



EXPLORE AI ACADEMY TEAM RR 11



RAPID-R ANALYTICS
Making data work...

PREDICTIVE ANALYSIS ON SPAIN ELECTRICITY SHORTFALL

25 MAY 2023



OUR TEAM

The formidable team of expert data scientists and data engineers



Omolaja Adodo

*Team leader
Data Scientist*



Chikezie Ezenna

Data Engineer



Sumaya Hassan

Data Scientist



Mark Kasavuli

Data Engineer



Akanni Peterson

Data Scientist



**Madubuko
Onyebuchi**

Data Scientist



ABOUT US

At RR Analytics, we specialize in transforming challenges into growth opportunities. With a focus on innovation and expertise, we help companies thrive in today's dynamic business landscape. Harnessing the power of advanced technology and data-driven insights, we provide tailored solutions that drive remarkable results.





TABLE OF CONTENTS

01

Introduction

02

Problem Statement

03

Exploratory Data Analysis (EDA)

04

Data Engineering

05

Model Evaluation

06

Conclusion



INTRODUCTION

- Population growth in Spain has led to intermittent power deficits.
- Energy ministry has shared 3-year data from 5 cities for analysis and predictive modeling.





PROBLEM STATEMENT

- Insufficient electricity supply poses a significant challenge to the productivity and economic growth of a nation.
- The impact of electricity shortages hampers various industries, stifles innovation, and impedes overall development.
- Resolving these shortfalls is crucial to unlocking the full potential of the nation's economy and fostering sustainable GDP growth.



Data Architecture Overview

- Training data consists of 46 numerical features and 3 categorical features.
- One column in the dataset has 2068 missing entries, accounting for 23.6% of the total data.
- The dataset comprises a total of 8763 entries.





Data Architecture Overview

The features presented capture mainly Wind, Rainfall, Temperature, Snow, Humidity, Pressure and Cloud Cover all in three hour intervals for the following cities:

- Barcelona
- Bilbao
- Madrid
- Seville
- Valencia



EXPLORATORY DATA ANALYSIS

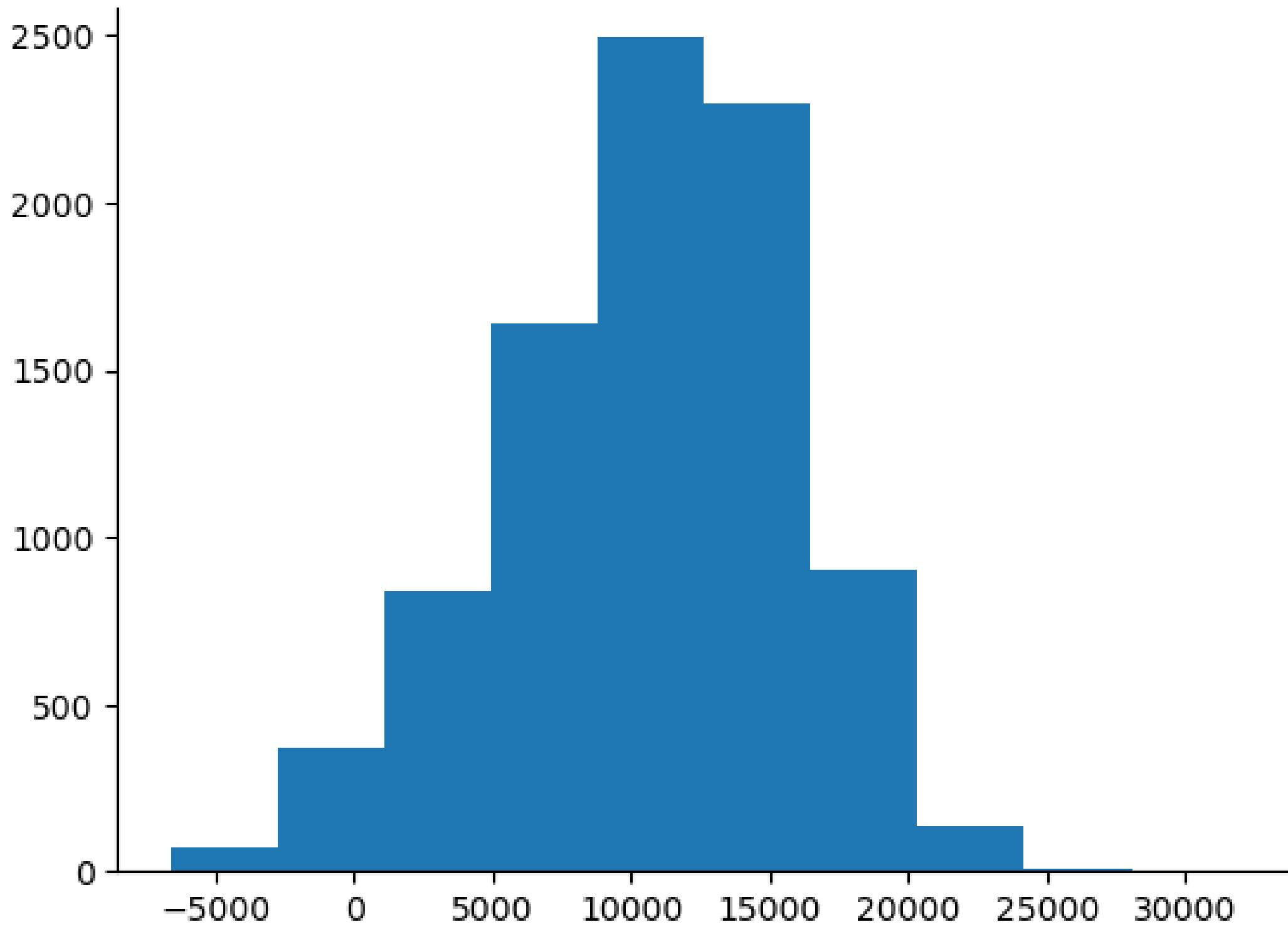
Presence of some features with some noticeable errors such as:

