dinamis seperti traffic atau cuaca.



Dampak pada Kepuasan Pelanggan

Estimasi yang meleset menurunkan kepuasan dan kepercayaan pelanggan terhadap layanan.



Tantangan dalam Pemilihan Model

Diperlukan evaluasi model machine learning untuk menemukan metode prediksi yang paling akurat.

METODOLOGY

- 01 Dataset: **Food Delivery Times.csv**
Struktur : 1000 baris, 9 kolom.
Source : kaggle

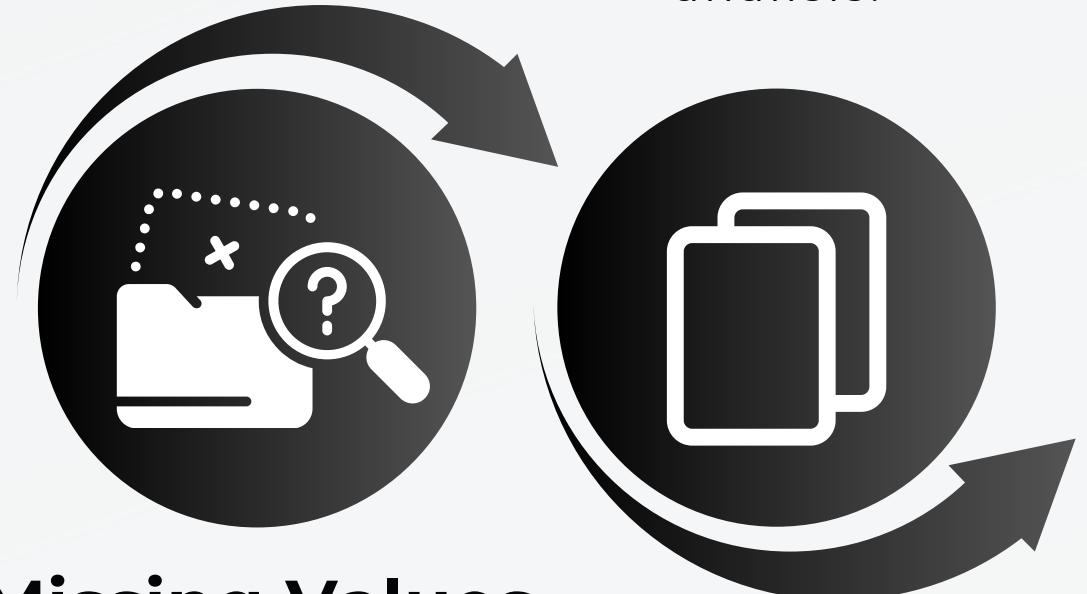
Jenis Kolom	Kolom
Numerik	Distance_km, Preparation_Time_min, Courier_Experience_yrs
Kategorikal	Weather, Traffic_Level, Time_of_Day, Vehicle_Type
Target	Delivery_Time_min (menit)



DATA CLEANING

Duplicate Check

Memastikan tidak ada baris duplikat yang bisa memengaruhi hasil analisis.

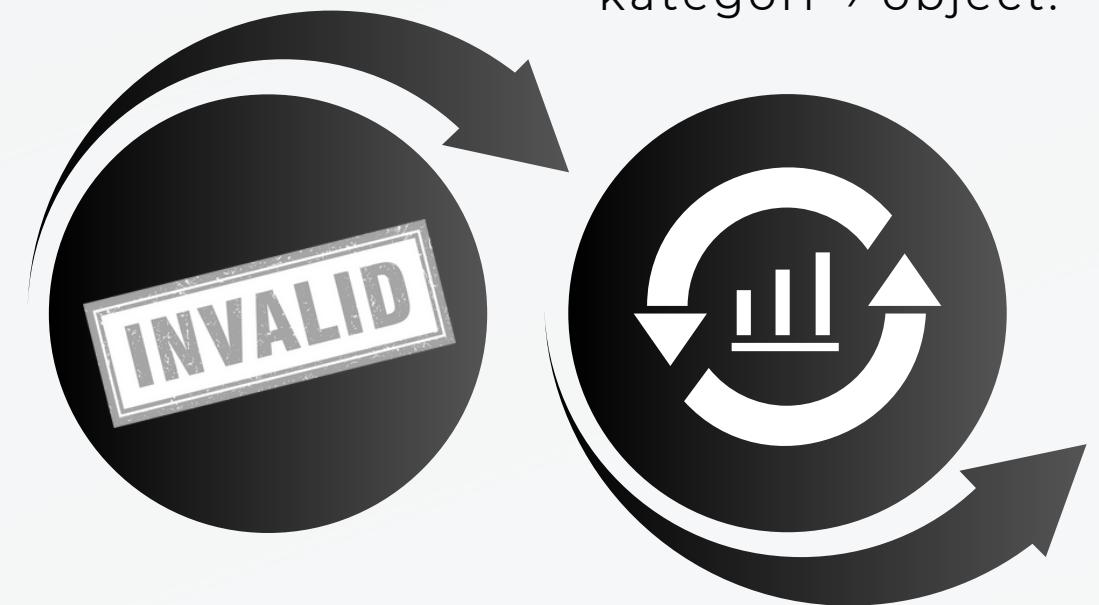


Missing Values Handling

Menghapus atau imputasi data hilang pada kolom Weather, Traffic_Level, Time_of_Day, Courier_Experience _yrs.

