

**College of Artificial Intelligence (El Alamein)**

Automated Sentiment Analysis of Restaurant Reviews.

*Natural Language Processing (IN321): Final Project Proposal*



Feb 2024

**Name: Eslam Mohamed**

**Registration number: 20102059**

**1. Project Description:**

Overview: This project aims to perform sentiment analysis on restaurant reviews using Natural Language Processing (NLP) techniques. The goal is to classify reviews as positive or negative, thereby helping restaurant owners understand customer sentiment.

Importance and Relevance: Sentiment analysis is crucial in NLP as it helps in understanding public opinion, improving customer service, and making data-driven decisions. This project contributes to the field by utilizing advanced NLP techniques, including transformers, to achieve high accuracy in sentiment classification.

NLP Tasks and Challenges: The project addresses key NLP tasks such as text preprocessing, tokenization, feature extraction, and model building using transformer-based architectures like BERT. The challenges include handling imbalanced data, preprocessing text data, and fine-tuning models for optimal performance.

**2. Problem Statement:**

Definition: The project aims to solve the problem of classifying restaurant reviews as positive or negative. This binary classification problem helps in automatically determining customer satisfaction from textual reviews.

Significance: Understanding customer sentiment is vital for restaurant owners to improve their services. This problem aligns with the course's learning objectives by applying NLP techniques to a real-world problem, demonstrating the practical utility of sentiment analysis.

**3. Methodology:**

Methods and Techniques: The project will utilize data preprocessing techniques, tokenization, and feature extraction methods like TF-IDF and Word Embeddings (Word2Vec, GloVe). Transformers like BERT will be used for model building.

Suitability: These methods are suitable as they are well-established for handling text data and extracting meaningful features. Transformers, specifically BERT, have proven effective in various NLP tasks, including sentiment analysis.

**4. Data:**

Data Sources: The primary dataset is the publicly available "Restaurant Reviews" dataset from Kaggle.

Dataset Description: The dataset contains 1000 reviews with two columns: 'Review' and 'Liked'. The 'Liked' column indicates whether the customer liked the food (1 for positive, 0 for negative).

**5. Implementation Details:**

**Data Preprocessing:**

Remove duplicates and irrelevant columns.

Tokenize the text data.

Handle missing values, if any.

**Exploratory Data Analysis (EDA):**

Understand the distribution of the 'Liked' column (positive, negative).

Visualize the data distribution.

Check for class imbalances.

**Feature Engineering:**

Convert text data into numerical format using techniques like TF-IDF and Word Embeddings (Word2Vec, GloVe).

Utilize transformers like BERT for feature extraction.

**Model Building:**

Split the data into training, validation, and test sets.

Choose a transformer-based model architecture.

Fine-tune the model on the training set.

Validate the model performance on the validation set.

**Model Evaluation:**

Evaluate the model on the test set using metrics like accuracy, precision, recall, and F1-score.

Analyze the results and identify areas for improvement.

**Model Deployment:**

Save the trained model.

Create a function to load the model and make predictions on new data.

Classify new reviews as 1 (positive) or 0 (negative).

**6. Evaluation Metrics:**

Performance Measurement: The model's performance will be evaluated using metrics like accuracy, precision, recall, and F1-score.

Evaluation Metrics: These metrics provide a comprehensive understanding of the model's performance, particularly in handling imbalanced data and ensuring both positive and negative sentiments are accurately classified.

**7. Expected Results:**

Expectations: By the end of the project, we expect to achieve a high-performing sentiment analysis model capable of accurately classifying restaurant reviews. The model will take a text input and predict whether the review is positive or negative. For example:

Enter a review text: "This restaurant exceeded all of my expectations! The food was outstanding, the service was impeccable, and the atmosphere was delightful. I can't wait to come back!"

Prediction: 1

Significance: These results will demonstrate the effectiveness of using transformers for sentiment analysis and provide valuable insights into customer opinions, helping restaurant owners make informed decisions.

**8. References:**

Restaurant Review Sentiment Analysis:

•Thelander, A., & Yu, H. (2017). Yelp reviews: Sentiment selection and aspect detection. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018) (https://link.springer.com/article/10.1007/s10462-022-10252-y).

•Liu, B. (2012, October). Sentiment analysis and opinion mining of social media data. In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1639–1648) (https://www.sciencedirect.com/science/article/pii/S1319157823003300).