- Barcelona pressure contains a max value of 3687.56 hPa with the highest recorded pressure is 1,085 hPa
- Valencia wind speed contains a max value 52 m/s while the highest recorded wind speed is 20 m/s



Exploratory Data Analysis

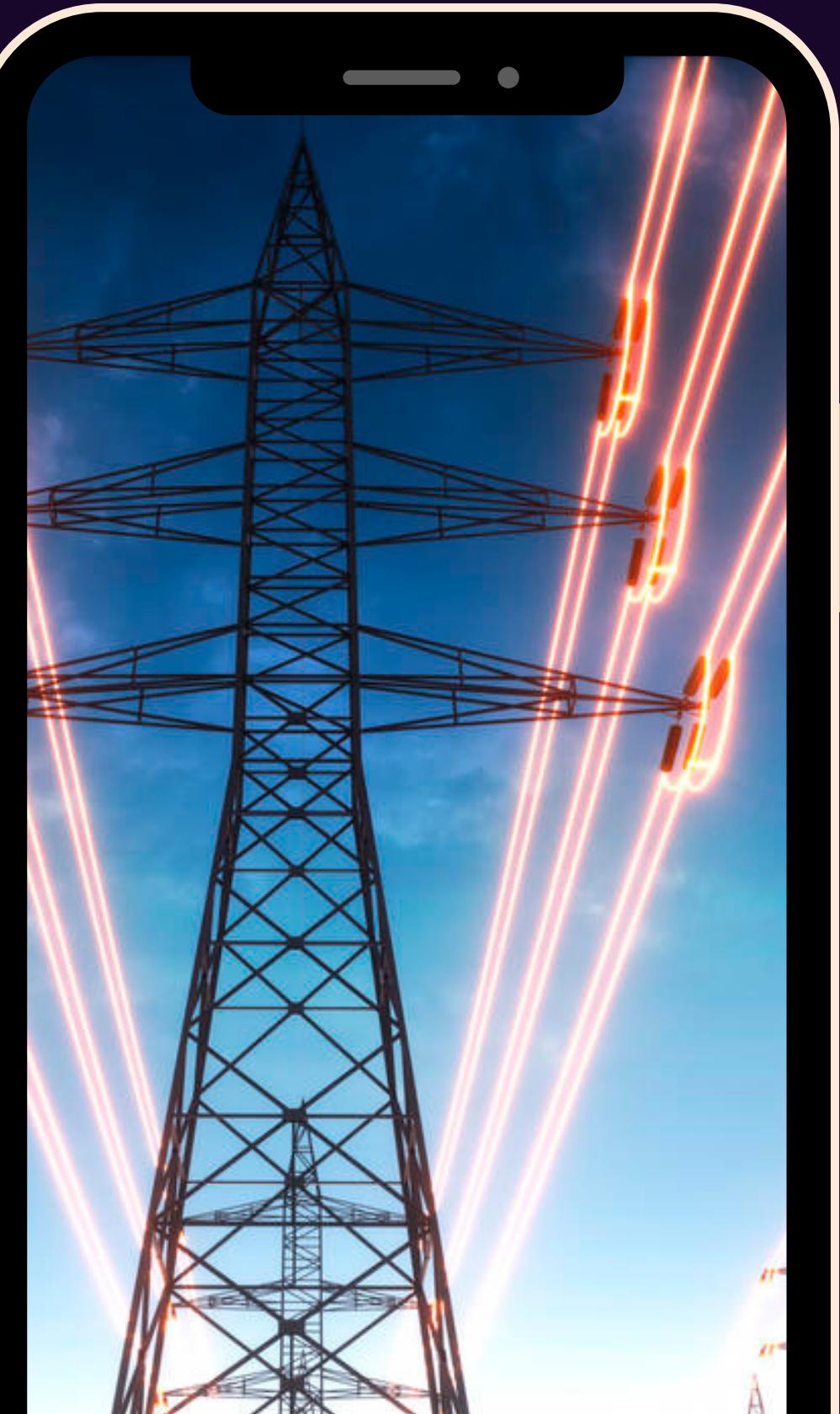
- The target variable was found to have a normal distribution with a skewness of -0.38.