Data Type Conversion

Mengubah tipe data sesuai kebutuhan:
numerik → float/int,
kategori → object.



Outlier & Invalid Data

Membuang data tidak logis seperti jarak ≤ 0 km atau waktu persiapan ≤ 0 menit.

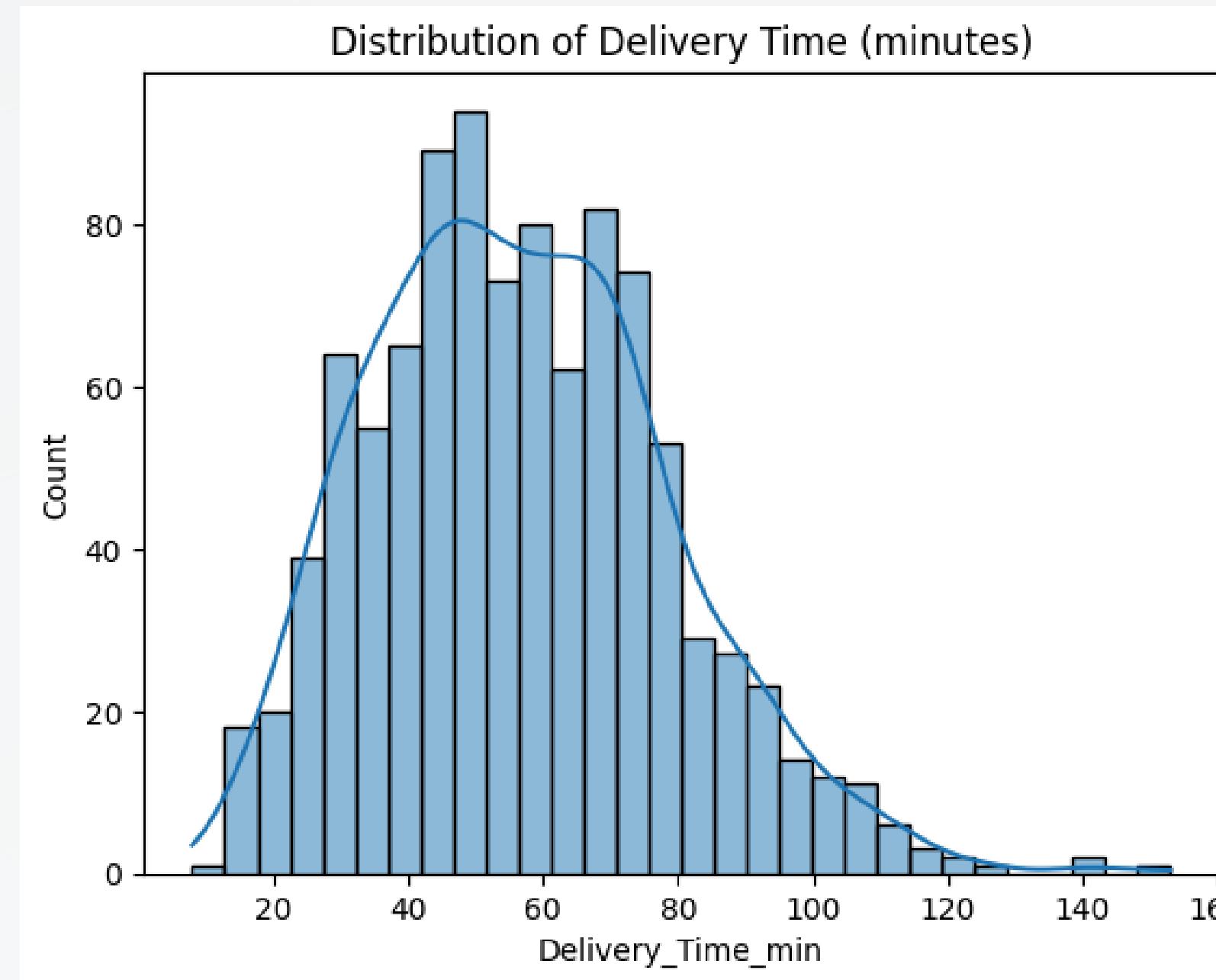
Preprocessing Ready

Dataset siap dipakai modelling
→ fitur numerik di-scale, fitur kategori di-encode.

