



BIKE SHARING DEMAND PREDICTION USING
MACHINE LEARNING

**FORECASTING DAILY BIKE RENTAL DEMAND TO
OPTIMIZE OPERATIONAL PLANNING AND
RESOURCE ALLOCATION**

Oleh: Dede Yudha N

OUTLINE

1. Background & Problem Statement
2. Executive Summary
3. Data Understanding
4. Exploratory Data Analysis
5. Data Cleaning & Feature Engineering
6. Modeling & Evaluation
7. Business Impact
8. Conclusion & Recommendations



EXECUTIVE SUMMARY

BACKGROUND

Capital Bikeshare adalah layanan berbagi sepeda di kawasan Metro DC dan sekitarnya, dengan ribuan sepeda dan ratusan stasiun. Pengguna bisa meminjam lewat aplikasi atau tiket, lalu mengembalikan di stasiun mana saja.

Tersedia sepeda klasik dan e-bike. Sistem ini menawarkan opsi penggunaan cepat, ekonomis, dan ramah lingkungan.

PROBLEM

Fluktuasi permintaan harian menyulitkan perusahaan mengatur alokasi sepeda, docking station, dan kebutuhan logistik. Tanpa prediksi, terjadi over-supply atau shortage.

STAKEHOLDER / USER

- Tim operasional
- Tim logistik & rebalancing
- Manajemen strategis

GOAL

Membangun model ML untuk memprediksi jumlah penyewaan harian secara lebih akurat sebagai dasar perencanaan operasional.

ANALYTICAL APPROACH

Regression modeling (multiple models dibandingkan via cross-validation).

METRIC

RMSE, MAE, dan R² sebagai metrik evaluasi akurasi model regresi.

