FEATURE EXPLORATION FOR BETTER PREDICTION

Presented by: Akila

TABLE OF CONTENT

Presentation Overview

Objective:

Enhance forecasting accuracy of telecom KPIs using time series modeling.

Key Steps:

- Identified gaps in raw data and missing features
- Applied advanced feature engineering (temporal, rolling, lag)
- Used Prophet for forecasting Avg. UE Number & DL PRB Utilization
- Implemented randomized train-test splitting to avoid bias
- Evaluated model with combined R^2 score

Result:

Achieved high forecast accuracy with $R^2 = 0.8489$, suitable for real-world deployment.

PROBLEM STATEMENT

- The input time-series data lacks deeper evaluation, limiting the discovery of hidden patterns that could enhance model accuracy.

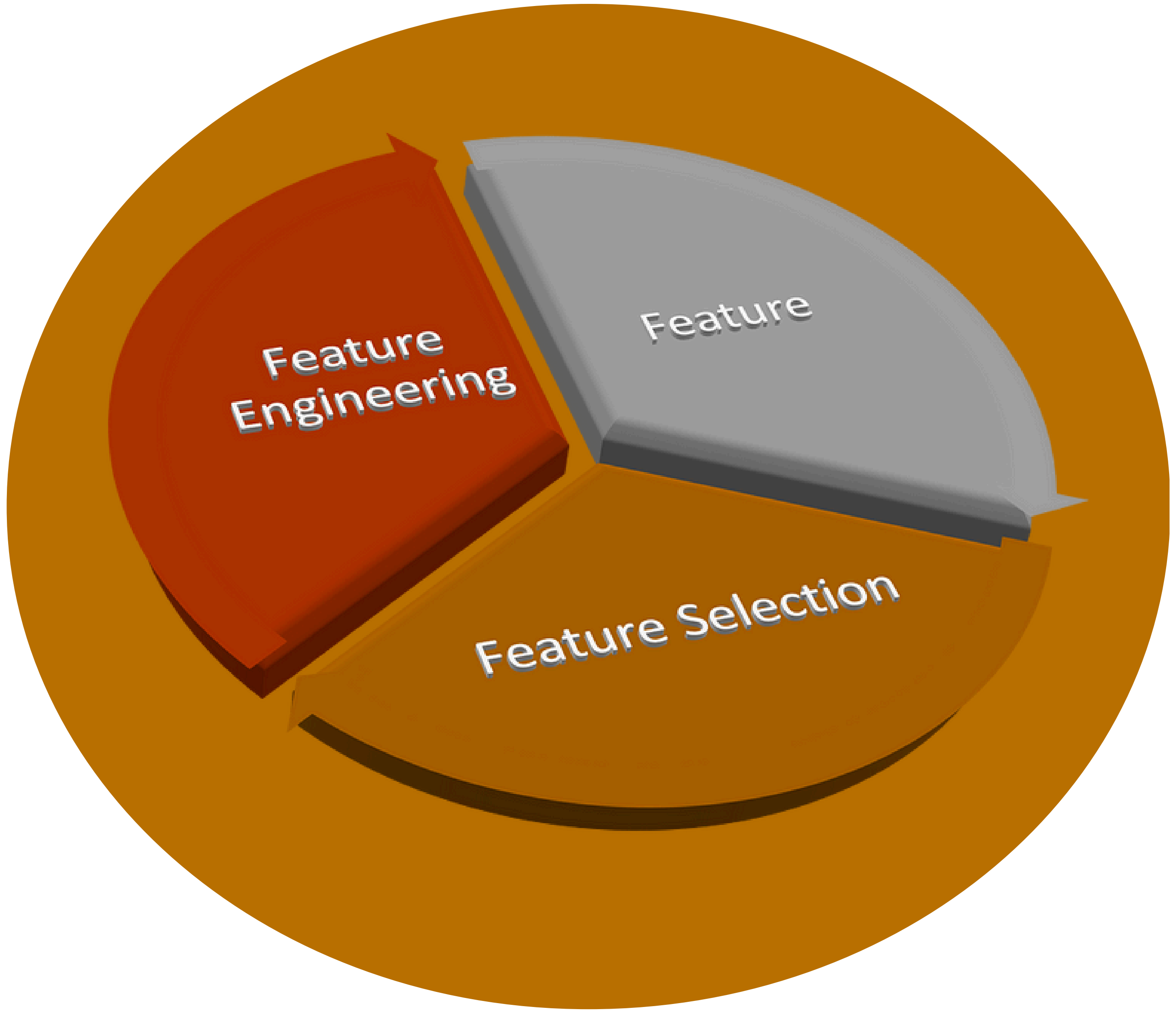
- Additional feature engineering opportunities remain unexplored, potentially affecting the predictive performance of the forecasting model.

SOLUTION PROPOSAL

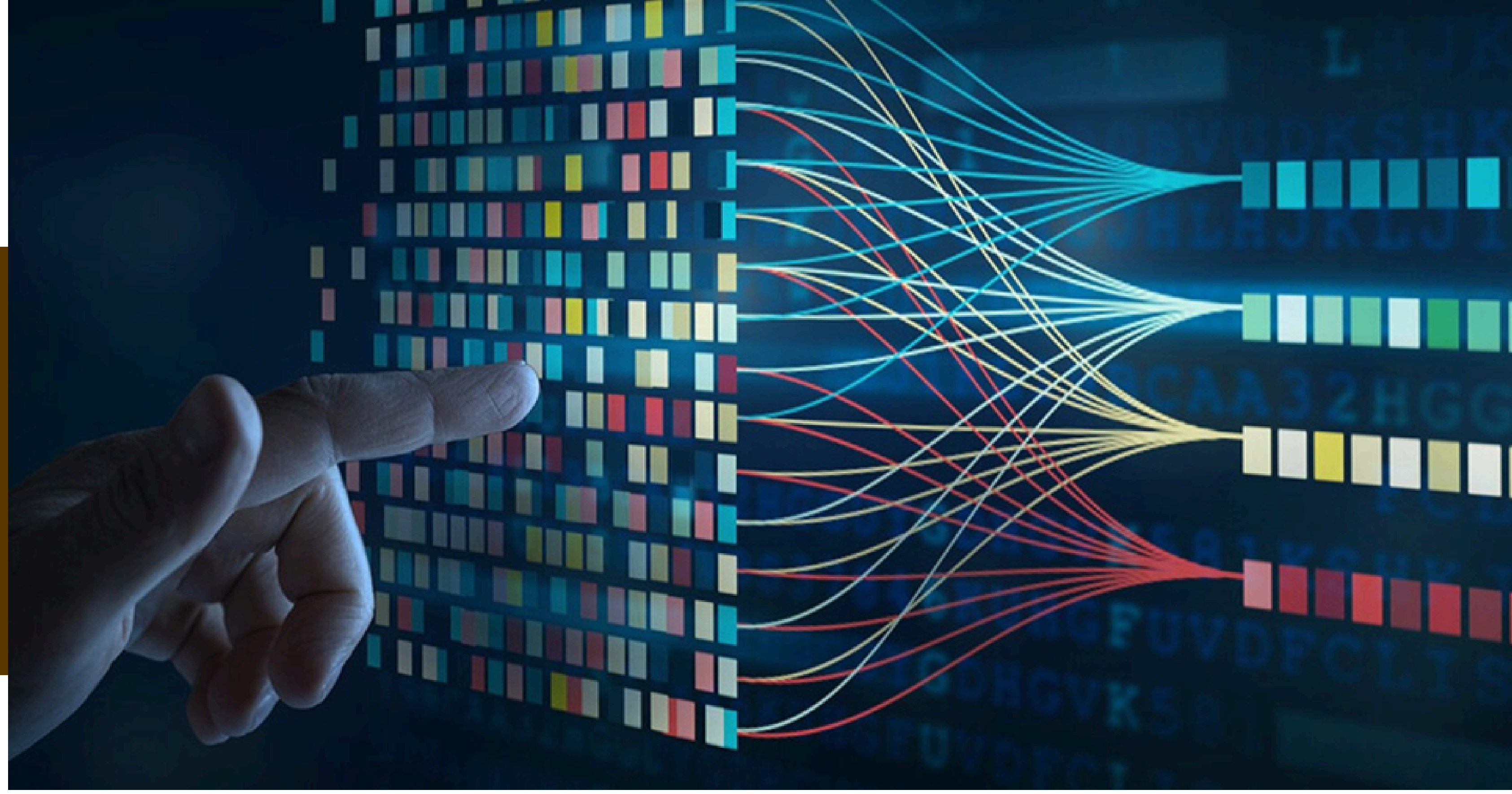
- Utilize Facebook's Prophet model tailored for time-series forecasting to handle seasonality, trends, and complex telecom KPI patterns.

