Electricity Demand and Price Foreeecasting

Presented by Raja Roy

19 July 2024
Internship Project

CONTENTS

Problem Statement

Objective

- Develop a machine Learning model to accurately forecast electricity demand and it’s Prices

Datasets Provided

- Weather Dataset:- Contains Details Such as Rain, snow and Temperature
- Energy Dataset:- Contains information on Prices, Demand and Energy Value

Analyze various machine learning models

- Linear Regression
- Random Forest
- Gradient Boosting
- Long Short-term memory

Conlusion

Make a Final Report with the best suitable model least error and also observe the important features related to the model

Data Introduction

Energy Dataset

- contents:- Data, Price, Load, generation of various energy sources (fossil fuel, renewable energy etc.
- Rows:- 35065
- Columns:- 27(with 6 all zero, 2 with missing Values

weather Dataset

- Contents:- Pressure, Humidity, snow rain etc.
- Rows:-178396
- Columns:-17
- Details: Details: Hourly weather data for 5 major cities in Spain.

-

Granuality

- Hourly data for both datasets from 31st Dec 2014 to 31st Dec 2018.

EXPLORATORY DATA ANALYSIS

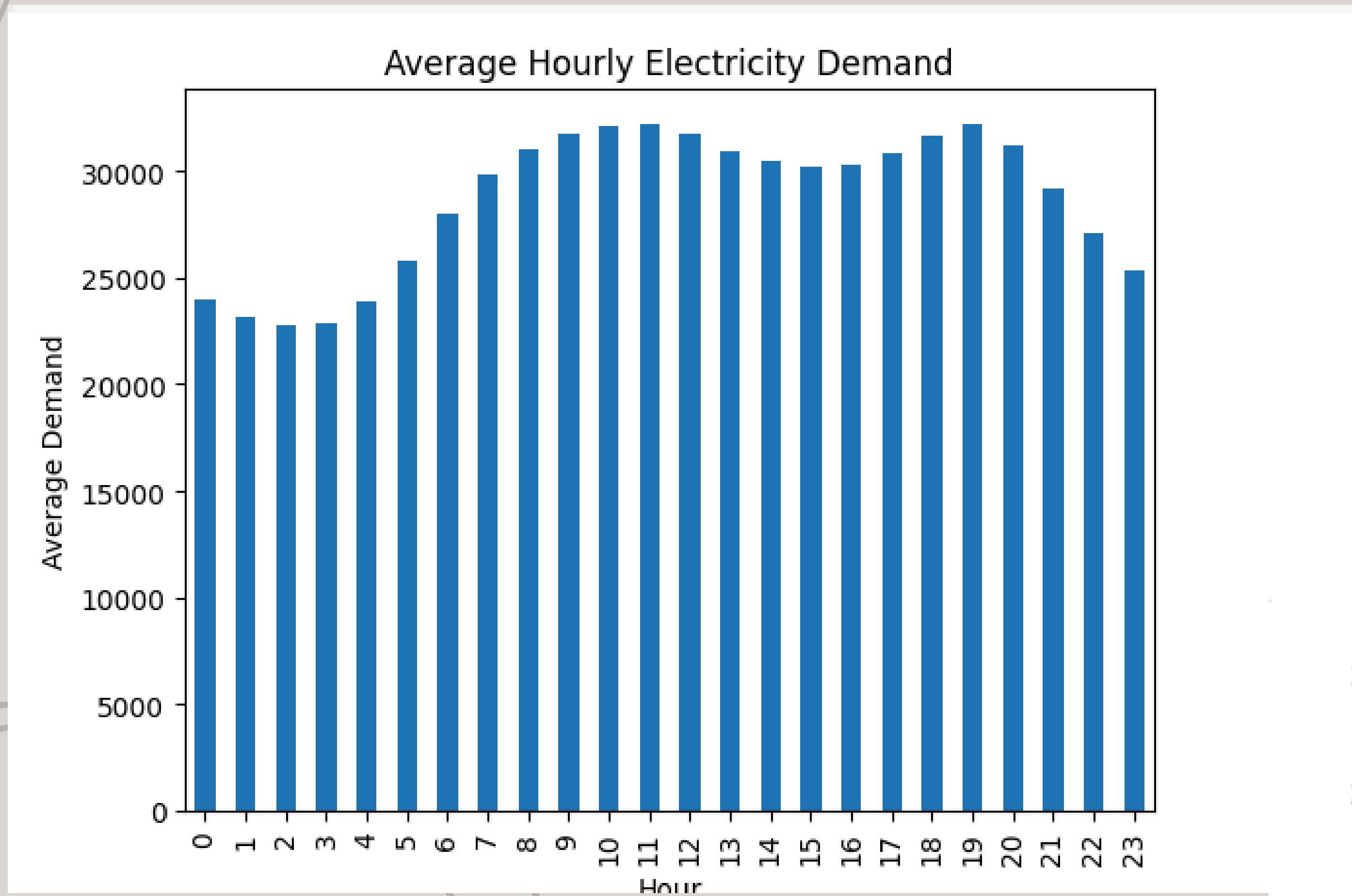
DATA PREPROCESSING

- First, we will remove the column “city_name”.
- Aggregate the weather data grouped it by timestamp (‘dt_iso’), and summarized various weather-related metrics. This aggregated data can now be used for further analysis, visualization, or reporting.

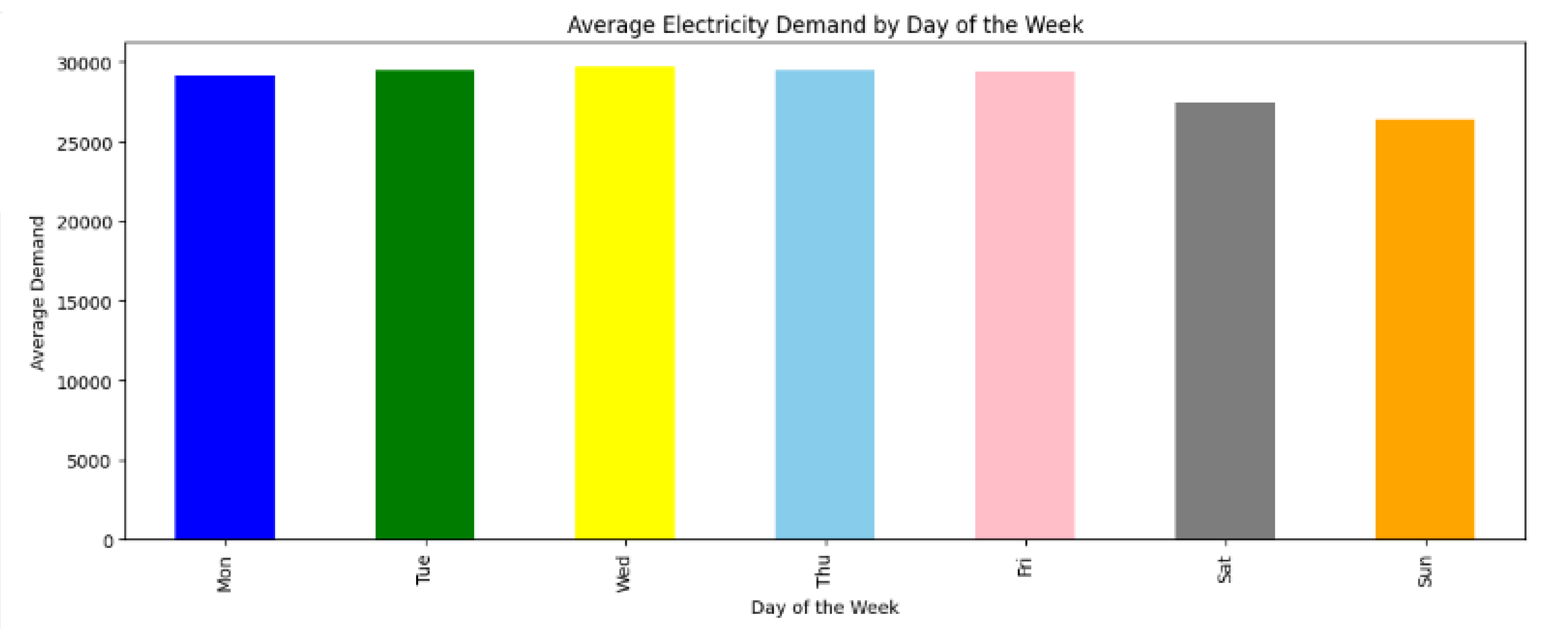
FEATURE ENGINEERING

- Detecting the outliers using the Interquartile Range (IQR) method.
- After detecting them in different columns like “wind_speed”, ”price_actual”. Replace these outliers with NaNs.
- Created some new columns like weekday columns – Sunday, Monday, Tuesday… and so on.
- Also created the months columns – January, February, March…. And so on.