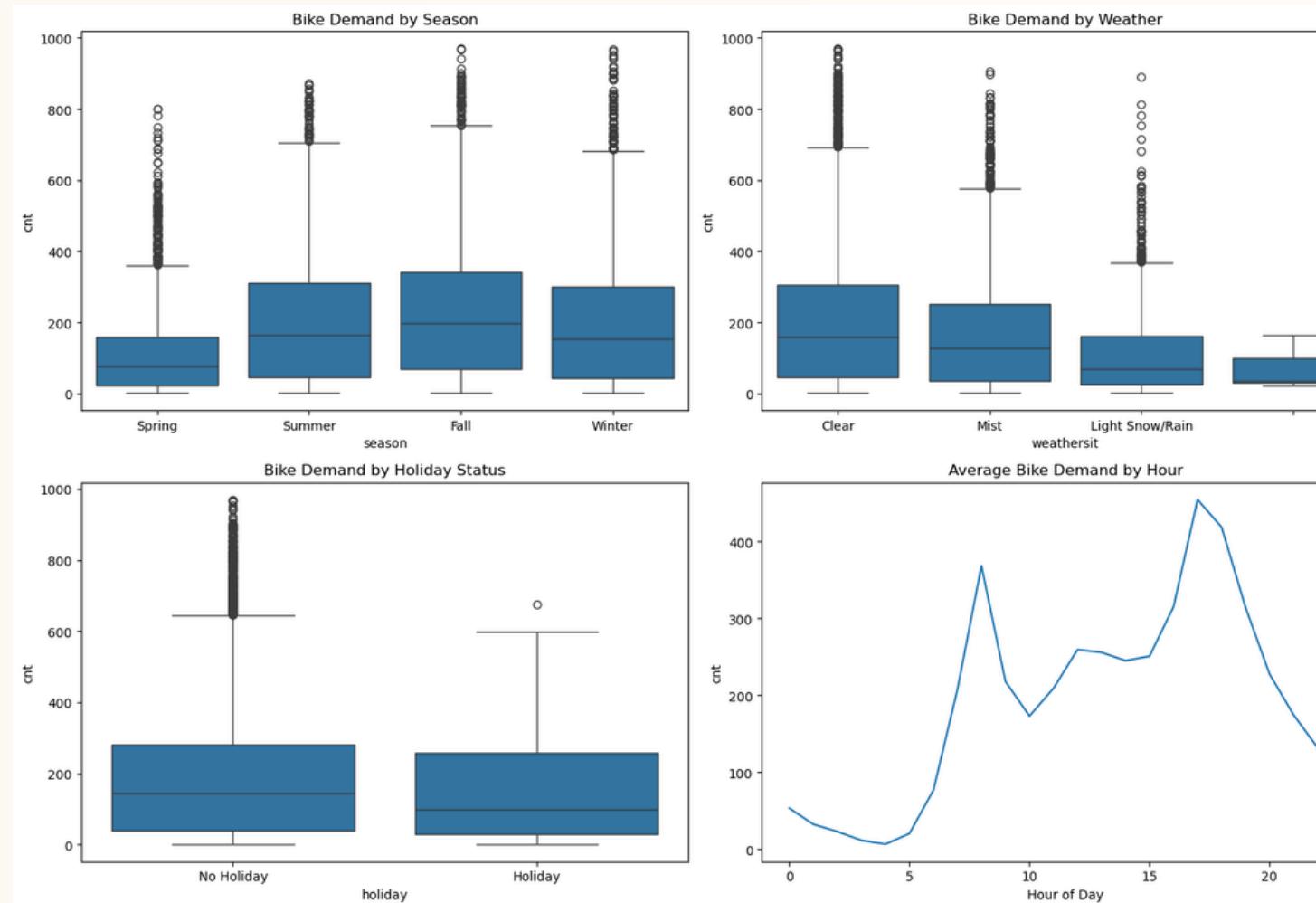
DATA UNDERSTANDING

Merujuk pada dataset Bike Sharing - UCI

Fitur Utama

- datetime — tanggal observasi
- season — musim (1: Spring, 2: Summer, 3: Fall, 4: Winter)
- holiday — status hari libur (0: bukan libur, 1: hari libur)
- workingday — status hari kerja (0: akhir pekan/libur, 1: hari kerja)
- weather — kondisi cuaca
- temp / atemp — suhu aktual & suhu persepsi (°C)
- humidity — kelembapan (%)
- windspeed — kecepatan angin
- casual / registered — jumlah pengguna kasual & pengguna terdaftar
- count — total penyewaan (casual + registered)

EXPLORATORY DATA ANALYSIS



Distribusi Target (count)

- Distribusi cenderung right-skewed.
- Peningkatan rental terlihat konsisten pada musim panas.

Hubungan Feature vs Target

Highlight temuan utama:

