Machine Learning Engineer Nanodegree

Capstone Project Report

Lee, Seang Mei

June 20, 2020

Title: Supply Chain Order Demand Forecasting

Domain Background

Enterprises are attaining double-digit improvements in forecast error rates, demand planning productivity, cost reductions and on-time shipments using machine learning today, revolutionizing supply chain management in the process.

Examples where data analytics and machine learning can be beneficial for supply chain management is within demand forecasting and warehouse optimization.

Accurate demand forecasting enables increased profitability, increased customer satisfaction, reduced inventory stockouts, reduced safety stock requirements and reduced product obsolescence costs.

Problem Statement

For this case, a company has gathered 7 years of their order demand data from 2011 to 2017. Like many other companies, this company wish to create more revenue by accurately master the order demand for coming months and thus make the right decision on their supply chain to fulfill the demand.

Datasets and Inputs

Data is getting from Kaggle. Historical Product Demand.csv - CSV data file containing product demand.

Data is uploaded to AWS S3 bucket. During the research of finding best model, output data in json file format can be stored in S3 bucket.

First, data preprocessing is kick off with checking the number of columns, data types, data size and then convert the date field to datetime as it is found that the date is stored as object. It will not be able to fit into the model later. After the date datatype conversion, proceed to find any null value and drop them. Only 1% of data contains null value and dropped from the data frame.

	Before	After
Warehouse	object	object
Product_Category	object	object
Date	**object**	**datetime64**
Order_Demand	int64	int64

Solution Statement

To produce a high accuracy demand forecast, we need to study it along a time series. As such, the machine learning solution for this problem will be using SARIMA - Seasonal Autoregressive Integrated Moving Average method for time series forecasting with univariate data containing trends and seasonality.

There is a function to create a time series which will consider leap year and non-leap year. Then charts are plotted to study the trend of order demand for a particular year.

