**Sentiment Analysis and Natural Language Processing for Marketing**

**From the project summary:**

In this Project, Company Team of a freshly launched startup that is introducing a new video game to the market. One of the key targets of a growth hacking team is to enhance the massive growth of early startups in a short time. To do so, it introduces strategies with the help of which one can acquire as many customers as possible with the lowest cost as possible. As part of your team’s growth hacking strategy, your boss wants to map the field of the video game market. She aims to find out how customers evaluate your competitors’ products, namely what they like and dislike in a video game. Knowing what makes a video game attractive to a gamer helps the marketing team articulate the message of your product more effectively.

To be able to find out what makes a video game worth buying according to gamers, you need to get deeper insights into the linguistic features of their utterances. As an NLP Specialist, your task will be to analyze customers’ reviews about video games. In order to carry out this task, you will employ different NLP methods. These methods will enable you to acquire a deeper understanding of customer feedback and opinion.

Your task as an NLP Specialist on the Growth Hacking Team is the following:

* Download the dataset of Amazon reviews.
* Create your own dataset from the Amazon reviews.
* Decide whether people like or dislike the video game they bought. Label each review with a sentiment score between -1 and 1.
* Check the performance of your sentiment analyzer by comparing the sentiment scores with the review ratings.
* Evaluate the performance of your sentiment analyzer and find out if you managed to correctly label the reviews.
* Try out other methods of sentiment analysis. Explore how people evaluate the video game they purchased by classifying the reviews as positive, negative, and neutral.

Summarize your results to the Head of the Growth Hacking Team. Based on your findings, list those things that are liked and those ones that are disliked about video games.

**Techniques Employed**

In order to get a deeper understanding of people’s opinion about video games, you will employ various NLP techniques. Here is a short list about what you will do and what techniques you will use.

* Sampling from imbalanced datasets using the imbalanced-learn package
* Enquiring about the sentiment value of the reviews with the dictionary-based sentiment analysis tools, which are part of NLTK, a natural language processing toolkit, used in Python.
* Finding out if your algorithm did a good job. Data evaluation with scikit-learn in Python.
* Analyzing the reviews with a state-of-the-art deep learning technique, namely with the DistilBERT model. To build this model, you will need to run Pytorch, transformers, and the simpletransformers packages.
* Evaluating your model and creating descriptive statistics in Python with scikit-learn library before reporting your results to your boss.
* Visualizing your findings about preferable and non-preferable words related to video games using Altair.

**Project Outline**

The project is made up of five steps, which are built on each other:

* Creating your dataset.
* Creating a dictionary-based sentiment analyzer.
* Evaluating your dictionary-based sentiment analyzer.
* Creating neural network-based sentiment analyzers.
* Reporting your results.

Both the steps and the techniques in this project model a real-life scenario. If you are employed as an NLP Specialist, you are likely to accomplish jobs like these ones.

**Dataset**

The Amazon review dataset can be downloaded from [here](https://nijianmo.github.io/amazon/index.html). Download the zipped json file of the category of video games 5 core, which can be found under the title Small subsets for experimentation. Once the download is complete, extract the file.