- Engineer comprehensive features such as temporal indicators (hour, day, weekend), rolling statistics, lag variables, and a load ratio metric to improve model accuracy.

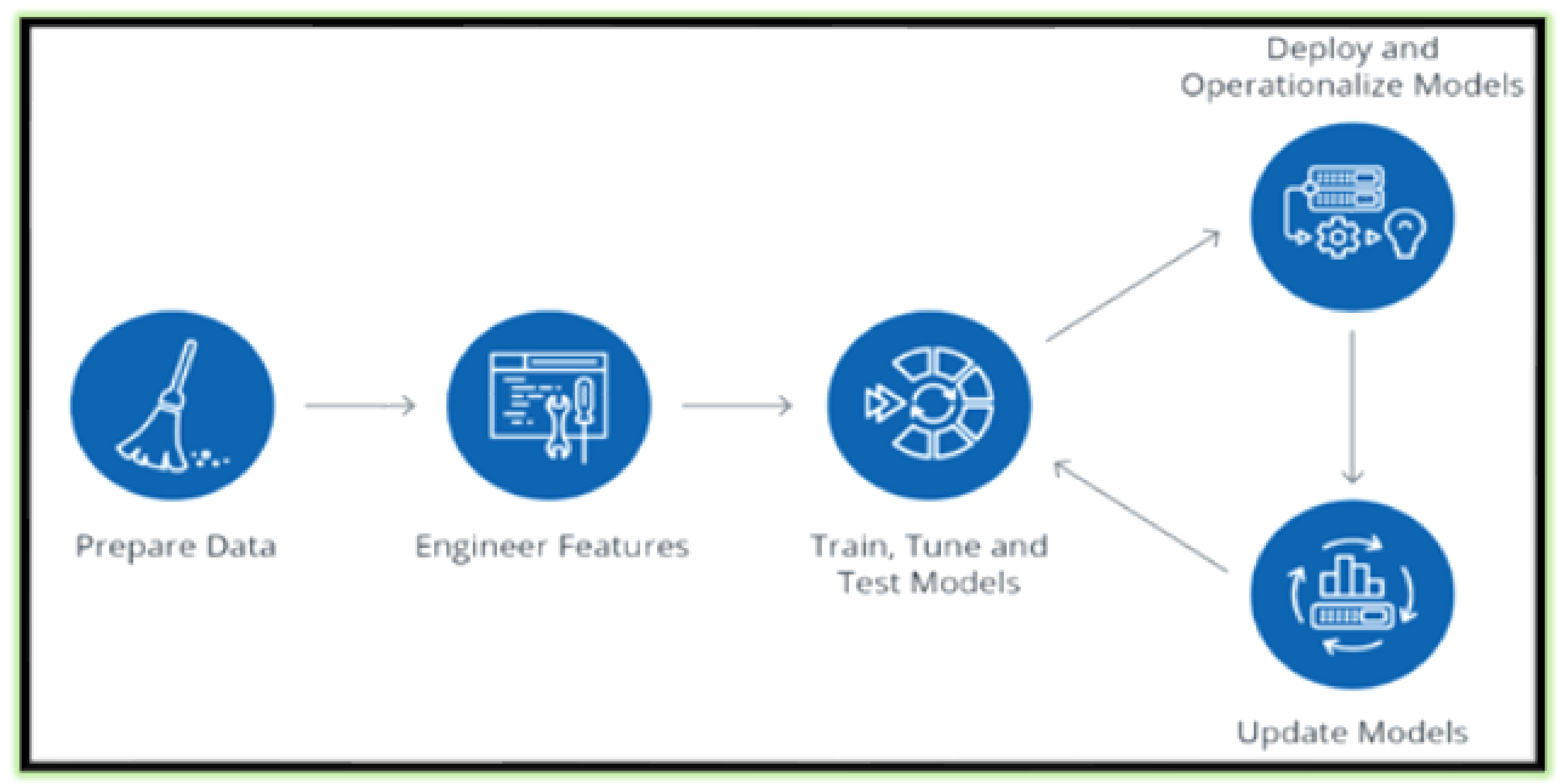
- Implement randomized train-test splitting to prevent temporal bias, ensuring robust and generalized model performance evaluation