VISUALIZATION

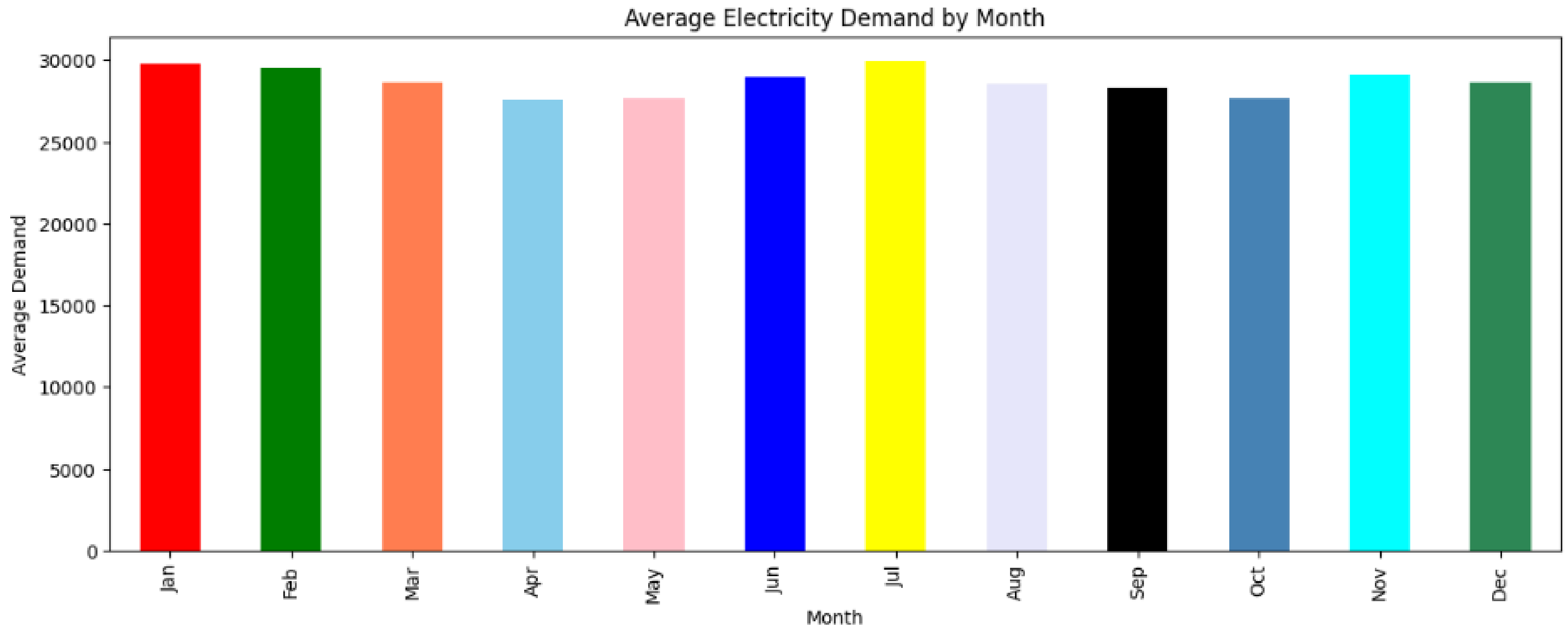


Peak and Non-Peak Hours



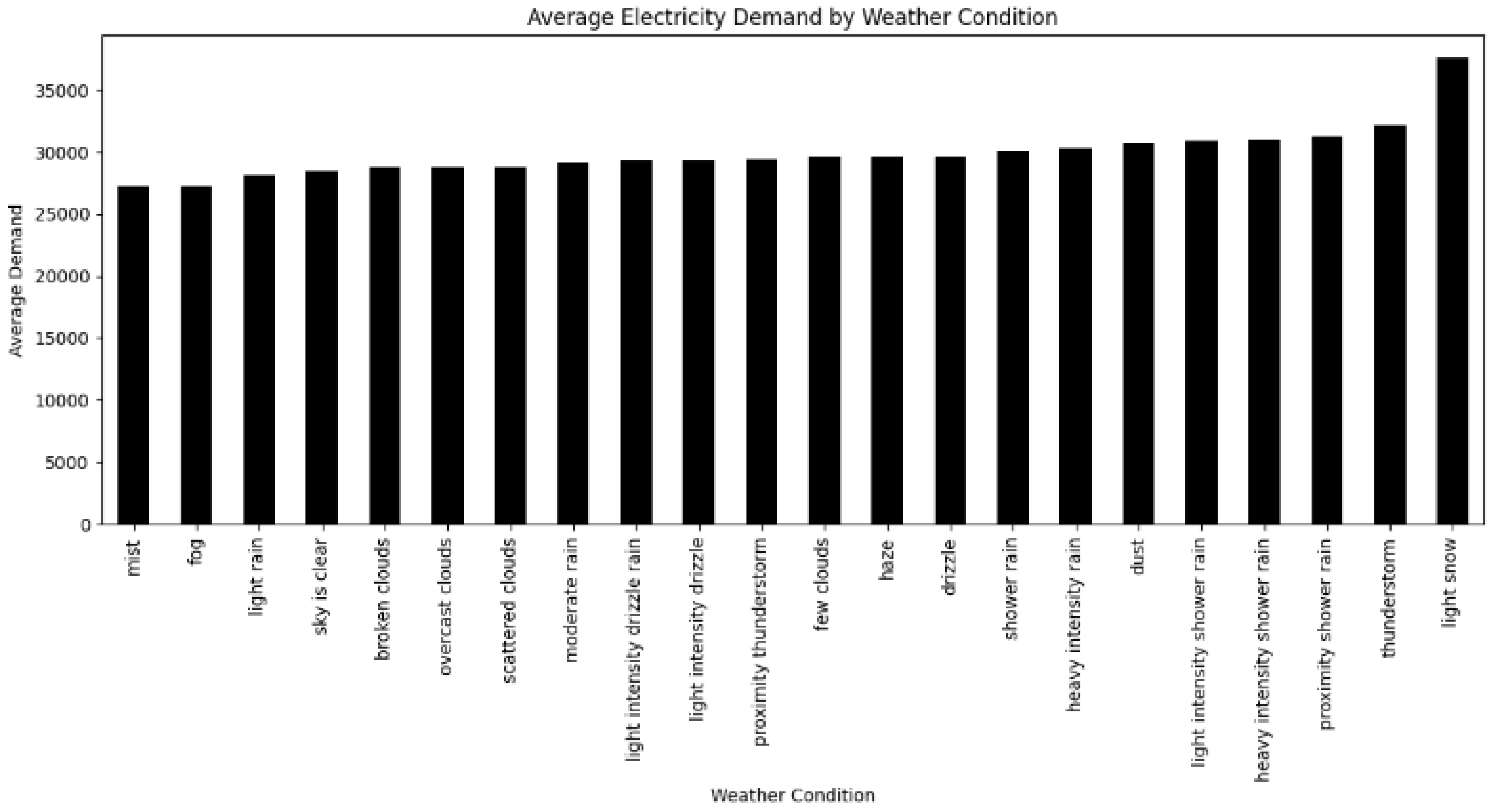
Average electricity demand by day of the week

