

# CLIMATE CHANGE ANALYSIS

---

CSS342 DATA SCIENCE AND DATA ENGINEERING





# MEMBERS



66090500424  
Amy Suganya Tripp



66090500430  
Thiraphat Panthong



66090500440  
Rachatapat Boonmahan



66090500449  
Nutthawara Sasrimuang

# CONTENTS

---

01

INTRODUCTION

02

DATA COLLECTION

03

DATA PIPELINE

04

DATA CLEANING AND PREPROCESSING

05

EXPLORATORY DATA ANALYSIS

06

PREDICTIVE MODELING

02

---



# INTRODUCTION



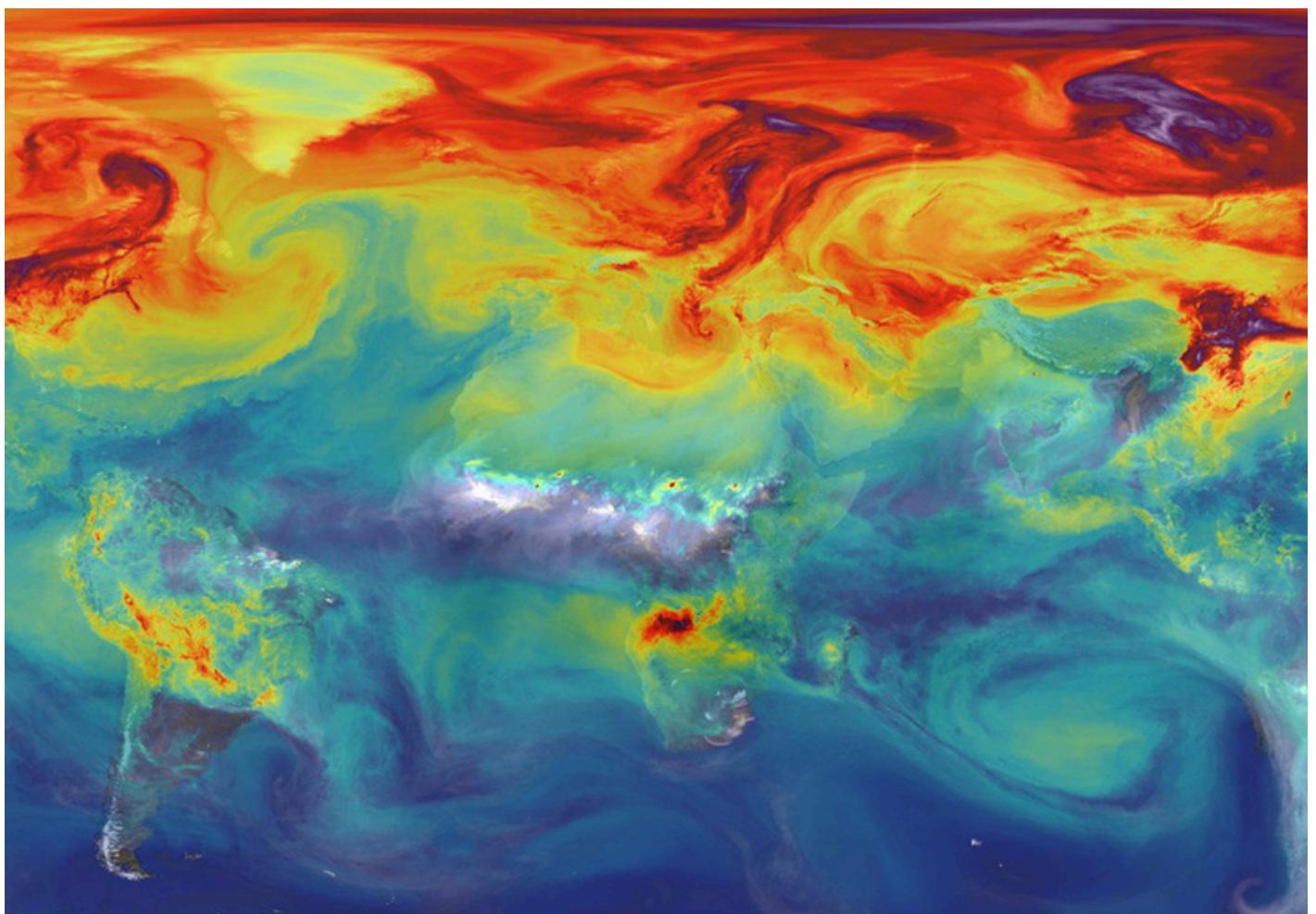


# INTRODUCTION

## BACKGROUND

---

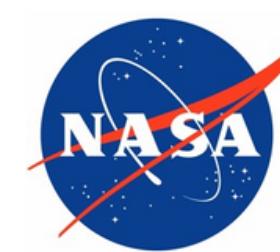
การพยากรณ์สภาพอากาศอย่างแม่นยำเป็นพื้นฐานสำคัญต่อการตัดสินใจเชิงปฏิบัติการในหลายภาคส่วน เช่น การจัดการพลังงาน การวางแผนการเกษตร และการลดความเสี่ยงจากภัยพิบัติ โดยเฉพาะตัวแปรอุณหภูมิที่ระดับ 2 เมตร (T2M) ซึ่งส่งผลโดยตรงต่อกิจกรรมของมนุษย์และโครงสร้างพื้นฐาน อย่างไรก็ตาม การเปลี่ยนแปลงของ T2M มีความซับซ้อนและมีความสัมพันธ์แบบ non-linear กับตัวแปรบรรยายอากาศอื่น ทำให้ Machine Learning เป็นหนึ่งทางเลือกที่เหมาะสม