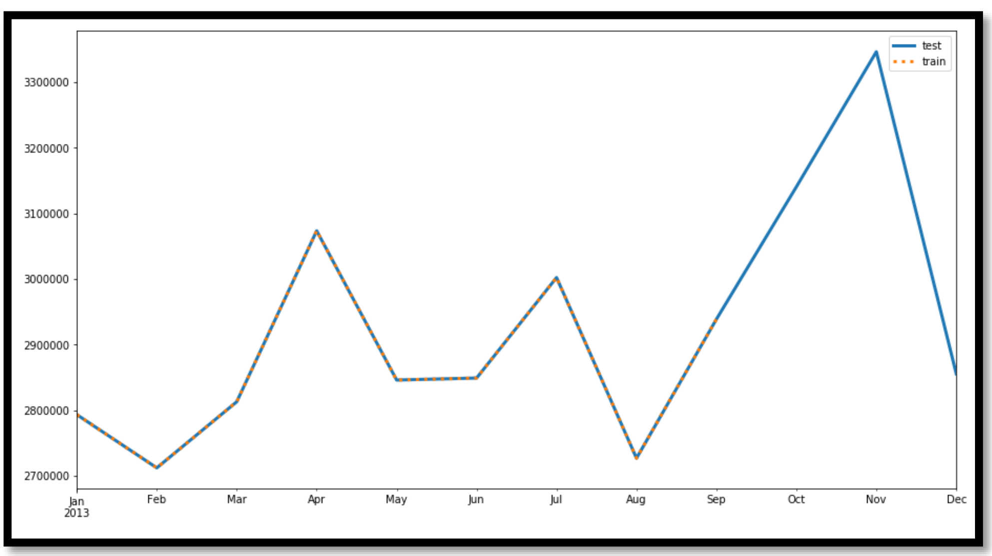

Fig1: Data show only 1 year of order demand

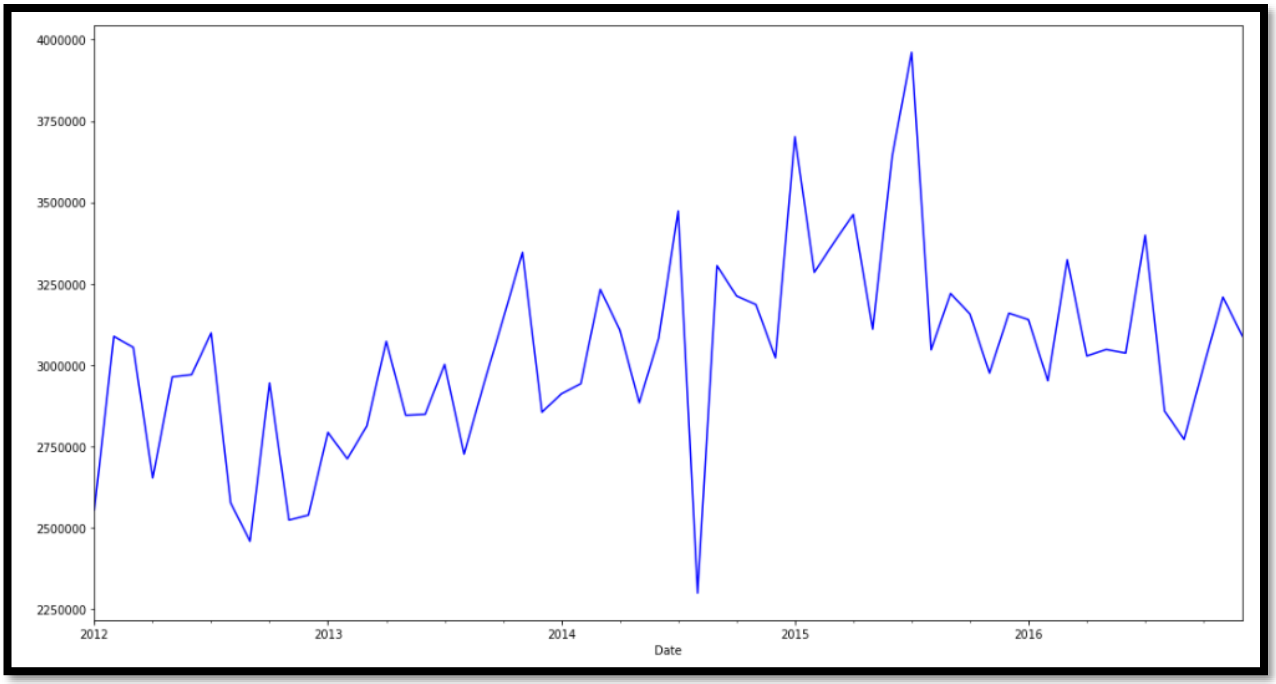

Fig2: plot the order demand data by year

To determine whether I should use ARIMA or SARIMA, a decomposition of data can help to check whether the data show a seasonal trend. From the third chart, I can basically decide that there is an obvious repetitive trend over the time. Therefore a SARIMA model should work with this dataset.

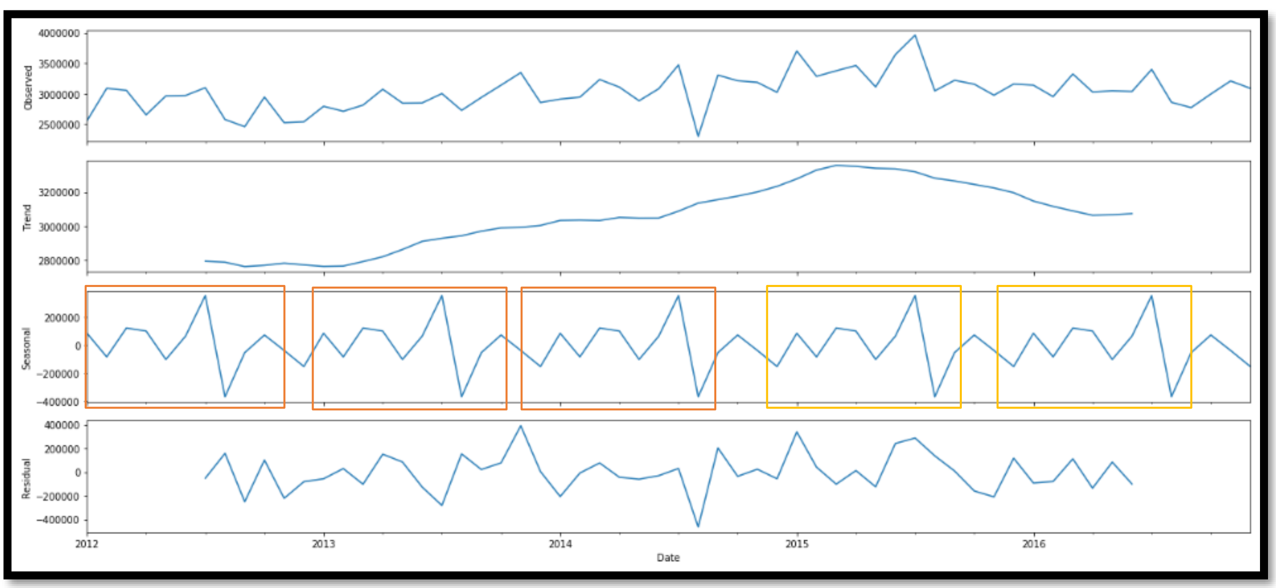

Fig3: plot the data to check decomposition

To train the data, there are three trend elements that require configuration.