Data Visulization

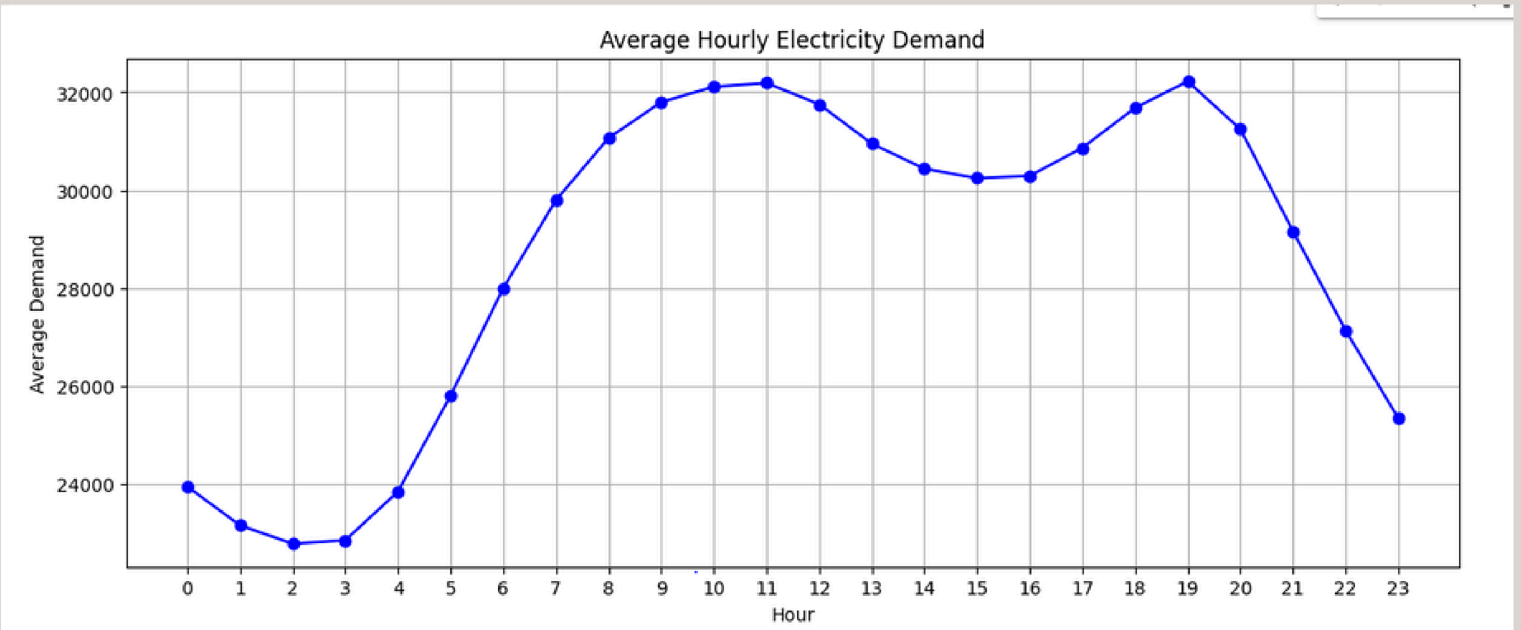


Bar Chart Comparing Electricity Demand by Month

Average electricity demand by weather condition

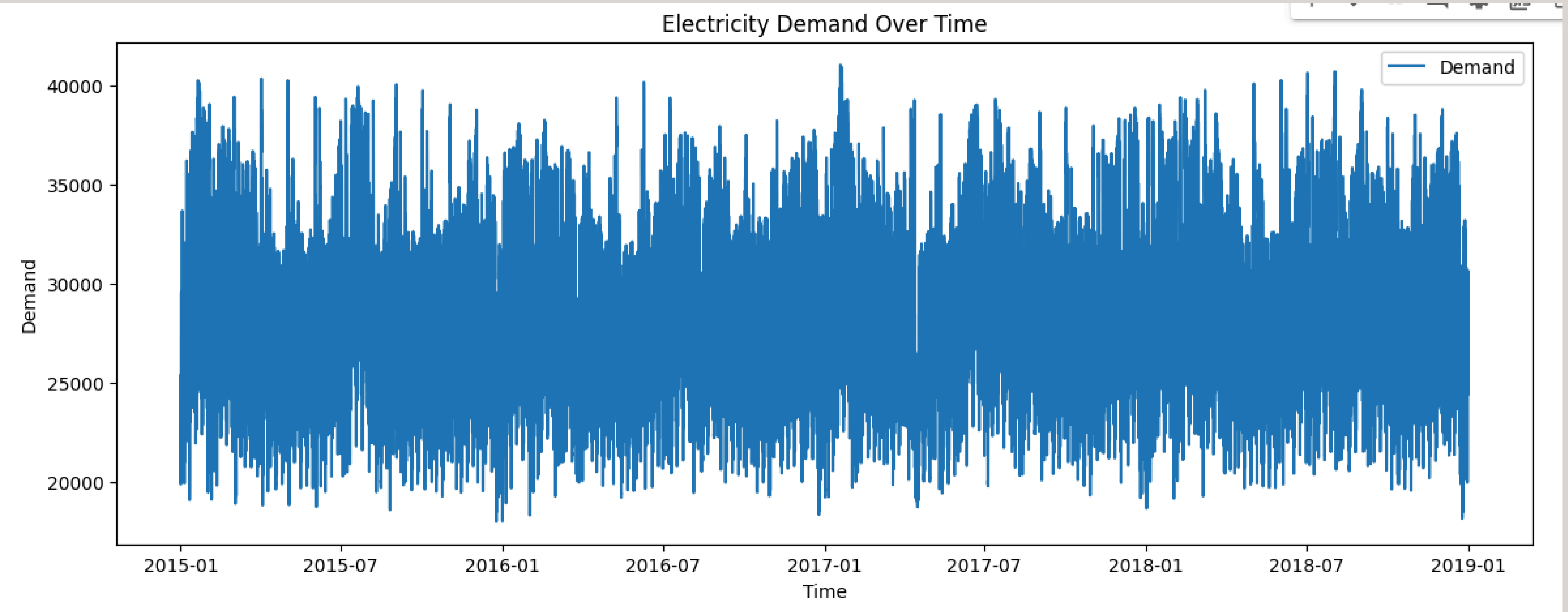


Data Visulization

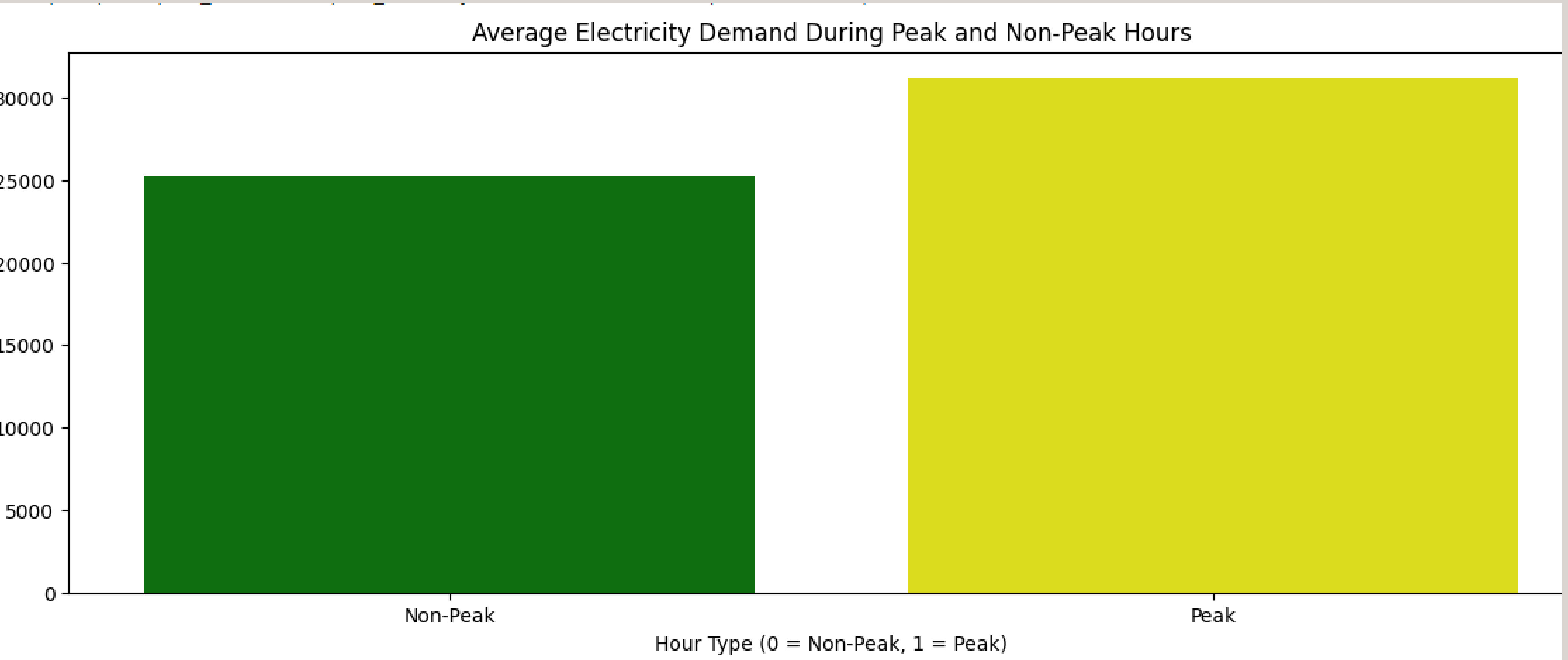


Average Electricity Demand hourly

Line Plot of Electricity Demand Over Time

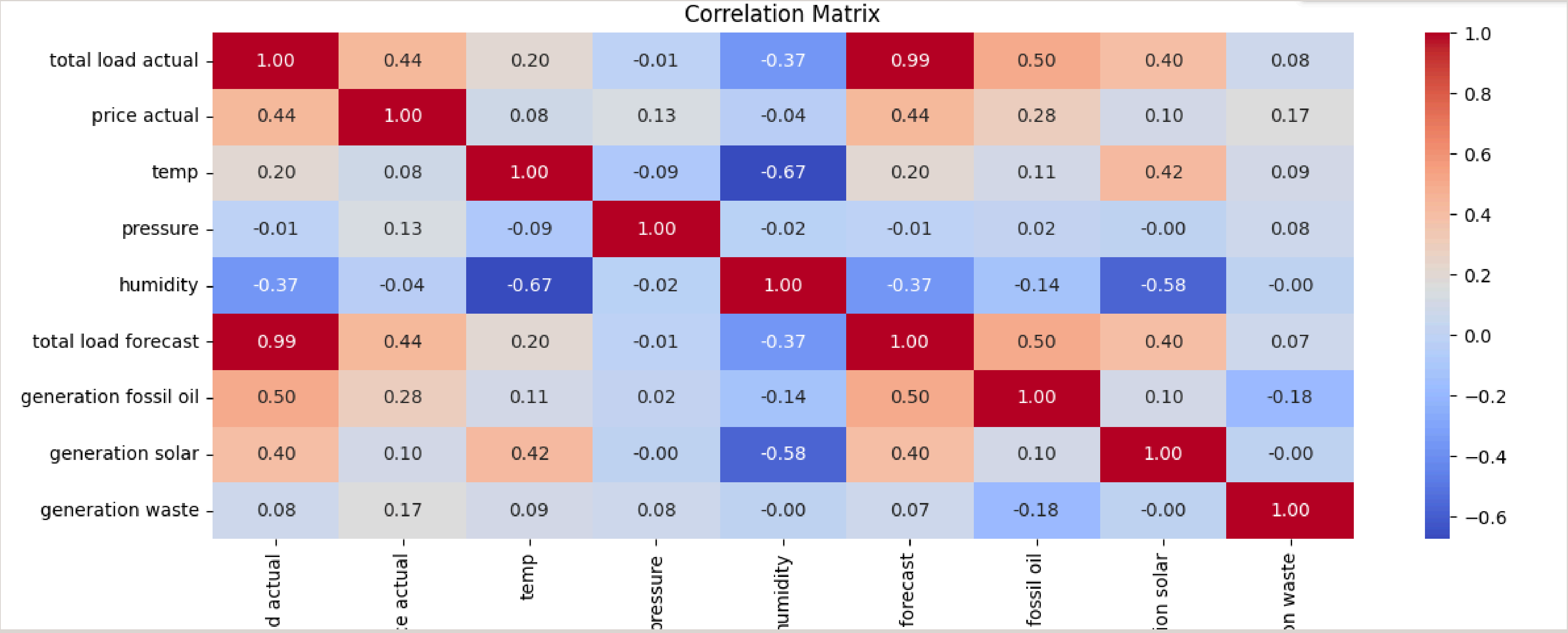


Data Visulization



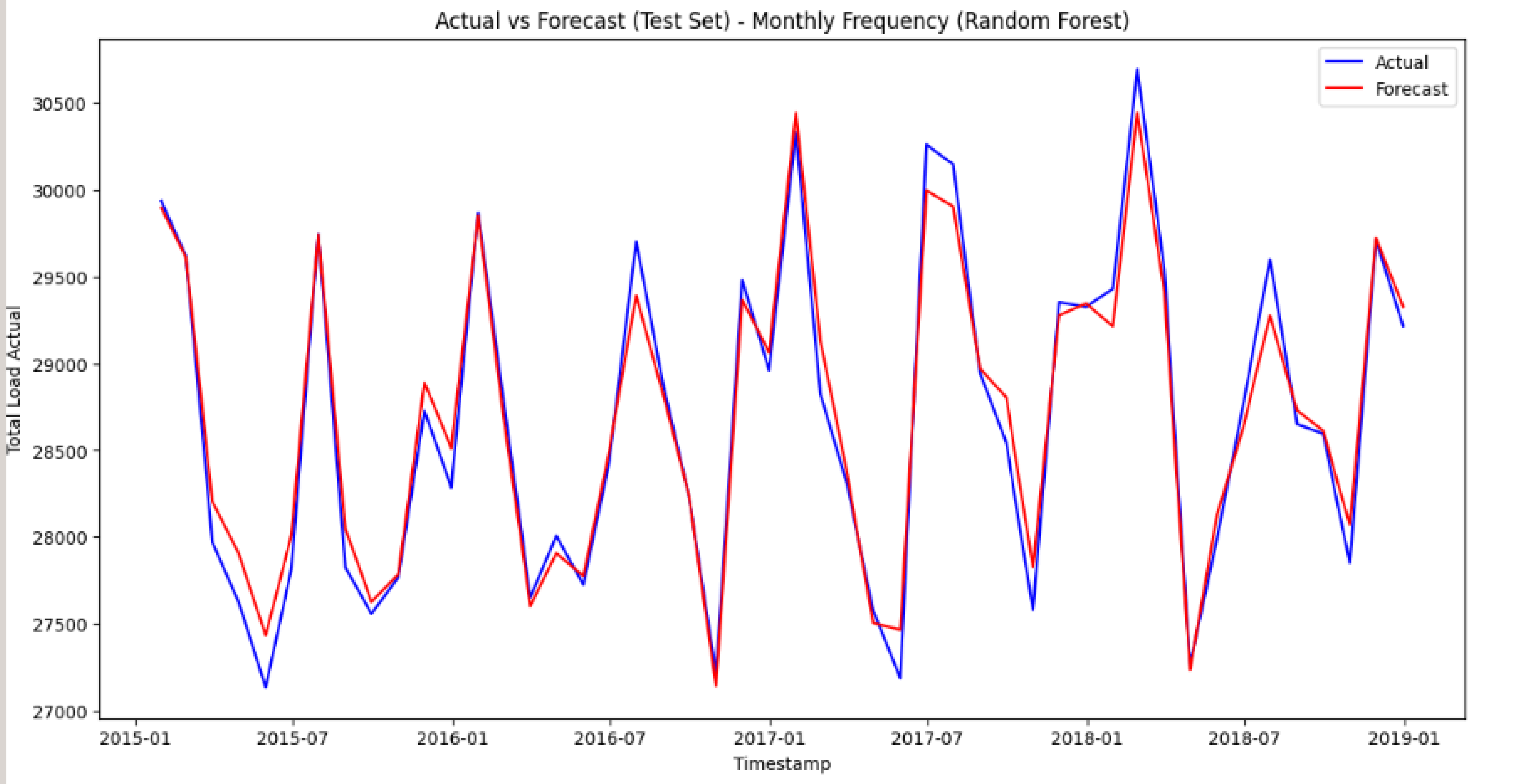
Average electricity demand during non Peak Hour

