



10 October 2025

PROJECT 1

SALES FORECASTING

AI INDONESIA - MACHINE LEARNING (GROUP A)

Taufiq Rizan

Andre Pranata

Adam Husain

OBJECTIVE

Membuat Sistem Machine Learning berbasis Artificial Intelligence (AI) untuk Melakukan peramalan penjualan periode mendatang berdasarkan data masa lalu.

WHY FORECASTING ?

peramalan mengubah ketidakpastian menjadi informasi yang dapat ditindaklanjuti, memungkinkan kita untuk beralih dari sekadar bereaksi terhadap masa depan menjadi aktif membentuk masa depan.



SCOPE - LIMITATION

Toko Grocery di negara Ecuador, Amerika Selatan.
menjual sekitar 33 kategory produk.

Kode Toko : 5

Family Product : Bread / Bakery

Data Periode : 1 Jan 2013 sd 15 Aug 2017

FRESH FOOD STOCK MANAGEMENT

Memandu perusahaan mempersiapkan bahan baku dan produk jadi untuk memenuhi permintaan customer
Expired Date
Bahan Baku dan Finish Good

FINANCIAL & RESOURCE BUDGETING

membantu menjadwalkan para staff penjualan secara tepat, menghindari biaya lebur yang tidak perlu atau kekurangan tenaga saat waktu sibuk

MARKETING STRATEGY

Memberikan data dan informasi bagi tim marketing, kapan waktu yang tepat untuk melakukan promosi atau aktivitas marketing lainnya



DATA SOURCE INFORMATION

ID	: ID transaksi (int)
Date	: tanggal transaksi (object)
Store_nbr	: kode toko (int)
Family	: kategori produk (object)
Sales	: total penjualan per hari (float)
onpromotion	: jumlah produk dipromosikan (int)
dcoilwtico	: harga minyak (float)
Entries	: 55.571



PROCESS DEVELOPMENT

Data Preprocessing

mengubah data mentah (raw data)
menjadi format yang bersih, terstruktur,
dan sesuai kebutuhan modeling

Evaluation

Mengukur seberapa baik kinerja
model dalam membuat prediksi.
perbandingan beberapa model



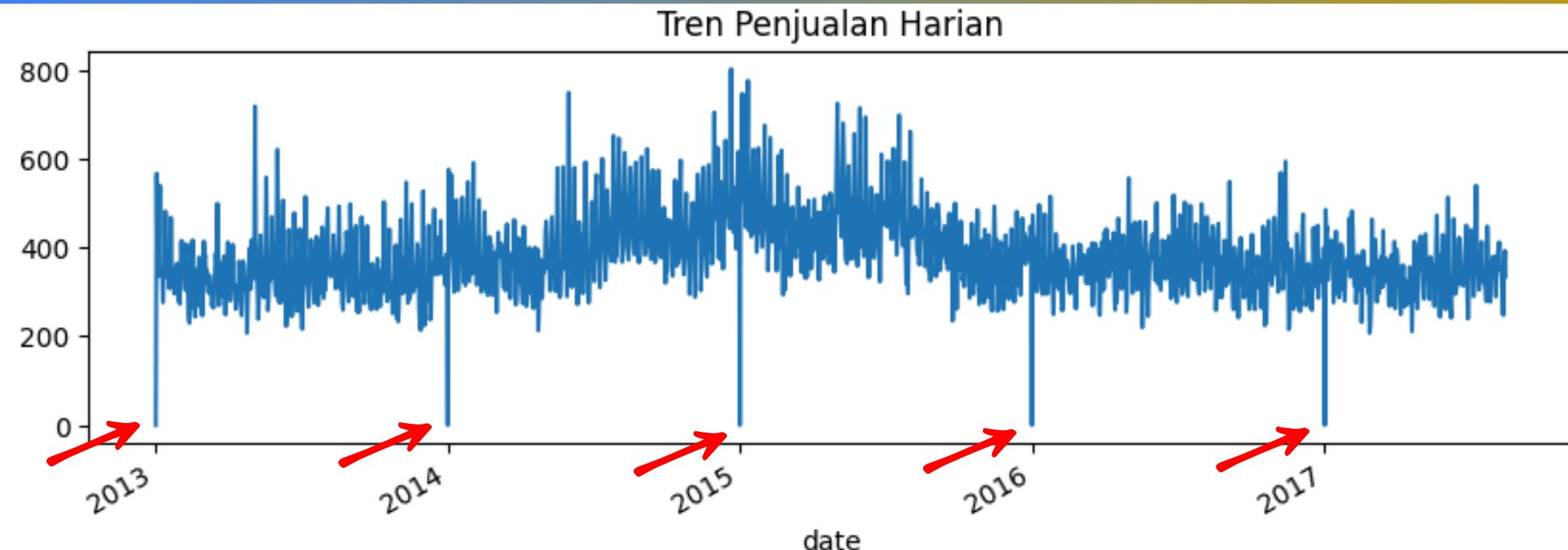
Exploration Data Analysis

Memahami struktur , karakteristik atau pola
Data secara deskriptif atau eksploratif