EXPLORATORY DATA ANALYST

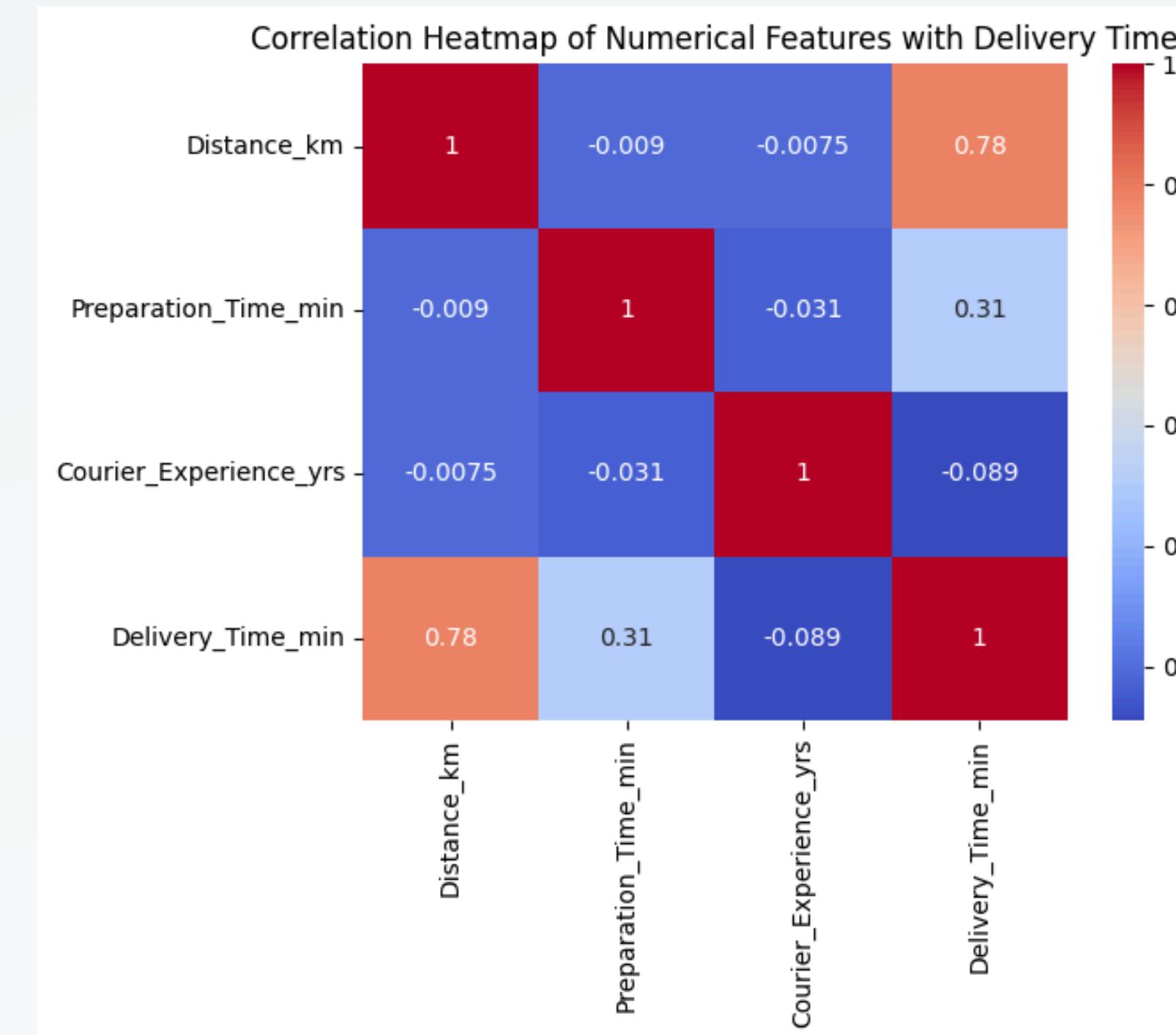


DISTRIBUTION OF TARGET VARIABLE



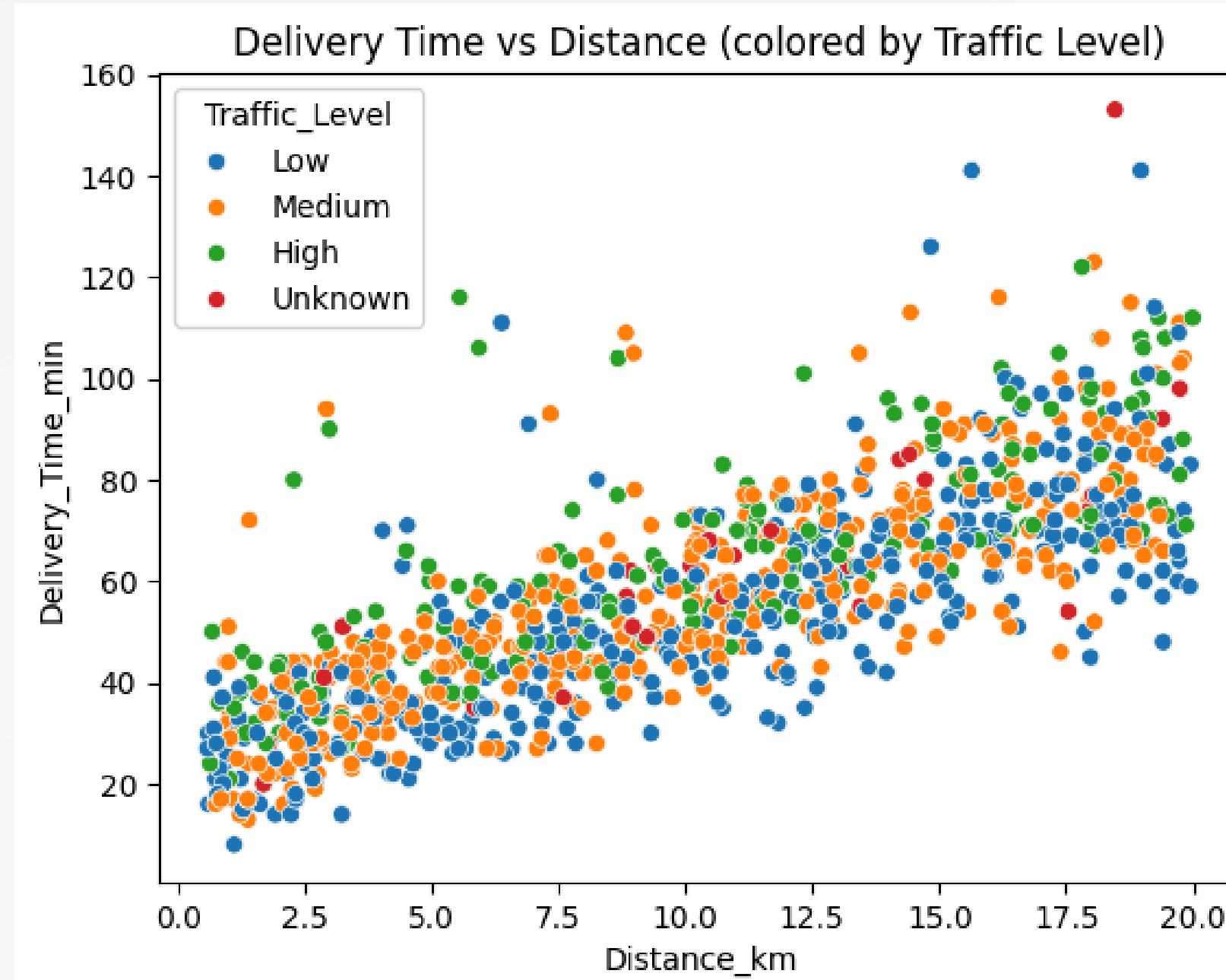
Mayoritas pesanan dikirim dalam 30–80 menit dengan puncak di 50–60 menit. Distribusi miring ke kanan menandakan adanya beberapa pengiriman yang sangat lama (outlier).

