## SECOND CAPSTONE PROJECT PROPOSAL

PREDICT RETURN DATE OF EMPTY CONTAINER FROM CUSTOMER

USING MACHINE LEARNING

**Statement of the Problem**

This study will work on the prediction of return date of empty container from customer using machine learning. My company is a global shipping, logistics / supply chain and is currently solving an optimization problem to minimize the cost of moving an empty container within North America and from North America to Asia. This prediction model will help to give an inventory outlook of equipment from current to future weeks.

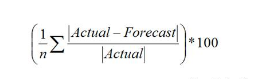
Specifically, this study aims to:

* Explore data from shipment, vessel schedule and cargo information then look for patterns
* Analyze data features from shipment, vessel schedule and cargo information to support this prediction model
* Apply machine learning techniques and modelling such as regression, random forest and gradient boost machines

**Dataset**

Data is will be from my current company and will comprise of shipment, vessel schedule and cargo information such as container size, container type, vessel arrival date, cargo weight and commodity description. Such information is confidential to the company so data masking will be done when publishing this study. Date range of the data can be from one year of data up to present covering about 120k shipments around the globe.

**Methodology**

For this study the following approach was used. First, extract and analyze the shipment, vessel schedule and customer data from my current company. Data cleaning, formatting and wrangling is done mainly using standard query language (SQL). Second, a machine learning algorithm script written in R and divides the dataset into training and test dataset at 80/20 percent proportion. An accuracy measurement is used to assess the performance of the three algorithms (MAPE or mean absolute percent error).