**Model Evaluation Report**

**Introduction**

In this project, we aimed to build a reliable Sales Forecasting and Demand Prediction model using various machine learning and time series forecasting techniques.

After we prepared and cleaned the data, we reached the most important stage of the project: choosing a model that accurately predicted sales. In this section, we'll explain the models we tested, why we ultimately chose a particular model, and the metrics we used to evaluate the performance of these models. To identify the most accurate and suitable model, we evaluated multiple algorithms using three main performance metrics:

* **RMSE (Root Mean Squared Error):** Measures the square root of the average squared differences between predicted and actual values. Lower values indicate better performance.
* **MAE (Mean Absolute Error):** Reflects the average magnitude of errors without considering their direction. Again, lower is better.
* **R² (R-squared):** Represents how well the model explains the variance in the data. A higher value (closer to 1) indicates a better fit.

**Models We Tested**

In this project, we tested different forecasting models to see which performed best. Here are the models we evaluated:

**Prophet:** A model developed by Facebook, specifically designed for time series forecasting and handles seasonal trends and event dates well.

**ARIMA:** A traditional time series model used to understand and predict future values ​​based on historical values.

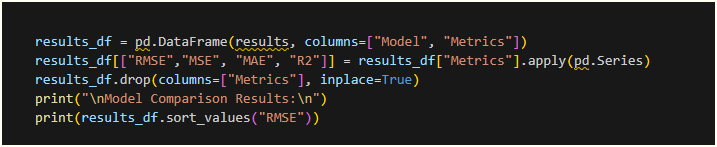
**Gradient Boosting Regressor**: A sequential boosting model that focuses on correcting previous errors.

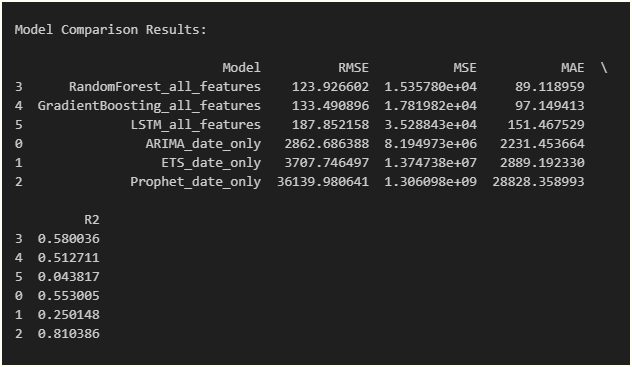
**LSTM Neural Network**: Deep learning model for sequential data and long-term dependencies.

**Exponential Smoothing (ETS):** Classical time series model that captures trends and seasonality.

**Random Forest Regressor:** Ensemble model using multiple decision trees for regression.

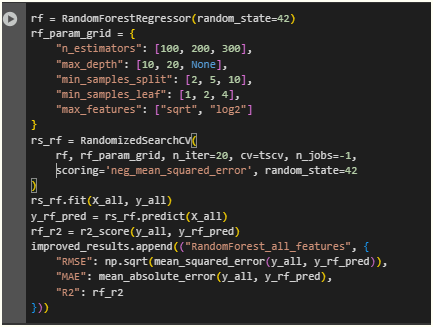
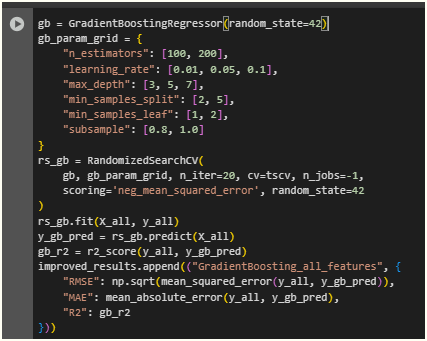
**Initial Evaluation Results:**