Exploratory Data Analysis

- Time, Sevilla Pressure and Valencia wind deg has an object datatype while the rest are numerical.
- Presence of high correlation between features.

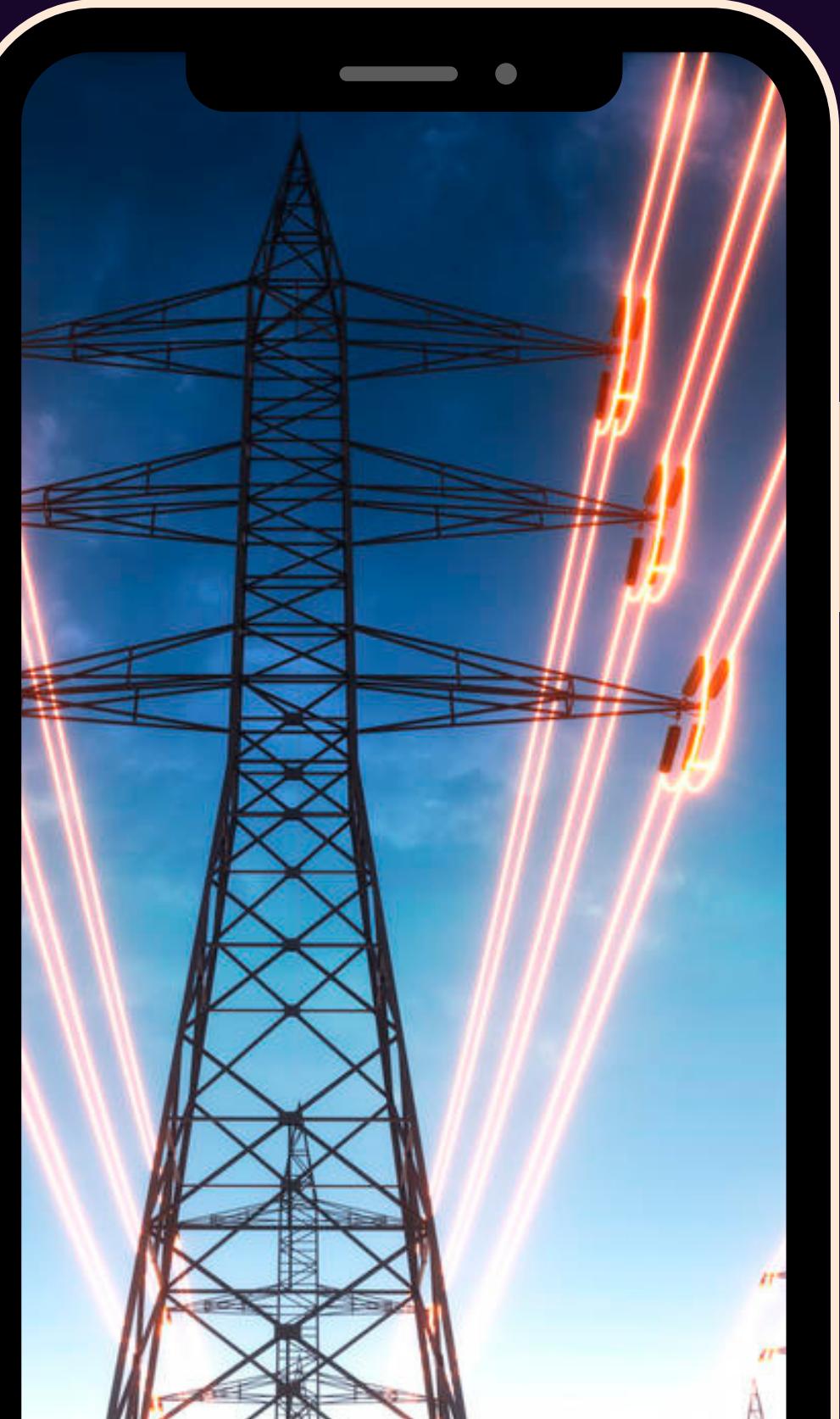


Presence of Correlation

- Feature Bilbao_temp_max and Bilbao_temp_min are direct reflection of the Bilbao_temp.
- The same applies to Barcelona_temp_max, Barcelona_temp_min, Madrid_temp_max, Madrid_temp_min, Seville_temp_min, Valencia_temp_min



DATA ENGINEERING



- Removal of correlated columns
- Replacement of missing values with mean
- Conversion of object datatypes to numeric datatypes
- Break down of the time feature
- Removal of unnecessary features
- Scaling the data.



MODEL EVALUATION

A photograph showing a group of people's hands reaching up from the bottom of the frame to hold several light-colored wooden puzzle pieces. The puzzle pieces are interlocking, symbolizing teamwork and completion. The background is bright and slightly blurred.

