Predicting Sentiment of Next Review

# **Gloria Moore Summer 2021** https://github.com/gloriaypradal/next\_review

Name of the repository for this project is: Next\_Review

# Which Domain?

Data to use is going to be collected from:

* <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews> ; From this website I am collecting Women’s Clothing E-Commerce dataset revolving around the reviews written by customers, based on tweets
* <https://towardsdatascience.com/a-beginners-guide-to-sentiment-analysis-in-python-95e354ea84f6> : Theory about the sentiment analysis in Python
* <https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk> : More theory about sentiment analysis
* <https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk> : Predicting sentiment of next comment
* <https://www.kaggle.com/andrewmvd/trip-advisor-hotel-reviews> : I would like to analyze and use the same model to predict hotel reviews from this data set, if I am building a model to predict sentiment of a review, I should be able to use it for any topic, this conclusion would come after check if I have a balanced data set
* <https://subscription.packtpub.com/book/data/9781789954043/1/ch01lvl1sec06/kick-starting-an-nlp-project> : I have never work on a NLP project all by myself, so I am reading and looking for a lot of theoretical information
* <https://www.kaggle.com/hosubjeong/bakery-sales?select=Bakery+Sales.csv> : In this website a group of data scientist made predictions on bakery sales, this project will be used as reference for codes and analysis examples.
* <https://www.kaggle.com/therohk/million-headlines> : A million of headlines, this data set will be also used to build the model
* <https://www.kaggle.com/sid321axn/amazon-alexa-reviews> : Amazon alexa review also will be used for this project
* <https://monkeylearn.com/blog/word-visualization/#:~:text=A%20word%20cloud%20(or%20tag,customer%20reviews%2C%20among%20other%20text>. : Theory about word count visualizations, this visualizations would be created also to check what words are more represented in these data set by topic

# Which Data?

The data to use will be obtained from Kaggle community, all data sets to use are on csv format and they have been all downloaded from the website and about to be read on Jupyter Notebook, a sample of one of the data sets are shown here

Graphical user interface, application, table, Excel

Description automatically generated

# Research Questions? Benefits? Why analyzes these data?

Question: What is the classification (Sentiment) of the next review?

I have never worked on NLP projects by myself before, motivated by my classmates’ comments on teams and after reading a lot of articles about what type of projects should a Data Scientist portfolio should have, NLP projects is one important that I am missing, I think that is vital for me now that I am finishing my program to get practice and apply what I have learned about text mining and NLP. Also I would fin interesting on creating a model that can predict any sentiment review? Independent on the topic related factor? Let’s see

# What Method?

I would ne using Python, and a lot of EDA and text mining to understand these data sets, I’ll be using Vader Lexicon for classification (labeling) of the data, Tableau and Python for visualizations. I’ll be using Classification models to predict classification of the next review. On a second part of this project, I would like to use unsupervised machine learning to classify these reviews and make predictions

# Potential Issues?

* Expertise on NLP and text mining
* Time for this project , only 2 weeks
* Format of the data set
* Knowledge and practice on unsupervised machine learning algorithms

# Concluding Remarks

This project for me is as important as finish this class, I chose my weaker knowledge topic to work on I have always feared NLP, but I think that if I get to finish this project, I would have applied what I learned in this program plus a bit more, I think that working on this topic will allow me to practice what I have not done much lately. To build a classification predictive model about sentiment of a review would be helpful for any company that would like to know their next marketing strategy, or maybe for those who predict the market, predict the sentiment of a tweet or a social media comment or news, would be helpful to use it in their tock market predictions, at the end, social medias and news shape the world

References

<https://towardsdatascience.com/unsupervised-sentiment-analysis-a38bf1906483>