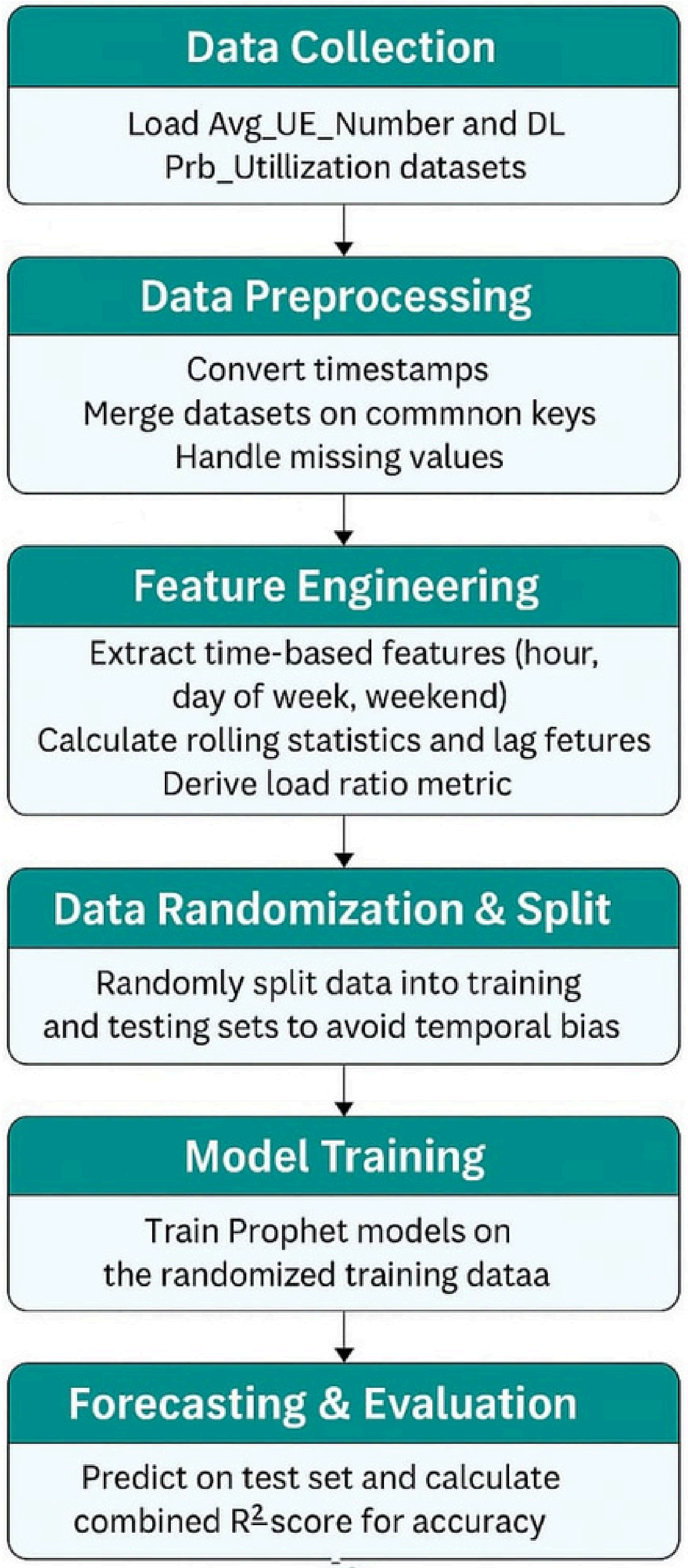
APPROACH



- Data Preparation & Features: Clean and merge data, add time-based features, rolling stats, lags, and load ratio to capture patterns.
- Randomized Split: Use random 80-20 train-test split to avoid time bias and improve model generalization.
- Prophet Modeling: Train Prophet models on smoothed data for each KPI, forecast, evaluate with R^2, and save results.



FLOWCHART


Data Collection
Load Avg_UE_Number and DL
Prb_Utillization datasets
Data Preprocessing
Convert timestamps
Merge datasets on commnon keys
Handle missing values
Feature Engineering
Extract time-based features (hour,
day of week, weekend)
Calculate rolling statistics and lag fetures
Derive load ratio metric
Data Randomization & Split
Randomly split data into training
and testing sets to avoid temporal bias
Model Training
Train Prophet models on
the randomized training dataa
Forecasting & Evaluation
Predict on test set and calculate
combined R2 score for accuracy

ALGORITHM

Data Ingestion and Preprocessing

- Load average UE and PRB utilization datasets
- Convert timestamps, remove malformed and null entries for consistency

Feature Engineering and Aggregation

- Merge datasets on common identifiers (Timestamp, NCI, gNB)
- Derive temporal features (hour, day of week, weekend flag)
- Calculate rolling statistics and lag variables
- Compute additional metric: PRB-to-UE load ratio

Data Preparation and Modeling

- Apply smoothing to reduce noise in time series
- Perform randomized train-test split to prevent temporal leakage
- Train separate Prophet models for UE count and PRB utilization

Forecast Generation and Evaluation

- Generate forecasts using test set timestamps
- Evaluate model performance using combined R^2 score
- Save forecasted output and accuracy metrics to CSV for reporting