Corelaion Matrix



Machine Learning

Random Forest



For Load

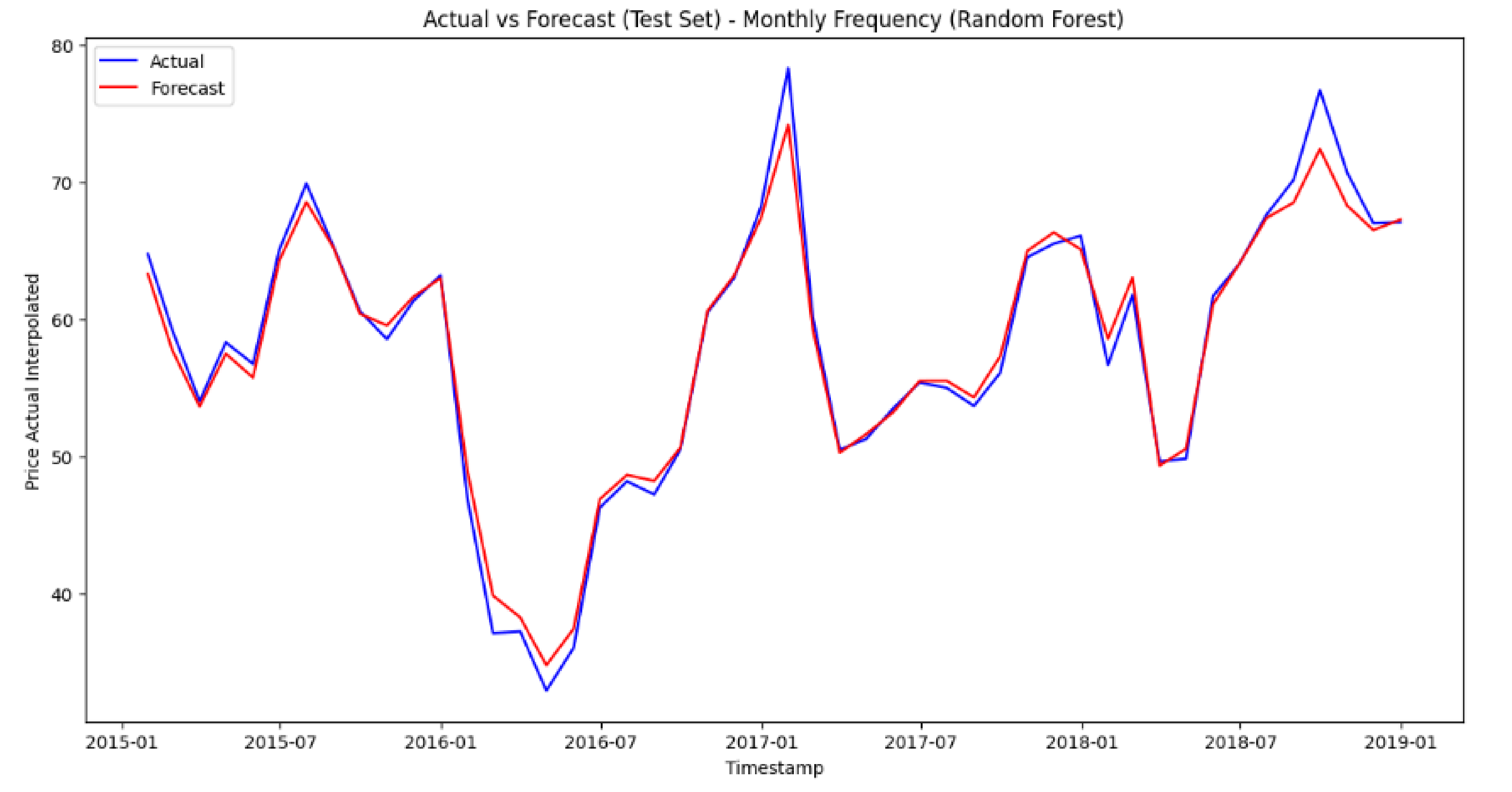
MAPE : 2.47%

MAE : 689.00

RMSE : 939.23

R2 : 0.96

Adjusted R2 : 0.96



For Price

MAPE : 5.67%

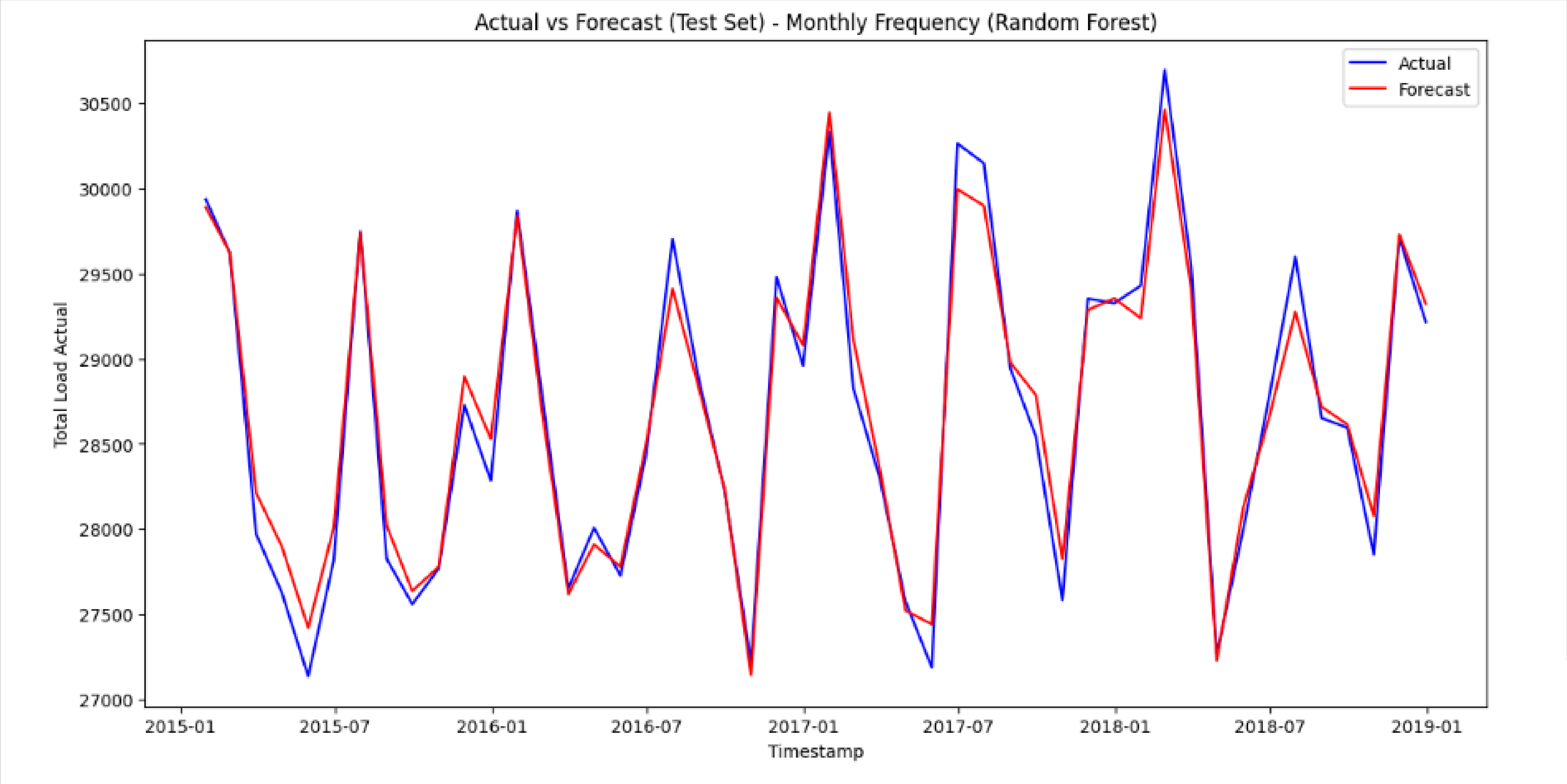
MAE: 2.90

RMSE: 4.18 R

R2 :0.91

Adjusted R2 : 0.96

Random Forest Hyper Tunning



For Load

MAPE : 2.48%

MAE : 692.70

RMSE : 942.00