They are the same as the ARIMA model; specifically:

p: Trend autoregression order.

d: Trend difference order.

q: Trend moving average order.

I get the best result with the following configuration:

Order: (0, 1, 2)

Seasonal Order: (1, 1, 0, 12)

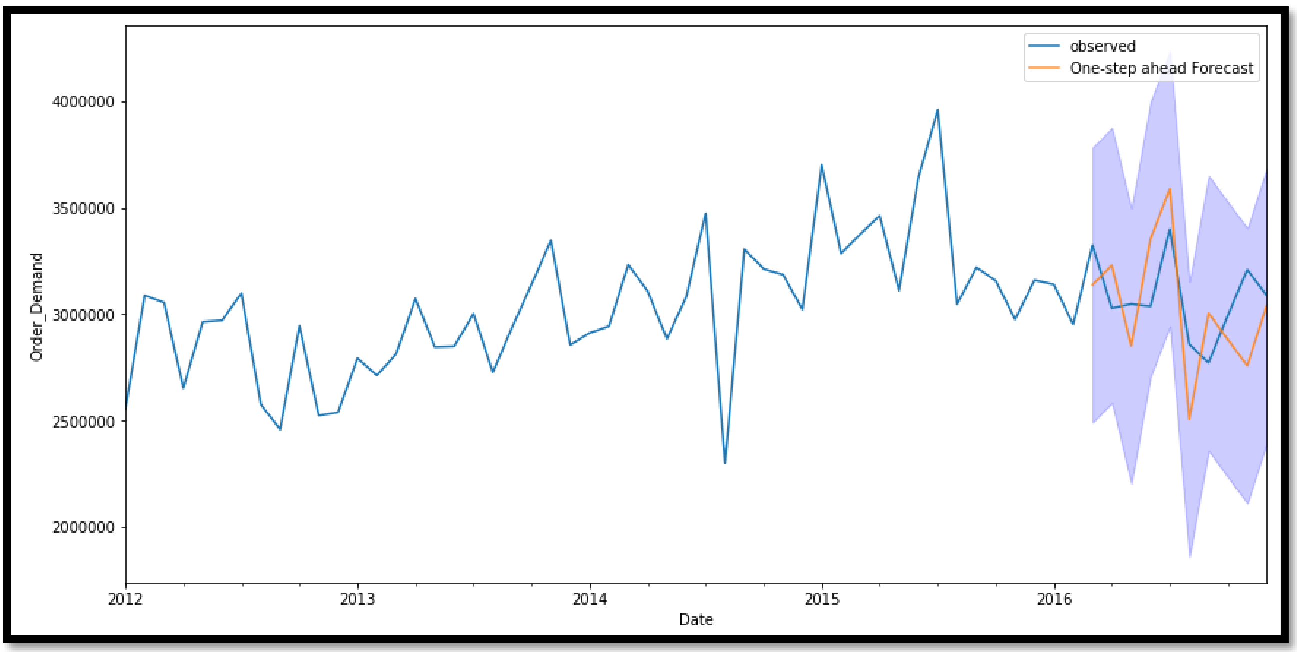

Fig4: fit the model

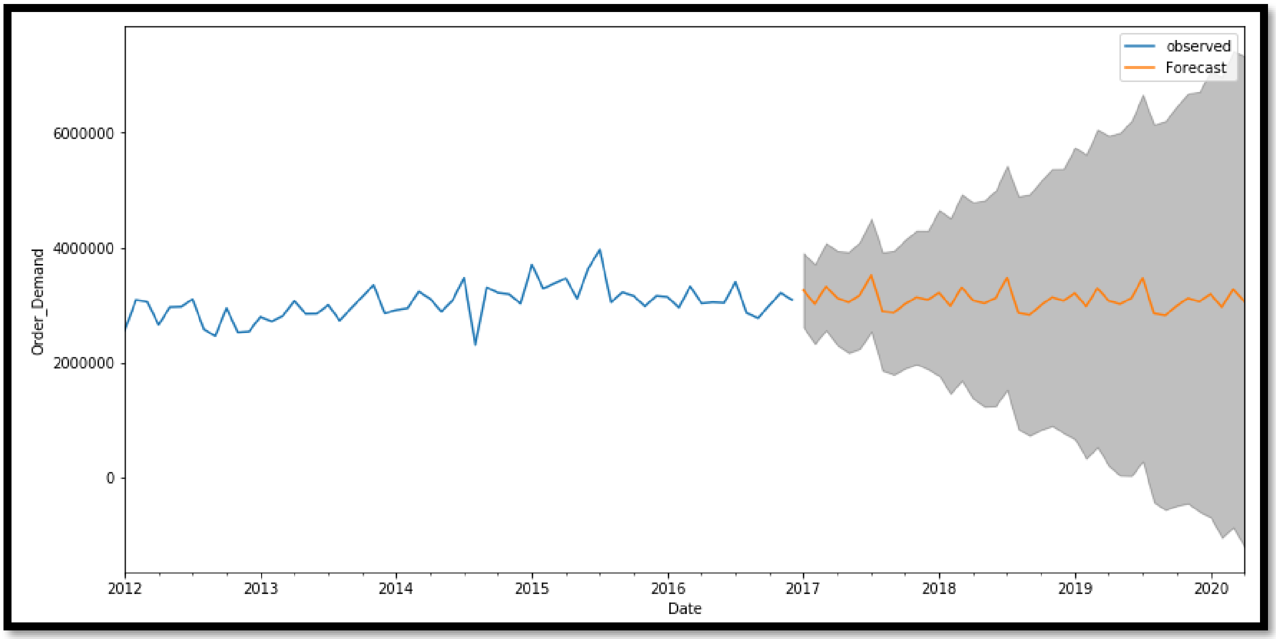

Fig5: forecast result for the next 3 years

Benchmark Model

Another widely used model for time series prediction is ARIMA - Autoregressive Integrated Moving Average.

ARIMA does not support time series with a seasonal component. The ARIMA model is then extended to SARIMA to support the seasonal component. This capstone project has chosen SARIMA as its forecast model.

Evaluation Metrics

Four commonly used metrics to evaluate a time series model are:

MAE (Mean Average Error)

RSS (Residual Sum Squares)

MSE (Mean Squared Error)

RMSE (Root Mean Squared Error

MSE and RMSE are calculated, though the number is pretty large, the model can be further optimized by fine tuning with different sets of parameters.

Project Design

In the beginning of the project, data is downloaded from S3 bucket. Next is to kick off data preprocessing.

Data preprocessing

1. Explore the datasets, understand the data volume, data types
2. Convert date to datetime format for time series model purpose
3. Drop any null value found
4. Check data skew, if it is low skew, it won't impact the result of prediction
