CSCE 5290: Natural Language Processing
Project Proposal

Team Details

Deepu Gondhi **Kaushik Apoori** **Manumitha Kamarapu** **Teena Poorna Sudheera Maganti**

Github link:https://github.com/kaushikapoori/Group_4_NLP

Title: Starbucks Sentiment Analysis on Aspect-Based Customer Reviews

1. Motivation

The reviews taken could contain important information about the features that matter most concerning product quality, price, and customer service. The conventional methods of sentiment analysis in this example would return an overall score without valuable feedback about coffee quality, service, and ambiance. This project tries to overcome this limitation by conducting Aspect-Based Sentiment Analysis on reviews related to Starbucks and spotting sentiments related to particular aspects mentioned in the reviews. This approach works out the problem of informational overload, as Starbucks will be provided with the granular view of the customers' perception of diverse aspects of their experience. These insights will surely help Starbucks in improving customer satisfaction and loyalty by indicating areas that need improvement, while at the same time highlighting the aspects that are going well.

2. Significance

The value of this project lies in the fact that it will enable detailed feedback regarding the aspects of concern and will, therefore, be very helpful for businesses such as Starbucks. By applying Aspect-Based Sentiment Analysis to the project, Starbucks will know which particular areas the customers appreciate or criticize, and improvements will be carried out in a more focused manner. These findings will thus enable Starbucks to take action upon matters determined through feedback, such as further enhancing the quality of service or the atmosphere. It will also serve to illustrate in real life the usefulness of Aspect-Based Sentiment Analysis in practical applications and, as such, help in the growth of NLP studies. The approach can be extended to other domains where customer feedback is crucial for the improvement of products or services and, therefore, further broaden the applicability and value of Aspect-Based Sentiment Analysis.

3. Objectives

The main goals of the project are:

- **Aspect Identification:** Develop a model to identify salient aspects of Starbucks customer reviews regarding, for instance, coffee quality, service, and ambiance.
- **Sentiment Classification:** Classify each of the sentiments expressed towards the identified aspects as positive, negative, or neutral.
- **Performance Measurement:** The models developed for aspect extraction and sentiment classification should be evaluated using metrics including accuracy, precision, recall, and F1-score, with targeted precision at levels of at least 80%.
- **Success Criterion:** The project is considered successful if the model outputs valid, relevant, and clear aspect-specific sentiments that can provide actionable insights for Starbucks.

4. Features

- **Aspect Identification:** Specific aspects being reviewed can also be detected, such as coffee quality, service, ambiance, and pricing. Helps to break down customer feedback into distinct categories, thus the capability to target various other elements of the customer experience.
- **Aspect-Level Sentiment Classification:** This assigns sentiment scores across the spotted aspects of reviews as positive, negative, or neutral. Provides a clear understanding of customer opinions about each aspect, enabling businesses to focus on improving areas with negative sentiments and reinforcing aspects with positive feedback.
- **Trend Analysis:** This will analyze the changes in sentiment and aspect frequency over time. It identifies patterns or changes in the trends of the customers' perceptions and allows firms to chart the effects of their changes or upgrades and predict future trends accordingly.
- **Summarization of Feedback:** Detailed reviews are summarized in short phrases, showing the key aspects and sentiments. Allows for fast understanding of the most crucial feedback without having to read through all reviews individually; thus, time is saved, and decision-making becomes more effective.
- **Visualization Tools Description:** Graphs and charts, or interactive dashboards showing results of analysis. Improves access to effective interpretation and communication of findings, thus allowing stakeholders to visually and interactively explore insights from the data.

Milestones

- Data Collection and Preprocessing
- Aspect and Sentiment Extraction
- Model Development for Aspect-Based Sentiment Analysis
- Evaluation

5. Dataset

Material for this project is based on the Starbucks Reviews Dataset, which was web-scraped from the ConsumerAffairs website-a comprehensive collection of consumer reviews and ratings for Starbucks, a renowned coffeehouse chain.

Attributes from the dataset:

- Name: Reviewer's name.
- Location - The location or city associated with the reviewer, if provided.
- Date: The date when the review was posted.
- Rating: The rating given by the reviewer out of five stars.
- Review: The textual content of the review is the actual text that depicts the experience and opinions of the reviewer.
- Image Links: Attach links to images with the review, if any.

Size: The dataset contains the **reviews_data.csv** file, which is **464.6 kB** in size, thus containing a considerable number of reviews. This would form a good basis for extracting aspects and their sentiments.

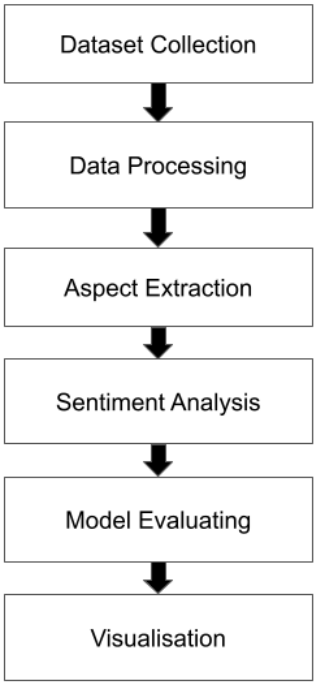
Preprocessing: Handling of Missing Values: In cases of missing or incomplete entry in name, location, date, ratings, and review, values should handle the missing value by either removing or imputing it accordingly.

Text Preprocessing: The preprocessing utilizes the standard NLP steps, like removing stop words, punctuation, special characters, and lemmatizing into base forms.

Timestamp Transformation: Normalize the dates to allow for temporal trend analysis.

6.Visualization:

The overall process of the project is as follows:



This is how the result looks like. The data shown below are example

Aspect Extraction Result

Review ID	**Category**	**Key Insight**
401	Coffee	Coffee is amazing
402	Service	Service is bad
403	Atmosphere	Ambiance is cool

Sentiment Analysis Result

Aspect	**Positive**	**Negative**	**Neutral**
Coffee	85%	10%	5%
Service	75%	15%	10%
Atmosphere	80%	10%	10%

We also provide graphs of the results for better understanding.