Modeling

melatih algoritma untuk menghasilkan
sebuah model yang mampu membuat
prediksi berdasarkan data baru

DATA ANALYSIS - BAKERY



✓ SALES TREND

Period : 1 Jan 2013 sd 15 Aug 2017
Average : 381.74
Max : 802.36
Std Dev : 87.28 → 23% dari average

WEEKEND SALES (Sabtu-Minggu)

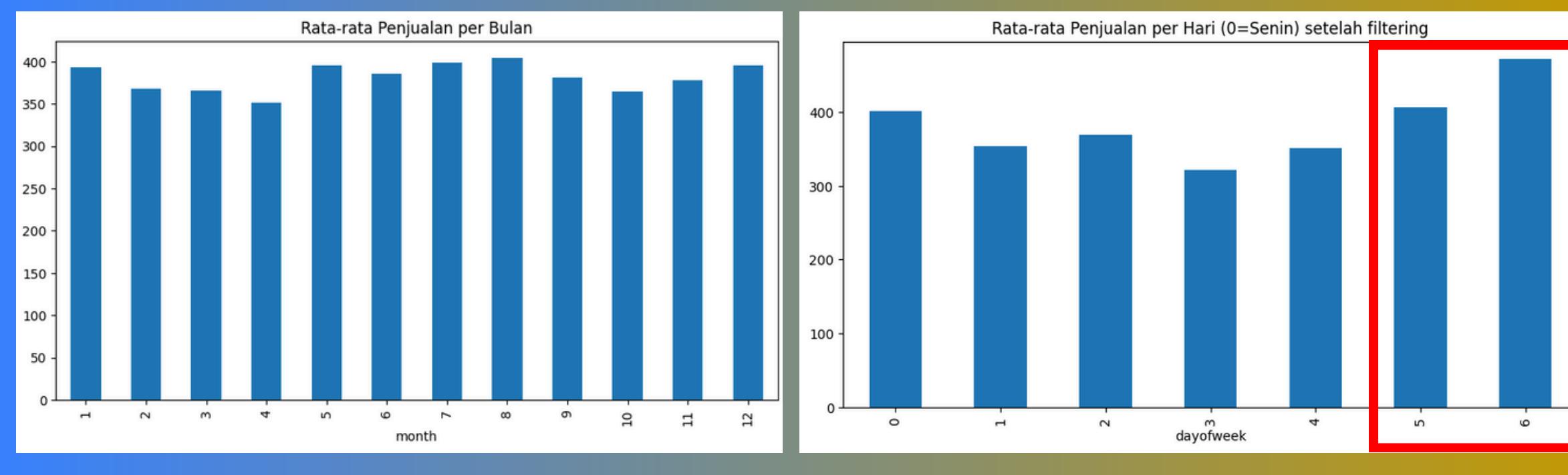
Average sales +15% : 438.67

MONTHLY SALES

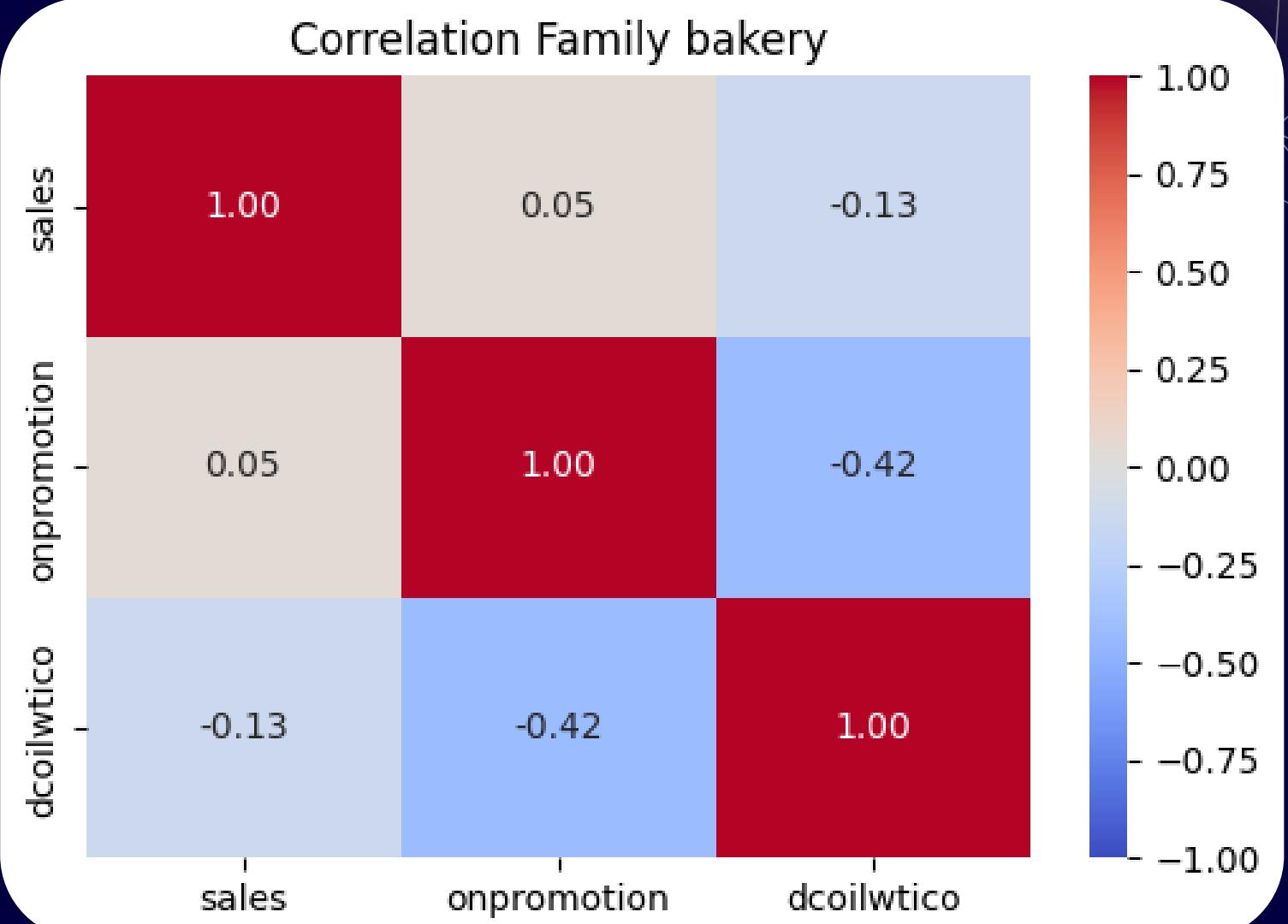
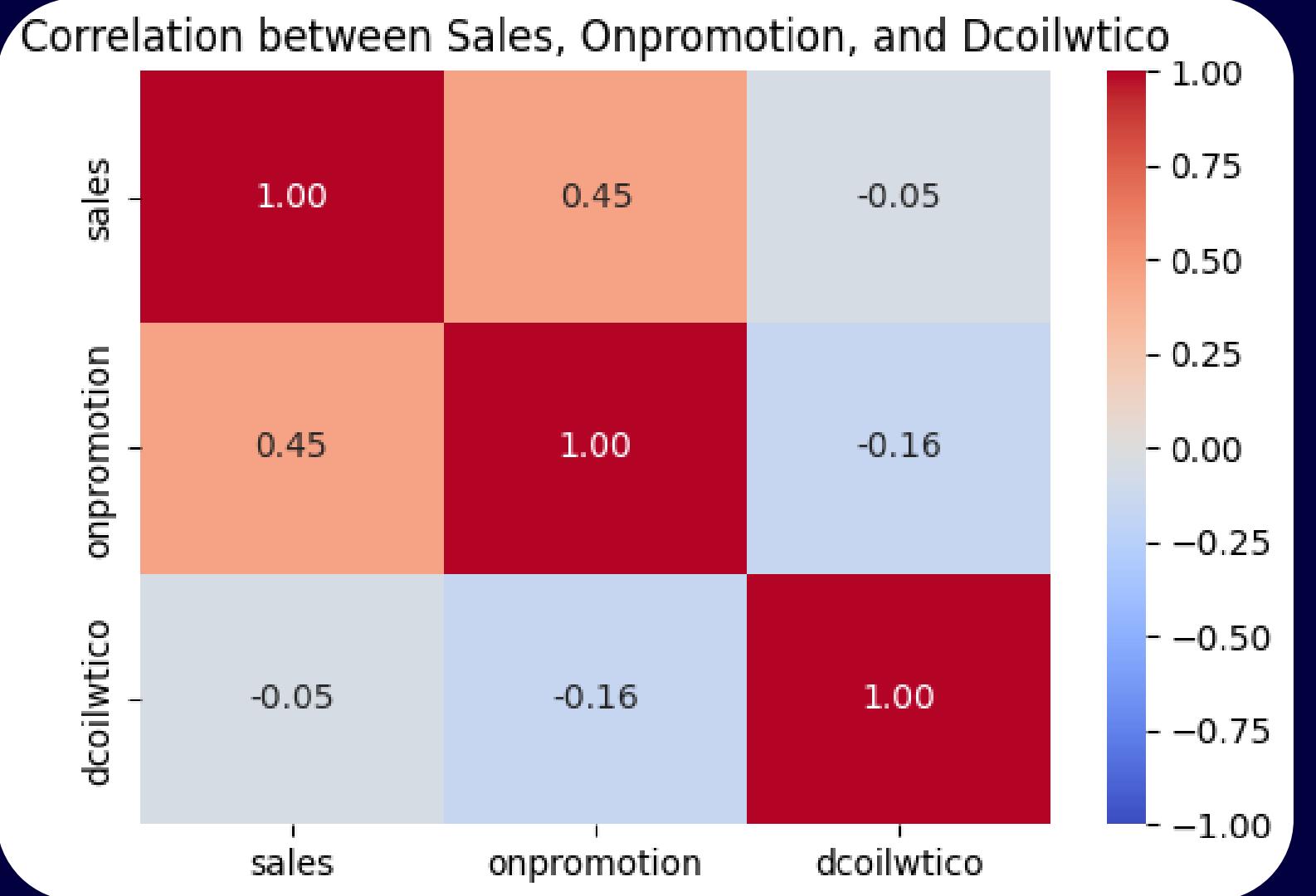
Highest sale (Aug) : 403.7 (+5.5%)
Lowest sale (Apr) : 351.3 (-8%)