# DATA PIPELINE

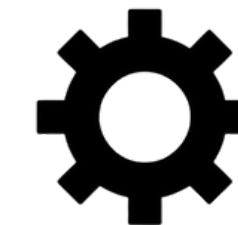


# DATA PIPELINE



NASA  
POWER API

20 params  
per batch



Extract Raw Batch  
Data to Cloud  
Lakehouse

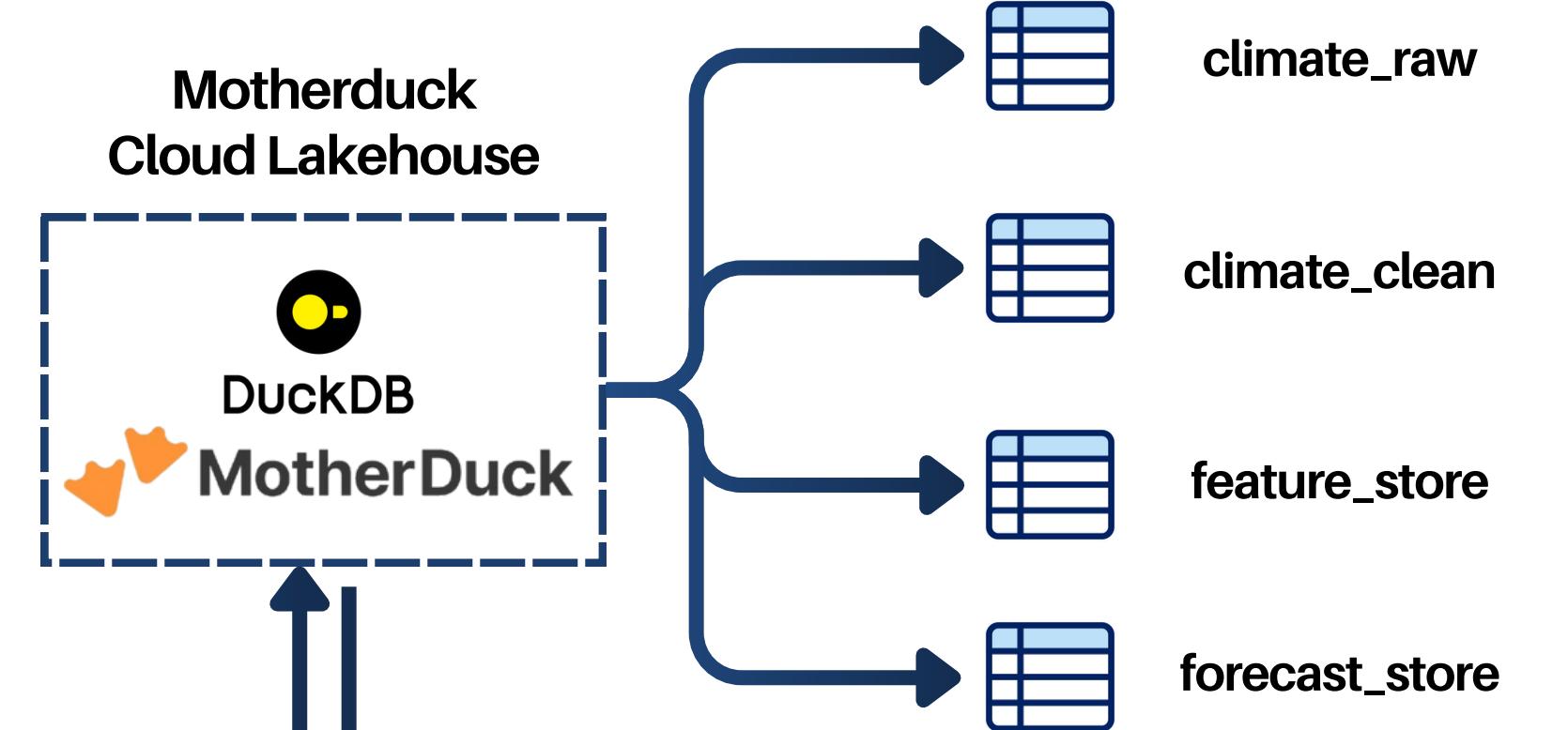


Data Cleaning  
Data Transformation  
Feature Engineering  
Feature Importance

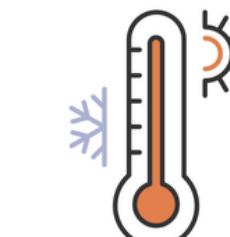
Data Preparation

Motherduck  
Cloud Lakehouse

DuckDB  
**MotherDuck**



Predict



7-Day Forecast  
Output

climate\_raw

climate\_clean

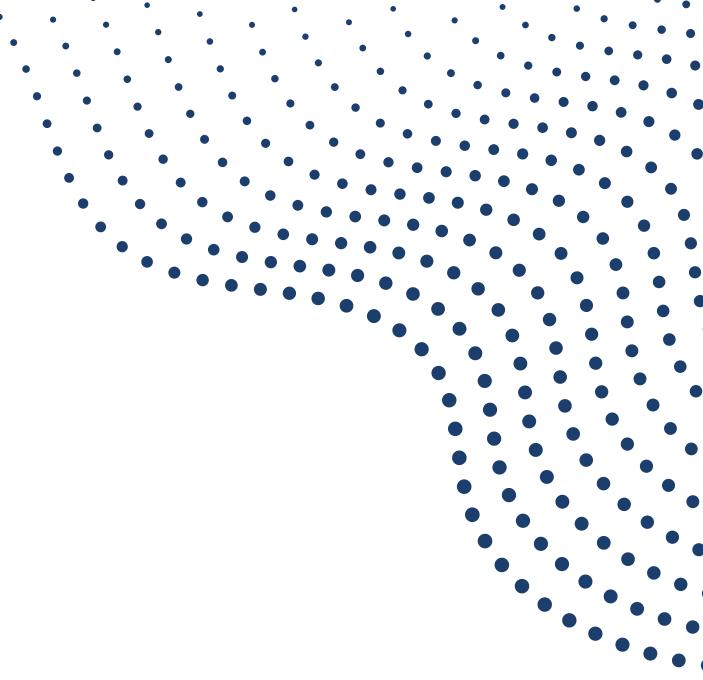
feature\_store

forecast\_store

Store  
Predictions

# DATA COLLECTION





# DATA COLLECTION

---

DATA SOURCE



NASA POWER API

(3 Days delay)

Meteorological data

- อุณหภูมิ T2M
- ความชื้น
- ค่ารังสีดวงอาทิตย์
- ความกดอากาศ
- ความเร็วลม
- ค่าการระเหยและพลังงานการคายระเหย