CORRELATION AMONG NUMERICAL VARIABLES



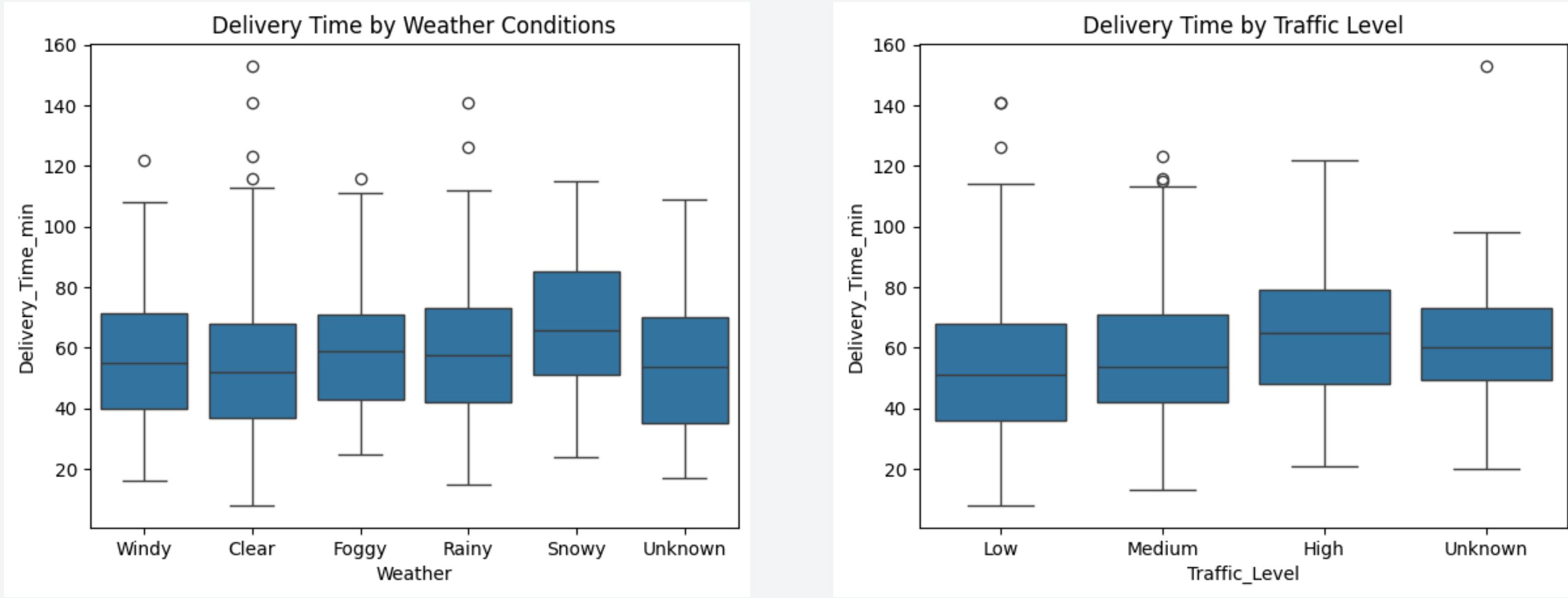
Jarak pengiriman memiliki korelasi paling kuat terhadap waktu antar ($r=0.78$), diikuti waktu persiapan ($r=0.31$). Pengalaman kurir hanya berpengaruh kecil.