✓ MISSING VALUE

missing value : 5 dari 1688 data (0.02%)
01/01/2013 01/01/2014
01/01/2015 01/01/2016
01/01/2017



DATA ANALYSIS



FAMILY : ALL

Sales vs Onpromotion (**0.45**)

Ada korelasi sedang (positif)

Sales vs Dcoilwtico (**-0.05**)

Tidak ada korelasi (mendekati 0)

CORRELATION

3 Variable

Sales
onpromotion
dcoilwtico

FAMILY : BAKERY

Sales vs Onpromotion (**0.05**)

Tidak ada korelasi (mendekati 0) → delete

Sales vs Dcoilwtico (**-0.13**)

Tidak ada korelasi (mendekati 0) → delete

DATA PREPROCESSING

DATA DEKOMPOSI

memecah data time series
menjadi komponen dasarnya
DATE - MONTH - YEAR

DATA FILTER

Memilih subset data yang
relevan berdasarkan kriteria
tertentu. (BAKERY)
DROP VARIABLE

MISSING VALUE

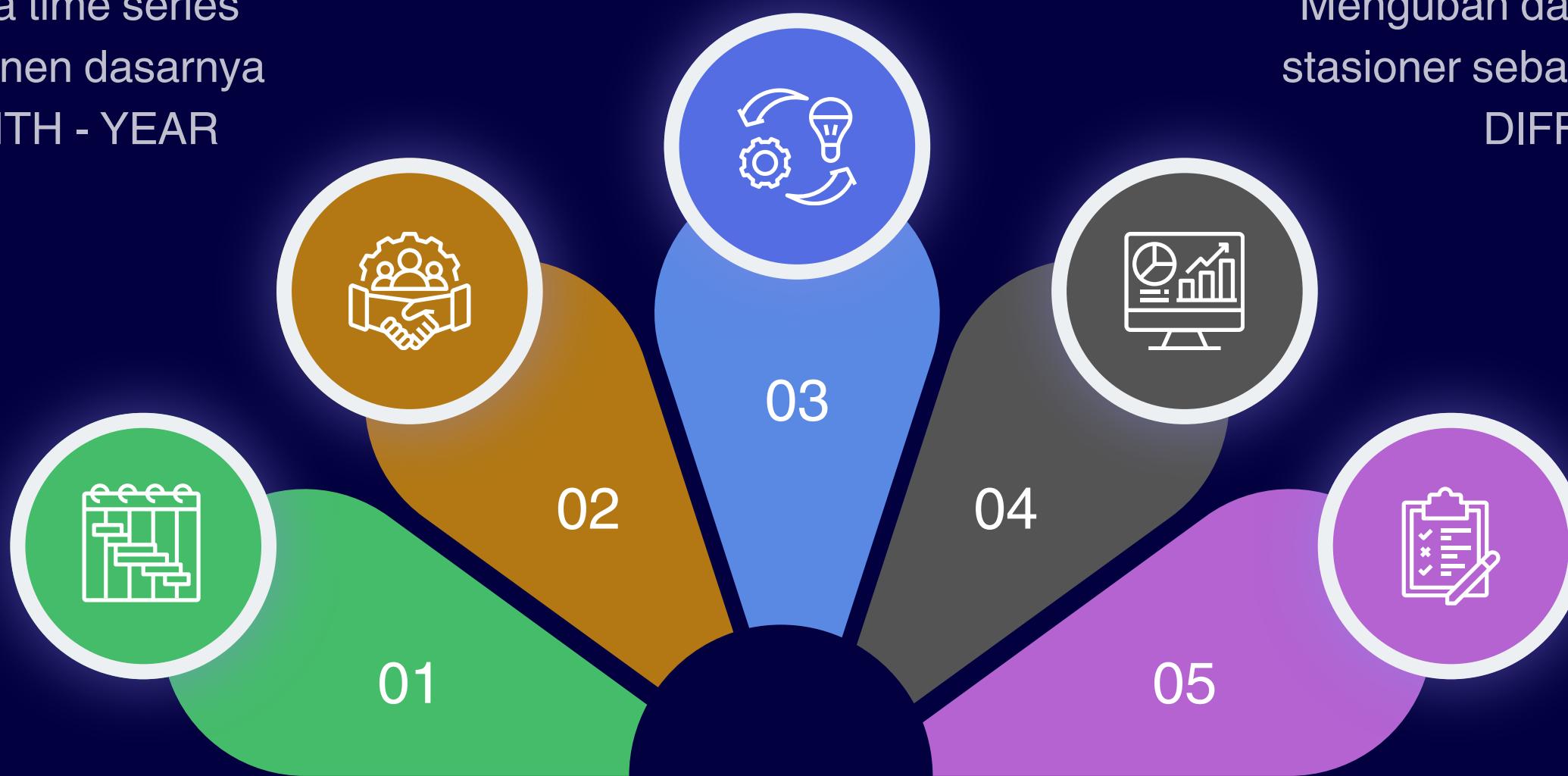
Mengidentifikasi dan menangani
missing values (data kosong)
Imputasi / Penghapusan data

