Test Assignment

Text Classification & Sentiment Analysis

**Objective** – This test assignment is for a summer intern @ Course5 in which, we are trying to gauge the development, data engineering and statistical skills of the candidates aspiring for a summer internship with the Course5 family, and a future in Data Sciences.

**Problem Statement** –

We would like you to build a BOT (web crawler) which crawls the OnePlus 6T Product Detail page (<http://tiny.cc/sck34y>) on the Amazon India website, and capture the following attributes of the products in the MongoDB NoSQL database. To build a crawler, we would like you to use Python as a language.

* Product Title
* Product Description
* Product Enlarge Image
* Product Price (With Exchange and Without Exchange)
* Product Colours
* # of Reviews
* Star Rating
* Technical Details
* Most Recent 100 Reviews

Once you extract the above data, we would like you to classify the downloaded reviews of the product in to following 5 buckets:

* Battery Life
* Picture Quality
* Value for Money
* Sound Quality
* Fingerprint

Once you are done with classification, we would like you to do the sentiment analysis around the classified reviews data under each bucket, and demonstrate, what % of reviews under each bucket are Positive, Negative and Neutral along with few sample examples for us to audit.

Lastly, you need to do the version controlling of your code by uploading it on GitHub.

**BONUS POINTS – Demonstrate the output of your work in a form of a dashboard.**

**Basis of Evaluation** –

* Quality of your crawler code
* How you have persisted the product attributes in the MongoDB
* Your approach for data cleansing and preparation
* Different algorithms you used for classification and the approach you followed for training and validation
* Accuracy, Precision and Sensitivity values for your best fit models
* Your approach for sentiment analysis and how accurately you classified the reviews under Positive, Negative and Neutral bucket