DATA RANDOMIZATION STRATEGY

Random Data Distribution

Randomized Train-Test Split

Data is shuffled and split to ensure diverse representation

Removes Temporal Bias

Avoids reliance on chronological order, improving model generalization

Enhances Forecast Reliability

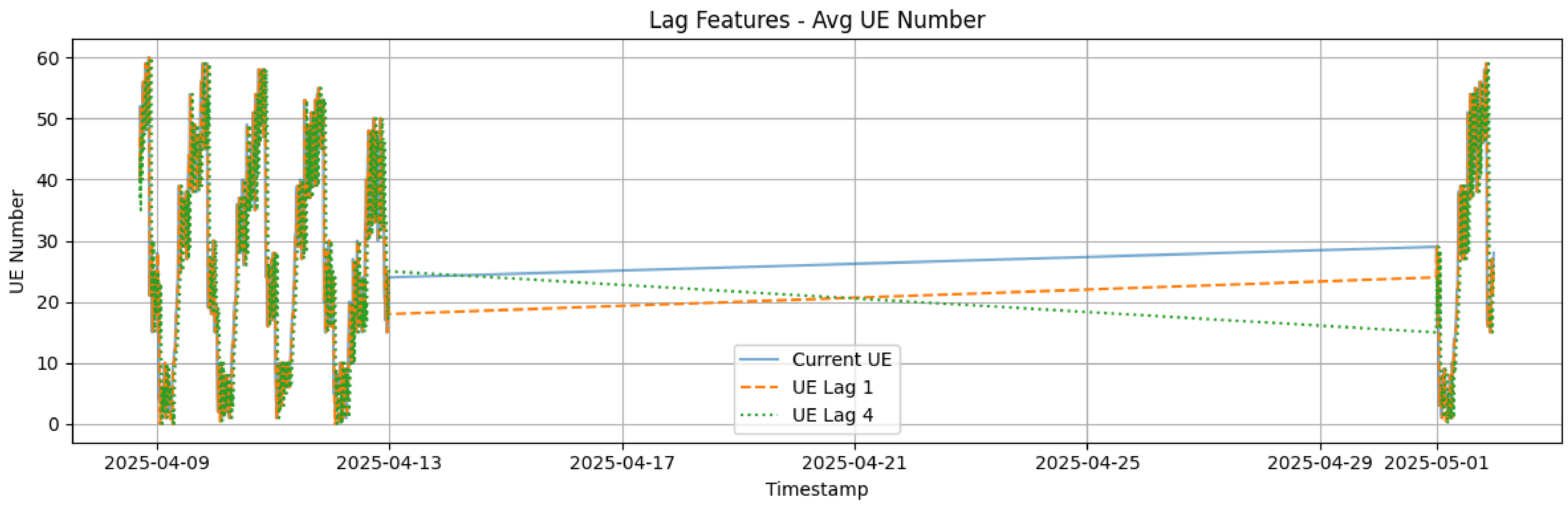
Allows the model to learn from varied conditions across the dataset.

Improved Evaluation Metrics

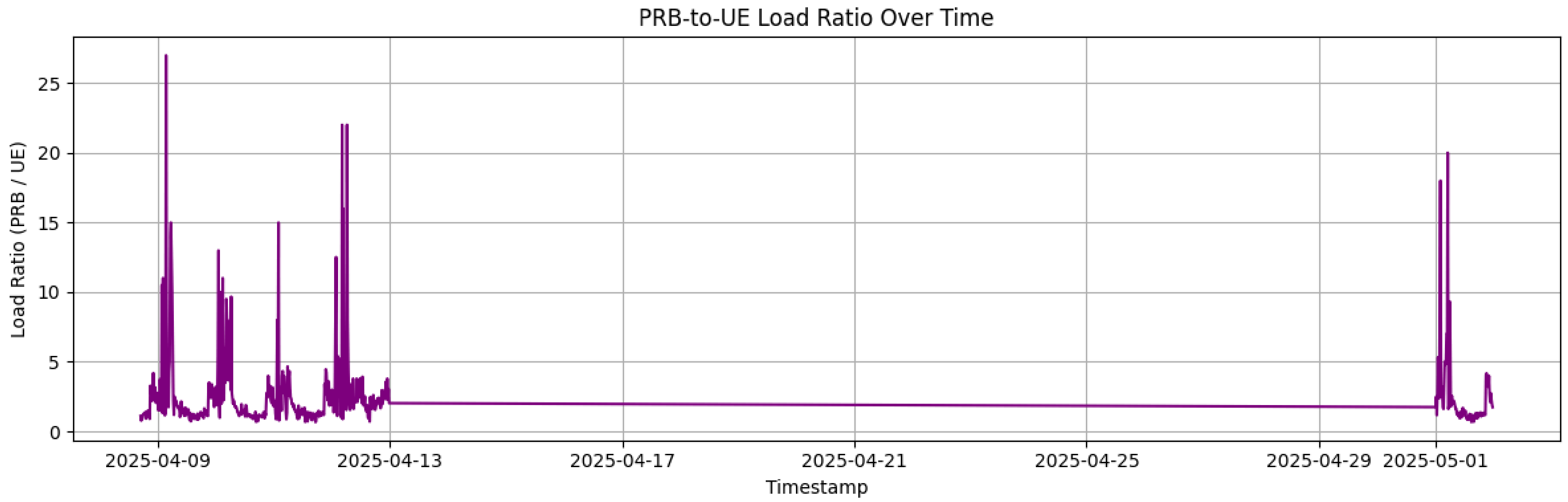
R^2 score better reflects true predictive performance across all time periods.

