**Capstone Project Proposal (Group - 1)**

**A machine learning approach to the classification of online product reviews**

**Problem Statement:**

Online product review is one of the main sources for Product Managers to understand how the product is performing in the market. In today’s India, nearly 224 milion people are online buyers (<https://www.statista.com/statistics/251631/number-of-digital-buyers-in-india/>). So, the reviews provided by them in the product page of the e-commerce website contribute significantly in the measurement of product performance.

But all the online reviews are not necessarily about the product. Buyers give reviews about their overall shopping experience which includes the seller’s service, delivery efficiency and, of course, the product quality. So, the if the product manager takes this wholistic review into account to evaluate the performance of the product, it will definitely fetch a wrong result. To, solve this problem, we propose to create a machine learning approach that will be efficiently classify the reviews and segregate the true product reviews from the bunch of reviews and help to provide the true picture of the performance to the product manager.

**Our approach to solve the problem:**

To solve the above mentioned business problem, my approach would consist of NLP based english text classification.

**Dataset:**

For building this model we shall get the data from e-commerce websites like Amazon or Flipkart or the online review websites like car-dekho.com. We will be using web-scraping tool like ParseHub or import.io to get data from those websites. As of now we have the data of 3000 reviews for ‘One plus 6T’ mobile phone from Amazon India website, nearly 300 reviews of Kindle Paperwhite from Amzon India website. We will be capturing more data from different websites as well.