**March 21: 2-3 Minute Presentation**

# **Otto Group Product Classification Challenge- Classify products into the correct category**

**Overview**

* The quality of our product analysis depends heavily on the ability to accurately cluster similar products
* The better the classification, the more insights we can generate about our product range
* Provided a dataset with 93 features for more than 200,000 products
* The objective is to build a predictive model which is able to distinguish between our main product categories.
* Submissions are evaluated using the multi-class logarithmic loss. Each product has been labeled with one true category. For each product, you must submit a set of predicted probabilities (one for every category). The formula is then,
* Logloss=−1N∑i=1N∑j=1Myijlog(pij),
* where N is the number of products in the test set, M is the number of class labels, logis the natural logarithm, yij is 1 if observation i is in class j and 0 otherwise, and pij is the predicted probability that observation i belongs to class j.
* The submitted probabilities for a given product are not required to sum to one because they are rescaled prior to being scored (each row is divided by the row sum). In order to avoid the extremes of the log function, predicted probabilities are replaced with max(min(p,1−10−15),10−15).

What data have you gathered?

* Test Data and Train Data

How did you gather it?

* Kaggle

What steps have you taken to explore the data?

* .Plot and .tail and .head

Which areas of the data have you cleaned, and which areas still need cleaning?

* Its clean already

What insights have you gained from your exploration?

* Summed one of the columns of data which was 1/15 of the rows of data- not sure if this means anything
* Plotted all the data and it was linear line

Will you be able to answer your question with this data, or do you need to gather more data (or adjust your question)?

* Yes I will be able to answer my question with the data

How might you use modeling to answer your question?

* Not sure yet, need to model a predictive algorithm to group the types of products