R2 : 0.96

Adjusted R2 : 0.96

For Price

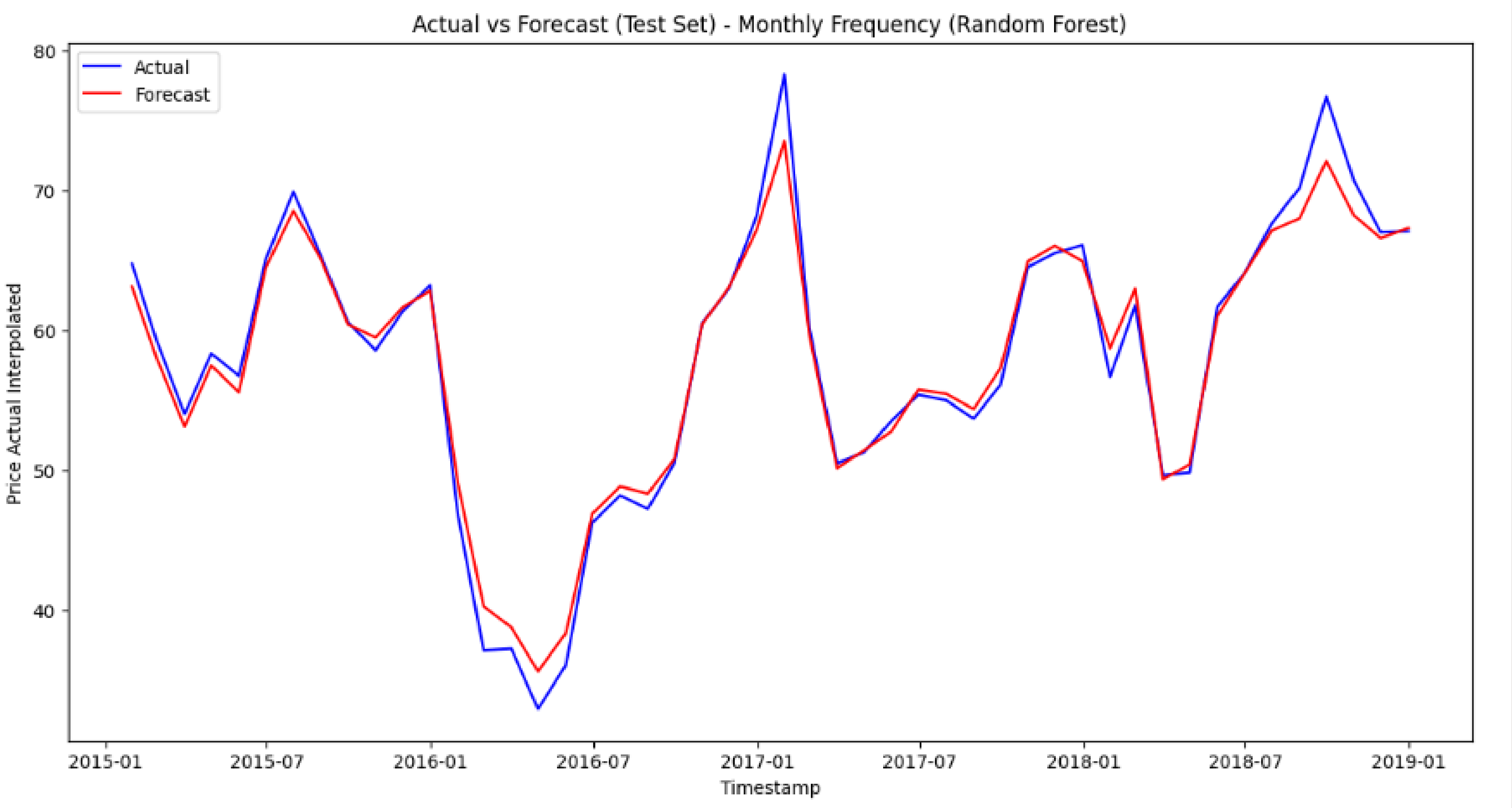
MAPE : 5.85%

MAE : 3.01

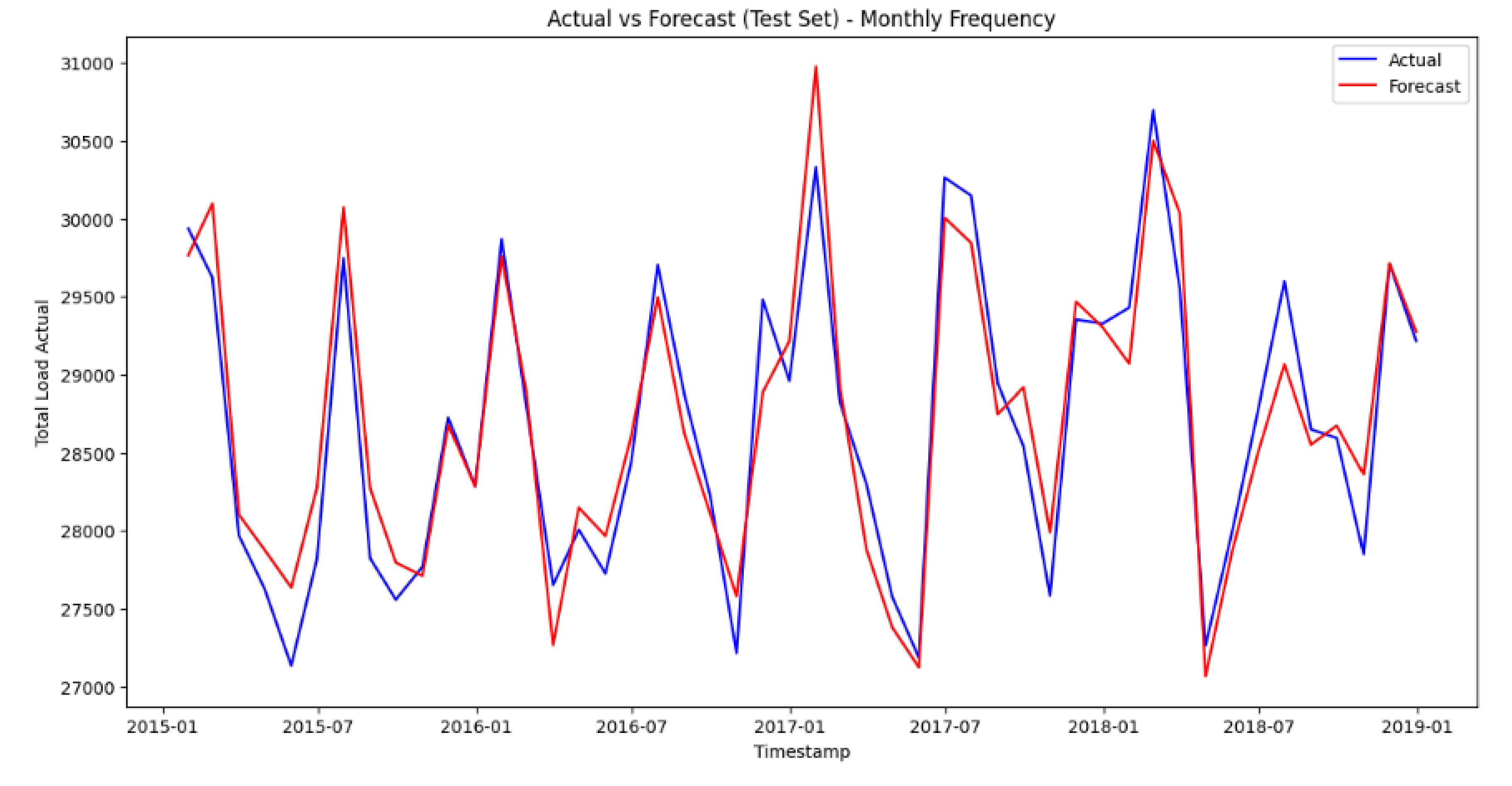
RMSE : 4.28

R2 : 0.91

Adjusted R2 : 0.96



Linear Regression

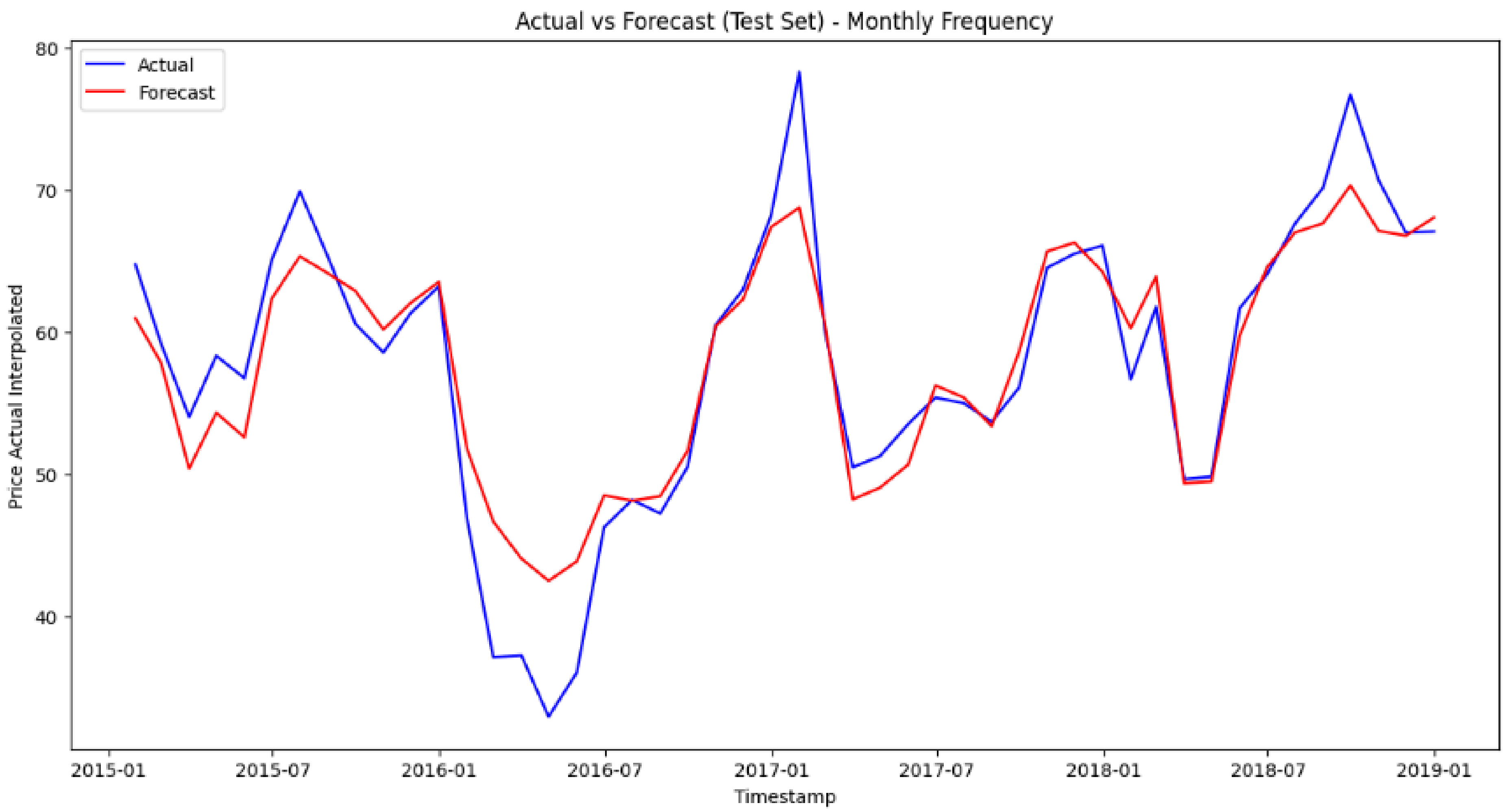


For Load

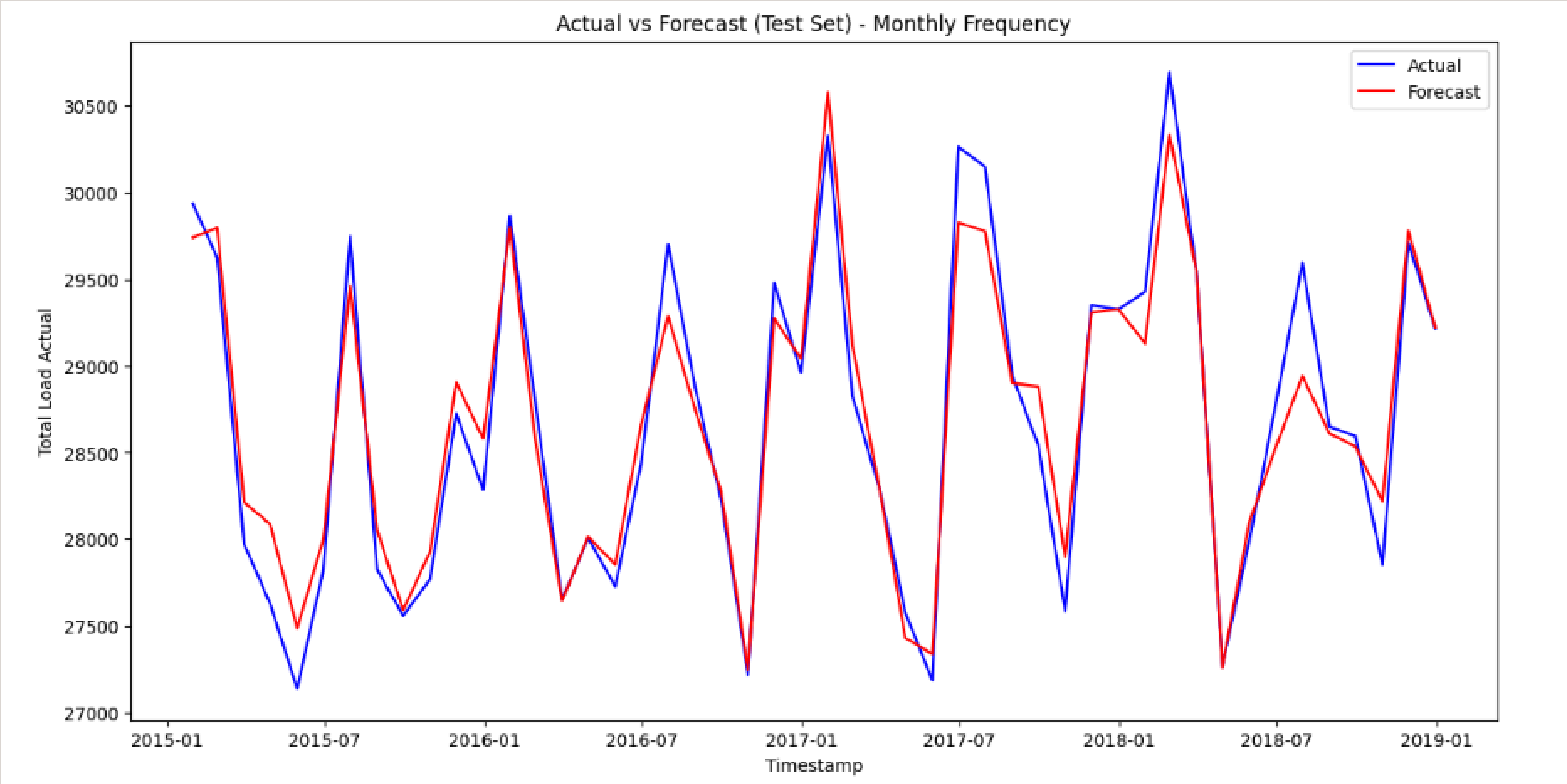
MAPE : 3.51%
MAE : 988.43
RMSE : 1270.09
R2 : 0.92
Adjusted R2 : 0.92

