**E-commerce Shipping Data**

**Abstract**

An international e-commerce company based wants to discover key insights from their customer database. They want to use some of the most advanced machine learning techniques to study their customers. The company sells electronic products

Content The dataset used for model building contained 10999 observations of 12 variables.This data of Product Shipment Tracking, answer instantly to your questions:

* What was Customer Rating? And was the product delivered on time?
* Is Customer query is being answered?
* If Product importance is high. having higest rating or being delivered on time?
* How do you plan on shipping/delivering the products?(sometimes,business owners want to creat tiered shipping rules that can get complex)

**Design**

Classifying statuses accurately via machine learning models would enable the E-commerce to take action to improve operations and maintenance planning of these operations, allocate resources more quickly to needed areas, and ensure shipment is accessible to as many people as possible.

**Algorithms**

\*Feature Engineering\*

1. Mapping latitude and longitude to 3-dimensional coordinates so nearby continuous values would also be close in reality

2. Converting categorical features to binary dummy variables

3. Combining particular dummies and ranges of numeric features to highlight strong signals and illogical values for waterpoint status identified during EDA

4. Selecting subsets of the total unique values for categorical features that were converted to dummies, according to the number of samples they were associated with and their contribution to certain statuses

**\*Model Evaluation and Selection\***

The entire training dataset of 59,400 records was split into 80/20 train vs. holdout, and all scores reported below were calculated with 5-fold cross validation on the training portion only. Predictions on the 20% holdout were limited to the very end, so this split was only used and scores seen just once.

The official metric for DrivenData was classification rate (accuracy); however, class weights were included to improve performance against F1 score and provide a more useful real-world application where classification of the minority class (functional needs repair) would be essential.

**Tools**

- Numpy and Pandas for data manipulation

- Scikit-learn for modeling