# DATA CLEANING AND PREPROCESSING





# DATA CLEANING AND PREPROCESSING

All Parameters Daily Data

Raw API

16K rows

Delete Delay Data

3 Day Delay Data

Missing Values

Drop column

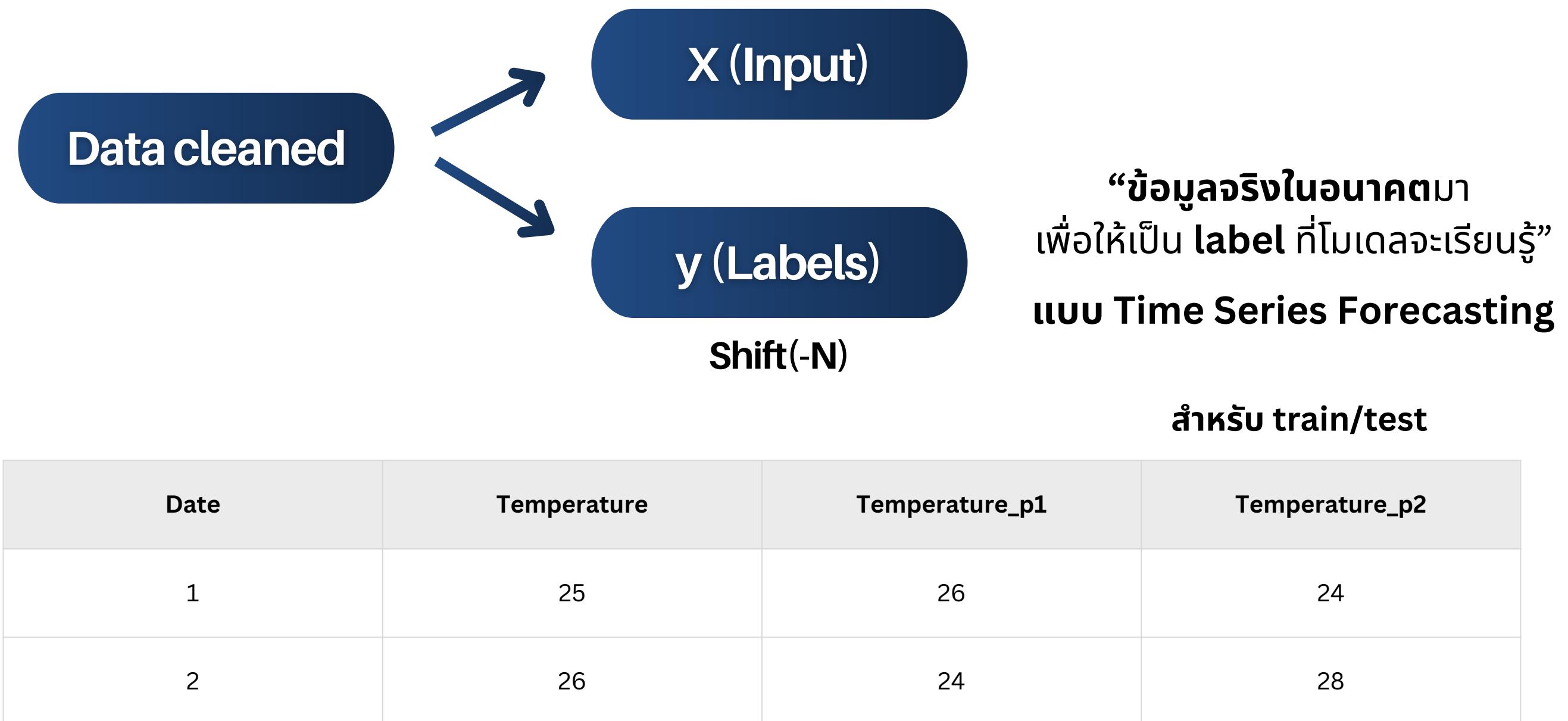
mean

-999 or NaN





# DATA CLEANING AND PREPROCESSING



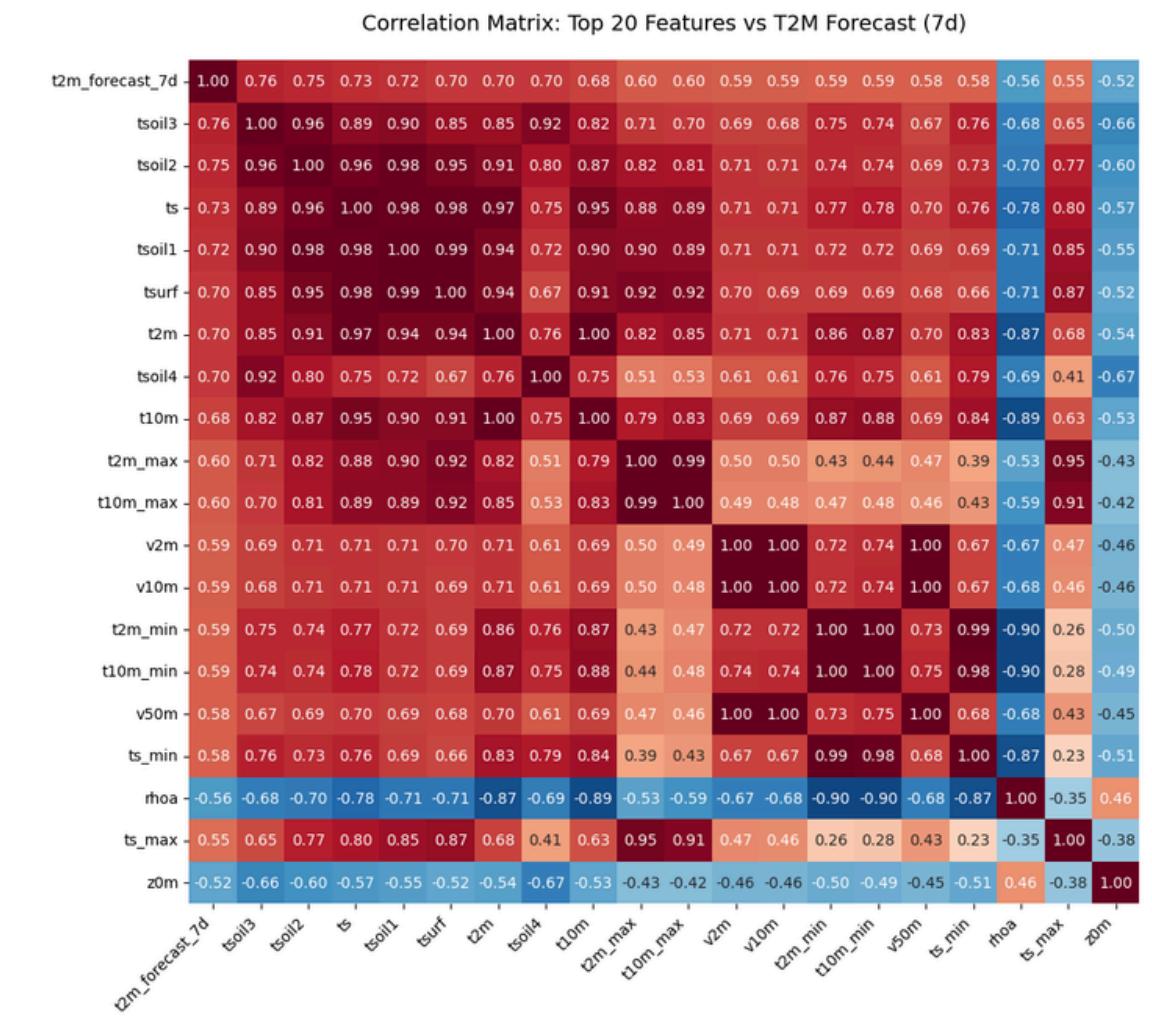
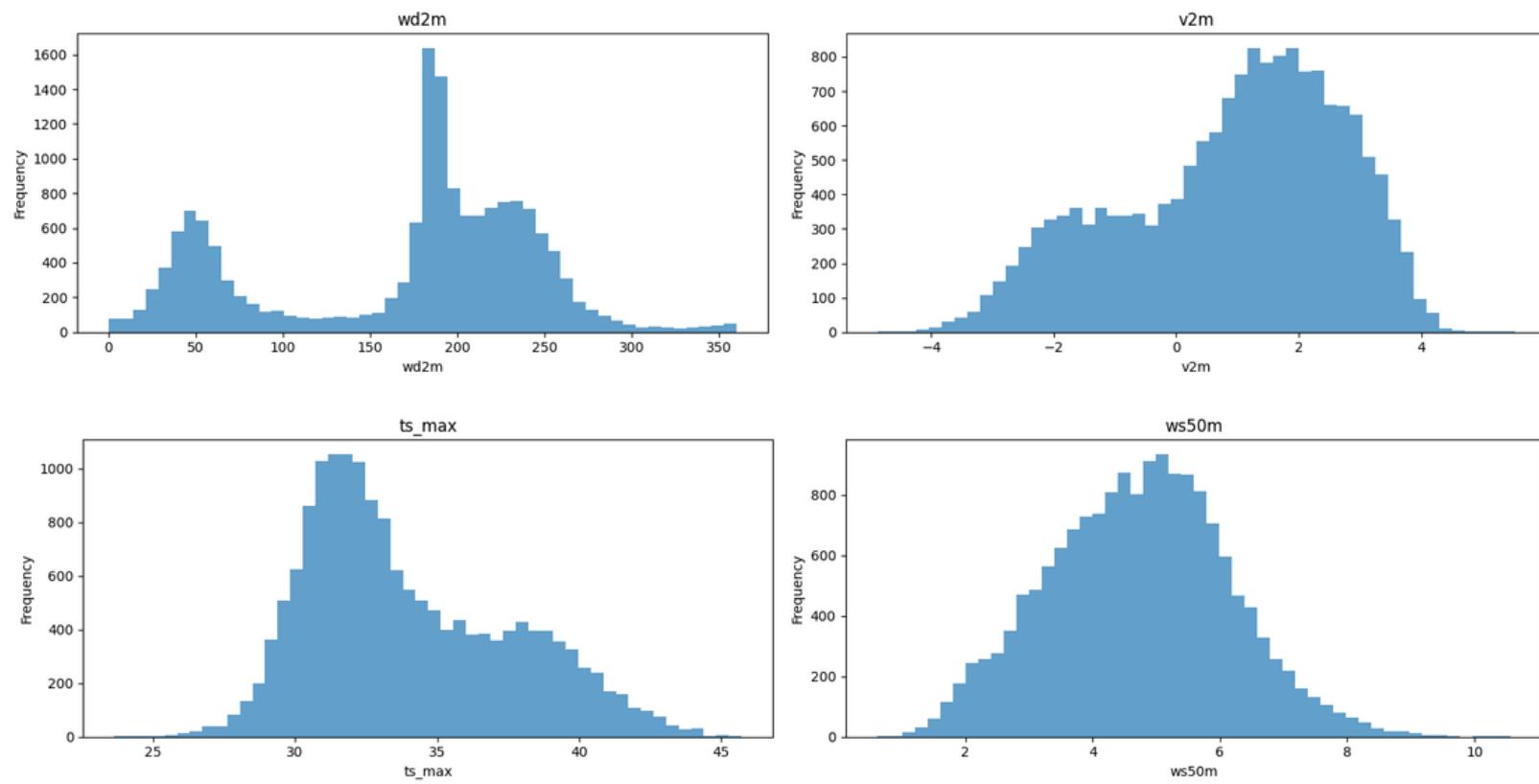
# **EXPLORATORY DATA ANALYSIS**



