## horizontal line

Sales Forecasting

18.05.2020

**─**

Ashish Singh

# Overview

Sales forecasting is the process of estimating future sales. Accurate sales forecasts enable companies to make informed business decisions and predict short-term and long-term performance. Companies can base their forecasts on past sales data, industry-wide comparisons, and economic trends.

# Goals

1. Predict achievable sales revenue
2. Efficiently allocate resources
3. Plan for future growth

# Specifications

Forecasting sales is a common task performed by organizations. This usually involves manually intensive processes using spreadsheets that require input from various levels of an organization. This approach introduces bias and is generally not accurate especially during the initial few weeks of a quarter. In fact that's the time when an accurate forecast has the most benefit after all there's little value in providing an accurate forecast in the last week of a quarter.

Since forecasts are data driven the solution allows users to also perform "What-If" analysis. This is a tool that allows sales leaders to determine impact of certain factors on sales numbers. This type of analysis helps them determine what types of levers they have access to and what impact, either positive or negative, they can have on the sales. This advanced What-If analysis is based on machine learning where the model gets executed every time a user interacts with the tool. Some of the variables used in this analysis are number of sales reps, average deal duration, average deal amount, percent win rate. For example a sales manager can see what happens if they increase recruiting or if determine impact of a discounting program they have been considering. This list of features is configurable and can include other factors that may be more meaningful to a company

# Data Set Description

## Train Set

1. Date
2. Store Id
3. Item Id
4. Sales

## Test Set

1. Date
2. Store Id
3. Item Id

# Machine Learning Model Proposal

1. Arima Model, DNN

ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

1. XGBoost Model

XGBoost is an implementation of gradient boosted decision trees designed for speed and performance.

The implementation of the model supports the features of the scikit-learn and R implementations, with new additions like regularization. Three main forms of gradient boosting are supported:

* **Gradient Boosting** algorithm also called gradient boosting machine including the learning rate.
* **Stochastic Gradient Boosting** with sub-sampling at the row, column and column per split levels.
* **Regularized Gradient Boosting** with both L1 and L2 regularization.

# 

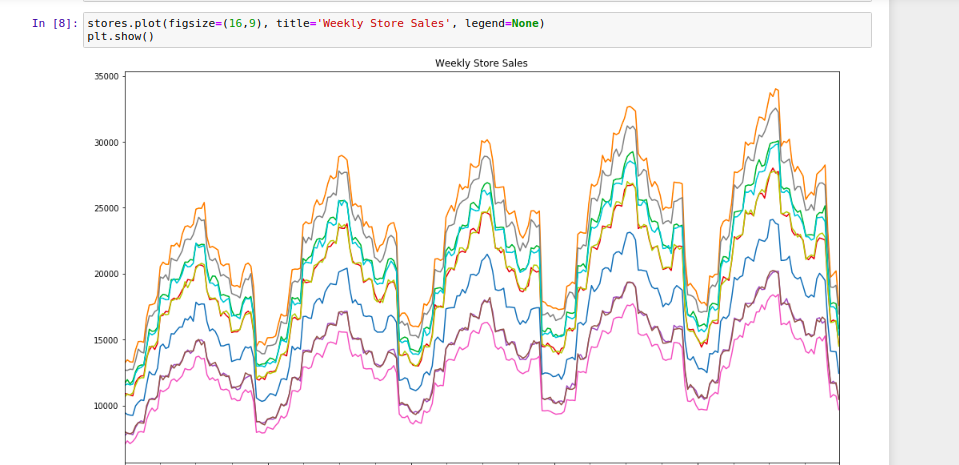
# 

# 

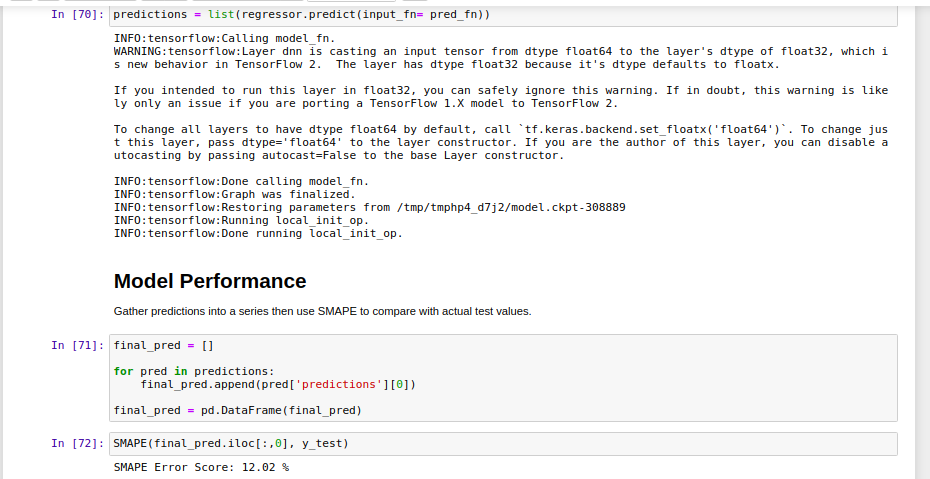
# 

# Preliminary Results

**Store-Sales Analyisis**



**Model Performance**



**Model Performance with XGBoost**