For Price

MAPE : 14.45%
MAE : 7.36
RMSE : 9.41
R2 : 0.55
Adjusted R2 : 0.55



Gradient Boosting



For Load

MAPE: 3.46%

MAE : 972.93

RMSE : 1253.70

R2 : 0.92

Adjusted R2 : 0.92

For Price

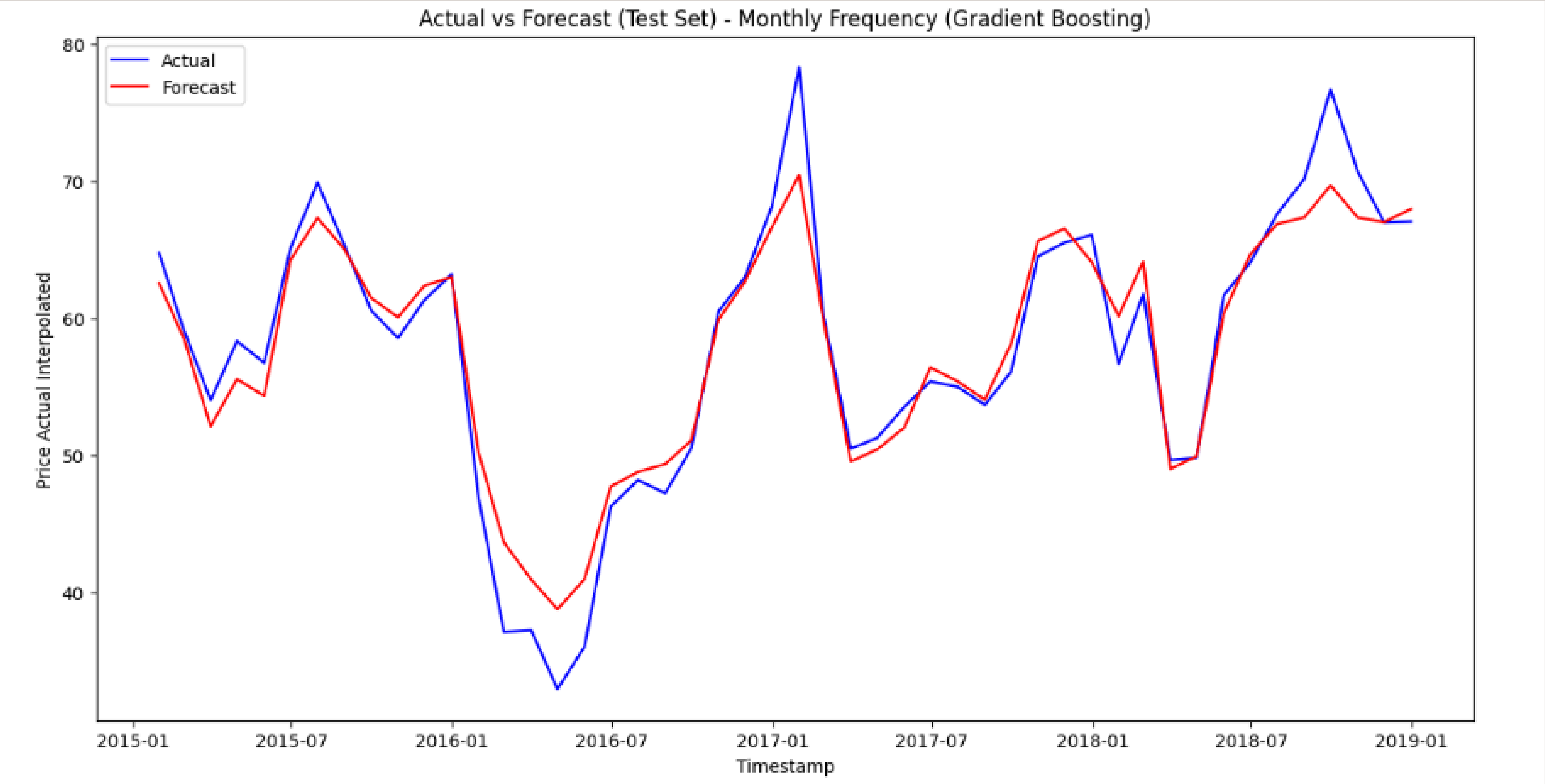
MAPE : 11.16%

MAE : 5.68

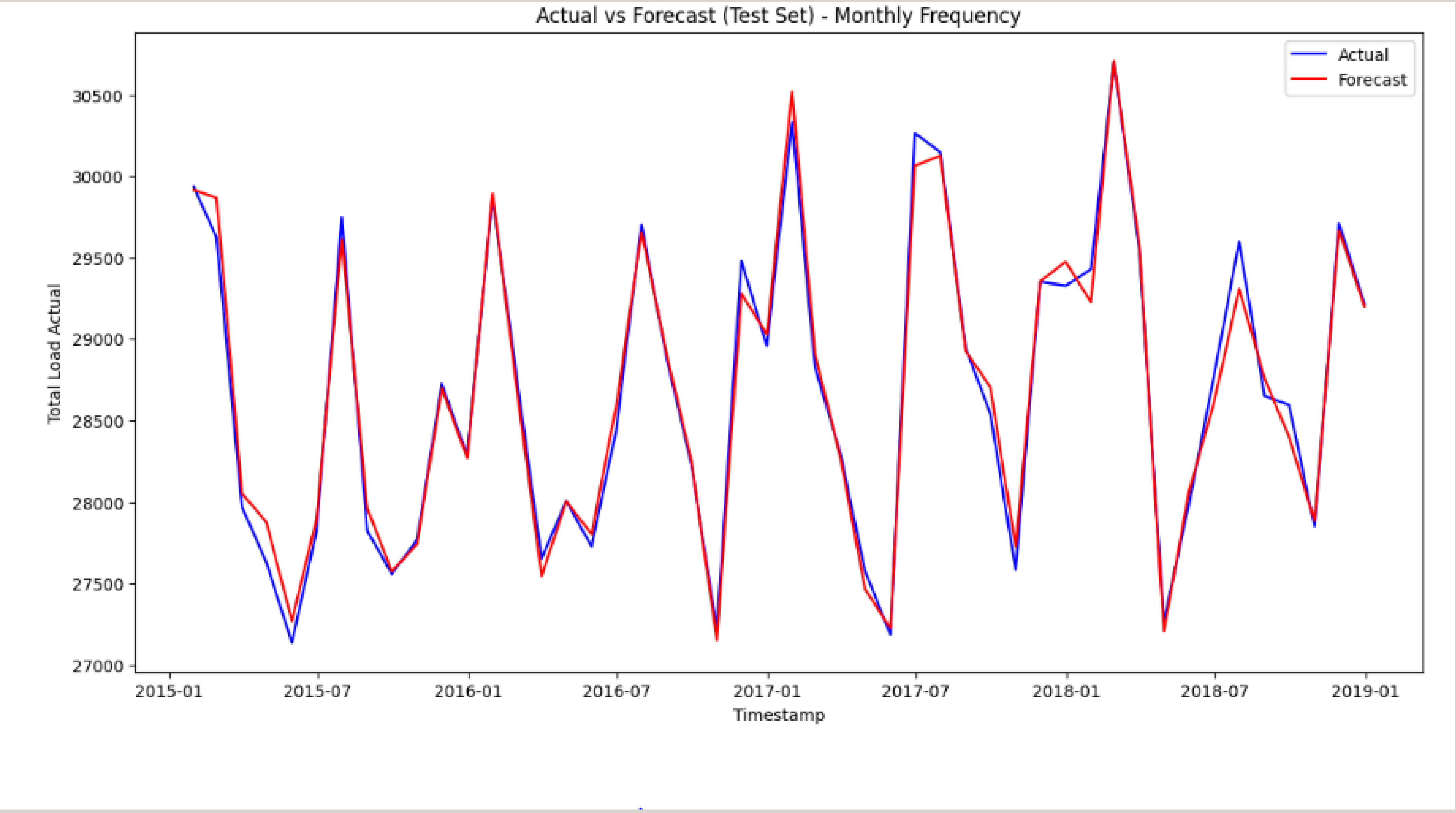
RMSE : 7.36

R2 : 0.72

Adjusted R2: 0.92



Gredient Boosting Tuning

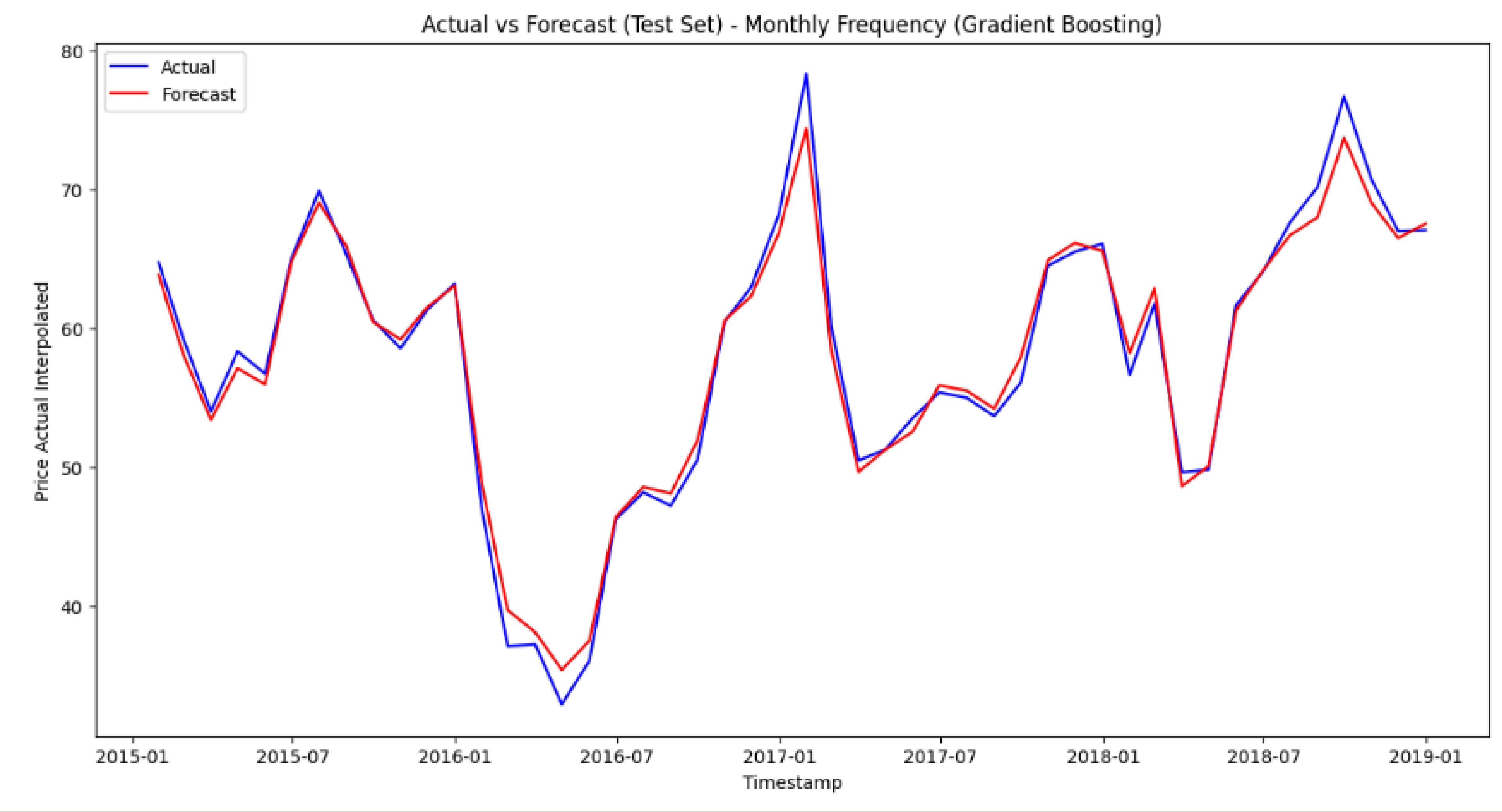


For Load

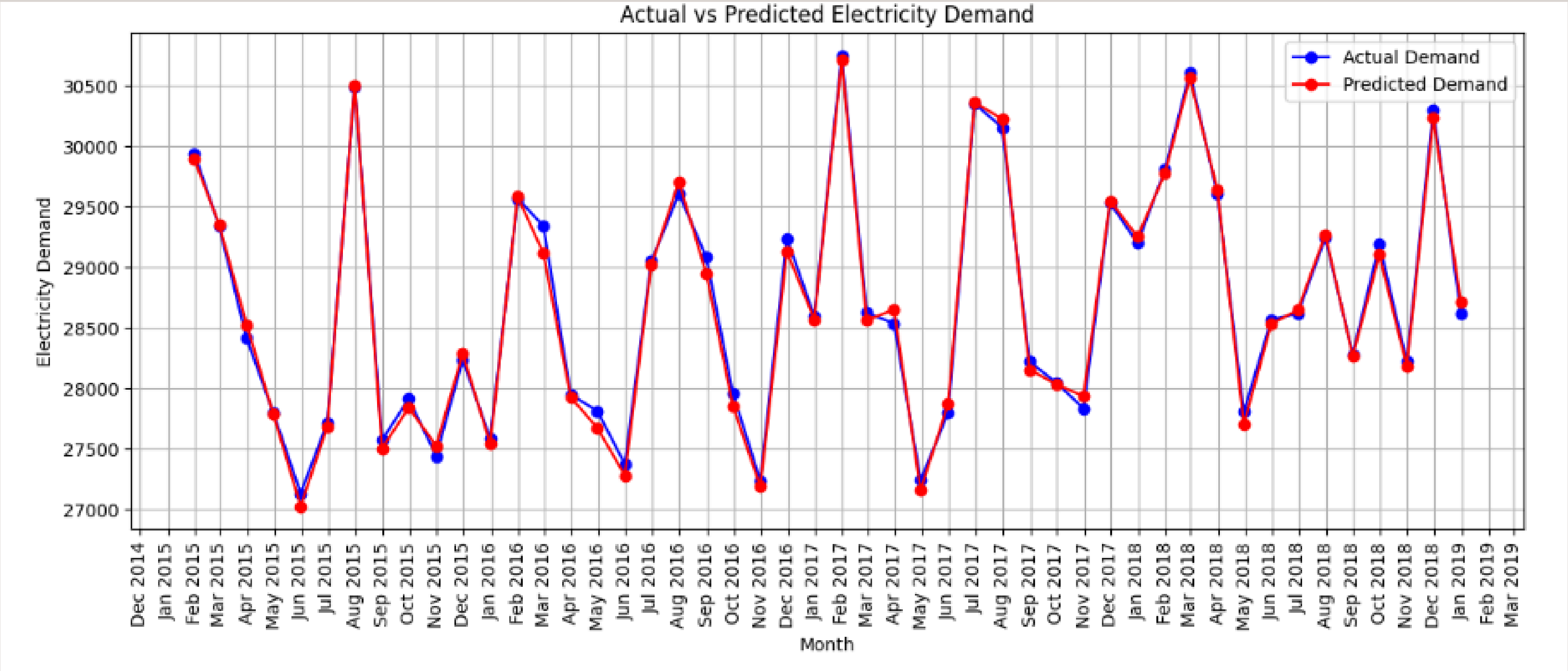
MAPE: 2.36%
MAE: 667.39
RMSE: 863.99
R2: 0.96
Adjusted R2 : 0.96

