

# SENTIMENT ANALYSIS OF YELP REVIEWS

Shreyasi Biswas, Deepanshu Malhotra, Shruti Ravichandran

## 1) Background/History

### 1.1. Literature Review of dataset or analysis method

Yelp is a widely used online review platform with a vast collection of over 200 million user-generated reviews covering various local businesses, such as restaurants and dog parks. Yelp's user-generated reviews and ratings can be helpful for both businesses and consumers, as businesses can use the feedback to improve their services, and consumers can make informed decisions. The sentiment analysis of Yelp reviews has been a popular research topic in recent years. Many studies have focused on developing machine learning models that accurately classify the sentiment of Yelp reviews. One such study by Xiong et al. (2018) used deep learning models to classify Yelp reviews as positive or negative. Another study by Ghose and Ipeirotis (2011) focused on the impact of online reviews on the sales of local businesses. The study analyzed over 40,000 Yelp reviews and found that a one-star increase in Yelp rating leads to a 5-9% increase in revenue for a local business. Overall, the sentiment analysis of Yelp reviews has been a valuable research area, with potential applications in business, marketing, and customer experience. While there have been successes in developing accurate sentiment analysis models, there is still room for improvement in addressing the limitations of past research.

### 1.2. Limitations of previous study or analysis method

One of the main limitations of sentiment analysis of Yelp reviews in the past has been the accuracy of the machine learning models used. Previous studies have shown that achieving high accuracy in sentiment analysis is challenging due to the complexity of human language and nuances in sentiment expression. Bias in training data and difficulty in identifying context and tone can also impact the results. Moreover, studies like Hu et al. (2018) have shown that sentiment analysis models struggle to classify ambiguous reviews containing sarcasm or irony. Reviews can contain variations of language, such as slang, abbreviations, and misspellings, making it difficult for sentiment analysis models to interpret accurately. Another limitation is the subjective nature of sentiment analysis, as different people may interpret the sentiment of a review differently, leading to subjective judgments about the accuracy of sentiment analysis models. These findings suggest a need for advanced models that can accurately interpret the context of reviews.

## 2) Proposed Project

### 2.1. Objective

The objective of this sentiment analysis project is to develop a machine learning model that accurately classifies the sentiment of Yelp reviews into one of five categories: positive,

somewhat positive, neutral, somewhat negative, or negative. The ultimate goal is to provide insights around brand perception for businesses that rely on customer feedback, allowing them to identify areas for improvement and enhance the overall customer experience.

## **2.2. Dataset(s)**

This project is using the Yelp Customer Review dataset from Yelp, which is a collection of user-generated reviews, ratings, and other metadata for local businesses in various cities. It includes over 200 million reviews covering a wide range of businesses, such as restaurants, bars, salons, and more. The project focuses on the subset of reviews that have been labeled with sentiment ratings from 1 to 5 by Yelp's machine learning algorithm. The data includes the business ID, date, review ID, stars given by the reviewer, the text of the review, the type of review, user ID, and the number of cool, useful, and funny votes given by other users. The reviews cover different categories of businesses, including restaurants, cafes, bars, hotels, spas, and many others.

## **2.3. Proposed Method(s) Applied**

The sentiment analysis of Yelp reviews involves two broad steps: text preprocessing and multiclass classification. In the text preprocessing stage, natural language processing (NLP) techniques will be used to prepare the reviews for analysis. The reviews will undergo tokenization, where they will be split into individual words or tokens. Stop words, such as "the" and "and," will be removed to reduce noise in the data. The remaining words will be lemmatized, where they will be transformed into their base form to reduce the number of unique words in the dataset. Finally, the reviews will be vectorized, where each review will be represented as a numerical vector. In the multiclass classification stage, classification models such as Naïve Bayes, SVM, and Random Forest will be trained on the processed reviews. These models will classify the reviews into five categories, ranging from positive to negative. By using these techniques, we can develop accurate and robust sentiment analysis models for Yelp reviews.

## **2.4. Evaluation Metric(s)**

The model performance will be evaluated by getting predictions on unseen data. The classification accuracy can be measured by looking at the confusion matrix and inferring the precision, recall, accuracy and F1-score from the classification report. The ROC curve can also be evaluated to pick the best classification threshold.

## **2.5. Expectation(s) of Results**

Based on past projects, the desired outcome is to build a robust machine learning model that accurately discerns the sentiment of customer reviews on Yelp with a minimum accuracy level of 70%.