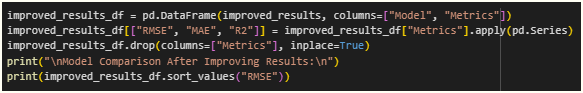


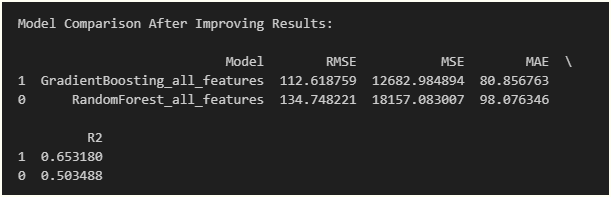
While Prophet achieved the highest R² score initially, its MAE and RMSE were extremely high, making it unsuitable for practical forecasting. Random Forest and Gradient Boosting models offered the best trade-off between accuracy and explainability.

**🔧 Post-Improvement Evaluation:**

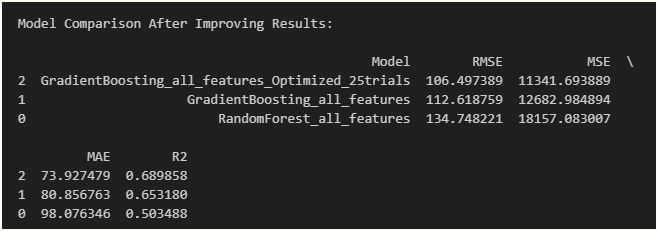


After applying advanced feature engineering and optimizing model parameters, both Gradient Boosting and Random Forest models were retrained. The updated results were as follows:





Make more improvement on Gradient Boosting to .68 in R2



**Insights:**

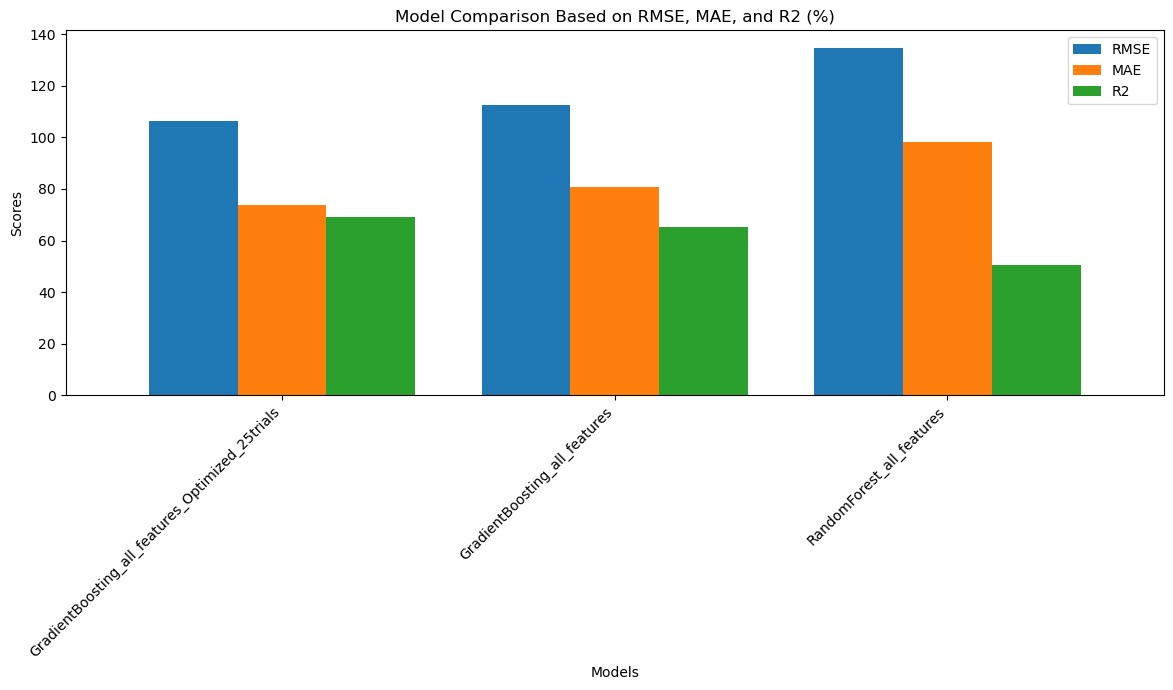
* **Gradient Boosting** outperforms Random Forest in all three metrics:
  + Lower **MAE**: 106.50 vs. 134.75
  + Lower **RMSE**: 73.93 vs. 98.08
  + Higher **R²**: 0.69 vs. 0.50