Multiple Linear
Regression

Decision Tree
Regression

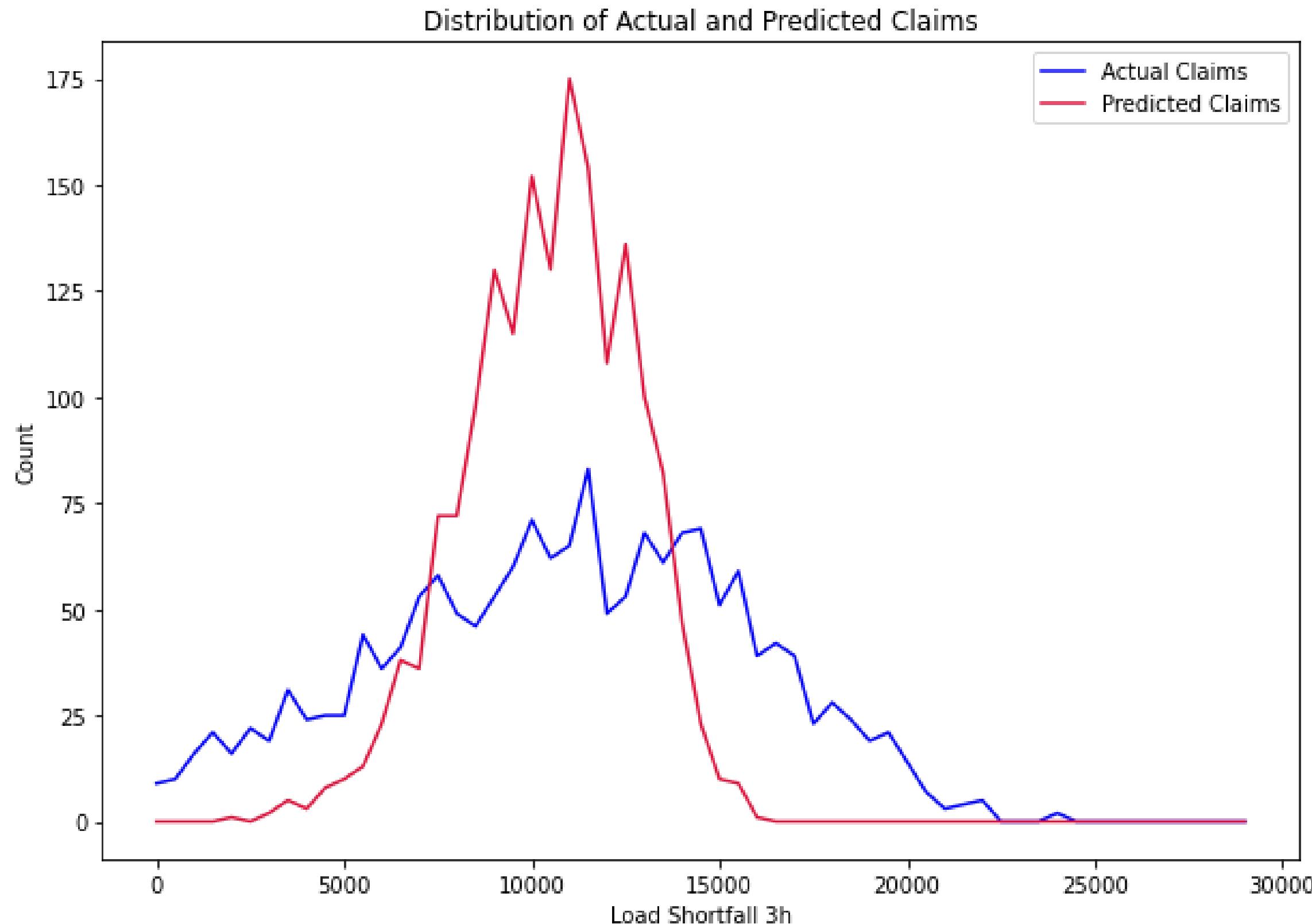
Random forest
Regression

XGBoost Regressor



Multiple Linear Regression Model

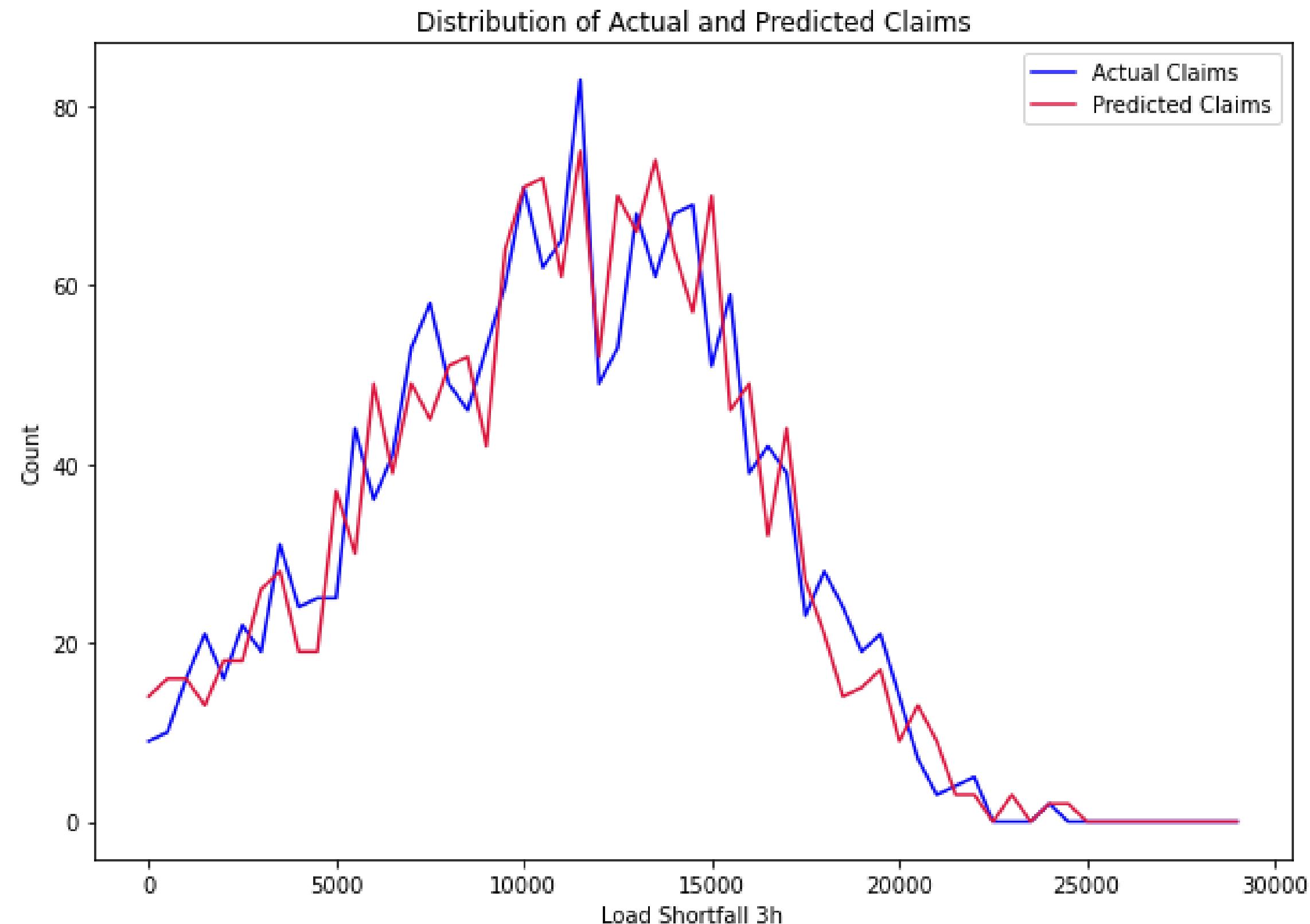
Parameter	Performance
