

# **Sentiment Analysis of Customer Reviews**

## **Conception Phase**

### **Description:**

The goal of this project is to make a tool for analyzing customer sentiment that can tell from reviews of a product whether customers feel positive, negative, or neutral about it.

### **Steps:**

1. Create a virtual environment using the YAML file/ requirements.txt and fill it with virtual dependencies for various Python libraries like Pandas, NLTK, NumPy, Scikit-learn, etc.
2. Create a code repository on GitHub (<https://github.com/>).
3. Download a dataset of customer reviews at [www.kaggle.com](http://www.kaggle.com).
4. To clean the data, perform operations like as tokenizing, lowering case, lemmatizing, stemming.
5. Apply Logistic Regression to a training and validation set. Sentiment analysis benefits from logistic regression since it is easy to train on huge datasets and produces accurate and reliable findings. Calculate the accuracy. Save the model in a pickle file.
6. Tf-IDF/term frequency-inverse document frequency is good for vectorize sentiment analysis because it shows how important each word is in a set of documents. Save as vectorizer.pickle file.
7. Perform sentiment analysis. Streamlit API user input collection because it is easy and simple to make. <https://streamlit.io/>
8. Predict positive, negative, or neutral sentiment analysis by loading the model and vectorizer.
9. Show and write the prediction result.
10. Improvements are made iteratively by training and testing the model with new data.

### **Software for processing:**

Ms Word, exported to Pdf, UML editor: Lucidchart, Python 3.9 with Streamlit API.

UML Activity Diagram:

