* ***APPROACHES***

In total we have 15 columns, which are:

* uniq\_id
* crawl\_timestamp
* product\_url
* product\_name
* product\_category\_tree
* pid
* retail\_price
* discount
* image
* is\_FK\_Advantage
* description
* product\_rating
* overall\_rating
* brand
* product\_Specification

Clearly, out of these attributes; uniq\_id, crawl\_timestamp, pid, image, is\_FK\_Advantage, product\_rating, overall\_rating don’t affect the classification. Therefore, we will be considering other attributes.

In our approach:

1. Features = {*“description”*}

In this case we only have 1 Independent Variable [“description”] and 1 Dependent Variable [“Category”].

1. Features = {}

The following are the methods used:

* Decision Tree
* Random Forest
* Gradient Boosting (XG Boost)
* AdaBoost (Not Used)
* ***RESULTS***

Refrences

[Decision Tree Algorithm, Explained - KDnuggets](https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html)

[python - How to deal with SettingWithCopyWarning in Pandas - Stack Overflow](https://stackoverflow.com/questions/20625582/how-to-deal-with-settingwithcopywarning-in-pandas)

[Topic Modelling into a Category Tree | by Anand P V | Towards Data Science](https://towardsdatascience.com/topic-modelling-into-a-category-tree-acafad0f0050)