# EXPLORATORY DATA ANALYSIS

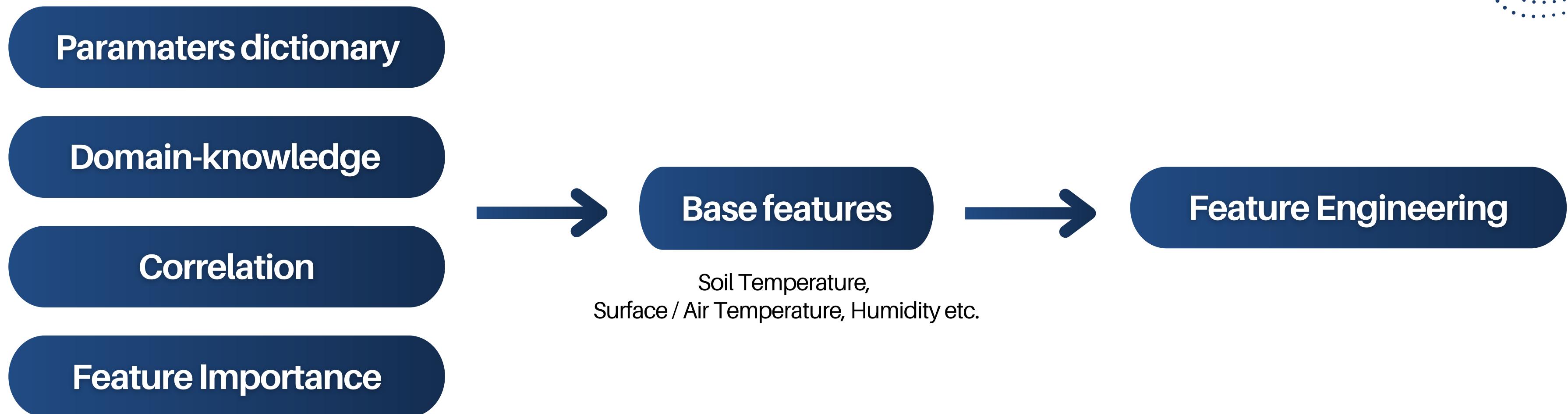
Heatmap

Data Distribution





# FEATURE SELECTION & FEATURE ENGINEERING

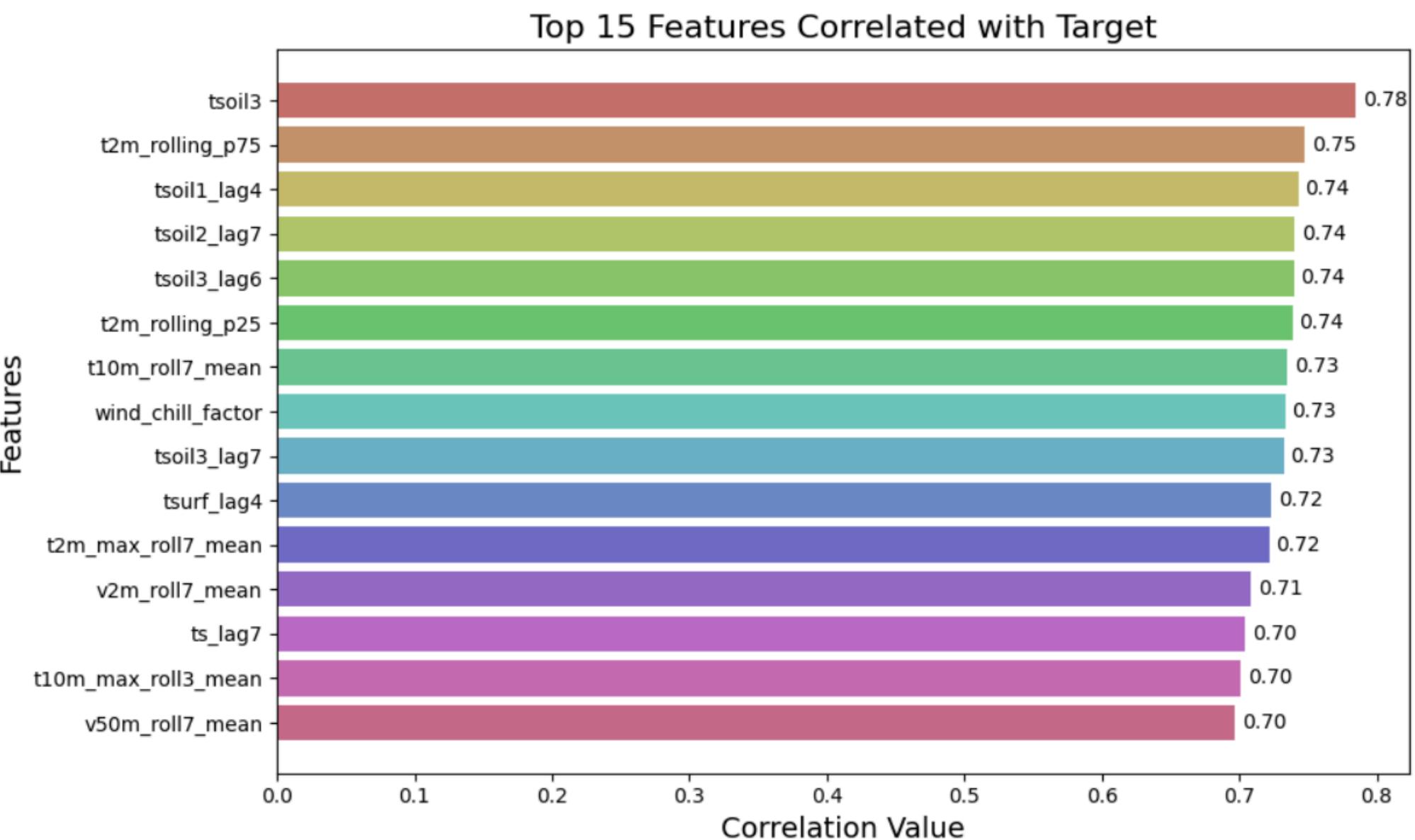




# FEATURE SELECTION

## Feature Importance ( LightGBM )

กราฟนี้แสดงอันดับความสำคัญของฟีเจอร์ 15 ตัวแรกที่มีผลต่อการพยากรณ์อุณหภูมิ T2M โดย LightGBM ซึ่งช่วยระบุว่าปัจจัยด้านกดุกาล ค่าการระเหย



15





# FEATURE ENGINEERING

---

## Base feature engineering

Seasonal Features

วัฏจักร

month\_sin, month\_cos, doy\_sin, doy\_cos

Lag Features 7 Days

ค่าที่ผ่านมา

v50m\_lag1-7, tsoil1-7, gwettop\_lag1-7, tsurf\_lag1-7

Rolling Features 3, 7 Days

สภาพสะสม

rain\_roll7\_max,  
tsurf\_roll3\_mean, tsurf\_roll3\_std, tsurf\_roll7\_mean, tsurf\_ roll7\_std, tsoil1\_roll3\_mean





