Machine Learning Mockup

We will be diving into a large dataset to perform predictive analysis that determines whether a future delivery will be late and also what features most impact the time to ship.

Given our dataset and problem, it has been determined a supervised learning model should be applied to group records into either the label “Shipment Late” and “Shipment on Time”. Next we will run multiple machine learning argorithms that we believe can properly fit the model. To gauge how well our model is performing we will create confusion matrices, calculate the balanced accuracy score for each model tested as well as the precision and sensitivity. Given these summary statistics we will then choose the model that we believe will give us the best results. After choosing the model of best fit we will look at what features have the highest impact on the output of our model and remove features that effectively act as noise or is causing overfitting.