STATIONARITY DATA

Mengubah data time series menjadi
stasioner sebagai prasyarat modeling
DIFFERENCING

DATA SPLITTER

Membagi dataset dengan
tujuan Training 80% dan
Testing 20%



MODELING AUTO ARIMA - SARIMA

DIFFERENCING (D = 1)

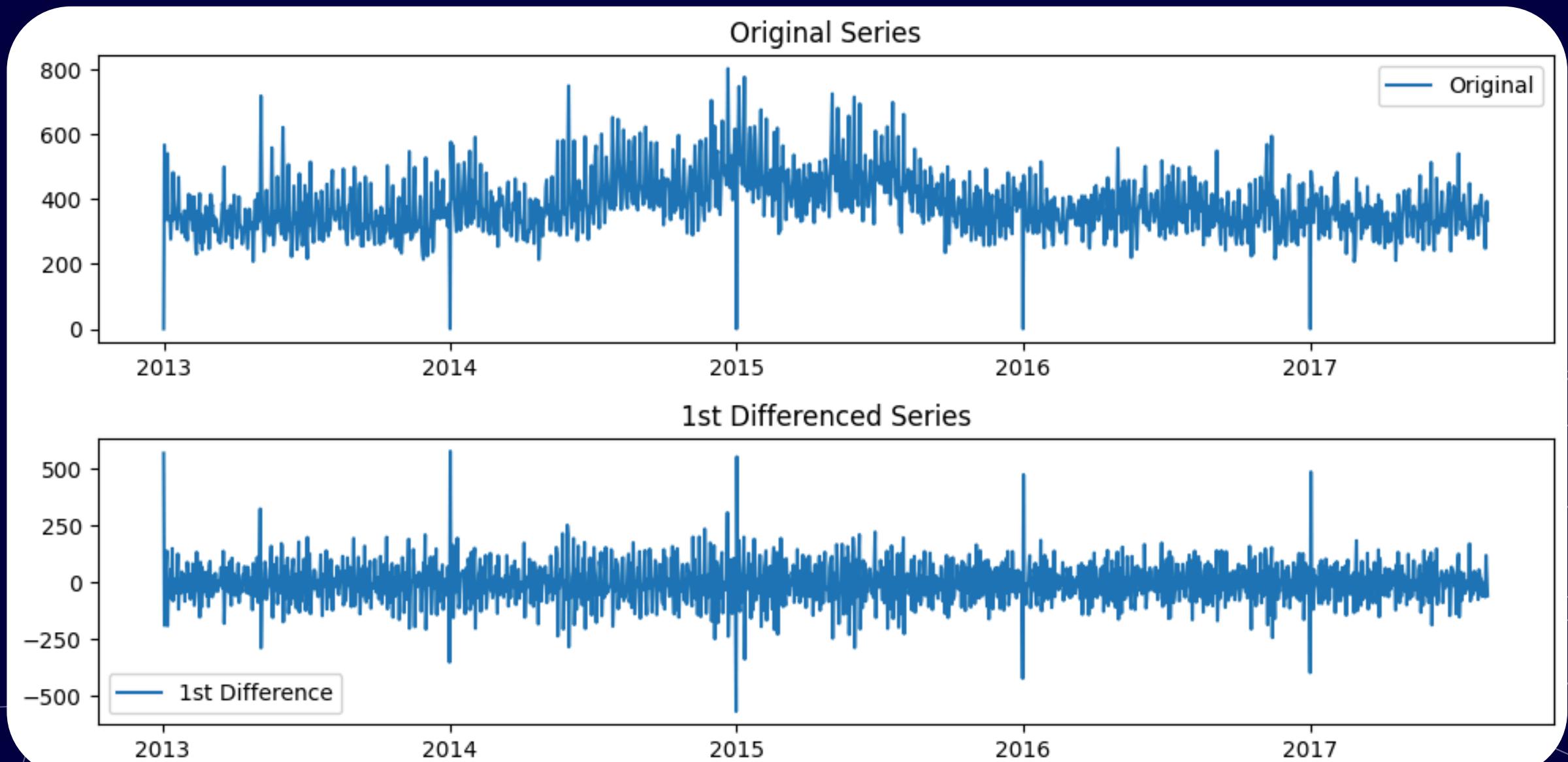
Jika data stasioner sejak awal, maka D = 0