Mean Squared Error	23529618.387
Mean Absolute Error	3870.351
r2_Score	0.165





Decision Tree Regression Model

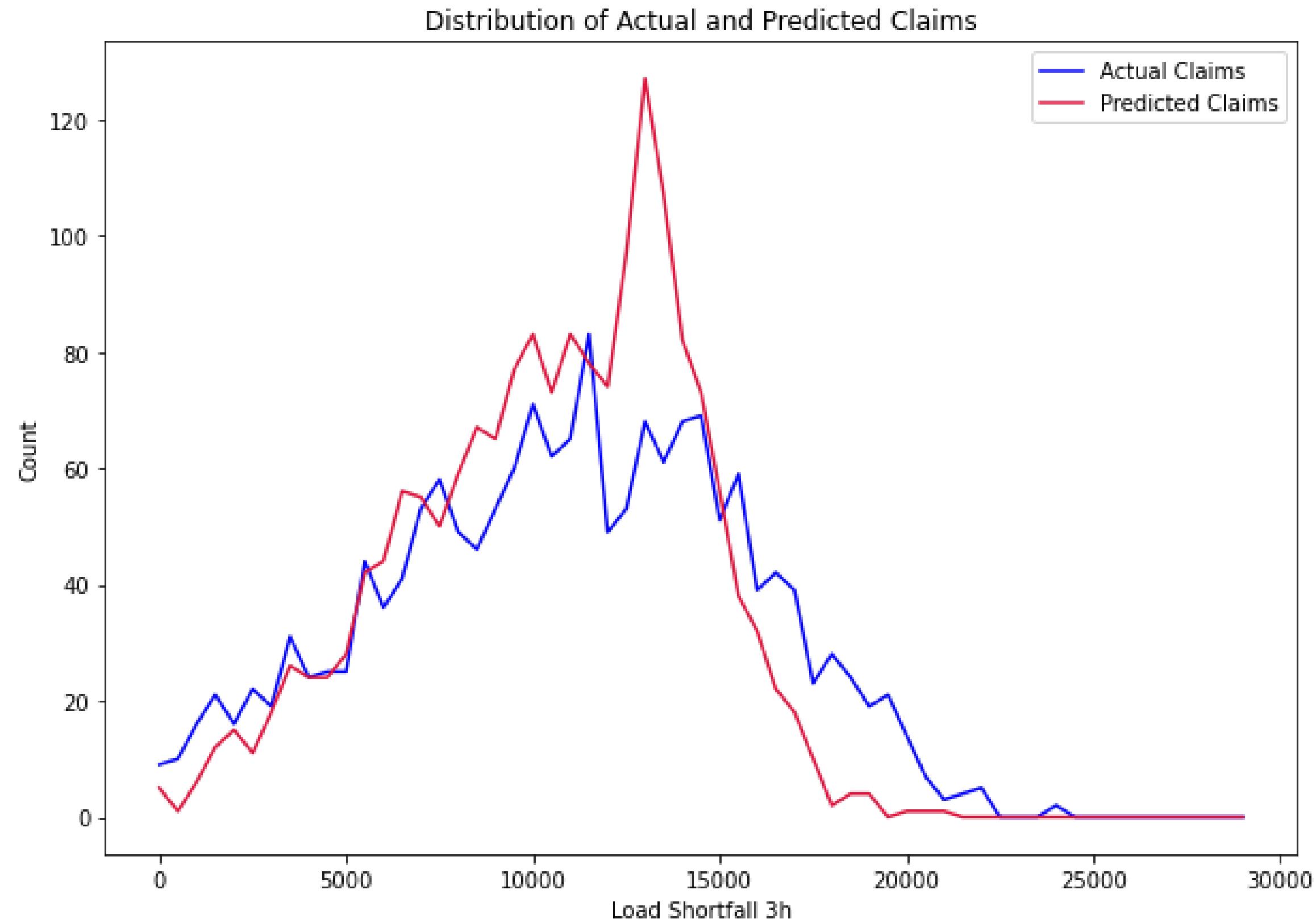
Parameter	Performance
Mean Squared Error	19268549.559
Mean Absolute Error	3237.070
r2_Score	0.316