RELATIONSHIP BETWEEN DISTANCE AND DELIVERY TIME



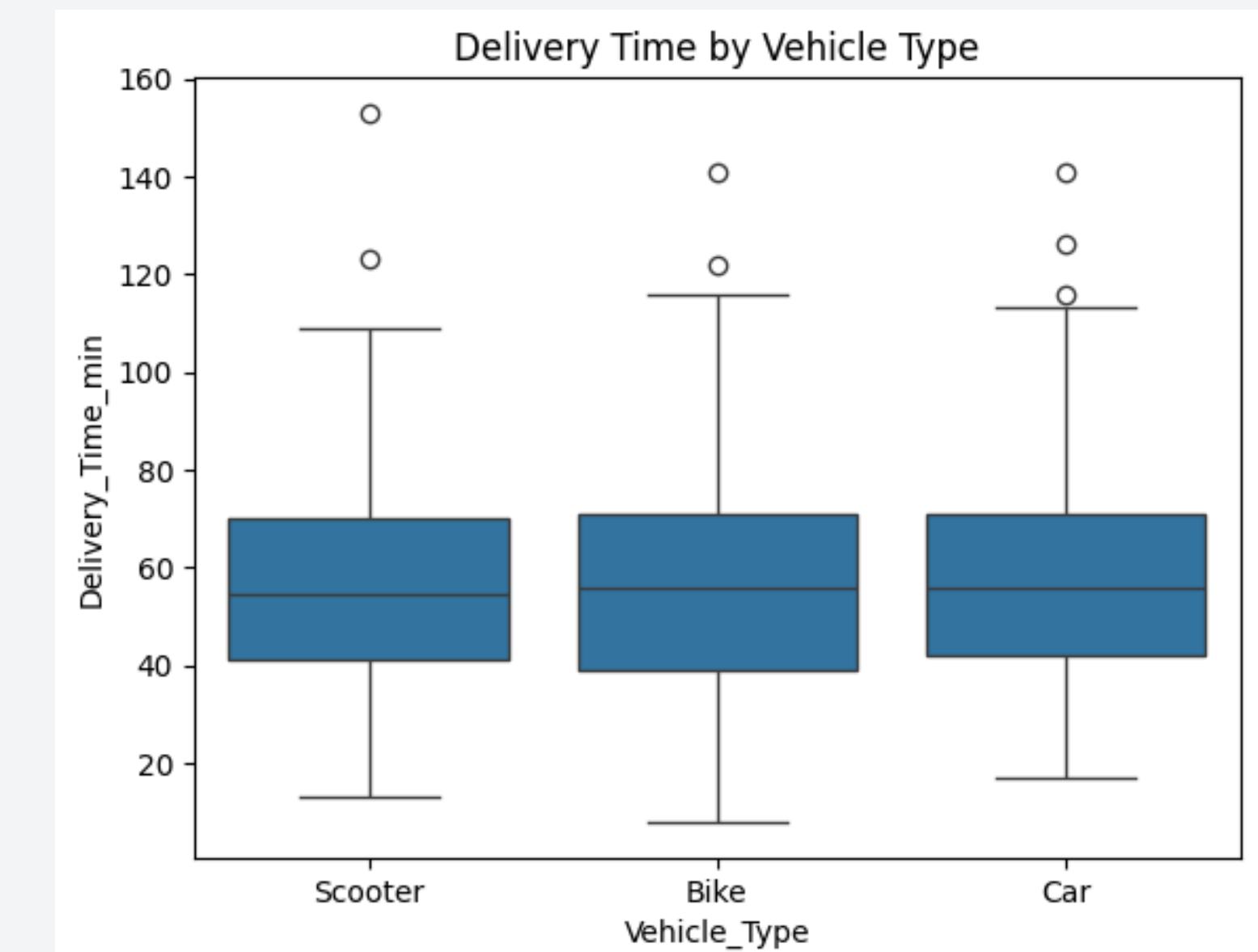
Semakin jauh jarak, semakin lama waktu pengiriman. Kondisi high traffic memperlambat waktu antar pada jarak yang sama, sehingga traffic menjadi faktor penting dalam prediksi.

