Three Ideas – Capstone 3

1. SMS Spam Collection:
   1. The SMS Spam Collection v.1 is a public set of SMS labeled messages that have been collected for mobile phone spam research.
   2. It has one collection composed by 5,574 English, real and non-enconded messages, tagged according being legitimate (ham) or spam.

This corpus has been collected from free or free for research sources at the Internet:

* A collection of 425 SMS spam messages was manually extracted from the Grumbletext Web site. This is a UK forum in which cell phone users make public claims about SMS spam messages, most of them without reporting the very spam message received. The identification of the text of spam messages in the claims is a very hard and time-consuming task, and it involved carefully scanning hundreds of web pages. The Grumbletext Web site is: <http://www.grumbletext.co.uk/>.
* A subset of 3,375 SMS randomly chosen ham messages of the NUS SMS Corpus (NSC), which is a dataset of about 10,000 legitimate messages collected for research at the Department of Computer Science at the National University of Singapore. The messages largely originate from Singaporeans and mostly from students attending the University. These messages were collected from volunteers who were made aware that their contributions were going to be made publicly available. The NUS SMS Corpus is available at: <http://www.comp.nus.edu.sg/~rpnlpir/downloads/corpora/smsCorpus/>.
* A list of 450 SMS ham messages collected from Caroline Tag's PhD Thesis available at <http://etheses.bham.ac.uk/253/1/Tagg09PhD.pdf>.
* Finally, we have incorporated the SMS Spam Corpus v.0.1 Big. It has 1,002 SMS ham messages and 322 spam messages and it is available at: <http://www.esp.uem.es/jmgomez/smsspamcorpus/>.

**Reference Link : https://www.dt.fee.unicamp.br/~tiago/smsspamcollection/**

1. Amazon Product Reviews (Electronic\_Products User Ratings)

* Online E-commerce websites like Amazon, Filpkart uses different recommendation models to provide different suggestions to different users.
* Amazon currently uses item-to-item collaborative filtering, which scales to massive data sets and produces high-quality recommendations in real time.
* This type of filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list for the user.
* In this project we are going to build recommendation model for the electronics products of Amazon.
* The dataset here is taken from the below website.

Source - Amazon Reviews data (<http://jmcauley.ucsd.edu/data/amazon/>) The repository has several datasets. For this case study, we are using the Electronics dataset.

Attribute Information:

● userId : Every user identified with a unique id (First Column)

● productId : Every product identified with a unique id(Second Column)

● Rating : Rating of the corresponding product by the corresponding user(Third Column)

● timestamp : Time of the rating ( Fourth Column)

**Reference Link :**

**https://www.kaggle.com/saurav9786/amazon-product-reviews**

# Credit Card Customers (Predict Churning customers)

* A manager at the bank is disturbed with more and more customers leaving their credit card services.
* They would really appreciate if one could predict for them who is gonna get churned so they can proactively go to the customer to provide them better services and turn customers' decisions in the opposite direction
* I got this dataset from a website with the URL as <https://leaps.analyttica.com/home>.
* I have been using this for a while to get datasets and accordingly work on them to produce fruitful results. The site explains how to solve a particular business problem.
* Now, this dataset consists of 10,000 customers mentioning their age, salary, marital\_status, credit card limit, credit card category, etc. There are nearly 18 features.
* We have only 16.07% of customers who have churned. Thus, it's a bit difficult to train our model to predict churning customers.

**Reference Link :**

**https://www.kaggle.com/sakshigoyal7/credit-card-customers**