Random forest Regression Model

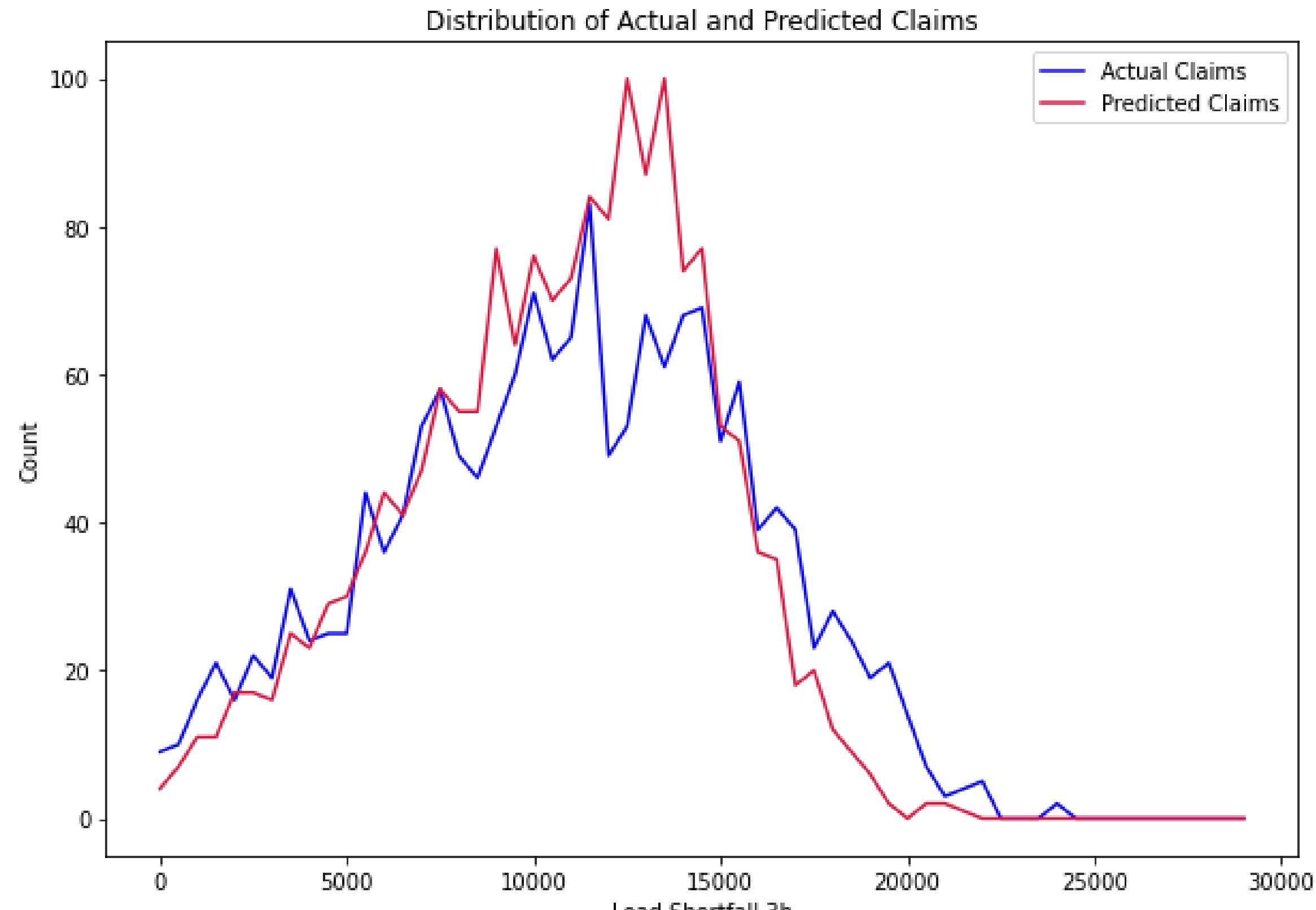
Parameter	Performance
Mean Squared Error	10032377.446
Mean Absolute Error	2491.397
r2_Score	0.644





