Quantitative Sales Forecasting Using Time Series and Causal Models

Omkar Maurya
(24M1536)

Abstract

This report documents the application of various quantitative forecasting methods to predict product sales. The goal of this project is to explore and apply different time series and causal models on a dataset containing historical sales records of products from multiple stores.

1 Introduction

In a personal endeavor to learn about time series analysis and forecasting, this project explores various quantitative forecasting methods. The methods discussed in this report can be applied to forecast product sales, along with a detailed explanation of the different metrics used to evaluate the forecasts.

Goal: The goal of this project was to apply various quantitative methods, such as Time Series Models and Causal Models, to forecast the sales of the products available in the dataset obtained from Kaggle.

Models covered in this report include:

- Seasonal Naive Model
- Holt-Winters Model (Triple Exponential Smoothing)
- ARIMA and Seasonal ARIMA Models
- Linear Regression Model

2 Methodology

2.1 Dataset Description

The dataset contains historical sales records of 10 stores and 50 products, spanning from the year 2013 through 2017.

2.2 Models Implemented

2.2.1 Seasonal Naive Model

The Seasonal Naive Model assumes that the sales of a product will be the same as the previous season's sales. This simple model serves as a baseline for comparison with more complex models.

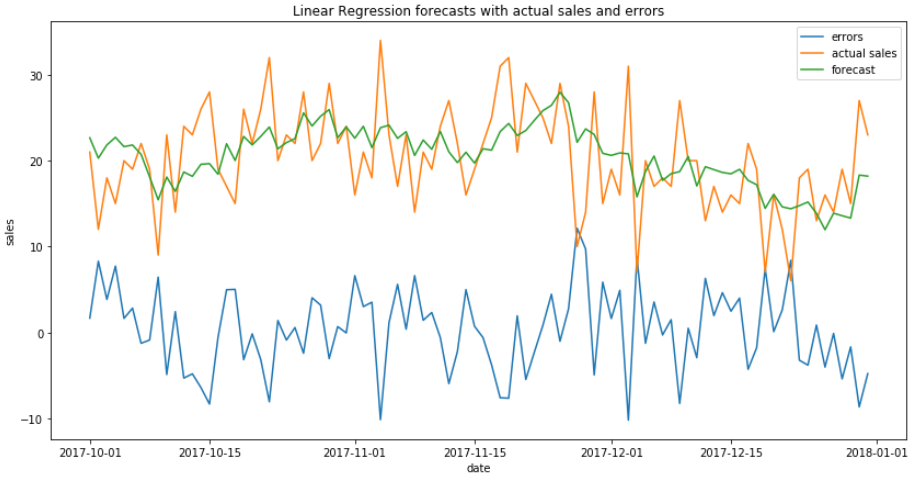

Figure 1: Seasonal Naive

2.2.2 Holt-Winters Model (Triple Exponential Smoothing)

The Holt-Winters Model extends exponential smoothing by adding a trend and seasonal component. It is well-suited for data with seasonality and trends.

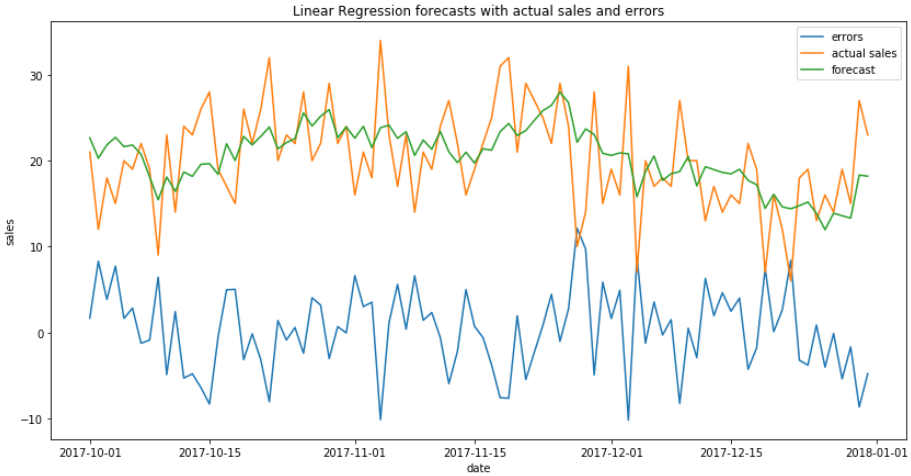

Figure 2: Holt-Winters Model

2.2.3 ARIMA and Seasonal ARIMA Models

ARIMA (AutoRegressive Integrated Moving Average) models are used to describe time series data that can be made stationary by differencing. Seasonal ARIMA (SARIMA) models extend ARIMA by explicitly modeling seasonal components.

