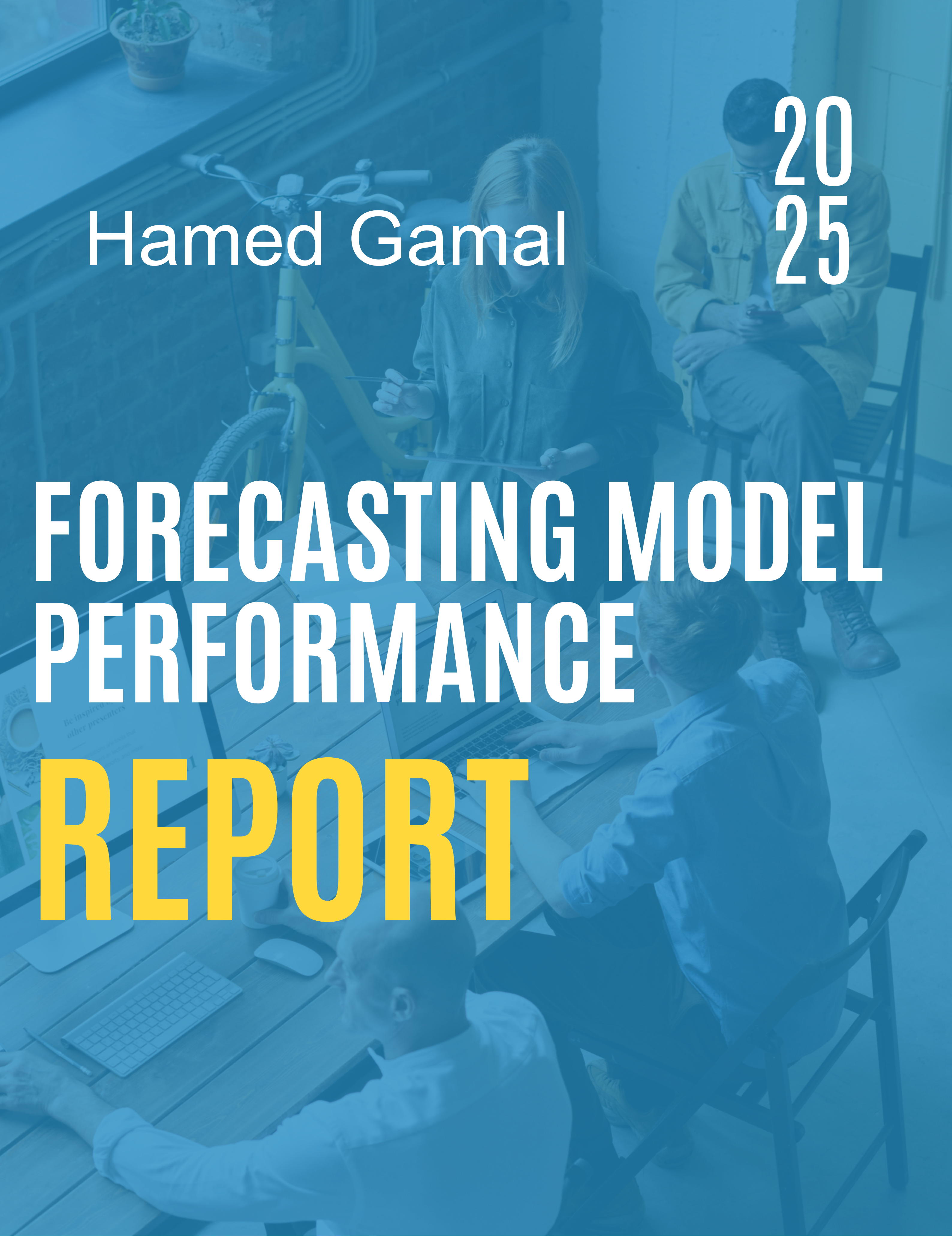

Hamed Gamal

2025

# FORECASTING MODEL PERFORMANCE REPORT

# Table of Contents

# Objective & Key Evaluation Metrics

## Objective

The purpose of this report is to evaluate and compare the performance of various forecasting models applied to a time series dataset. Model performance was assessed using standard statistical and regression metrics. The ultimate goal is to identify the most suitable model for production deployment based on accuracy, generalization capability, and interpretability.

## Key Evaluation Metrics

| METRIC | DEFINITION |
|---|---|
| RMSE (ROOT MEAN SQUARED ERROR) | The square root of the average of squared differences between predicted and actual values. Lower is better. |
| MAE (MEAN ABSOLUTE ERROR) | The average absolute difference between predicted and actual values. Lower is better. |
| MAPE (MEAN ABSOLUTE PERCENTAGE ERROR) | The average percentage error between predicted and actual values. Useful for interpretability. |
| $R^2$ SCORE (COEFFICIENT OF DETERMINATION) | Indicates how well the model explains the variance in the data. Ranges from $-\infty$ to 1.0. |
| EXPLAINED VARIANCE SCORE | Measures the proportion of the variance in the target variable that is explained by the model. |

# Model Performance Summary

| Model | RMSE | MAE | MAPE (%) | R² Score | Explained Variance |
| --- | --- | --- | --- | --- | --- |
| **Random Forest** | 7.94e-06 | 1.62e-07 | 3.16e-05 | 0.9999999999 | 0.9999999999 |
| **CatBoost** | 0.00026 | 0.00020 | 0.0309 | 0.9999999321 | 0.9999999321 |
| **XGBoost** | 0.00282 | 0.00196 | 0.2960 | 0.9999920557 | 0.9999920558 |
| **ElasticNet** | 0.70577 | 0.62929 | 65.13 | 0.50399 | 0.50399 |
| **LSTM** | 1.03641 | 0.89639 | 91.85 | -0.0694 | -0.0026 |
| **WaveNet** | 1.00227 | 0.92572 | 99.41 | ~0.0000 | 0.0000 |
| **ARIMA** | 1.00211 | 0.92722 | 100.02 | ~0.0000 | ~0.0000 |
| **SARIMA** | 1.00254 | 0.92903 | 100.77 | -0.00085 | -0.00072 |
| **Prophet** | 1,213,916.66 | 1,193,747.99 | 157,369,051.6 | -1.44e+12 | -4.74e+10 |

# Model Descriptions and Analysis

## Random Forest

An ensemble learning method that aggregates predictions from multiple decision trees. It showed near-perfect metrics, indicating likely **overfitting**, possibly due to using the same dataset for both training and testing. Further validation (e.g., cross-validation or out-of-sample testing) is required before production use.

## CatBoost

A gradient boosting model that handles categorical features automatically and is optimized for high accuracy and speed. It achieved **exceptionally low error rates** and a near-perfect $R^2$, making it a **top choice** for deployment with high trustworthiness.

## XGBoost

A well-established and scalable gradient boosting algorithm. It produced excellent results with a high $R^2$ score and minimal errors, making it a reliable option, especially when model interpretability and speed are important.

## ElasticNet

A linear model that combines L1 and L2 regularization. While it performed moderately well, its inability to model non-linear relationships limits its effectiveness for complex forecasting problems.

## LSTM

A deep learning model designed to capture long-term dependencies in sequences. Its performance was **weak**, with negative $R^2$ and high error metrics, possibly due to limited data, underfitting, or insufficient tuning.

## WaveNet

A deep convolutional neural network originally used in speech synthesis. Despite its architectural complexity, it **failed to outperform** simpler models, likely due to overfitting or lack of data.

# Model Descriptions and Analysis

## ARIMA

A classical statistical model tailored for stationary time series. It failed to capture the variance in the data and showed **poor generalization**, with nearly zero explanatory power.

## SARIMA

An extension of ARIMA incorporating seasonality. Its performance was slightly worse than ARIMA, indicating that the dataset may not have strong seasonal components or that model tuning was inadequate.

## Prophet

Developed by Meta (Facebook) for business-oriented time series forecasting. It **completely failed** on this dataset, producing extremely high errors and a highly negative $R^2$, indicating a total breakdown in predictive capability.

# Recommended Models for Production

**CatBoost** is the best performer across all key metrics. Its combination of **accuracy**, **robustness**, and **low complexity in tuning** makes it ideal for production deployment.

**XGBoost** also offers excellent accuracy and flexibility and can be a strong backup or complementary model, particularly when computational efficiency and interpretability are required.

**Note:** Although Random Forest appears to perform perfectly, its results strongly suggest overfitting and must not be used in production without rigorous validation using unseen data.

2025

With every step we take in this journey of learning, your support is the light that makes it all worthwhile. From the heart — thank you.

Hamed Gamal