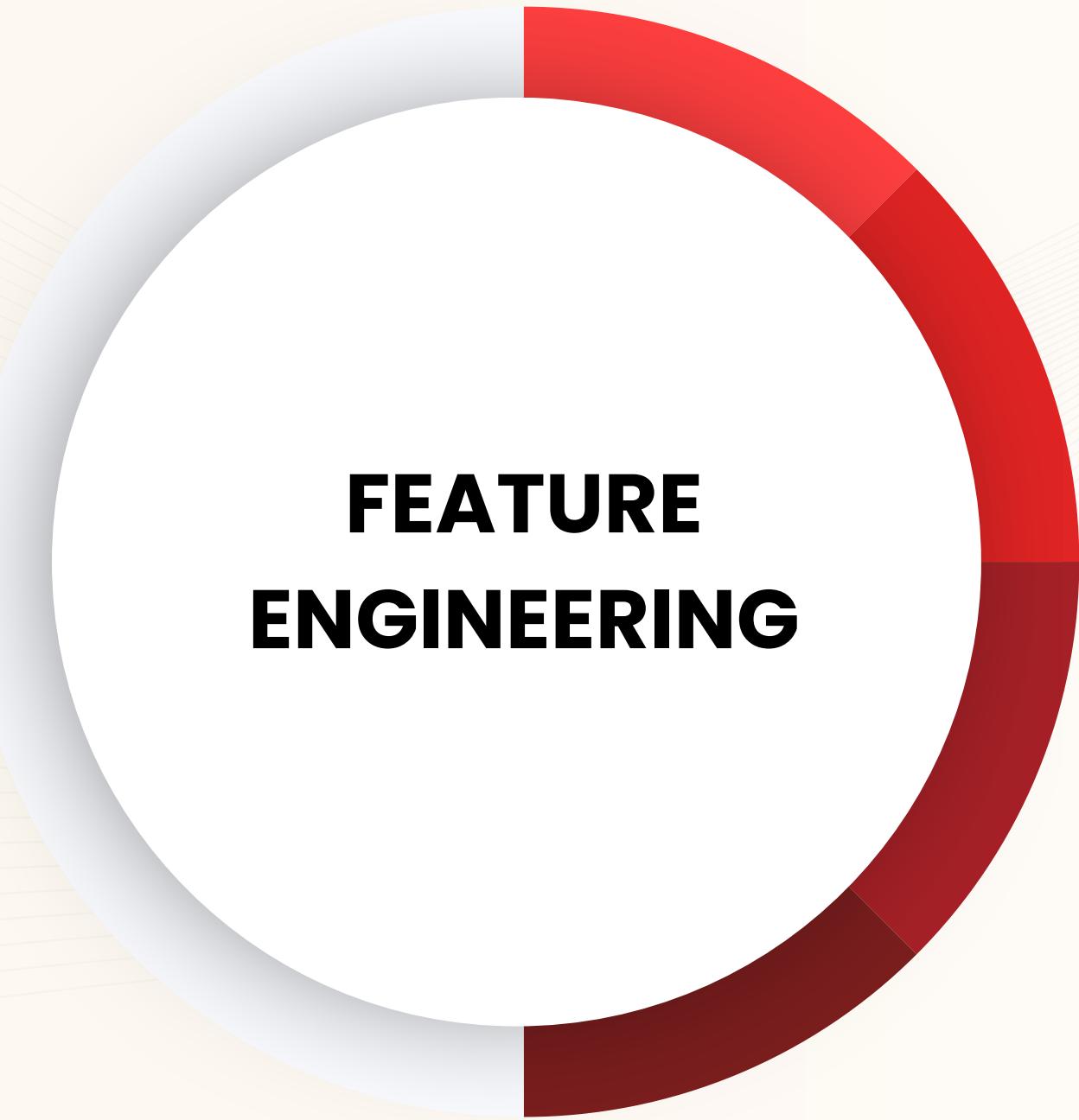
- Suhu berkorelasi positif dengan jumlah rental. Semakin hangat, semakin banyak penyewaan.
- Kelembapan tinggi menurunkan permintaan.
- Cuaca buruk (hujan/salju) sangat menurunkan penyewaan.
- Hari kerja memiliki pola berbeda: jam sibuk lebih tinggi pada weekday dibanding weekend.

DATA CLEANING

Menghapus
duplikasi

Memastikan
tidak ada
missing values

Encoding
kategori (season,
weather, holiday)



FEATURE ENGINEERING

Ekstraksi fitur waktu dari datetime
(year, month, day, weekday)

Transformasi log pada target (opsional,
jika distribusi terlalu skewed)