Jika satu kali differencing cukup, maka D = 1

DATA SPLITTER

Data Training 80% = 1350

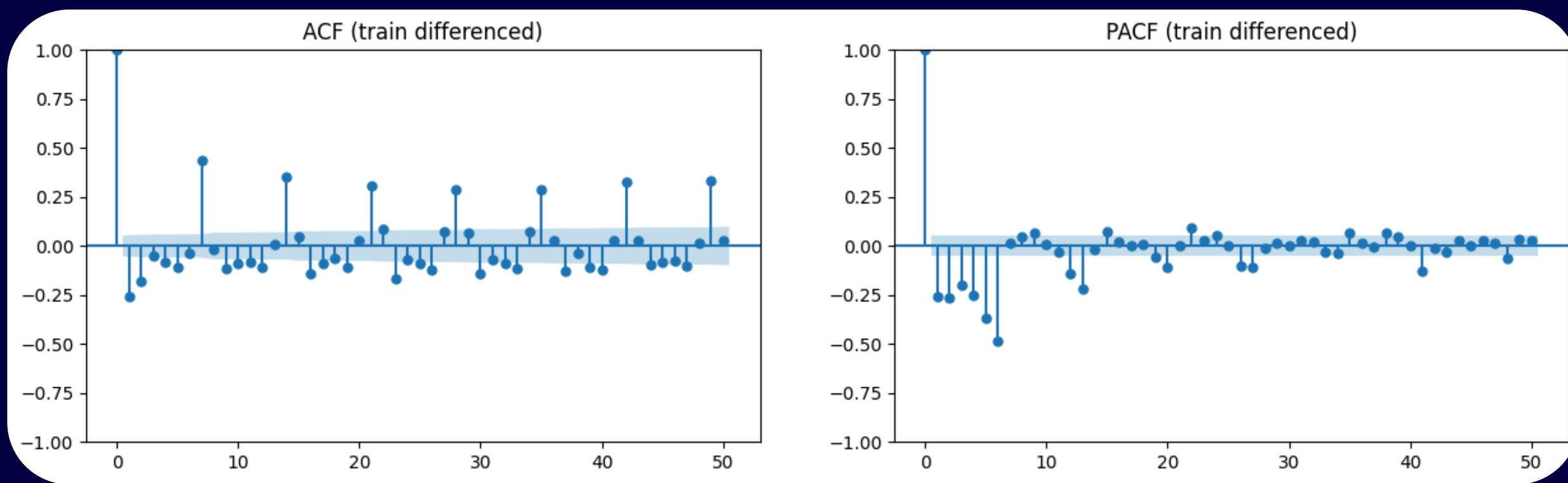
Data Testing 20% = 338



MODELING AUTO ARIMA - SARIMA

DIFFERENCING (D = 1)

ACF AND PACF PLOT



MODEL SARIMA

P = range (0, 3)

D = [0, 1]

Q = range (0, 3)

S = 7

BEST MODEL SARIMA

(p , d , q) (P , D , Q) m

(2 , 1 , 2) (0 , 1 , 2) 7

AIC = 144761

AUTO REGRESSIVE (P)

Berdasarkan tabel PACF

p = range (0, 6)

MOVING AVERAGE (Q)

Berdasarkan tabel ACF

q = range (0, 6)

BEST MODEL ARIMA

(p , d , q) \rightarrow (5, 1, 5)