# FEATURE ENGINEERING

---

## More feature engineering

Temperature  
Volatility Features

ความผันผวนของอุณหภูมิ

t2m\_volatility\_3d, t2m\_volatility\_7d, t10m\_volatility\_3d

Weather Pattern  
Features

t2m\_anomaly, t2m\_monthly\_avg, t2m\_vs\_seasonal\_avg

Cross-variable  
Interactions

temp\_gradient\_surface, temp\_gradient\_soil, temp\_humidity\_ratio

Gradient = ความเร็วใน  
การเปลี่ยนของตัวแปร

Interactions = ปฏิกิริยา  
ระหว่างตัวแปร

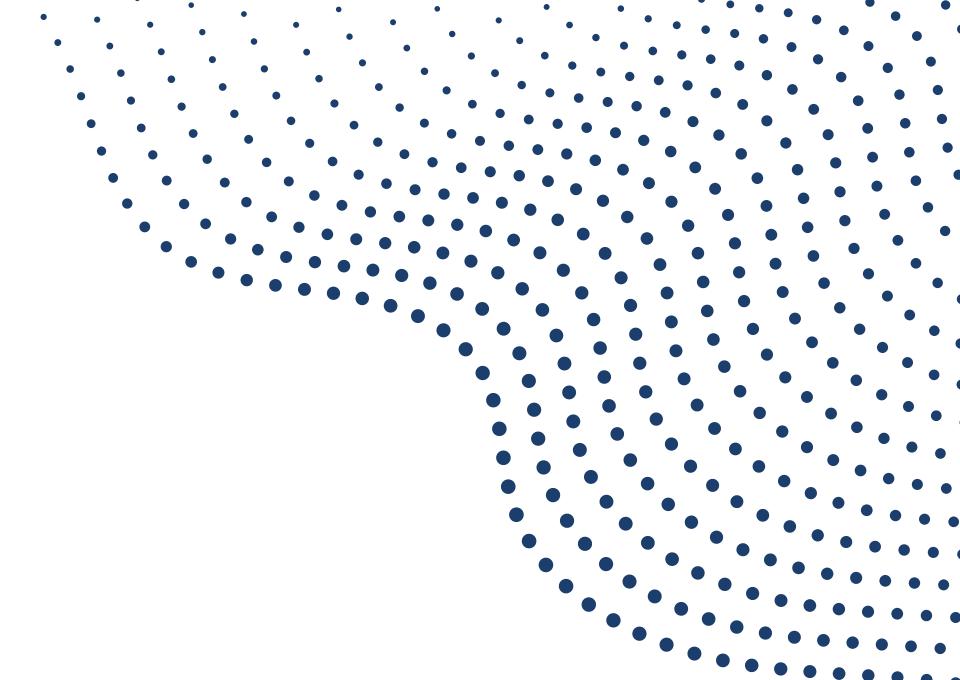
# PREDICTIVE MODELING



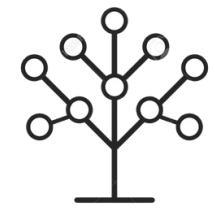


# PREDICTIVE MODELING

---



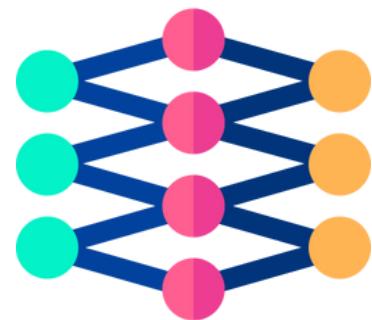
## Tree-Based Model



Random Forest

*XGBoost*

## Deep Learning Model



LSTM

## Hyperparameter Tuning



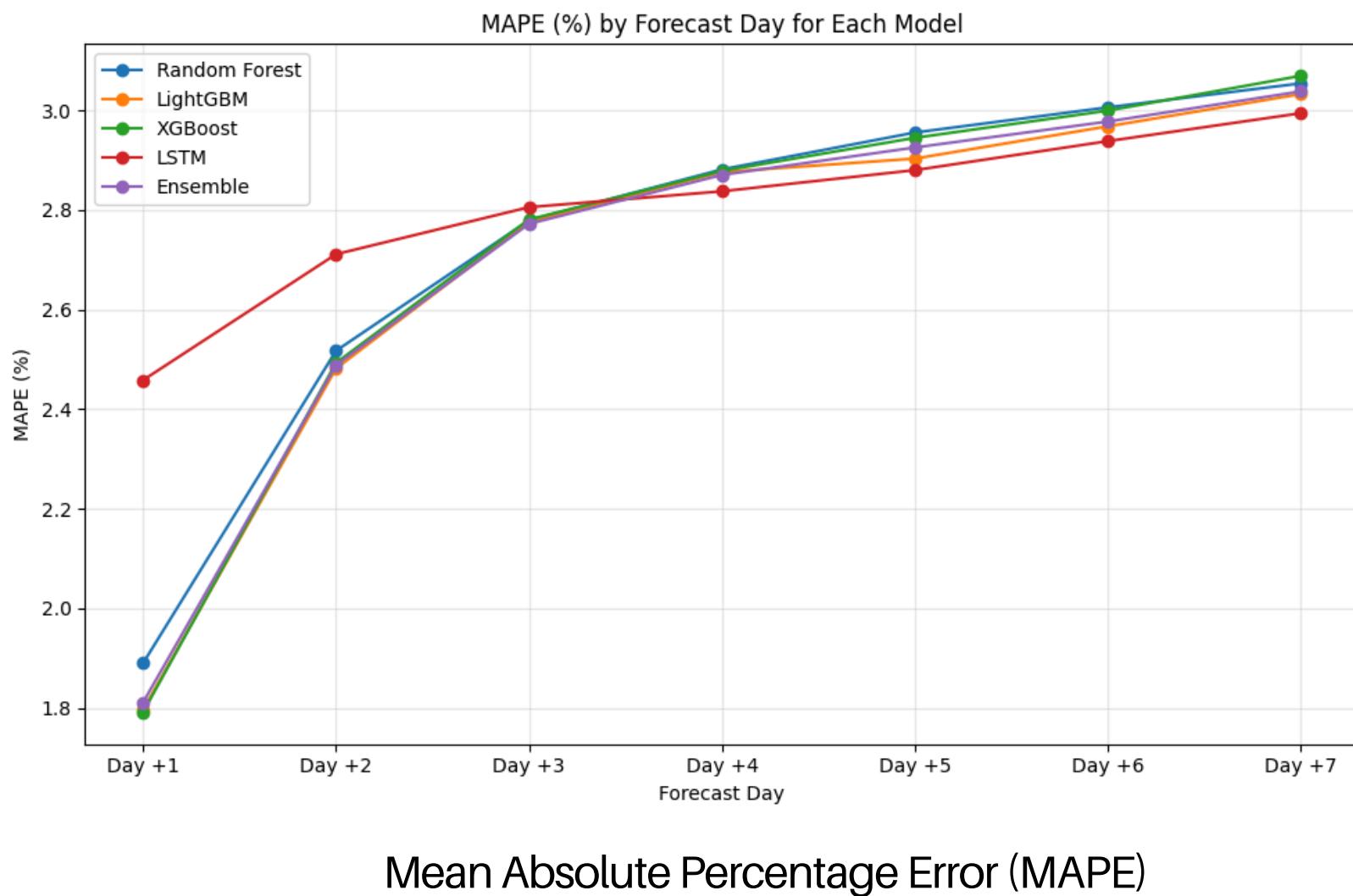
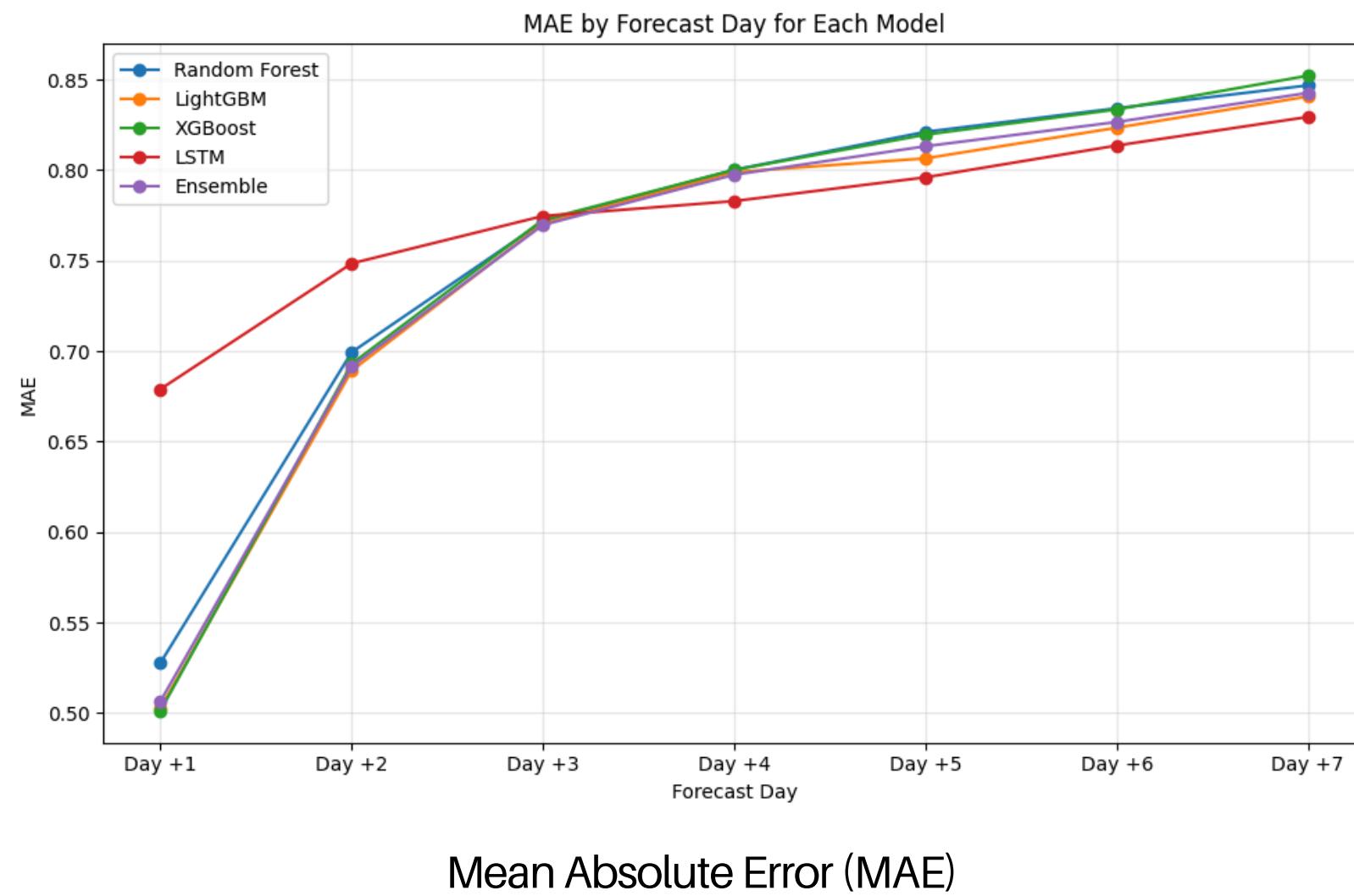
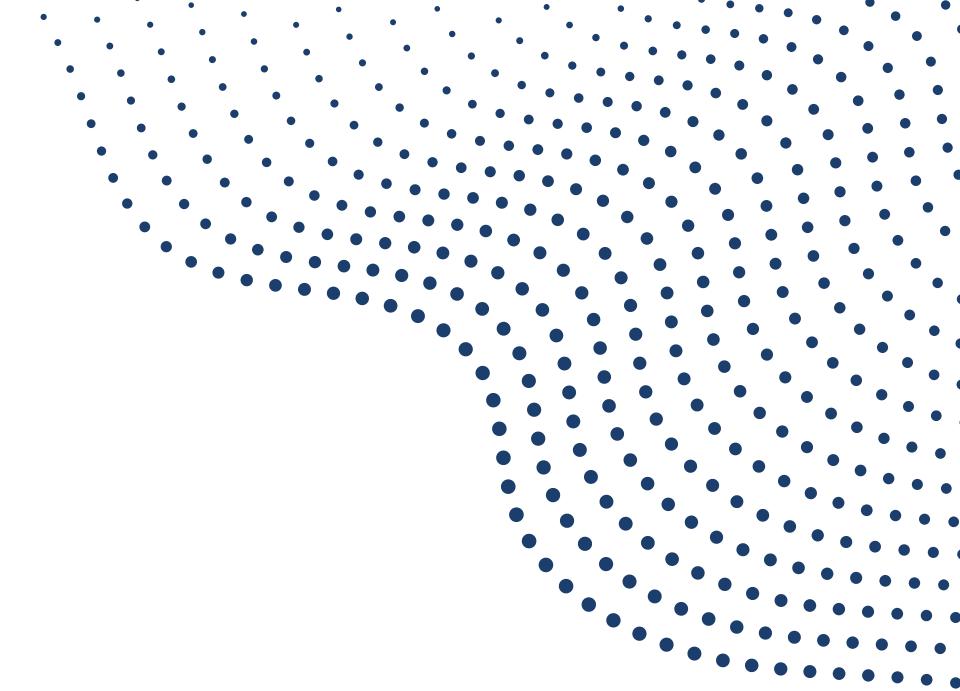
O P T U N A