5. Get the meaningful data range especially oldest data may not be complete

Data train and test

Create time series and break the datasets to training set and testing set. Fit into the model.

Prediction

Study the plot and check accuracy by running MSE and RMSE.

Conclusion

Conclude the model and forecast for the next 3 to 4 years with the prediction from the machine learning.

References:

Using Machine Learning for Supply Chain

1. https://www.forbes.com/sites/louiscolumbus/2019/04/28/how-to-improve-supply-chains-with-machine-learning-10-proven-ways/#645a66ed3f3c
2. https://www.forbes.com/sites/louiscolumbus/2018/06/11/10-ways-machine-learning-is-revolutionizing-supply-chain-management/#1e3b03ba3e37
3. https://www.blumeglobal.com/learning/machine-learning/
4. https://towardsdatascience.com/artificial-intelligence-in-supply-chain-management-predictive-analytics-for-demand-forecasting-80d2d512f155
5. https://towardsdatascience.com/https-medium-com-h-javedani-how-smart-are-your-supply-chain-predictions-daf5a154ac6d
6. https://en.wikipedia.org/wiki/Inventory_optimization

Data source

7. https://www.kaggle.com/felixzhao/productdemandforecasting
8. https://www.kaggle.com/imsanjoykb/forecast-order-demand-and-visualization

MSE explanation

9. https://www.freecodecamp.org/news/machine-learning-mean-squared-error-regression-line-c7dde9a26b93/

Techniques for forecasting

10. http://rstudio-pubs-static.s3.amazonaws.com/15285_1260a766696d4960b9c63442a399831e.html