**CSCE 5290: Natural Language Processing**

**Project Proposal**

**Title: Aspect Based Sentiment Analysis**

**GitHub Link:** [**https://github.com/DominicArgento/CSCE-4290-Team-9**](https://github.com/DominicArgento/CSCE-4290-Team-9)

**Members: Dom Argento, Sabrina Salazar, Sahardeed Ahmed**

**1. Motivation**

The goal of this assignment is to identify certain aspects of a business that can effect the increase or decrease currency. This evaluation can improve the overall business for the company. Its significant because seeing a clear insight of each effect can develop a better understanding the communities reactions for certain approaches

**2. Significance**

Based on the knowledge of how other enhancers have on another business is very effective in its own way. For example Yelp reviews can significantly impact a business by affecting the business reputation and decrease the overall attraction for their product. Having that in mind, its extremely important to make a good reputation at first. However knowing this can cause stress on any business.

**3. Objectives**

At the end of the project, the overall goal is to see how a business can increase revenue and improve their profit margins.A guide that can be given is by understanding a customers experience with a company and analyzing their reactions. If that is completed then then the business can apply changes to improve reviews. Improving reviews will soon improve on currency because that will bring more customer attraction.

**4. Features**

The technical characteristics we plan to integrate into this project can be associated most with word recognition, phrase processing and understaning, data cleansing and decision making. Word recognition will be understanding what words hold what value in the context of reviewing this item or business. Phrase processing will be attempts at retaining some reviews that may be written differently than a typical good or bad review, they may say pizza was hot or pizza wasn’t my favorite instead of “great” or “disgusting”. Data cleansing will involve removing irrelevant spacing and symbols that may be contained within the review(s) before consolidation on searched keyword(s). Lastly, we plan to integrate some logic that will be able to determine if the overall review(s) of said items or businesses within the dataset is overwhelmingly positive, positive, neutral, negative, or overwhelmingly negative.

We plan to develop a system fully capable of extracting data from Yelp’s reviews and analyzing sentiment towards any keyword(s) within their dataset. We will have a fully functioning UI that will be visually appealing for users and will include a search bar that also has filters for level of review.

Our milestones will include the following:

* GitHub Creation
* Aspect-based Sentiment Analysis Research
* Dataset Collection
* Data Processing
* Model Construction
* Front-end Development
* Code Implementation
* Testing and Refinement
* Deliver Application

**5. Dataset.**

We plan to utilize a dataset from an existing website that allows for users to review and comment about certain products, locations and services. Yelp's diverse dataset provides us with a comprehensive foundation for analyzing customer sentiments from all over the country. The dataset encompasses over 6.9 million reviews from 1.9 million users visiting 150,000 businesses and also includes over 200,000 photos, providing a multi-faceted view of user experiences. The actual files provided by Yelp are two rather large zipped files that contain several JSON files within them, they include one file for the review data and one for the photos associated with them. We will thoroughly filter through the dataset, searching for any reviews that may conflict with this project due to various reasons, ensuring that any information utuilized is suitable for analysis.

**6. Visualization**

Table 1: Objectives of the Project

| **Objective** | **Description** |
| --- | --- |
| Identify aspects affecting business revenue | Determine specific aspects of a business that can impact the increase or decrease in currency, providing insight for overall business improvement. |
|  |  |
| Analyze customer reactions | Understand customers' experiences with the company and analyze their reactions to make changes that improve reviews and attract more customers. |
| Improve business reputation | Implement changes based on customer feedback to improve reviews, thereby enhancing the business's reputation and attracting more customers. |

Table 2: Features of the Project

| **Feature** | **Description** |
| --- | --- |
| Word recognition | Identify the significance of words in the context of reviews for businesses or items. |
| Phrase processing | Process reviews that may deviate from typical positive or negative expressions, such as nuanced phrases or descriptions. |
| Data cleansing | Remove irrelevant spacing and symbols from reviews before consolidating and analyzing them for keywords. |
| Sentiment analysis | Determine the overall sentiment of reviews (overwhelmingly positive, positive, neutral, negative, or overwhelmingly negative). |
| UI development | Create a visually appealing user interface with a search bar and filters for different levels of review. |