

# ENHANCED SENTIMENT ANALYSIS ON YELP RESTAURANT REVIEWS

## 1 TEAM MEMBERS

- (1) Akhil Pavan Sai Machavaram
- (2) Naga Vaishnavi Marupaka

## 2 TASK ALLOCATION

### 2.1 Akhil Pavan Sai Machavaram

*2.1.1 Main Responsibilities.* Model architecture design and implementation, including advanced transformer models and hybrid model approach. Integration of model explainability tools (LIME, SHAP).

*2.1.2 Secondary Responsibilities.* Research and implement cross-lingual support using multilingual models like mBERT or XLM-RoBERTa. This includes ensuring the dataset accommodates multilingual reviews or augmenting it with translated versions, and fine-tuning the selected model to effectively perform sentiment analysis across different languages.

### 2.2 Naga Vaishnavi Marupaka

*2.2.1 Main Responsibilities.* Dataset preparation and augmentation, including text data augmentation and aspect-based sentiment analysis preparation. Development of the interactive web application.

*2.2.2 Secondary Responsibilities.* Evaluation metrics and model performance analysis.

## 3 DESCRIPTION OF THE PROJECT

Our project is an extension and development of an existing project aiming to create and develop an advanced sentiment analysis system to analyze Yelp restaurant reviews more comprehensively and accurately, including support for multiple languages. The system will leverage state-of-the-art NLP techniques, including advanced transformer models and hybrid architectures, to understand the nuanced sentiments expressed in reviews across different languages. It will also incorporate multimodal data analysis to consider images and text together, aspect-based sentiment analysis for granular insights, model explainability for transparency, and key feature of our project is a user-friendly web application that allows users to submit reviews in various languages and receive sentiment predictions along with explanations for the model's decisions.

## 4 DATASET

The primary dataset for this project will be Yelp restaurant reviews. This dataset will include textual reviews and associated ratings, as well as restaurant images where available. We plan to enrich the dataset with aspect-based annotations for a subset of reviews to facilitate aspect-based sentiment analysis. To support multimodal analysis, we will also curate a set of reviews with corresponding images. To ensure comprehensive coverage and diversity in sentiment analysis, we will utilize the following datasets:

## 4.1 Yelp Open Dataset

*4.1.1 Description.* Offers a wide array of data from Yelp, including millions of reviews, business attributes (such as location, category, and operational hours), and user information. It's particularly valuable for restaurant reviews and analysis.

*4.1.2 Why Use It.* Direct relevance, a large volume of data, includes both textual reviews and business metadata. Ideal for training models to understand a broad range of sentiments expressed in restaurant reviews and to perform aspect-based sentiment analysis.

## 4.2 Zomato Restaurant Data

*4.2.1 Description.* Zomato provides an API that can be used to gather data on restaurants, including user reviews. While not a dataset per se, the API allows for the collection of a dataset tailored to specific requirements.

*4.2.2 Why Use It.* Global coverage, including cities and countries not extensively covered by Yelp, allows for diverse and localized sentiment analysis. This dataset will enable our model to understand and analyze sentiments from a wide geographical and cultural spectrum, enhancing its applicability and robustness.

## 5 TIMELINE

- Week 1-2 (2/15 - 2/28): Dataset collection from Yelp and Zomato, initial preprocessing, and development of model architecture.
- Week 3-4 (3/1 - 3/14): Complete basic model training, implement text data augmentation, and start aspect-based analysis features.
- Week 5-6 (3/15 - 3/28): Refine model with augmented datasets, focus on cross-lingual support, and begin web application development.
- Week 7-8 (3/29 - 4/11): Finalize model training, complete web application development, and integrate model explainability.
- Week 9 (4/12 - 4/14): Final testing, bug fixes, and preparation for presentation and demo.
- Demo Week (Last week before final exam): Conduct a comprehensive demo of the system, showcasing its capabilities, user interface, and explainability features.

## 6 REFERENCES

[1] Nicholas Renotte, May 27, 2021. <https://github.com/nicknochnack/BERTSentiment/blob/main/Sentiment.ipynb>