AIC = 14903

MODEL EVALUATION

ARIMA

MODEL

$p, d, q = 5, 1, 5$

RMSE

56.69

MAE

41.57

SMAPE

12.22%

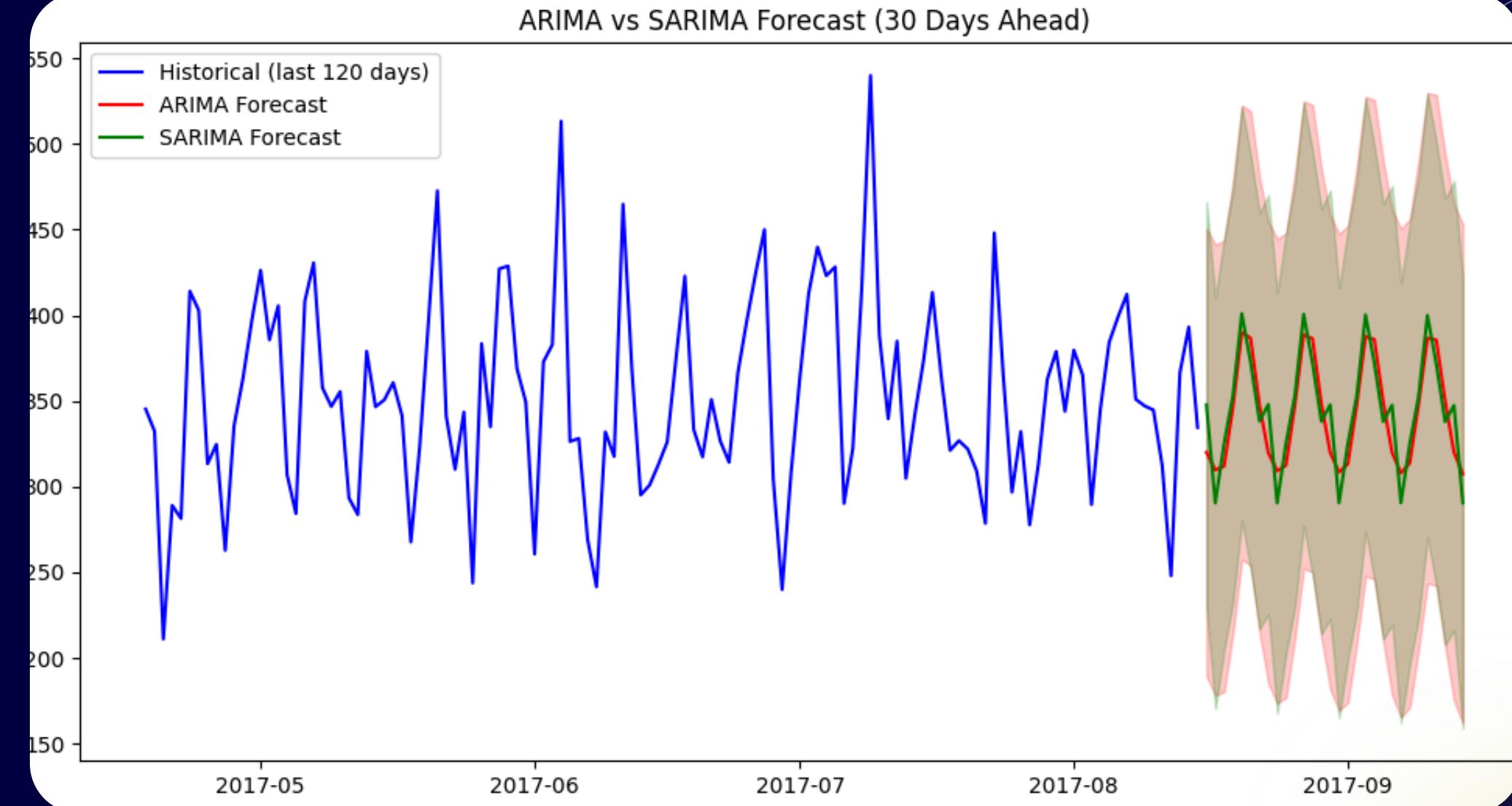
SARIMA

$(p,d,q) (P,D,Q)m$
 $(2,1,2) (0,1,2) 7$

56.56

40.28

11.91%



FORECASTING

	ARIMA	SARIMA
16/8/2017	321.51	376.33
17/8/2017	311.38	332.34
18/8/2017	315.71	351.47
19/8/2017	354.97	297.16
20/8/2017	394.32	322.29

