Machine Learning Engineer Nanodegree

Capstone Proposal

Lee, Seang Mei

June 6, 2020

Proposal

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.

With 1.05 million data volume for that 7 years, the goal of this capstone project is to study the trend of order demand by applying machine learning time series model and to produce with a prediction of order demand for the next few years.

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.

The dataset contains 1.05 million data with 4 columns. Data contains dates and order demand for several products that spread across multiple warehouses.

The training set data will take order demand from 2011 until 2015 and 2016 data will served as test data.

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.

Benchmark Model

A 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 proposal will use ARIMA as the benchmark model and apply SARIMA if a seasonal trend is clearly displayed by plotting a decomposition to dataset.

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

In this proposal, the mean squared error, or MSE, will be used to calculate the average of the squared forecast error values. The smaller number of MSE is better to fit the model. Even though zero MSE would mean perfectly predict training data, however, the ideal MSE isn't zero. Thus, I will evaluate MSE and RMSE to be a balance number between overfit and underfit for future data.

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