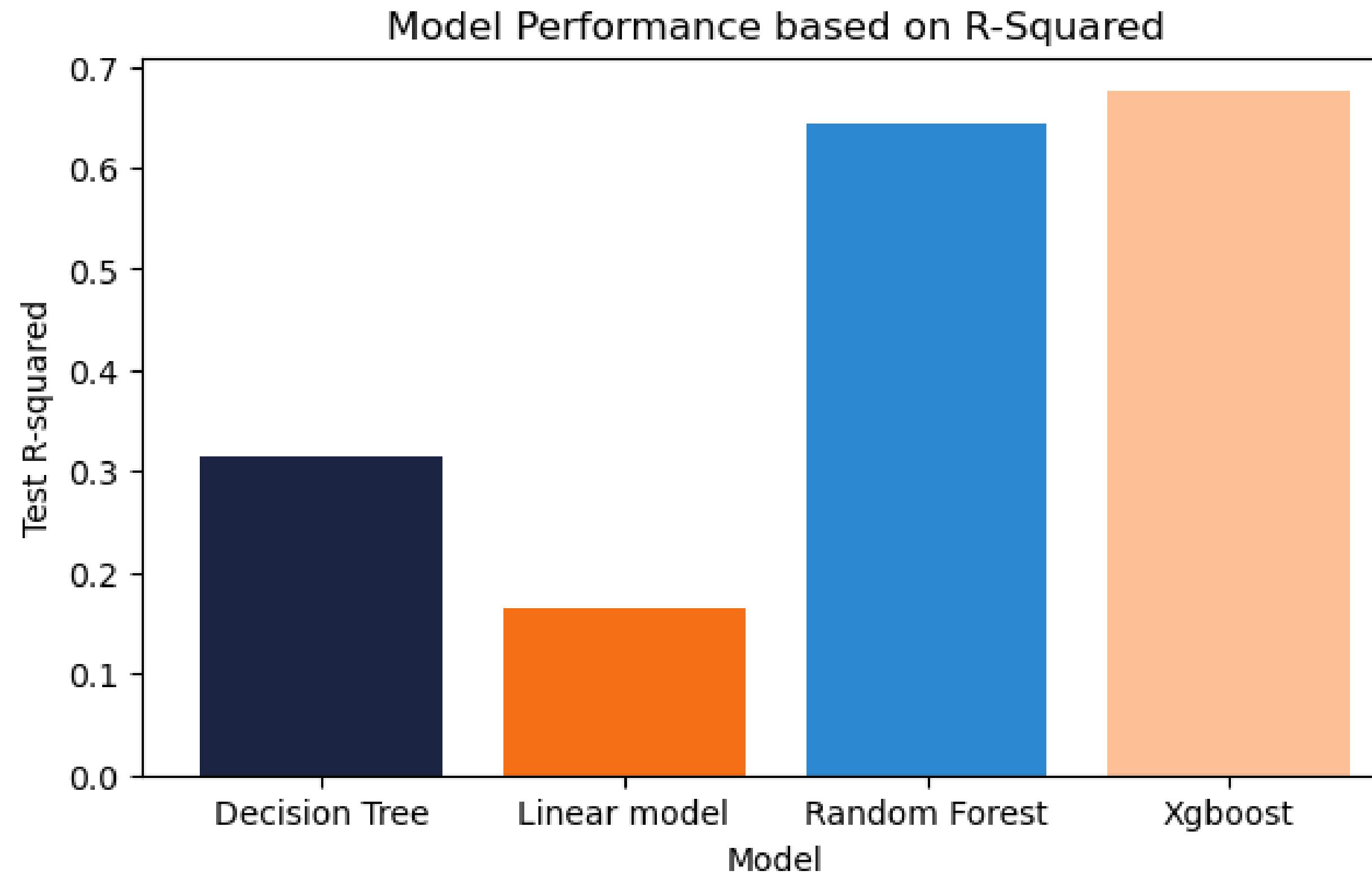
XGBoost Regressor Model

Parameter	Performance
Mean Squared Error	9148735.795
Mean Absolute Error	2371.413
r2_Score	0.675