Normalisasi numerik untuk model
tertentu

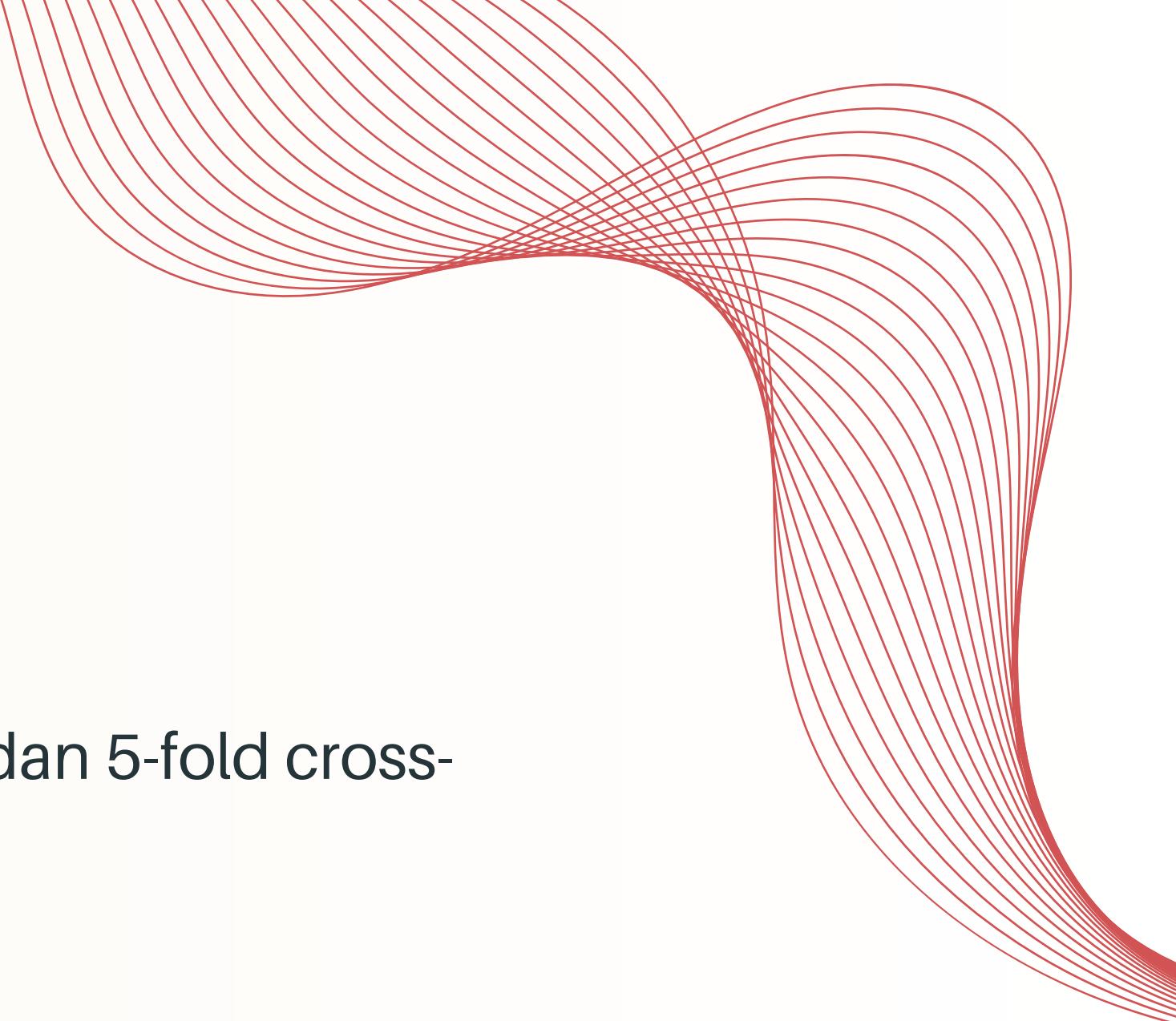
Pembuatan interaction features jika
diperlukan

MODELING & EVALUATION

MODEL CANDIDATES

Model yang diujikan menggunakan preprocessing Pipeline dan 5-fold cross-validation:

- Linear Regression
- Decision Tree Regressor
- Random Forest Regressor
- K-Nearest Neighbors Regressor
- XGBoost Regressor



MODEL SELECTION (BASED ON CROSS-VALIDATION)

Berdasarkan hasil evaluasi:

XGBoost memberikan performa terbaik

- RMSE terendah $\approx 98-102$
- MAE $\approx 66-68$
- R^2 tertinggi $\approx 68-69\%$
- Stabil (std kecil antar fold)