IMPACT OF WEATHER & TRAFFIC LEVEL ON DELIVERY TIME



Cuaca buruk (Rainy/Snowy) dan kondisi high traffic memperpanjang waktu pengiriman. Clear weather & medium traffic menghasilkan pengiriman lebih cepat dan stabil.

IMPACT OF TIME OF DAY & VICHICLE TYPE ON DELIVERY TIME



Pengiriman pada pagi dan siang hari cenderung lebih lama karena aktivitas tinggi, sedangkan malam lebih cepat dan stabil. Dari sisi kendaraan, scooter dan bike lebih efisien menembus macet, sementara mobil sedikit lebih lama karena terhambat traffic dan parkir.

FEATURE ENGINEERING

Proses ini menambah dan mengolah fitur agar faktor penting seperti jarak, traffic, cuaca, dan konteks pengiriman lebih jelas tercermin dalam model prediksi.

Flag Unknown

menandai data kategori yang hilang/“unknown” agar model tetap mengenali pola.

Ordinal Mapping Traffic Level

mengubah kategori traffic (Low, Medium, High) menjadi nilai ordinal sesuai tingkat kepadatan.

Interaction: Distance × Traffic

membuat variabel interaksi untuk menangkap efek jarak yang berbeda saat kondisi traffic bervariasi.

Durasi Awal (Prep + Distance)

menambahkan waktu persiapan makanan dengan jarak pengiriman sebagai estimasi awal durasi.

Konteks Pengiriman (Weather, Courier Experience, Peak Hour)

menambahkan fitur terkait cuaca buruk, pengalaman kurir, dan waktu sibuk untuk memperkaya prediksi.

MODELLING

Tiga algoritma dipilih untuk membandingkan performa prediksi waktu antar: LightGBM, Random Forest, dan XGBoost. Model ini mewakili metode boosting dan ensemble yang efektif menangkap pola kompleks dalam data.

LightGBM

Model boosting berbasis pohon yang dioptimalkan untuk kecepatan dan efisiensi.

Random Forest

Model ensemble berbasis pohon keputusan yang menggabungkan banyak pohon untuk meningkatkan akurasi prediksi.

XGBoost

Model boosting populer yang dirancang untuk performa tinggi. Mampu menangani data non-linear dan interaksi fitur dengan baik

MODEL EVALUATION

Test Prediction		RMSE	MAE	R2
Model				
Random Forest		9.601	6.696	0.794
XGBoost		9.236	6.316	0.810
LightBGM		9.513	6.896	0.789

- XGBoost memiliki performa baik dengan akurasi ($R^2 = 0.81$)
- Random Forest cukup stabil ($R^2 = 0.79$), namun sedikit lebih tinggi error.
- LightGBM memiliki akurasi paling rendah di antara ketiga model ($R^2 = 0.78$).

MODEL EVALUATION

Train Prediction		RMSE	MAE	R2
Model				
Random Forest		8.127	4.846	0.867
XGBoost		8.365	5.347	0.859
LightGBM		9.533	6.122	0.817

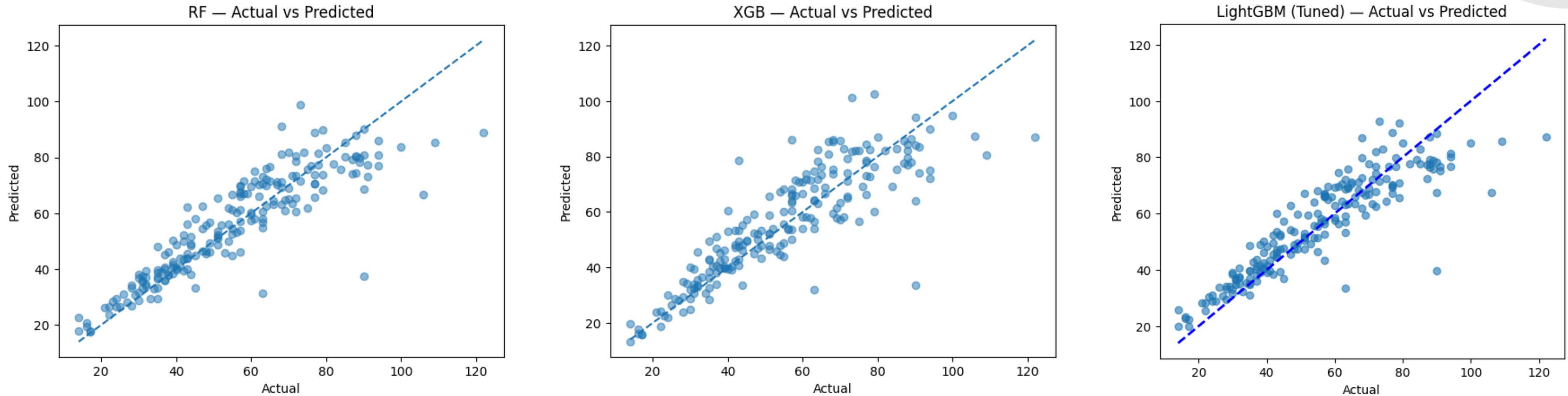
- Random Forest memberikan hasil, ($R^2 = 0.867$) dengan error paling kecil.
- XGBoost juga cukup baik ($R^2 = 0.859$) namun sedikit lebih tinggi error.
- LightGBM masih bekerja baik ($R^2 = 0.817$), tetapi akurasi lebih rendah dibanding dua model lainnya.

