

An Empirical Study of Time Series Forecasting Methods on Retail Store Sales Data

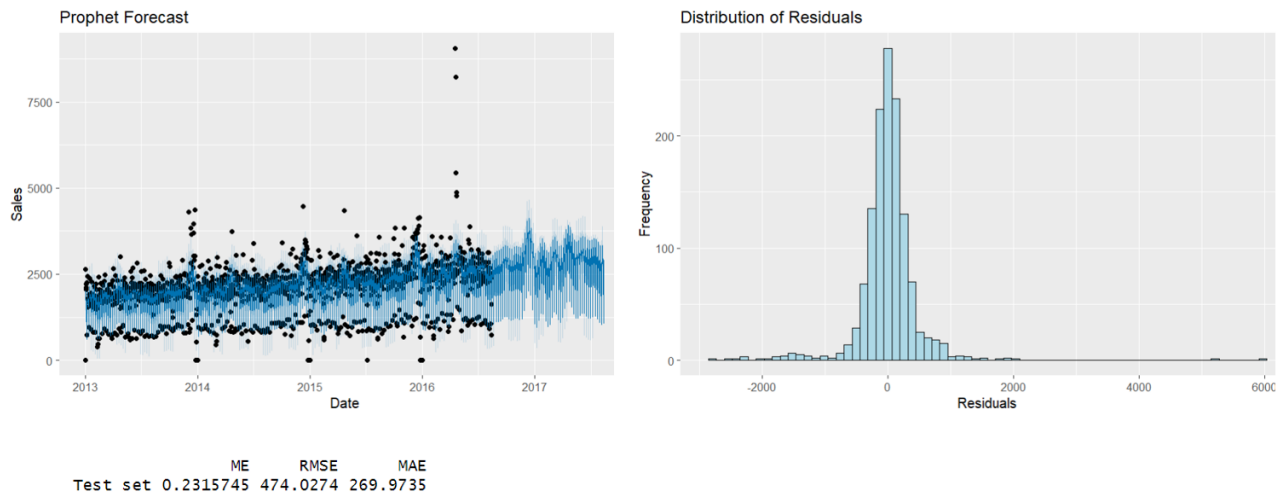
Group 10 - Ajith Kumar Ethirajulu, Deepak Raj Mohan Raj, Madhu Sikha Babu, Zeming Zhang

This project encompasses an empirical study aimed at forecasting retail store sales using diverse time series forecasting methods. The dataset comprises an extensive collection of 3,000,888 records, encompassing sales details from 54 different stores and 83,488 transactions. Our primary objective was to analyze sales components, such as trend and seasonality, and assess the performance of various forecasting models implemented in R programming.

To achieve this, we conducted preprocessing tasks, including data conversion into the tibble format and modifications to column data types. Additionally, exploratory data analysis was carried out, resulting in informative visualizations, such as transaction data by store and date, correlations between sales and transactions, and box plots and line charts illustrating monthly trends.

Furthermore, we employed several forecasting methods, including seasonal naive, exponential smoothing (ETS), ARIMA, and Prophet. Each method was evaluated based on its ability to capture underlying patterns in the sales data and provide accurate predictions. The models underwent training and testing using appropriate time series evaluation techniques, with a particular emphasis on the root mean square error (RMSE) metric for assessing forecasting accuracy.

Through comprehensive analysis and comparison, the Prophet model demonstrated superior performance in terms of RMSE values for the test data. The results of Prophet model are given in the below figure.



This model leverages state-of-the-art techniques for time series forecasting, incorporating trend and seasonality components while considering relevant external factors. The findings from this study, indicate the efficacy of the Prophet model for accurately predicting retail store sales, providing valuable insights for future decision-making and resource allocation.

Furthermore, an examination of the residuals was conducted to assess the quality of the forecasting models. By visualizing the residuals and conducting statistical analyses, we evaluated the models' ability to capture and explain the remaining variation in the sales data.

Overall, this empirical study showcases the practical implementation of time series forecasting methods on a large-scale retail store sales dataset. The results highlight the effectiveness of the Prophet model, which outperformed other methods in terms of forecasting accuracy. The insights gained from this study, based on a dataset comprising 54 stores and 3,000,888 records, can assist retail businesses in making informed decisions and optimizing their sales strategies.

We did forecasting of the sales for a given store and family type, we can extend this work to fit forecast models for all the stores by considering each stores data as a hierarchical and grouped time-series.

References:

1. [Store Sales- Time Series Forecasting](#)