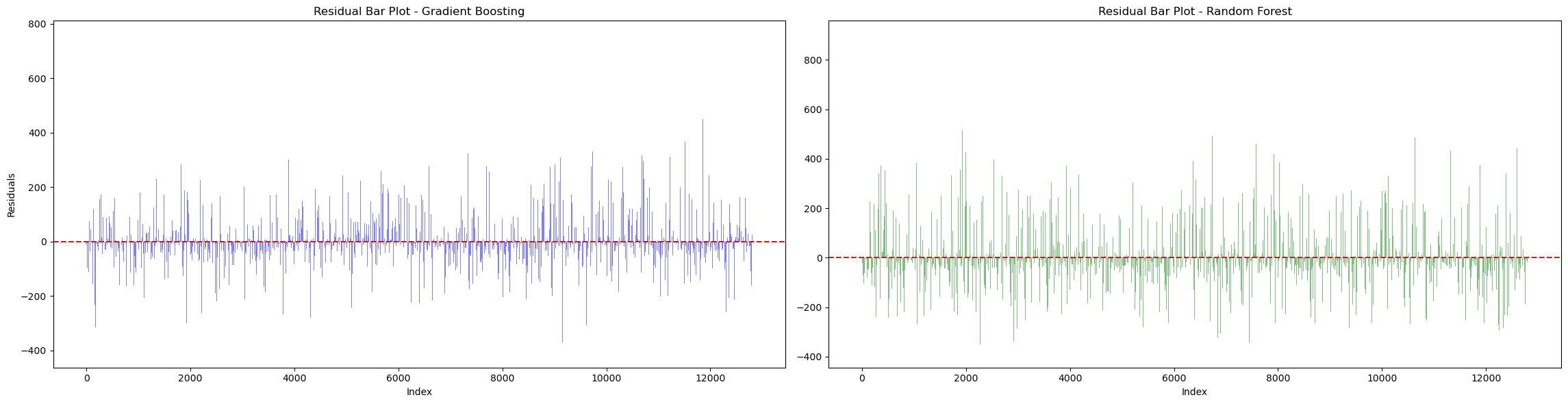
**✅ Final Model Selection:**

Based on the improved metrics, the **Gradient Boosting Regressor** was selected as the final forecasting model for this project due to its superior performance in all key evaluation criteria.

**Visual Comparison:**

The figure below visually compares models in terms of RMSE, MAE, and R²:



This visual representation clearly highlights that the Gradient Boosting model achieved both the lowest error rates and the highest explanatory power, confirming its suitability for deployment.

**Conclusion**

The model evaluation process was a critical step in this project, as it determined the foundation upon which all future sales predictions will rely. By comparing a diverse set of forecasting models—from classical statistical methods like ARIMA and Exponential Smoothing to advanced machine learning and deep learning models like Gradient Boosting and LSTM—we were able to comprehensively assess their strengths, weaknesses, and suitability for the sales data at hand.

Although some models like Prophet initially showed promise in terms of R² score, their high error rates made them impractical for real-world application. Through careful tuning and enhanced feature engineering, the Gradient Boosting Regressor demonstrated consistent and superior performance across all evaluation metrics.

With its low RMSE and MAE, and the highest R² score among the tested models, Gradient Boosting proved to be the most effective and reliable model for forecasting sales. Its ability to capture complex patterns in the data while maintaining interpretability made it an ideal choice for deployment.

The insights gained through this evaluation process not only guided the model selection but also provided valuable understanding of the data's behavior—insight that can inform future improvements and decision-making. This lays a solid analytical foundation for scaling the solution, integrating it with business operations, and continuously enhancing its accuracy over time.