FORECAST ACCURACY EVALUATION

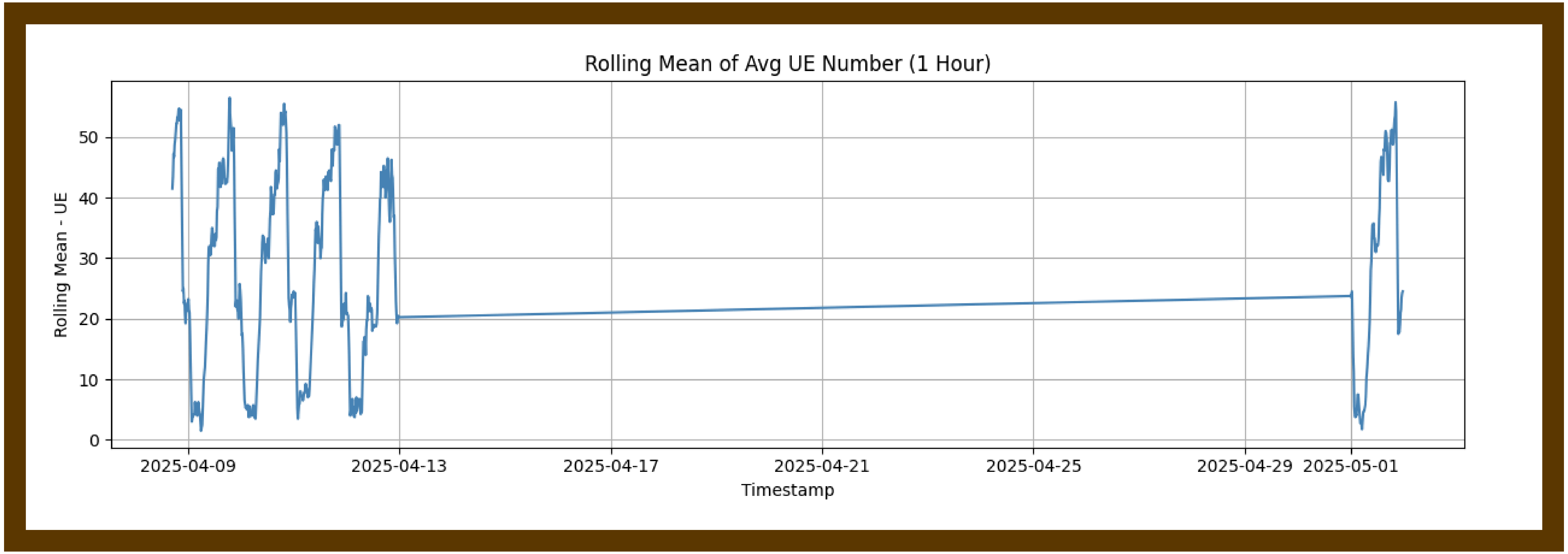
- Consistent Accuracy: The R^2 score remains stable at 0.8565, demonstrating that the model generalizes well, even with shuffled (randomized) data.
- Robust Model: Your use of advanced feature engineering and Prophet's modeling capability is effective across varied data splits.
- No Overfitting: Randomizing avoids temporal bias, and the model still performs reliably — a sign of low overfitting risk.
- Production-Ready Quality: An R^2 score near 0.85 for time-series forecasting in telecom KPIs (which are noisy and volatile) is strong and suitable for operational deployment

