**CSCE 5290: Natural Language Processing**

**Project Proposal**

**Group 25**

**Title: Customer Review Summarization for Products**

**Team Members:**

1. Geethika Lankimalla
2. Karthik Undi
3. Chandana Chevuturi

**Motivation:**

The volume of user-generated content, especially in the form of consumer reviews, is growing along with the e-commerce sector. When looking to make an informed purchase, buyers may find the volume of these evaluations to be intimidating and time-consuming. The effort to summarize customer reviews for items aims to solve the enormous amount of internet reviews that are available for different products. The goal of customer review summarizing is to simplify and make clearly readable important information from many reviews, enabling customers to make well-informed purchase decisions. The potential effects this problem may have on customers and businesses highlight how important it is. By saving consumers time and effort while sorting through lengthy reviews, it improves their entire purchase experience. Better customer involvement and satisfaction translate into potential improvements in revenue and loyalty for organizations.

**Significance:**

Reaching the objective of efficient customer review summary is crucial as it caters to the changing requirements and anticipations of customers and companies in the online market. It is critical to be able to condense and provide pertinent information in an understandable fashion as e-commerce grows and the influence of online evaluations on purchase decisions grows. By doing this, the initiative helps to make purchasing for all parties involved more transparent, effective, and enjoyable.

* **Time Efficiency for customers:** In the quick-paced digital world of today, customers frequently have limited time while doing online product research. By offering simple summaries, customer review summarization simplifies the process and enables customers to rapidly understand the main ideas without having to spend a lot of time going through several evaluations. For time-pressed consumers looking to make well-informed decisions quickly, this efficiency is vital.
* **Competitive Edge for Businesses:** Businesses that use customer review summaries stand to benefit from a competitive edge in the marketplace. Offering customers an easy-to-use and effective means of obtaining pertinent information presents a business as creative and customer-focused. This may build brand loyalty, draw in additional business, and improve the perception of the brand.
* **Improved Decision Making:** Consumers can evaluate product features, benefits, drawbacks, and overall satisfaction levels more quickly when they have access to condensed evaluations. As a result, decisions are made with greater assurance and there is a less chance of consumer regret.

**Objectives:**

* Develop a robust summarization algorithm which involves developing an algorithm that can recognize important elements and aspects stated in reviews, evaluate, and understand the opinions represented in text, and produce brief overviews.
* To produce a thorough overview, NLP methods will be used in this project to extract information from multiple modalities inside customer evaluations.
* Here, the summarization algorithm's accuracy and efficacy would be evaluated using measures like precision, recall, and F1 score, which would then be used to inform ongoing iterations and enhancements.
* NLP is used in the creation of a user-friendly user interface for the summarizing tool. To include natural language understanding into the tool and make it easy to use, it must be integrated into e-commerce systems.
* Refine the model iteratively in response to usage and performance data.

**Features:**

**Technical Characteristics:**

* This project will involve developing and putting into use cutting-edge natural language processing (NLP) algorithms that can comprehend the context, attitudes, and semantics of customer evaluations and reviews. The core of the summarizing process will be these algorithms, which will extract important details while maintaining the original meaning.
* Text summarization, named entity identification, and sentiment analysis models will be trained using supervised and unsupervised machine learning approaches. With the use of labelled datasets, these models will have the ability to precisely evaluate and summarize customer evaluations.
* This project stands out because it uses methods for managing multimodal data, such as text, photos, and videos included in customer evaluations. NLP algorithms will be modified to retrieve relevant information from many modalities, enhancing the summarizing procedure and offering more thorough insights.
* The purpose of the project is to provide metrics and a thorough assessment framework for evaluating the efficacy and performance of the summarization algorithm. The precision, recall, and F1 score are some of the metrics that will be used to gauge the summary correctness and direct future iterations.

**Deliverables and milestones for this project may include:**

* Firstly, we will develop NLP algorithms for named entity recognition, sentiment analysis and text summarization.
* Then we will train and fine-tune the machine learning algorithms using labelled datasets.
* Next the implementation of multimodal data processing techniques for extracting information from videos, text, and images will be done.
* Further, we will design and develop user-friendly interface for summarization tool.
* Then we will establish an evaluation framework and metrics for assessing the performance of the text summarization algorithm.
* Improvement and refinement of the algorithm based on user feedback and evaluation results will be done.

**Dataset:**

This dataset contains reviews on excellent cuisine from Amazon. The information covers more than a decade, encompassing all ~500,000 reviews until October 2012. Reviews come with ratings, plain text, and details about the product and users. Reviews from every other Amazon category are also included.