# PREDICTIVE MODELING

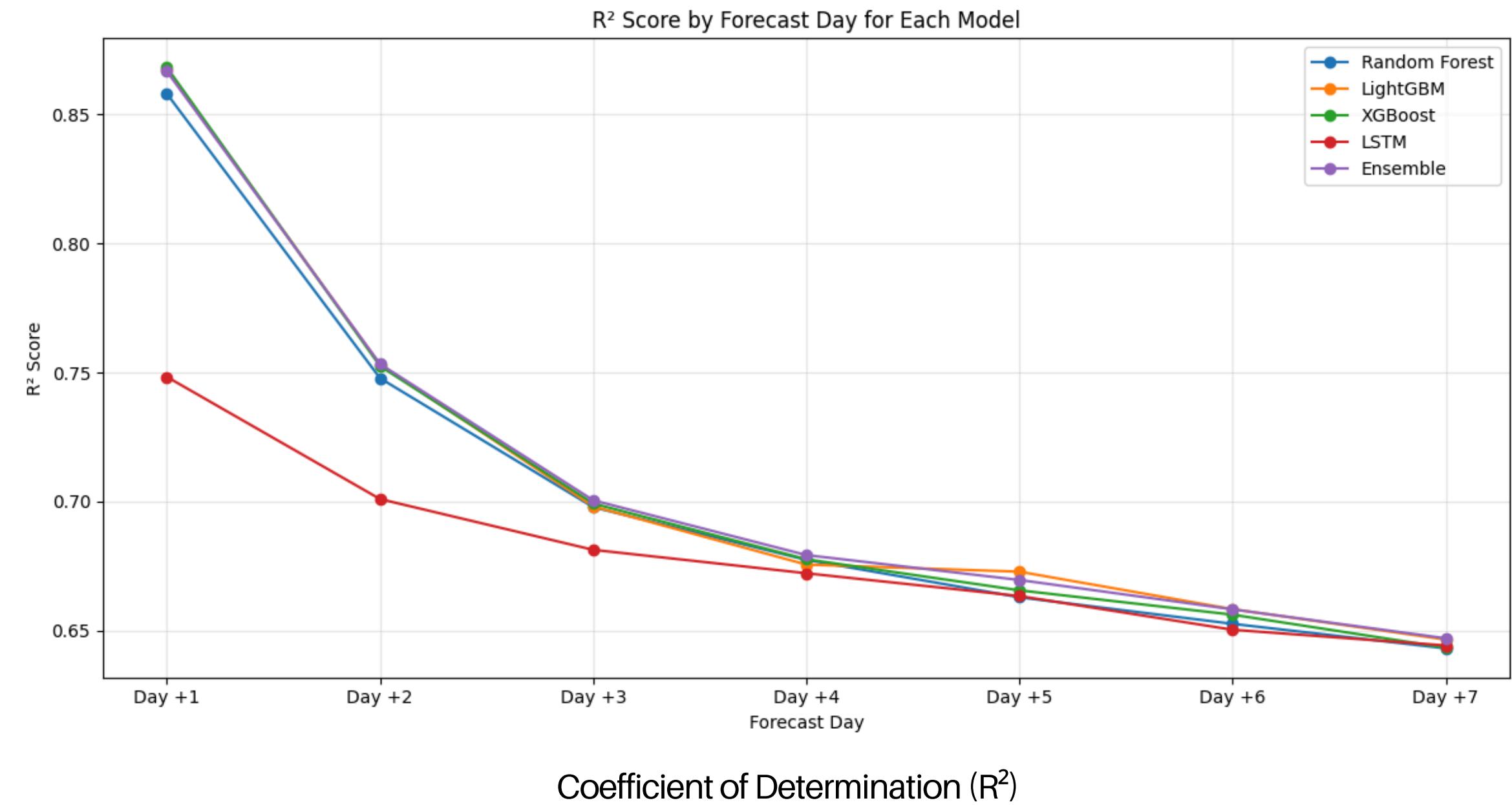
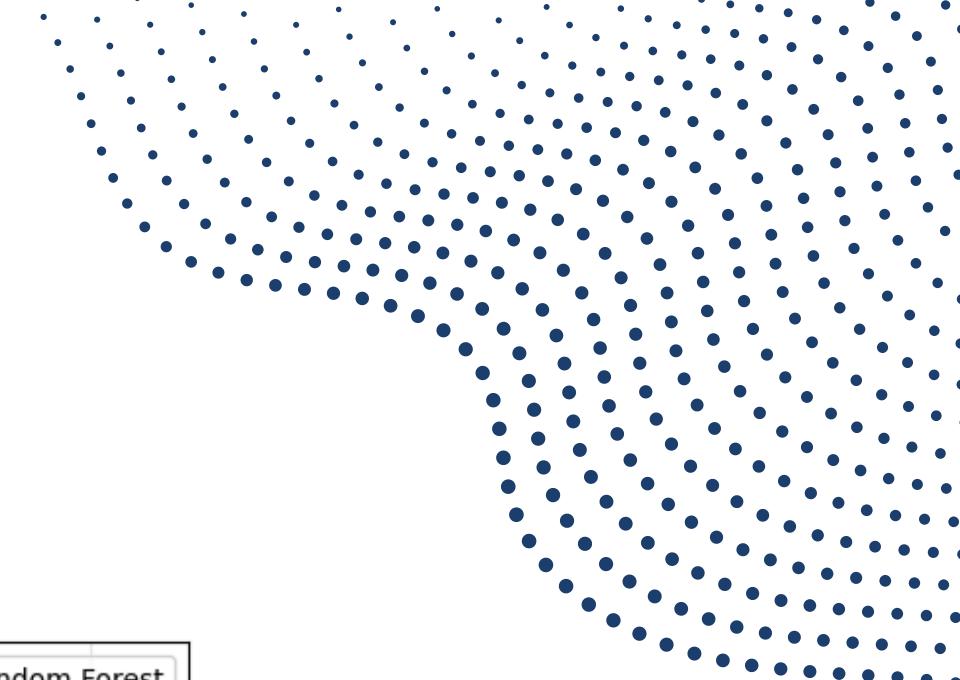
---