MODEL EVALUATION

Cross Validation V5			
Model	RMSE	MAE	R2
Random Forest	11.256 ± 1.588	7.460 ± 0.648	0.738 ± 0.053
XGBoost	12.201 ± 1.427	8.031 ± 0.531	0.693 ± 0.048
LightGBM	11.078 ± 1.458	7.306 ± 0.598	0.747 ± 0.046

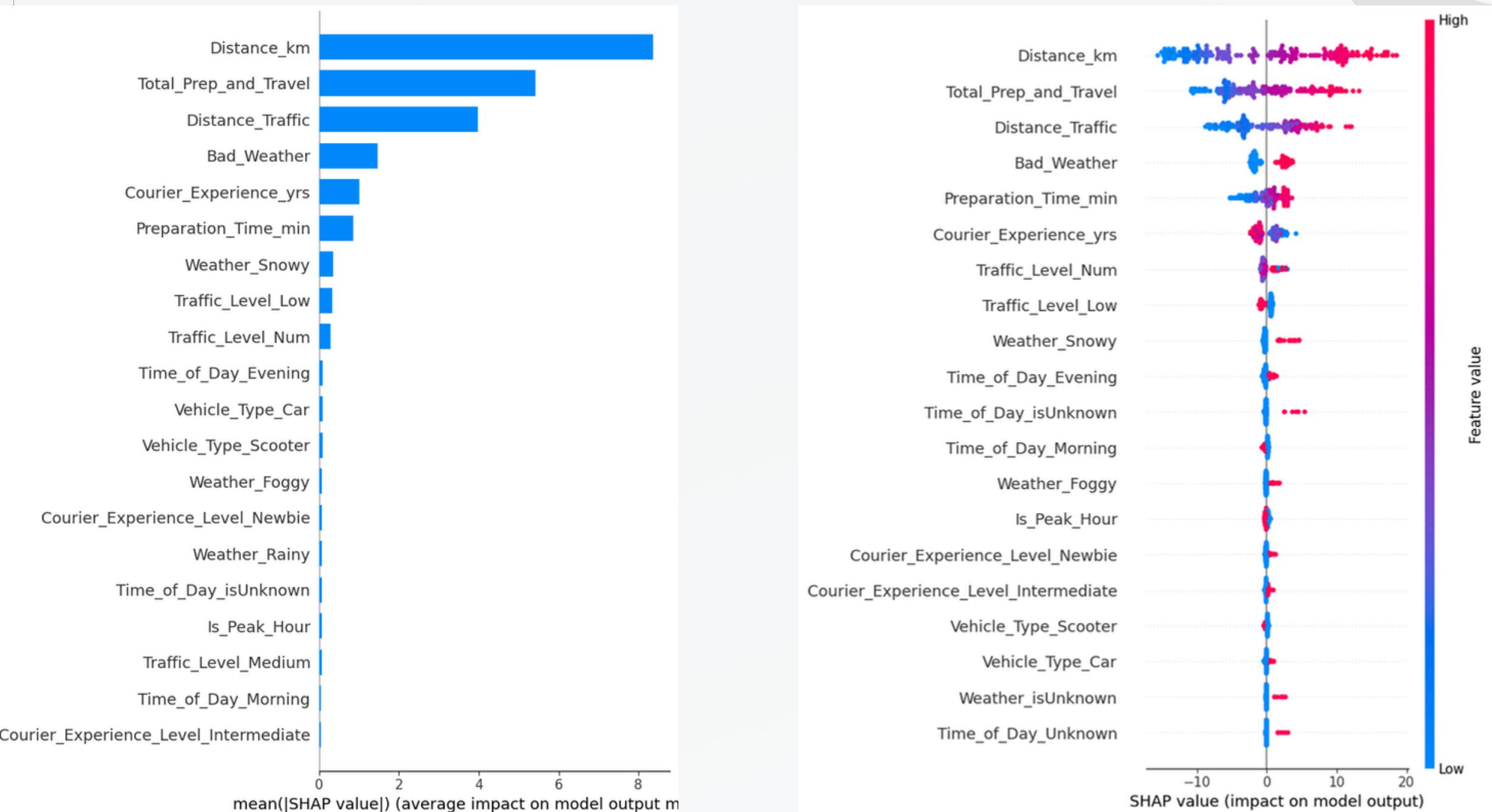
- Random Forest memberikan hasil, ($R^2 = 0.867$) dengan error paling kecil.
- XGBoost juga cukup baik ($R^2 = 0.859$) namun sedikit lebih tinggi error.
- LightGBM masih bekerja baik ($R^2 = 0.817$), tetapi akurasi lebih rendah dibanding dua model lainnya.