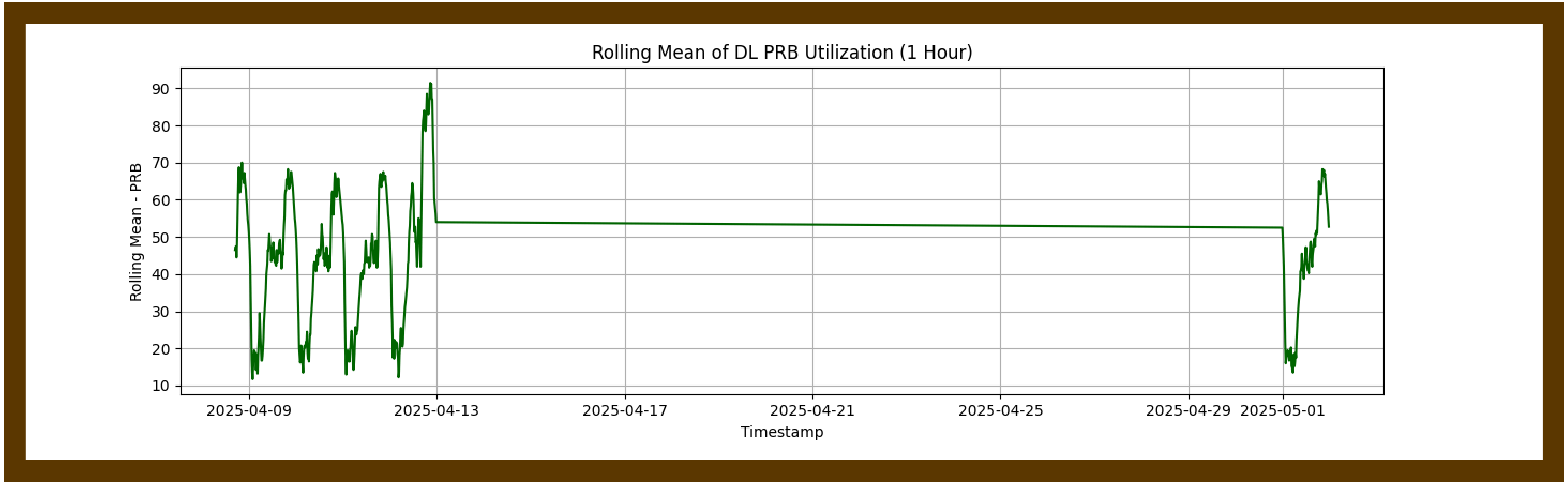


Lag Features - Avg UE Number
UE Number
60
50
40
30
20
10
0
Current UE
UE Lag 1
UE Lag 4
2025-04-09
2025-04-13
2025-04-17
2025-04-21
2025-04-25
2025-04-29
2025-05-01
Timestamp



PRB-to-UE Load Ratio Over Time
Load Ratio (PRB / UE)
25
20
15
10
5
0
2025-04-09
2025-04-13
2025-04-17
2025-04-21
2025-04-25
2025-04-29
2025-05-01
Timestamp


Rolling Mean of Avg UE Number (1 Hour)
Rolling Mean - UE
0
10
20
30
40
50
2025-04-09
2025-04-13
2025-04-17
2025-04-21
2025-04-25
2025-04-29
2025-05-01
Timestamp


Rolling Mean of DL PRB Utilization (1 Hour)
Rolling Mean - PRB
10
20
30
40
50
60
70
80
90
2025-04-09
2025-04-13
2025-04-17
2025-04-21
2025-04-25
2025-04-29
2025-05-01
Timestamp

CONCLUSION

Achieved high forecasting accuracy (R^2 Score: 0.8489) using Prophet with enhanced feature engineering and randomized data splitting

The model effectively captures temporal patterns and shows strong potential for scalable, real-world telecom KPI prediction

THANK YOU