Data comprises:

Reviews from October 1999 to October 2012: 568,454

There are 256,059 users.

74,258 products

With over 50 reviews, 260 users

**Dataset Attribute Information:**

Id - Row Id

ProductId - Unique identifier for the product

UserId - Unqiue identifier for the user

ProfileName - Profile name of the user

HelpfulnessNumerator - Number of users who found the review helpful

HelpfulnessDenominator - Number of users who indicated whether they found the review helpful or not

Score - Rating between 1 and 5

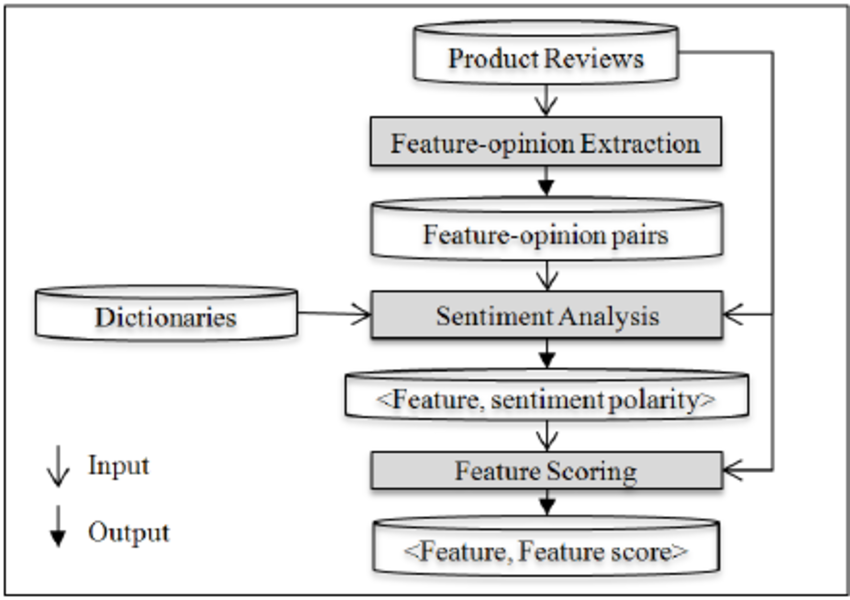
Time - Timestamp for the review

Summary - Brief summary of the review

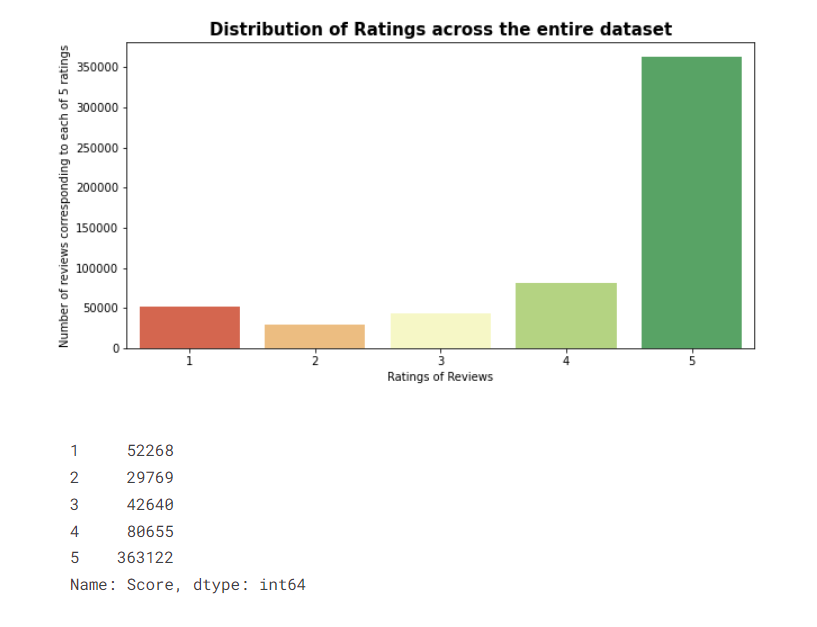
Text - Text of the review

**Visualization:**

This is our review summary method's system architecture. First, features and opinion terms are taken out of product reviews. The retrieved features and the associated opinion words are used to create feature-opinion pairings. After that, a sentiment analysis procedure is used to categorize the relationship between feelings for every pair. Positive and negative sentiment dictionaries are employed in the sentiment analysis process to establish the polarity of an opinion word. A novel feature scoring procedure is used to determine the scores for every product feature once the sentiment polarity of each pair is determined.



This is the basic distribution of across the entire dataset.

****