For Price

MAPE : 6.91%
MAE : 3.62
RMSE : 4.75
R2 : 0.88
Adjusted R2 : 0.88



Long Short Term Memory



For Load

MAPE: 1.83%

MAE: 512.73

RMSE: 680.12

R2: 0.98

Adjusted R2: 0.9

For Price

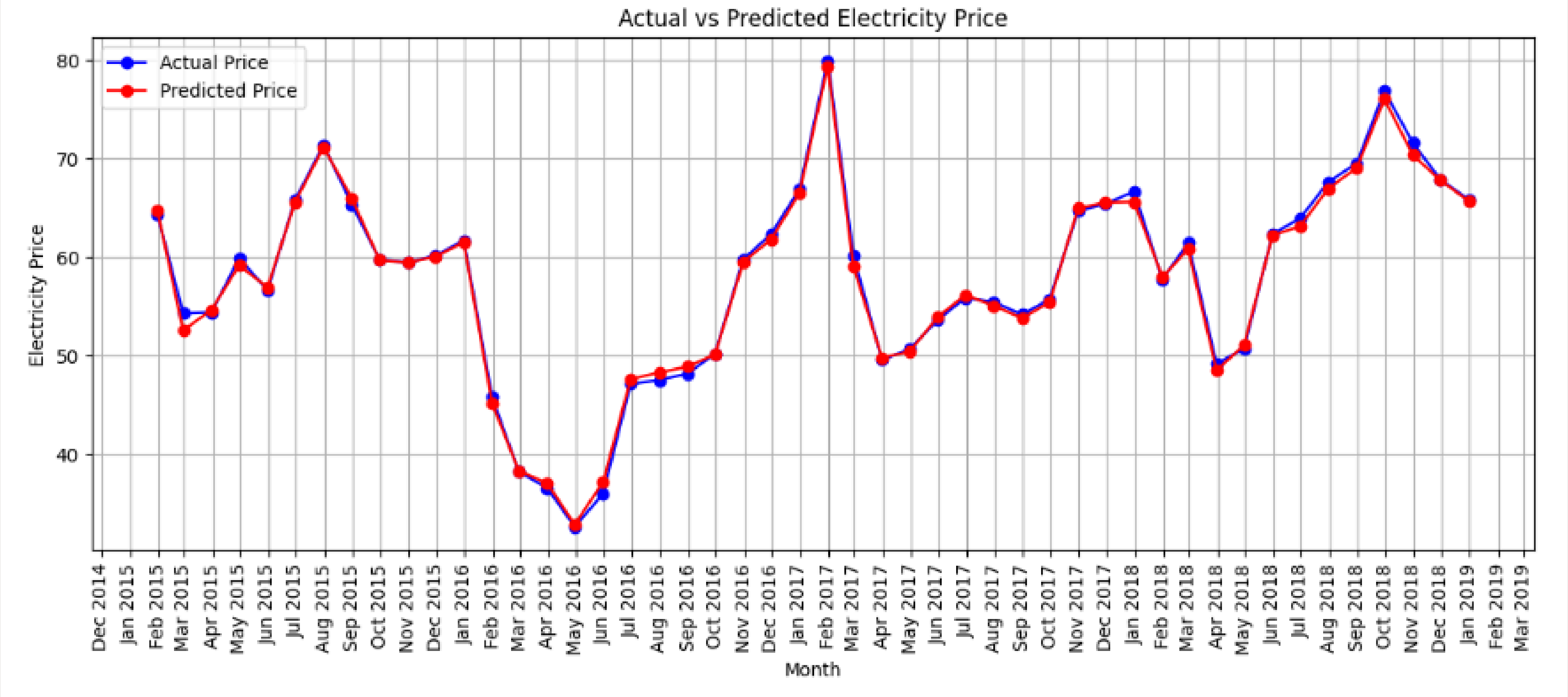
MAPE: 5.14%

MAE: 2.71

RMSE: 3.76

R2: 0.93

Adjusted R2: 0.98



Campare Both Table

Performance Table for Demand model

	Model	MAPE	MAE	RMSE	R^2	Adjusted R^2
0	Linear Regression	0.035078	988.427519	1270.093525	0.922603	0.922465
1	Decision Tree	0.038433	1078.999287	1546.341479	0.885273	0.885068
2	Random Forest	0.024693	689.001917	939.227423	0.957675	0.957600
3	Random Forest(tuning)	0.024813	692.695315	942.003647	0.957425	0.957349
4	Gradient Boosting	0.034585	972.933374	1253.700469	0.924588	0.924453
5	Gradient Boosting(tuning)	0.023629	667.386833	863.987165	0.964185	0.964121
6	Long Short-Term Memory	0.019154	536.838324	719.495328	0.975162	0.975161

For Load:-
LSTM is the Best Model ,Because it’s has least MAPE, MAE, RMSE and Perfect R2 and Adjusted R2.

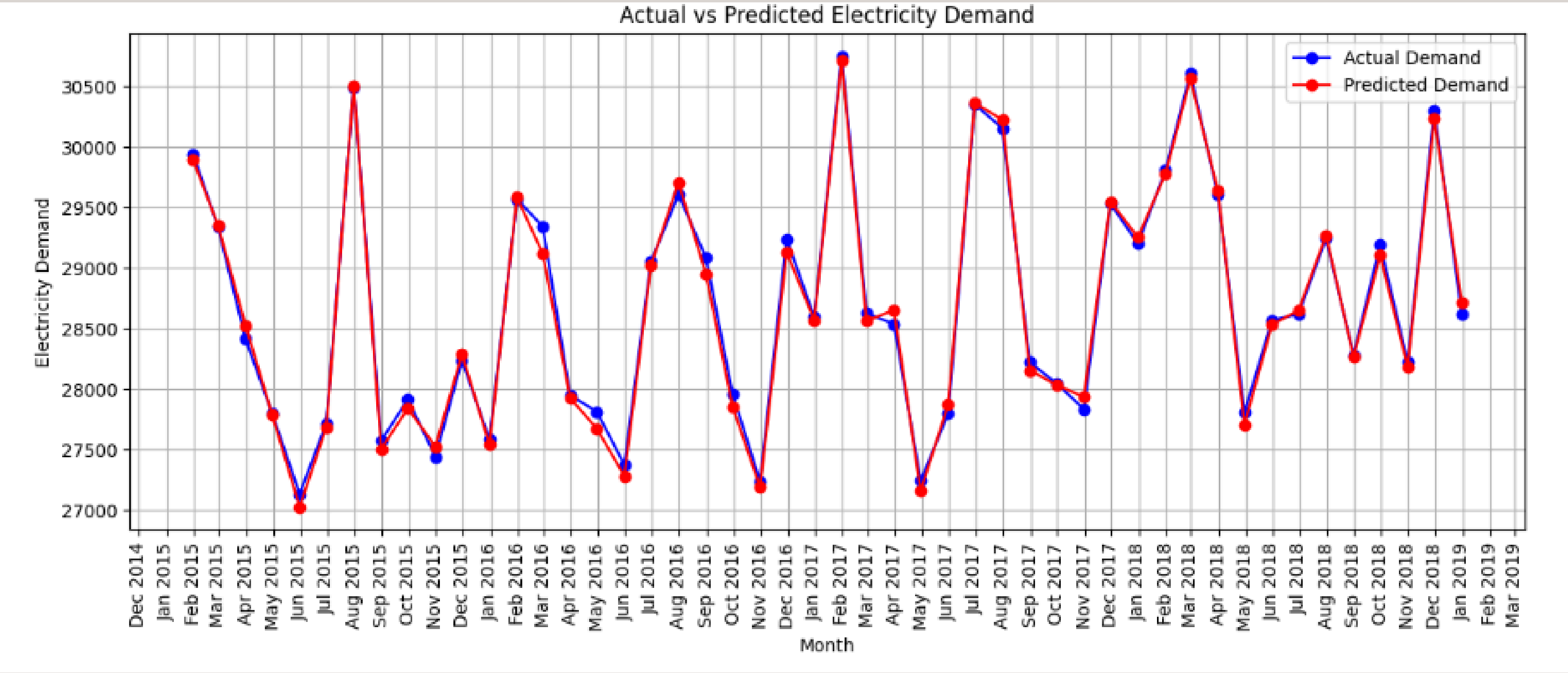
Performance Table for Price model

	Model	MAPE	MAE	RMSE	R^2	Adjusted R^2
0	Linear Regression	0.144510	7.359438	9.405719	0.547238	0.546429
1	Decision Tree	0.080660	4.242130	6.900054	0.756336	0.885068
2	Random Forest	0.056651	2.904602	4.184541	0.910385	0.957600
3	Random Forest(tuning)	0.058543	3.006241	4.282012	0.906161	0.957349
4	Gradient Boosting	0.111616	5.682892	7.356066	0.723065	0.924453
5	Gradient Boosting(tuning)	0.069100	3.619701	4.754859	0.884293	0.884086
6	Long Short-Term Memory	0.049218	2.611849	3.571432	0.934721	0.975161

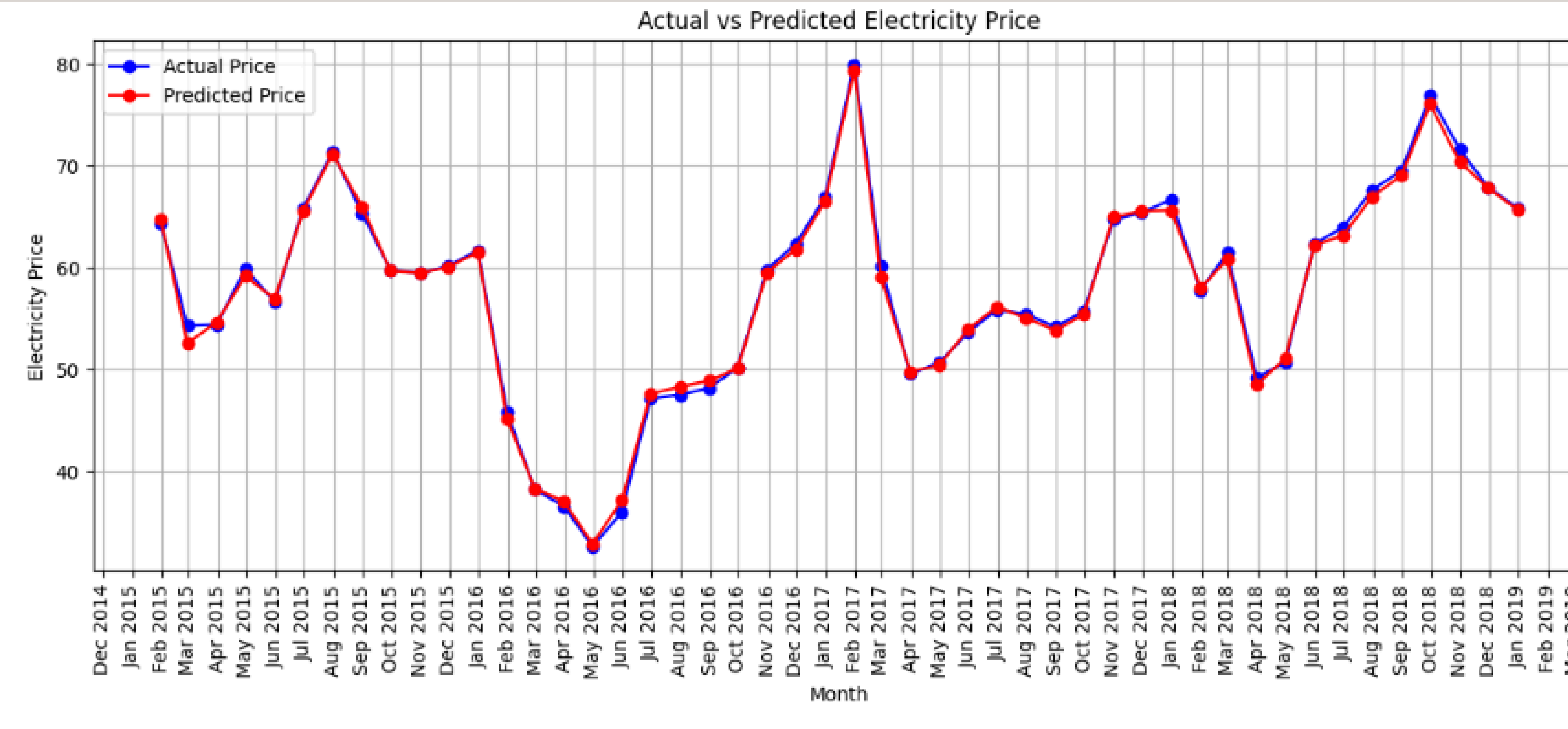
For Price:-
LSTM is the Best Model ,Because it’s has least MAPE, MAE, RMSE and Perfect R2 and Adjusted R2.

CONCLUSION

- For both Price and Load model the MAPE value is lowest for LSTM model with hyperparameter Tuning, Hence it is selected for Model Prediction.
- This is the Actual VS Predicted plots for both Price and load



Load



Price

CONCLUSION

- the LSTM model with hyperparameter tuning is highly scalable due to its ability to handle large datasets, efficient training and inference processes, and adaptability to dynamic resource allocation.

- This makes it a robust solution for predicting load and price in a wide range of operational contexts.

- LSTMs can be retrained incrementally as new data arrives, making it easier to update the model without needing to retrain from scratch.

- This ensures that the model remains up-to-date and scalable over time.

THANK YOU

By:- Raja Roy
19 July 2024