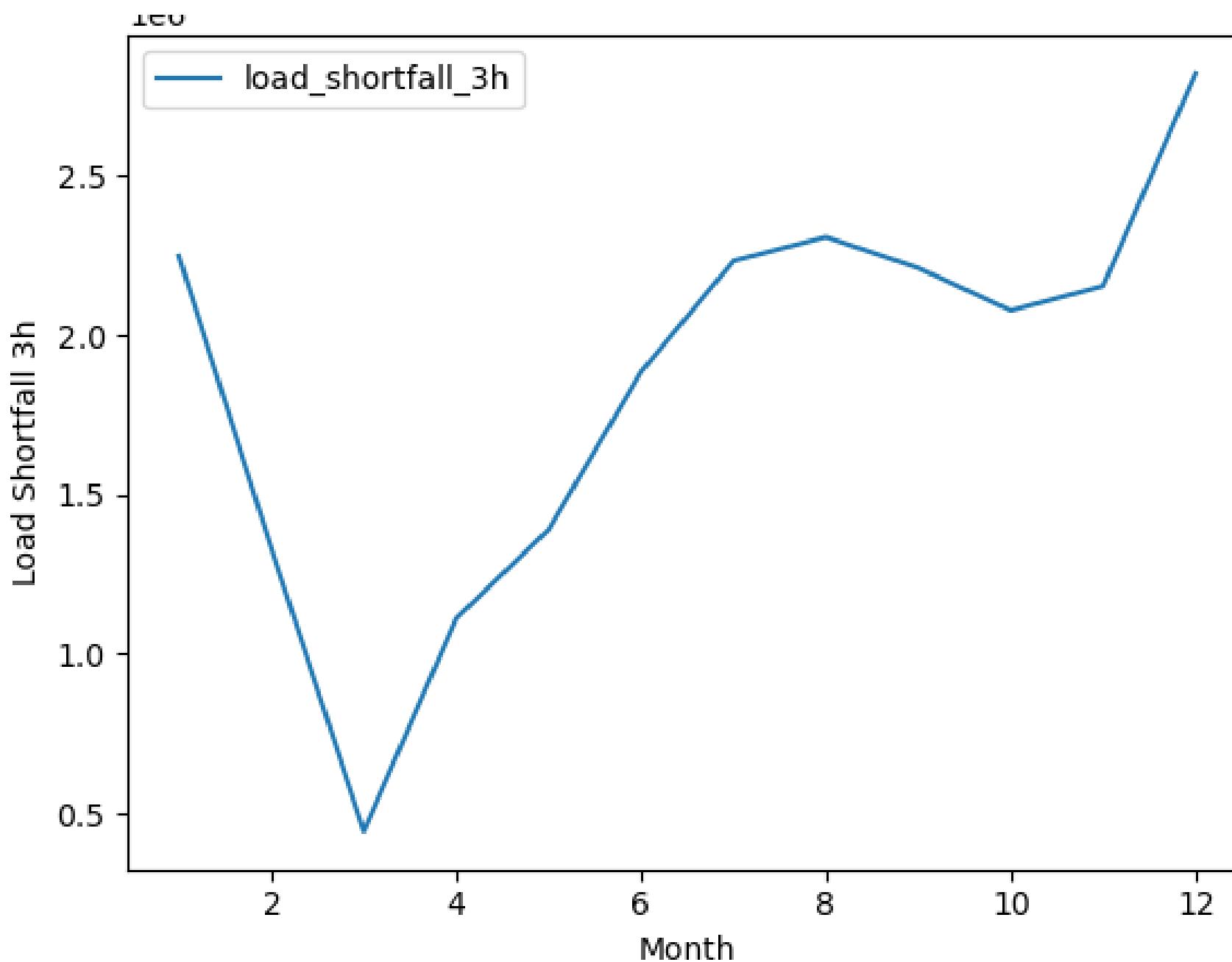
Model Performance





PREDICTIVE INSIGHT

LOAD_SHORTFALL PER MONTH:



- JANUARY THROUGH MAY: LITTLE LOAD SHORTAGE AND LOW DEMAND FOR RENEWABLE ENERGY.
- JUNE TO DECEMBER: SUBSTANTIAL LOAD GAP AND INCREASED DEMAND FOR RENEWABLE ELECTRICITY.



CONCLUSION

- The XgBoost Regressor proved the best model to be used for the prediction of the 3 hourly load shortfall with an accuracy score of 67%
- We urge that the Spanish government adopt renewable energy sources between June and December to ensure a reliable and demand-satisfying energy supply.



The background features a large, light blue triangle at the bottom right. Overlaid on it are several smaller triangles: a purple one pointing upwards from the bottom center, a light blue one pointing downwards from the bottom right, and a pink one pointing upwards from the bottom left. A thin horizontal bar, transitioning from light blue to light purple, runs across the top of the slide.

**THANK
YOU**

