

## MAIS 202 – Project Deliverable 1

**Project:** Aspect Based Text Sentiment Analysis

**The Data Set:** Amazon Reviews: Unlocked Mobile Phones (PromptCloud, 2016)

Link: <https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones>

I choose this dataset because it was large, with data on 400,000 phones. It felt doable in terms of figuring out the aspects of the analysis. There were also around a 100 kernels which means that I will have lots of resources to learn from!

**Methodology:**

**Data Processing:**

The dataset has been well-processed, even has a usability rating of 8.8 on Kaggle. However, some cleaning is needed as a little amount of the reviews aren't in English. These data points should either be deleted or translated. I have noticed that some of these non-English reviews were only one word so it might be worth it to translate them. The dataset has information on the product name, brand, price, rating, and reviews. The ratings could be used to test whether our model has reached the correct sentiment overall, for example if the rating is 5 and the model predicted a negative sentiment, something is wrong. That would be extremely useful!

**Machine Learning Model:**

I want to be able to predict sentiments in the reviews and to be able to categorize them by subject. For categorizing the subjects, I think I will specify them beforehand rather than let the machine figure it out by itself as to not overcomplicate the project. Maybe after I am done that would be a new dimension that I can add to my analysis. Overall, I will use semi-supervised learning. I thought of doing supervised learning using classification, however the dataset is too big for that. After doing some research, neural networks seem to be the most used method for sentiment analysis. I am hoping to use a combination of both methods.

**Final Conceptualization:**

I want to deliver a simple landing-page web-app, where the user is asked, "How do you feel about your phone?" or "Tell us about your phone". Then the website presents a summary analysis of the review.

## Bibliography

PromptCloud. (2016). *Amazon Reviews: Unlocked Mobile Phones*. Retrieved from Kaggle: <https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones>