Capstone Project Proposal
Udacity Machine Learning Engineer Nanodegree

Bryan Rosales
April 10th, 2021

1. Project Overview

Inventory forecasting is a challenge that many companies face and struggle to maintain an ideal inventory levels every day. Either being overstock or out of stock of any product are costly for a company. On one hand, overstock of an item means carrying more product than the company is able to sell and it is associated with over costs in storage, handling and cash flow which impact directly the profit. On the other hand, out of stock could be more critical since the company runs out of a specified item which is demanded by customer where not only the company is losing revenue, but also affecting customer satisfaction deriving a huge risk of increasing churn rate.

Classico Distributor LLC is a company founded in 2002, which is specialized to distribute liquor products in Arizona. The main business line is Wine and around 80% of the Wine portfolio is imported from Italy, yielding a 20% that is produced in the US mostly in Napa Valley, California. Classico purchases its inventory from local importers in the US, and only import 22% of the inventory sold.

During the last year, as a result of the Covid-19 pandemic crisis, global supply chain has been affected substantially and alcohol products have not been the exemption. As a consequence of the supply chain issue, Classico's importers have been experiencing low levels of inventory to supply demand which have placed Classico not only to lose sales opportunity, but customer exploring other competitors' portfolios. Therefore, Classico has decided to change its supply chain strategy, to increase the volume of purchases from Italy to its key products which will allow Classico to take more control of the inventory levels as capitalizing opportunities and profit.

2. Problem Statement

Classico portfolio contains around 420 items between wine, beer and spirits. However, the goal of this project is to create a forecasting solution to predict sales up to 4 months in the future using Machine Learning algorithms and techniques. Initially the analysis will be focus on top 5 best seller products, but extended to more products according to results. Forecasting 4 months of sales, Classico will

be able to planning containers in advance allowing allocate cash flow and most important to avoid running out of products with high demand.

3. Datasets and Inputs

The main dataset is extracted from Classico ERP (QuickBooks) which contains the following features:

- Sales Date
- Item Code
- Product Name
- Sold Quantity (Quantity shows in bottles)
- Sales Price ($USD)

In addition, to the main dataset, the variable temperature for Phoenix Metropolitan Area will be included. Arizona is a very seasonal state since it experiences very high temperatures (> 100°F) in summer time, which could be an important feature to the model. Temperature feature will be expressed in two columns as: Min and Max temperature monthly.

For the main dataset QuickBooks provides information in csv files which can be very easy to load and processing in a Pandas Data frame. In addition, the temperatures to use in the model will be obtained from https://w2.weather.gov/ and for simplicity only Phoenix temperatures will be included since majority of customers are located in Phoenix metropolitan area.

Moreover, the last feature to include in the final dataset is “holidays”. Since the data will be grouped by Month, this feature contains information about the number of holidays for a specific month. It is important because some peak in demand could be explained by holidays.

4. Solution

This project intent to solve the forecasting problem using Machine Learning algorithms and specifically PyTorch LSTM and Forecasting models. The initial idea is to start simple as possible with a LSTM network using the time series. Then, after evaluate results, I would decide to move forward to Forecasting Pytorch algorithm which will include more features like temperature and holidays.

5. Benchmark Model

Currently, the company is using a simple moving average of 4 weeks to estimate sales and order inventory. Forecasting model will be compare to such MA and results will be present to conclude which has a higher performance.

6. Evaluation Metrics

The solution will be measure using 2 metrics:

- MAPE (Mean absolute percentage error)
- RMSE (Root mean squared error)

7. Project Design

The pipeline to implement for the solution is as follow:

a. Data Preprocessing
   Include cleaning, imputations, transformation, scaling.

b. Split training and test datasets
   For training, I will use at least 24 data points, and testing dataset will be between 4 and 6 data points. Data will be monthly frequency.

c. Modelling
   Initially, I would start with a PyTorch LSTM model using only the time series, and then, depending on the result, moving to PyTorch forecasting model.

d. Prediction and Evaluation of the Model
   Using metrics defined and charts in order to evaluate the performance of predictions.

e. Deployment
   To solve this point, I will be using Flask and HTML template as front-end user app where the user will be able to see predictions.