Random Forest berada di posisi kedua

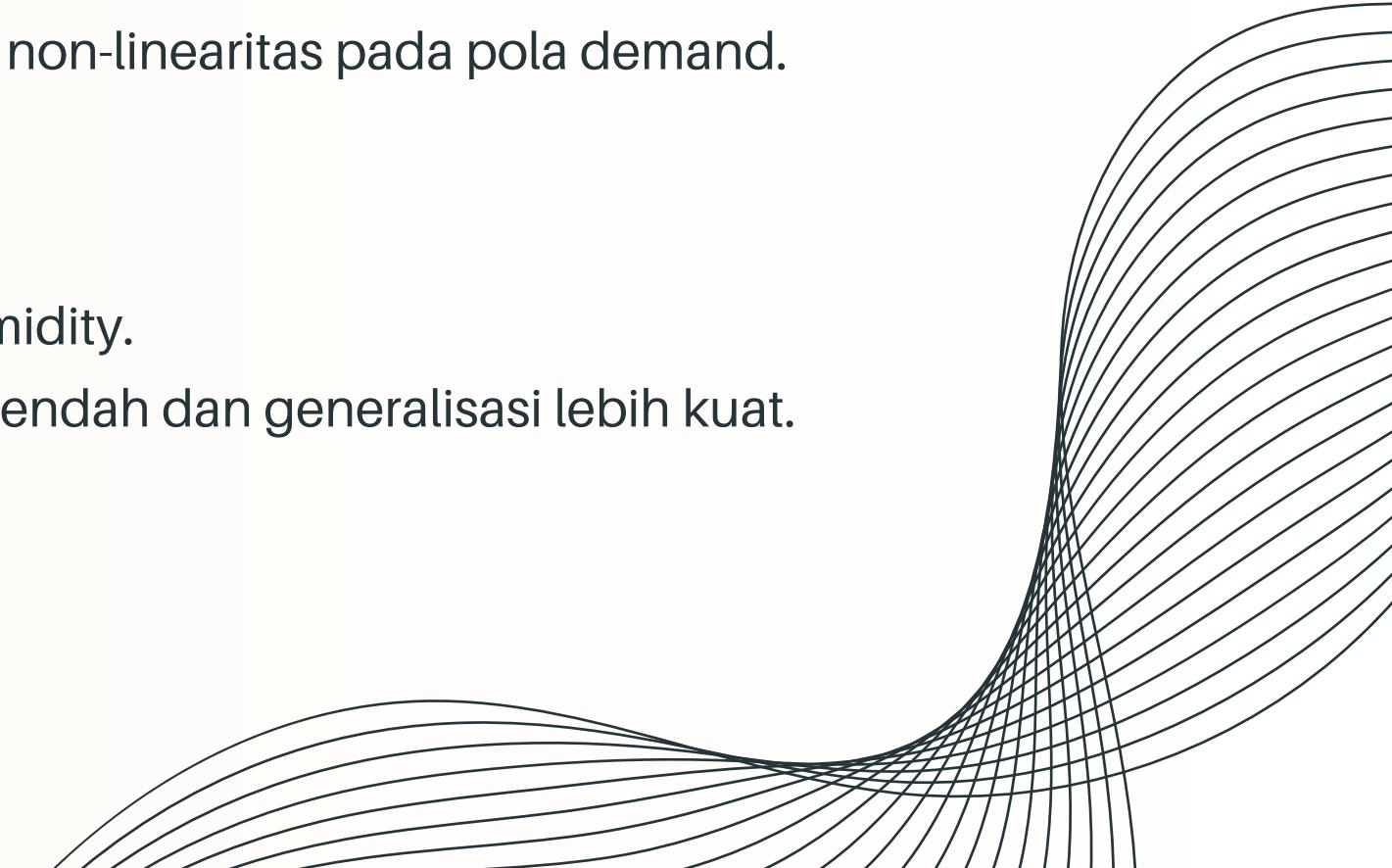
- RMSE $\approx 104-107$
- $R^2 \approx 64-65\%$

Linear Regression dan Decision Tree berada jauh di bawah karena tidak mampu menangkap non-linearitas pada pola demand.

Final Model yang dipilih: **XGBoost Regressor**

Alasan pemilihan:

- Mampu menangkap hubungan non-linear antara jam, musim, cuaca, temperatur, dan humidity.
- Boosting bekerja iteratif memperbaiki error residual sehingga menghasilkan RMSE lebih rendah dan generalisasi lebih kuat.
- Robust terhadap multikolinearitas (temp & atemp), outliers, dan variabel temporal.





HYPERPARAMETER UTAMA (SETELAH TUNING)

Hyperparameters optimal hasil

RandomizedSearchCV:

- n_estimators
- max_depth
- learning_rate
- subsample
- colsample_bytree

Tuning meningkatkan performa model:

Metric	Before Tuning	After Tuning	Improvement
RMSE	9.897	9.708	189
MAE	6.600	6.490	111
MAPE	9.244%	8.371%	873%
R²	6.857%	6.976%	119%

MODEL INTERPRETATION

Menggunakan feature importance:

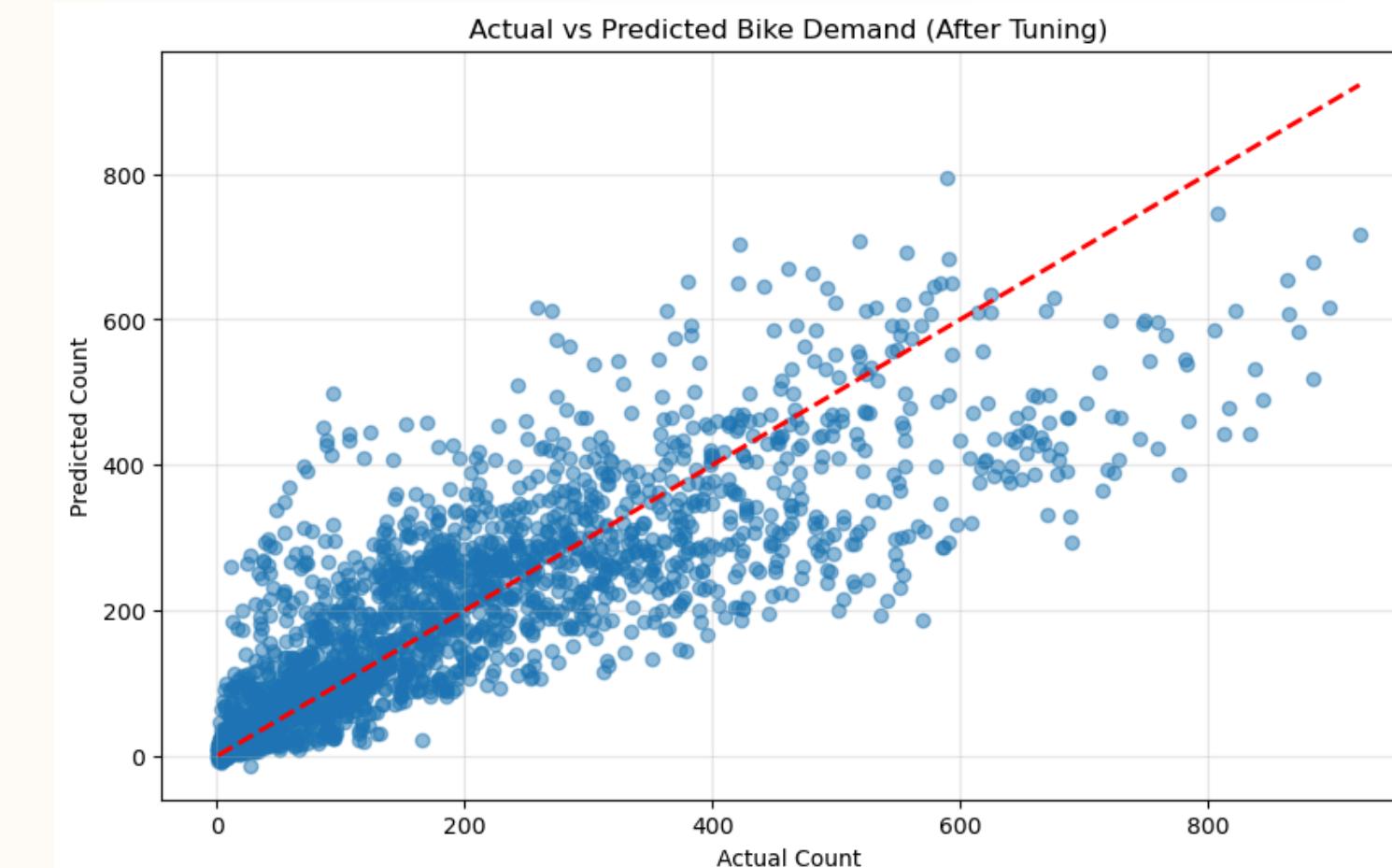
- Temperatur (temp & atemp)
 - Faktor paling dominan. Semakin hangat, demand meningkat kuat.
- Cuaca
 - Kondisi buruk (snow/rain/heavy mist) menurunkan prediksi secara signifikan.
- Kelembapan (humidity)
 - Tinggi = menekan permintaan (pengguna enggan bersepeda).
- Fitur waktu
 - Bulan, musim = memengaruhi pola musiman.
 - Jam = membentuk puncak demand morning-evening commute.

Interpretasinya konsisten dengan EDA dan perilaku nyata pengguna bike-sharing.

MODEL DIAGNOSTIC

Scatter actual vs predicted menunjukkan:

- Sebagian besar titik mendekati garis diagonal = prediksi cukup akurat.
- Overprediction/underprediction terjadi pada nilai ekstrem (count tinggi), wajar untuk demand dataset yang right-skewed.



LIMITATIONS

- Tidak menggunakan data cuaca real-time (hanya historical weather).
- Tidak memasukkan variabel peristiwa besar (festival kota, libur nasional spesifik).
- Polusi udara, traffic, dan mobilitas publik belum dimasukkan.
- XGBoost tetap bersifat "black-box" dibanding model linear.