MODELING DEEP LEARNING - LSTM

DATA PRE PROCESSING

Data Training 80% = 1350

Data Testing 20% = 338

Scaler = MinMax Scaler

PARAMETER LSTM

Units : 2 x 100

Dropout : 0.2

Epoch : 200

Batch Size : 32

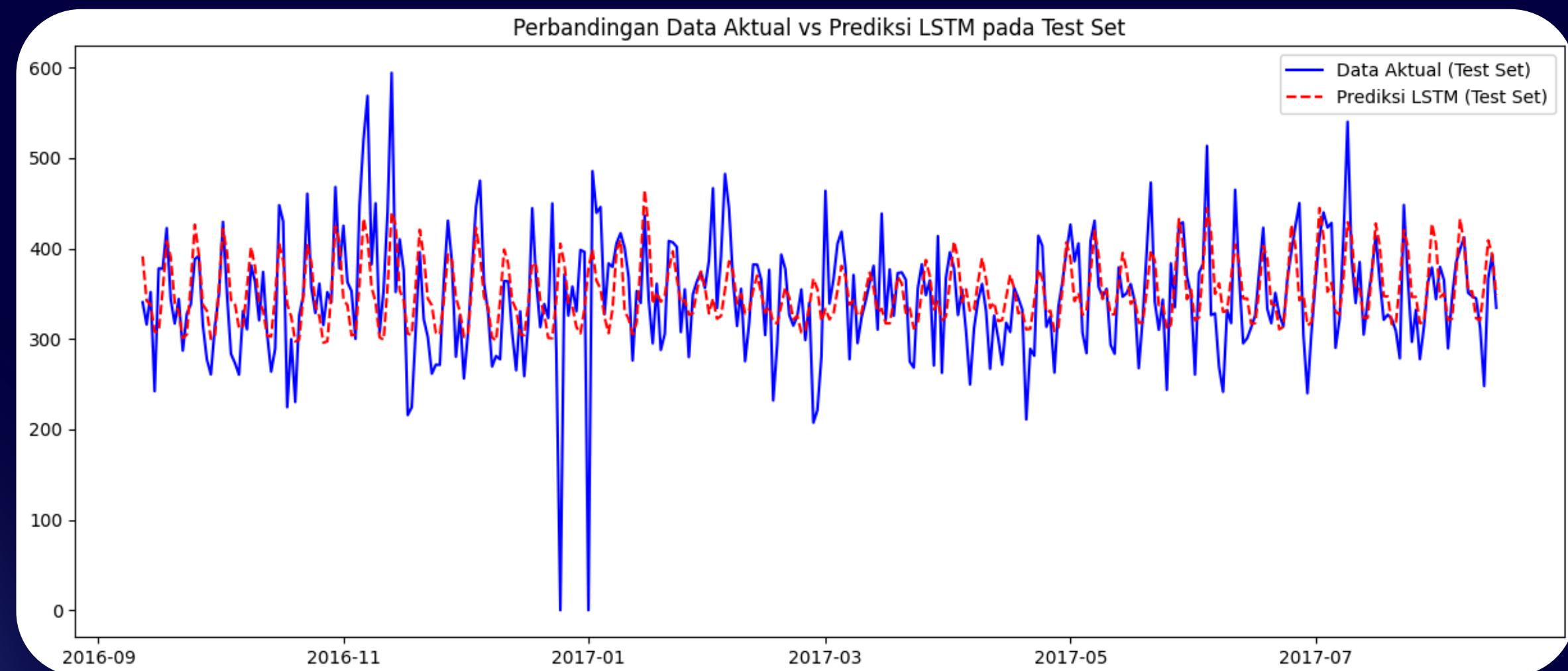
OPTIMIZER

Optimizer type : adam

loss function : MSE

callback : early stop

patience : 20



MODEL EVALUATION

	ARIMA	SARIMA	LSTM
RMSE	56.69	56.56	56.48
MAE	41.57	40.28	39.29
SMAPE	12.22%	11.91%	11.88%

FORECASTING

	ARIMA	SARIMA	LSTM
16/8/2017	321.51	376.33	345.27
17/8/2017	311.38	332.34	318.74
18/8/2017	315.71	351.47	316.29
19/8/2017	354.97	297.16	352.42
20/8/2017	394.32	322.29	403.34

CONCLUSION AND FUTURE IMPROVEMENT



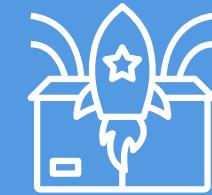
WEEKLY SEASONAL

Perlu melakukan persiapan ketika weekend karena ada kenaikan permintaan customer

perbedaan error model ARIMA, SARIMA dan **LSTM** tidak signifikan. sehingga bisa menggunakan pertimbangan teknis lainnya



BEST MODEL



MAINTAIN MODEL

Semakin Banyak Data baru maka model perlu di maintain sehingga akurasinya tetap terjaga



OTHER VARIABLE

explore variable lain yang memiliki korelasi ke sales Bakery Sehingga dapat meningkatkan akurasi modeling

10 October 2025

THANK YOU FOR YOUR ATTENTION

Taufiq Rizan

Andre Pranata

Adam Husain