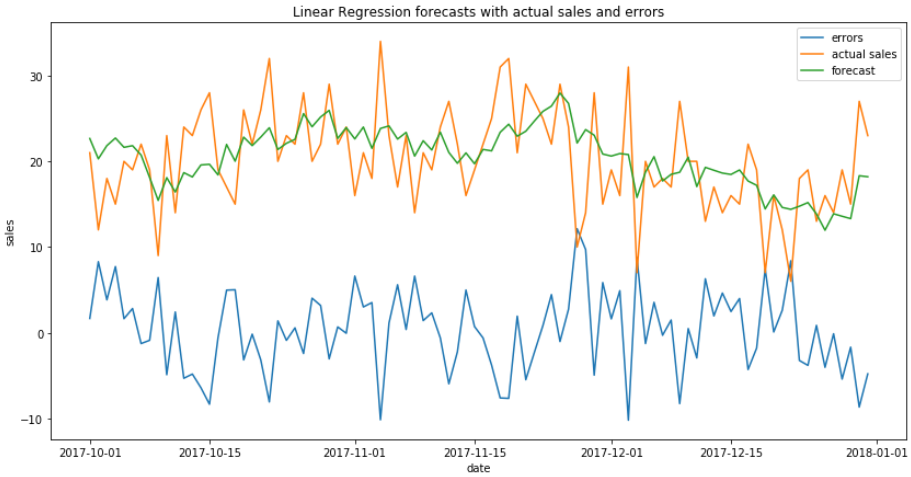

Figure 3: ARIMA and Seasonal ARIMA Models

2.2.4 Linear Regression Model

Linear Regression is a causal model that uses independent variables to predict the dependent variable, in this case, product sales. Feature selection techniques such as SelectKBest were used to choose the most relevant features.

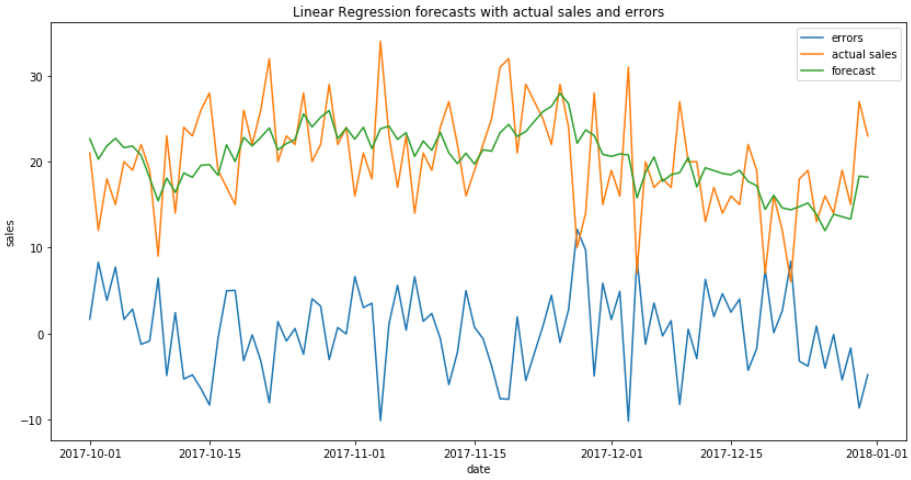

Figure 4: Linear Regression Model

3 Results

The performance of the models was evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results showed that Linear Regression performed better in capturing the trend and seasonality in the data.

4 Conclusion

In this project, several quantitative methods were applied to forecast product sales. The Linear Regression was found to be the most effective for this dataset, highlighting the importance of considering both trend and seasonality in sales forecasting.