# Project Proposal

## “Competitive Intelligence: Dynamic Pricing”


### Team Members:

Veeresh Mallikarjun Mahashetti, Shreyans Jasoriya, Hitesh Verma


### Problem:

Nowadays, E-commerce platforms like Mercari have made it very simple & trustworthy for people to trade with each other and allow people to buy and sell their used products. The problem with this business model is arbitrary pricing, wherein similar products may have a stark difference in price. Using historical data, and the recent advancements in Machine Learning, it is now possible to estimate the price of a product to help users with the pricing of their products.

A historical dataset was made available by Mercari on Kaggle containing the following attributes: name, item_condition_id, category_name, brand_name, shipping, item_description, price. Leveraging Data Mining and Machine Learning, we aim to build an Estimator, which when given the product attributes would be able to give an estimate price for the product.

Dataset link: (https://www.kaggle.com/c/mercari-price-suggestion-challenge)

### Approach:

Our first step would be to do data preprocessing for initial transformations, aggregations and data cleansing. Once this is done, we would proceed with Data Wrangling to prepare the data for our next step, that is, Exploratory Data Analysis. After such an analysis, we would apply Feature Engineering, where it is possible to derive and extract features suitable for our Machine Learning models.

Since our task is a regression problem, as a part of Data Modelling, we will use Linear Regression as our baseline model. This will give us a better intuition about how the various independent variable (ex. brand) affect the selling price of the item and see how it performs on the dataset. We will then compare the performance of the baseline model with other models like Random Forest Regression, XgBoost, clustering approaches starting with K-Means. We might also look at Deep Learning approaches.

### Final Deliverables:

Our Baseline goal is to use supervised/unsupervised ML techniques to predict the prices of the product. We will use RMSE as our evaluation metric. Our secondary goal is to generate analytics regarding product prices using Data Mining. In addition to that, we would like to observe how different model implemented stack up against each other, how could (if possible) the performance of a model be increased and why does one model perform better than another.