MODEL EVALUATION ACTUAL VS PREDICT



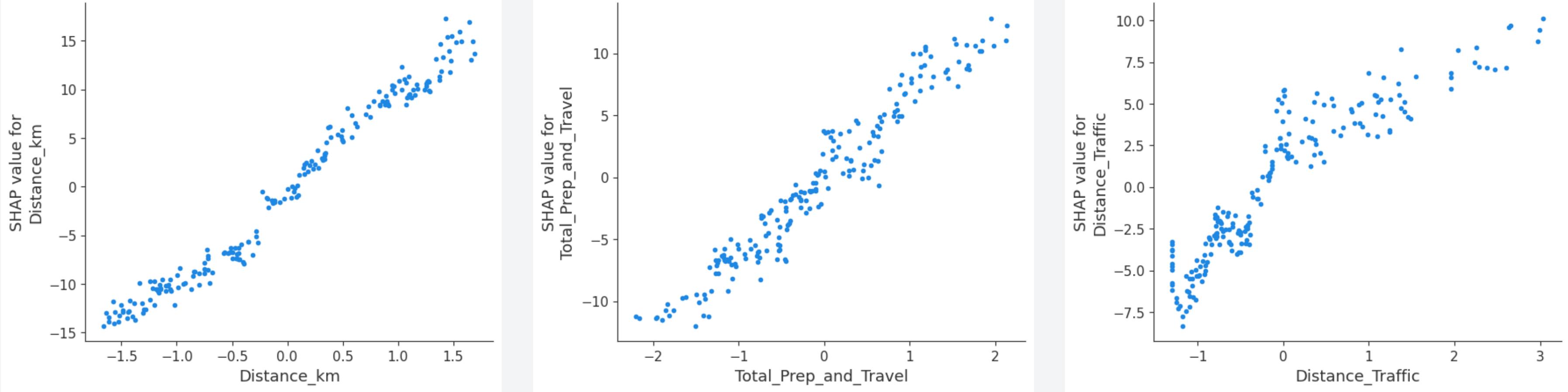
- Random Forest → Pola cukup rapi mendekati garis diagonal, ada beberapa sebaran error pada nilai tinggi.
- XGBoost → Titik prediksi lebih konsisten mengikuti garis diagonal.
- LightGBM → Cukup akurat, nilai lebih mendekati garis diagonal

GLOBAL IMPORTANCE LIGHTGBM



Faktor utama yang memengaruhi waktu pengiriman adalah jarak (Distance_km) dan durasi total persiapan + perjalanan (Total_Prep_and_Travel).

SHAP DEPENDENCES FEATURE LIGHTGBM



- $\text{Distance_km} \rightarrow$ Semakin jauh jarak, semakin besar kontribusi positif terhadap waktu pengiriman. Polanya sangat kuat dan linier.
- $\text{Total_Prep_and_Travel} \rightarrow$ berbanding lurus dengan keterlambatan. Semakin lama durasi, semakin tinggi dampaknya terhadap waktu antar.
- $\text{Distance_Traffic} \rightarrow$ Interaksi jarak dengan traffic meningkatkan variasi. Pada kondisi macet, pengaruh jarak semakin tajam, sehingga memperlambat pengiriman lebih signifikan.

CONCLUSION

Dari hasil pengujian, tiga model yang dibandingkan yaitu Random Forest, XGBoost, dan LightGBM menunjukkan performa yang cukup baik dalam memprediksi waktu pengiriman. XGBoost menghasilkan nilai akurasi tertinggi pada data uji dengan R^2 sebesar 0.810 serta MAE terendah, menjadikannya model paling andal dalam memberikan prediksi yang mendekati data aktual. LightGBM tetap memberikan hasil yang stabil dan efisien, sementara Random Forest menunjukkan performa kompetitif meski sedikit di bawah dua model lainnya. Berdasarkan kombinasi akurasi dan kemampuan generalisasi, XGBoost dipilih sebagai model akhir untuk digunakan dalam prediksi waktu pengiriman.

RECOMENDATION BUSINESS ACTION

- Gunakan XGboost sebagai model utama untuk memperkirakan waktu pengiriman agar estimasi lebih konsisten.
- Integrasikan hasil prediksi ke dalam aplikasi driver/customer untuk meningkatkan kepuasan pelanggan.
- Manfaatkan insight faktor penting (jarak, traffic, cuaca, jam sibuk) untuk mengoptimalkan penugasan kurir.
- Lakukan pemantauan berkala dan retraining model agar performa tetap terjaga seiring perubahan kondisi operasional.

THANK YOU

FOR YOUR ATTENTION

September 2025



Project



imadebayusatriawardhana@gmail.com



I Made Bayu Satria Wardhana

