**Product Backorder Prediction – Capstone Project 1**

**Problem Statement**:

Part backorders is a common supply chain problem, wherein a customer places an order for a product that is temporarily out of stock. The percentage of items backordered and the number of backorder days are important measures of the quality of a company's customer service and the effectiveness of its inventory management.

A company can manage its inventory more efficiently using a prediction on the backorder risk for the products. Goal here is to use the past data and metadata around the backorders, and provide a prediction on the potential products for backorders.

**Client:**

Sigma Retails Ltd is a leading online store providing products ranging from clothing, home improvements to grocery. Recently, client has been finding it difficult to manage the backorders, resulting into increasing customer issues and a decline in customer satisfaction.

Client is looking for ways to improve backorders handling. With the help of this analysis, a reasonable prediction on the products that can go on backorder is expected. Such a prediction could immensely help client to plan for a more effective stocking and backorder handling.

**Dataset:**

Dataset consists of the historical data around the backorders. It has 23 features and ~ 40,000 observations.

Dataset can be found at: <https://www.kaggle.com/tiredgeek/predict-bo-trial>

Feature details:

* sku - Random ID for the product
* national\_inv - Current inventory level for the part
* lead\_time - Transit time for product (if available)
* in\_transit\_qty - Amount of product in transit from source
* forecast\_3\_month - Forecast sales for the next 3 months
* forecast\_6\_month - Forecast sales for the next 6 months
* forecast\_9\_month - Forecast sales for the next 9 months
* sales\_1\_month - Sales quantity for the prior 1 month time period
* sales\_3\_month - Sales quantity for the prior 3 month time period
* sales\_6\_month - Sales quantity for the prior 6 month time period
* sales\_9\_month - Sales quantity for the prior 9 month time period
* min\_bank - Minimum recommend amount to stock
* potential\_issue - Source issue for part identified
* pieces\_past\_due - Parts overdue from source
* perf\_6\_month\_avg - Source performance for prior 6 month period
* perf\_12\_month\_avg - Source performance for prior 12 month period
* local\_bo\_qty - Amount of stock orders overdue
* deck\_risk - Part risk flag
* oe\_constraint - Part risk flag
* ppap\_risk - Part risk flag
* stop\_auto\_buy - Part risk flag
* rev\_stop - Part risk flag
* went\_on\_backorder - Product actually went on backorder. This is the target value.

**Problem solving steps:**

* Understanding the backorder system
* Exploratory Data Analysis:
  + Understanding the dataset
  + Understanding the features, its relations and distributions
  + Identifying the approach for handling missing values and cleaning the data
  + Feature selection
  + Creating and analyzing various classification models
  + Using visualizations, wherever required, to analyze features and model accuracy
  + Model tuning and selection

**Deliverables:**

* Well documented python code depicting below steps:

- Data wrangling

- Visualization

- Model building

- Model selection

- Prediction on test data

* Update GitHub with iPython notebook narrating above process
* Report (pdf) to share with client
* Slide Deck