# PREDICTIVE MODELING

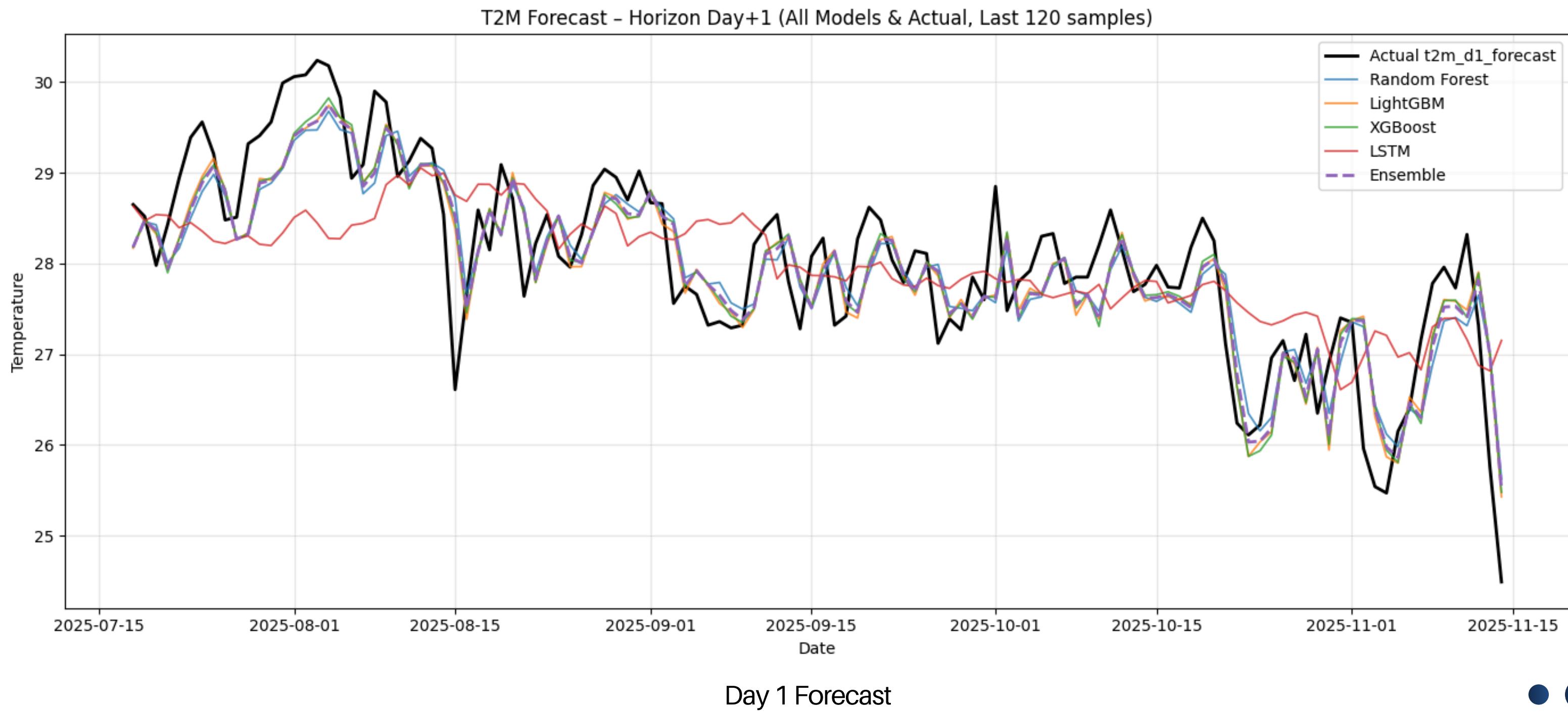
---





# PREDICTIVE MODELING

---

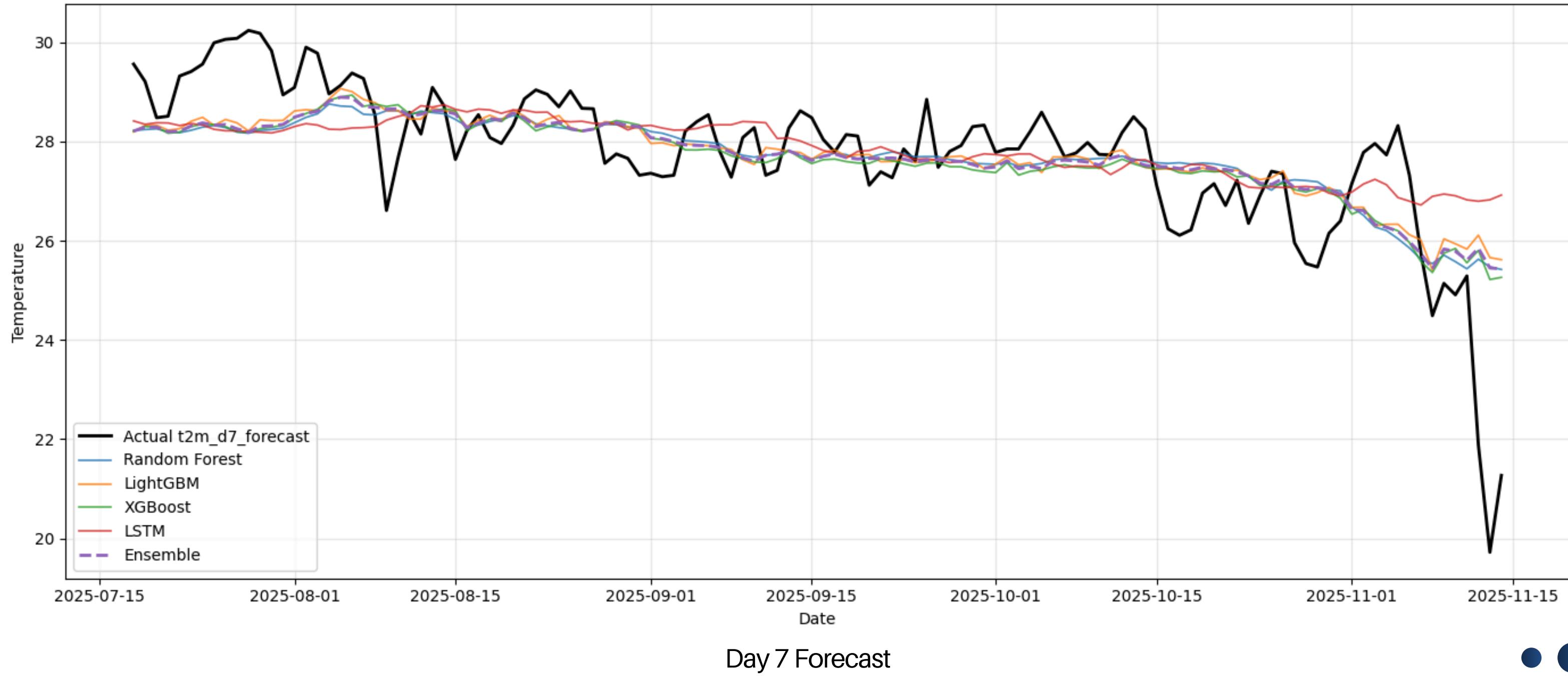




# PREDICTIVE MODELING

---

T2M Forecast - Horizon Day+7 (All Models & Actual, Last 120 samples)

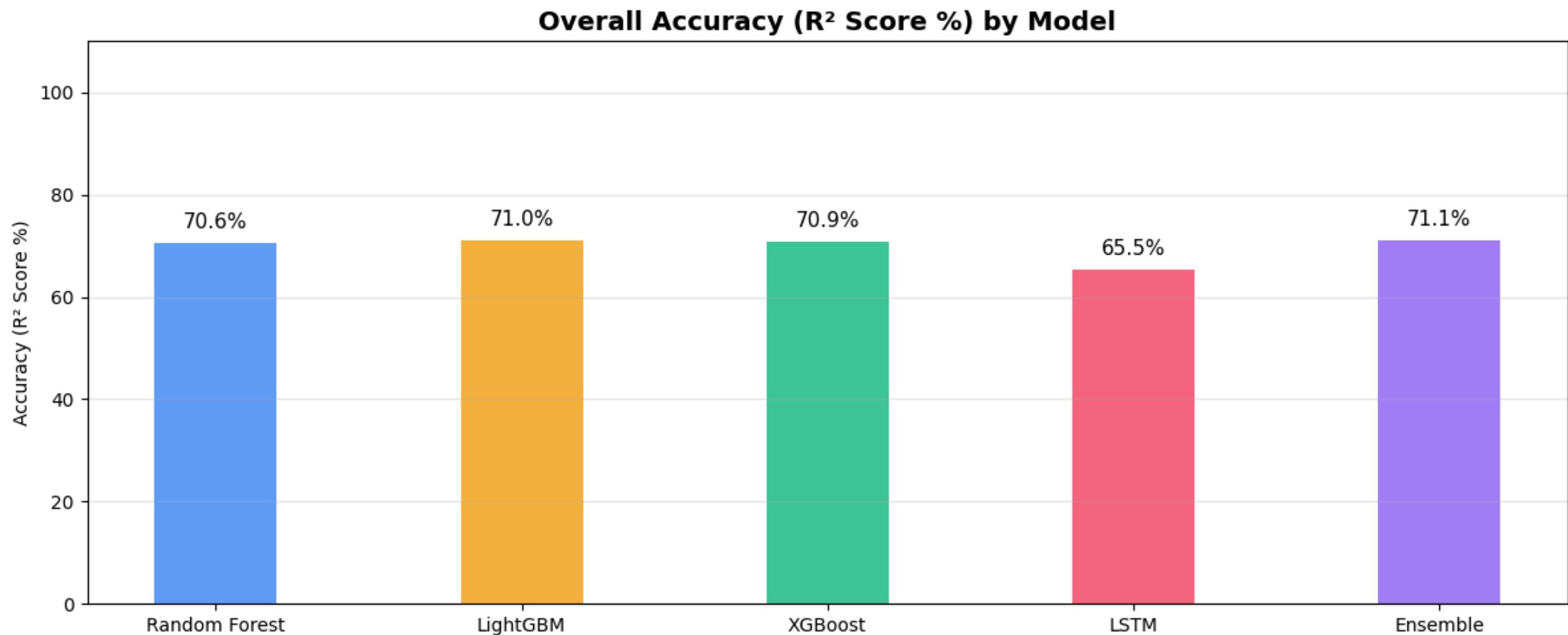




# PREDICTIVE MODELING

## RESULT

ผลจากค่า  $R^2$  แสดงว่าโมเดล Tree-based มีความแม่นยำสูงกว่า LSTM และการใช้ Ensemble ให้ผลตีที่สุดที่ราว 71.1%



# DEMO



# THANK YOU

---