FUTURE IMPROVEMENTS

- Menambahkan data eksternal: polusi, traffic, public mobility index.
- Menggunakan model time-series hybrid (XGBoost + LSTM).
- Menggunakan SHAP untuk interpretasi model yang lebih transparan.
- Melatih model per-musim atau per-zone (jika dataset diperluas).

ANALISIS DAMPAK BISNIS

ASUMSI

- Armada 500 sepeda, 20 stasiun, harga sewa \$3/jam.
- Biaya: redistribusi \$2/trip, maintenance \$0.5/sepeda/hari, staf \$20/jam.
- Base demand ≈ 200 sewa/jam, utilisasi awal 60%.

TOTAL MANFAAT & ROI (ANNUAL)

- Total manfaat tahunan: \$167,767
- Biaya implementasi (tahun 1): \$15,000
- Net benefit T1: \$152,767 → ROI ≈ 918%
- Payback period ≈ 1 bulan

SUMBER PENGHEMATAN UTAMA (ANNUAL)

- Penurunan stock-out: +\$26,280
- Optimasi utilisasi (idle ↓): +\$13,687
- Efisiensi redistribusi + staf: +\$27,100
- Retensi pelanggan & perpanjangan: +\$35,000
- Pendapatan tambahan dari utilisasi: +\$65,700

PERBAIKAN KPI

- Stock-out: 20% → 8% (↓ 60%)
- Fleet utilization: 60% → 75% (↑ 25%)
- Rentals/day: 300 → 360 (↑ 20%)
- CSAT: 65% → 82% (↑ 17 pp)
- Churn: 15% → 8% (↓ 47%)
- Redistribution cost: ↓ ~51%

RISIKO & SENSITIVITAS

- Skenario Best: manfaat +20% → ROI ≈ 1,242%
- Skenario Base: seperti di atas → ROI ≈ 918%
- Skenario Worst: manfaat -20% → ROI ≈ 795%
- Risiko utama: akurasi prediksi pada ekstrem demand, integrasi data real-time.

MANFAAT TAK TERUKUR

Reputasi & akuisisi naik, keunggulan kompetitif, budaya data-driven, dasar untuk dynamic pricing dan ekspansi.

REKOMENDASI ROLLOUT

Bulan 1-2: Pilot di 5 stasiun high-demand.

Bulan 3-4: Rollout ke 20 stasiun.

Bulan 5-6: Fine-tune, monitoring drift; bulan 7+: skala/ekspansi.

CONCLUSION

- Demand bike sharing dipengaruhi kuat oleh jam, suhu, cuaca, dan musim.
- Pola commuting (7-9 AM & 5-7 PM) adalah pendorong utama permintaan.
- Cuaca buruk serta suhu ekstrem terbukti menurunkan demand secara signifikan.
- Model XGBoost memberikan akurasi tertinggi (RMSE terendah, R^2 tertinggi).
- Model mampu meningkatkan efisiensi operasional dan mengurangi stock-out.
- Prediksi harian dapat digunakan untuk perencanaan operasional dan strategi musiman.

RECOMMENDATIONS

Operational & Inventory Strategy

- Fokus redistribusi pada jam peak (7-9 AM, 5-7 PM).
- Alokasi armada menyesuaikan cuaca dan suhu harian.
- Lakukan seasonal adjustment (lebih banyak sepeda pada Summer/Fall).
- Gunakan monitoring real-time untuk mendeteksi potensi stock-out.

Fleet Optimization

- Optimalkan kapasitas di stasiun high-traffic.
- Kurangi armada aktif saat cuaca ekstrem untuk efisiensi maintenance.

Pricing & Customer Strategy

- Terapkan dynamic pricing pada peak hours dan cuaca ideal.
- Promosi pada off-peak hours untuk meratakan demand.

Model & Technology Improvements

- Retraining berkala (monthly/quarterly).
- Tambahkan fitur eksternal: cuaca real-time, event kota, traffic.
- Bangun dashboard operasional & alert system untuk forecasting hourly demand.



CONTACT |

[linkedin.com/naeyanika](https://www.linkedin.com/in/naeyanika)



naeyanika@gmail.com



62 812-8817-2775

