| TIGER ANALYTICS |
| --- |
| Springboard Project: Review Insights on Customer Product Reviews |
|  |
|  |
| **Vaishnavi Kandala,Pushvinder Rohtagi** |

| **April 2023** |
| --- |

# Background:



Client C, a leading beverage company, collects feedback from their customers, they want to gain actionable insights which can lead to better customer experience. The feedback is collected as a free text.

This project is to help Client C understand their customer and identify areas of improvements.

# Sub-case study 1: Analyzing Feedbacks for Attributes



This case study aims to gain insights from customer reviews about the product attributes.

# Data:



Client C will provide the following data for the project:

* **Customer review** data includes reviews and manually tagged attributes and sentiments. Assume that the labels are ground truth.

# Deliverables:



## NLP Data Preparation

* Assess data sufficiency and quality
* Handle missing or problematic data points
* Clean and preprocess text for modeling

## Exploratory Data Analysis (EDA)

* Understand dataset through review analysis and keyword identification and identify key insights
* Visualize data using various techniques
* Clean and preprocess text for comparison
* Perform EDA with visualizations (word clouds, histograms, heat maps) and text cleaning
* Create summaries and insights based on EDA results
* Research various text embedding techniques and pre-trained models
* Evaluate and visualize pre-trained model performance

## Build models to extract the aspects and their respective sentiments from the reviews

* Clean and process text data for modeling
* Utilize TigerNLP library wherever possible
* Fine-tune pre-trained models calculate performance metrics
* Fine-tune and compare models on labeled data; assess coverage on unlabeled data
* Test various pre-trained models, compare results, and justify the best model selection
* Analyze and report model performance using appropriate metrics

# Sub-case study 2: Analyzing Feedbacks for Attribute Types



This case study aims to gain insights from customer reviews about the product attributes and their corresponding types.

# Data:



Client C will provide the following data for the project:

* **Customer review** data contains customer reviews, manually tagged attributes and attribute types.Assume that the labels are ground truth.

# Deliverables:



## NLP Data Preparation

* Assess data sufficiency and quality
* Handle missing or problematic data points
* Clean and preprocess text for modeling
* Ensure processed text is free of issues and has a continuous index
* Divide text into smaller sentences for improved model coverage

## Exploratory Data Analysis (EDA)

* Understand dataset through review analysis and keyword identification and identify key insights
* Visualize data using various techniques
* Clean and preprocess text for comparison
* Generate summaries and insights based on EDA results
* Research various text embedding techniques and pre-trained models

## Build models to extract the aspects and their respective aspect types from the reviews

* Clean and process text data for modeling purposes
* Utilize TigerNLP library wherever possible
* Evaluate pre-trained models on dataset
* Fine-tune pre-trained models on labeled data and calculate performance metrics
* Compare performance of different models
* Retrain and compare models on labeled data; assess coverage on unlabeled data and present findings
* Select and justify the best model
* Analyze and report model performance using appropriate metrics

# Business Outcomes



* Analyze patterns in product attributes and their types
* Leverage the outputs to enhance customer experience, marketing strategies, and operational efficiency
* Pinpoint areas for future development and improvement

# Reading material:



* Courses under learning paths [DS NLP L2 Training in](https://tigeranalytics.udemy.com/learning-paths/3981264/) Tiger Udemy cover NLP topics that are required for this project.

# Deliverable Best Practices



* Use a private GitHub repository for development activities
* Develop using Python and TigerNLP library
* Create a new conda environment for the project
* Follow Tiger coding standards, add test cases, and document using Sphinx
* Include relevant analyses in a notebook with markdown comments and create a README with markdown guidelines. Share final codes and notebooks with mentors before presentation. The presentation should cover the objective, EDA insights, model results, conclusion, and potential next steps. Answer all questions in the presentation.
* You should use the [TigerNLP](https://tigeranalytics-code-templates.readthedocs-hosted.com/projects/tigernlp/en/dev/tigernlp.html) library for deliverables. You can refer [here](https://tigeranalytics-code-templates.readthedocs-hosted.com/projects/tigernlp/en/dev/tanlp_solutions/installation.html) for the installation steps for TigerNLP. For more information on how to proceed, refer to the documentation of [NLP Solutions](https://tigeranalytics-code-templates.readthedocs-hosted.com/projects/tigernlp/en/dev/tanlp_solutions.html).