

R&D Report: Time Series Analysis and Model Selection for Forecasting Unit Sales

Introduction: This report provides an overview of the research and development (R&D) process conducted to select an appropriate model for forecasting unit sales on Amazon. The primary goal of this project is to predict the number of units sold for various items using historical sales data. After evaluating various time series models, we selected **Prophet**, an additive model developed by Facebook, for its effectiveness in handling time series data with daily observations, missing values, and potential seasonality effects.

1. Overview of Time Series Models Considered:

Several time series forecasting models were considered during the R&D phase, including:

ARIMA (Autoregressive Integrated Moving Average):

- Pros: Well-suited for univariate time series with trend and seasonality, widely used, interpretable.
- Cons: Requires data stationarity, complex model selection, does not handle missing data well.

Exponential Smoothing (ETS):

- Pros: Good for data with trends and seasonal patterns, relatively simple.
- Cons: Limited in handling more complex seasonality and lacks flexibility in modelling external regressors.

Random Forest and Gradient Boosting:

- Pros: Can capture complex interactions between features, robust to overfitting.
- Cons: Not inherently designed for time series data, requires extensive feature engineering.

LSTM (Long Short-Term Memory) Networks:

- Pros: Effective for capturing long-term dependencies, good for sequential data.
- Cons: Requires a large amount of data, computationally expensive, difficult to interpret.

Prophet:

- Pros: Handles missing data, seasonality, and holiday effects well, simple and intuitive parameters, easy to fit and tune, handles outliers and missing data effectively.
- Cons: Designed primarily for daily data, may not perform well with data that does not exhibit clear seasonal patterns.

2. Detailed Analysis of Prophet:

Why Prophet?

- Ease of Use: Prophet is designed for business analysts and data scientists, making it easy to implement and interpret without requiring extensive time series expertise.

- Handling of Missing Data: Unlike ARIMA and other classical models, Prophet can handle missing data and large gaps between observations, which is common in sales data.
- Additive Model: Prophet uses an additive model where non-linear trends are fit with seasonality and holidays, making it suitable for sales data that often exhibit such patterns.
- Scalability: Prophet is scalable and can handle large datasets efficiently, which is crucial for this project involving multiple items with potentially thousands of records.
- Interpretability: The model provides intuitive and interpretable parameters such as growth rate, seasonality's, and holiday effects, making it easier to understand and communicate the results.

Model Implementation:

- Data Preparation: Data was pre-processed to handle missing values and create time-based features (e.g., day of the week, month).
- Model Fitting: The Prophet model was trained on the historical sales data, with hyperparameters tuned to optimize forecasting performance.
- Evaluation: The model was evaluated using the Mean Squared Error (MSE) on a validation set to ensure it generalizes well to unseen data.

3. Model Evaluation and Results:

- Training Performance: The model was trained on the historical sales data, capturing the overall trend and seasonality effectively.
- Validation Performance: On the validation set, Prophet demonstrated strong performance, with an MSE of , outperforming other models considered during the R&D phase.
- Interpretability: The model provided clear insights into the factors influencing sales, such as seasonal peaks and the impact of specific holidays.

4. Conclusion:

Based on the research conducted and the performance metrics obtained, Prophet was selected as the final model for forecasting unit sales. Its ability to handle seasonality, holidays, and missing data, combined with its ease of use and interpretability, made it the most suitable choice for this project.

Future Work:

- Hyperparameter Tuning: Further tuning of Prophet's parameters could be conducted to improve accuracy.
- Incorporation of External Factors: Future iterations of the model could include external regressors like economic indicators or marketing campaigns to enhance predictive accuracy.