

# Capstone Project Literature Review: Yelp Sentiment Analysis

Joshua Afleje, Nicole Feil, Andrew Liu, and Jack Oebker

Arizona State University, Tempe, AZ 85281, USA

February 7, 2025

## Abstract

Online reviews play a crucial role in shaping consumer behavior and business success. Platforms like Yelp, Google Reviews, and TripAdvisor significantly influence purchasing decisions. Research indicates that over 90 percent of consumers read online reviews before visiting a business, and 84 percent trust online reviews as much as personal recommendations. Despite its widespread use, Yelp’s system has significant limitations, including bias in reviews, difficulty capturing nuanced sentiment, and an over-reliance on numerical star ratings. Advances in aspect-based sentiment analysis (ABSA), machine learning models, and fake review detection techniques have improved how sentiment is understood in large-scale datasets like Yelp. This literature review synthesizes past research on sentiment analysis, explores methodologies for aspect-based classification, sentiment-driven clustering, and fake review filtering, and discusses how consumer behavior is impacted by Yelp reviews.

## 1 Introduction

Yelp, a prominent online review platform, hosts millions of user-generated opinions covering restaurants, healthcare providers, and various local businesses. These reviews provide valuable insights for both consumers and business owners, shaping purchasing behavior and informing service improvements. However, Yelp’s star rating system reduces complex, multi-faceted reviews into single numerical values, failing to capture the depth and nuances of consumer sentiment.

Recent advances in sentiment analysis have attempted to bridge this gap by moving beyond simple sentiment classification to more granular, aspect-based approaches. Researchers now aim to extract actionable insights by identifying specific features—such as food taste, service quality, and ambiance—that influence positive or negative feedback (?, ?).

Despite the usefulness of Yelp, researchers have identified key challenges such as bias in data, fake review filtering inaccuracies, and cultural differences in sentiment expression. Studies have shown that Yelp’s filtering algorithm disproportionately flags extreme sentiment reviews—whether genuine or not—introducing bias into overall ratings (?, ?). Further, cross-cultural differences in review expression lead to inaccuracies in sentiment classification models, which were primarily trained on Western reviews (?, ?).

### 1.1 Background

#### 1.1.1 The Emergence of Feature-Based Sentiment Summarization

Sentiment analysis initially focused on classifying entire reviews as positive or negative, oversimplifying multi-aspect consumer feedback. However, Hu & Liu (2004) (?, ?) introduced aspect-based sentiment analysis (ABSA), which identifies specific product or service attributes and evaluates their associated sentiment. ABSA has since been widely used in Yelp studies to enable businesses to understand which aspects drive customer satisfaction or dissatisfaction.

### 1.1.2 Cultural and Linguistic Bias in Sentiment Analysis

Nakayama & Wan (2019) ( ?, ? ) found that cultural norms shape how users express sentiment on Yelp. Their research highlights that Western users prioritize service and ambiance, whereas Japanese users focus on food quality and price fairness. This cultural bias suggests that sentiment models trained on English reviews may misinterpret non-Western reviews, leading to misclassification errors in Yelp sentiment analysis.

### 1.1.3 Fake Review Detection and Yelp’s Filtering Algorithm

Mukherjee et al. (2021) ( ?, ? ) examined Yelp’s fake review filtering system and found that:

- Yelp’s algorithm flags extreme sentiment reviews, even when they are legitimate.
- Behavioral cues (reviewer activity, rating patterns) outperform linguistic analysis in fake review detection.
- Crowdsourced fake reviews differ significantly from real-life spam, meaning models trained on artificial datasets often fail when applied to Yelp.

These findings indicate that hybrid models combining linguistic and behavioral features are essential for improving fake review detection.

### 1.1.4 Yelp’s Impact on Non-Restaurant Domains

Chen & Lee (2023) ( ?, ? ) analyzed Yelp physician reviews and found that sentiment strongly influences patient choice. Their research showed that a one-star increase in rating resulted in a 1.9% revenue increase and a 1.2% rise in patient volume. This validates the real-world impact of sentiment analysis beyond restaurants, showing that user-generated sentiment can shape healthcare decisions.

## 2 Methodologies

### 2.1 Aspect-Based Sentiment Analysis (ABSA)

Aspect-based sentiment analysis (ABSA) provides a structured approach to extracting key features from Yelp reviews and assigning sentiment labels to those aspects ( ?, ? ). Deep learning techniques, such as BERT-based transformers, have significantly enhanced ABSA performance.

### 2.2 Sentiment-Driven Business Recommendations

Currently, Yelp suggests businesses based on categories and location, without factoring in sentiment-driven preferences. Clustering businesses based on user sentiment trends can offer a more personalized recommendation system ( ?, ? ). This can be achieved using vector-based similarity models, which compare businesses based on review sentiment patterns rather than just business type.

### 2.3 Synthetic Review Generation

Generative models, such as GPT-based architectures, can create synthetic reviews at different star levels. These synthetic reviews, conditioned on real sentiment trends, allow users to better understand a business’s strengths and weaknesses without manually reading thousands of reviews.

## 2.4 Conclusion

Yelp sentiment analysis has evolved from simple star ratings to complex aspect-based insights. Research has demonstrated that businesses are affected by review sentiment, particularly in healthcare and service industries ( ?, ?). However, challenges such as fake review filtering, cultural biases, and oversimplification of sentiment remain significant obstacles ( ?, ?).

To enhance Yelp’s effectiveness, researchers have explored deep learning models for sentiment classification, clustering techniques for better business recommendations, and synthetic review generation to create structured summaries. The continued integration of ABSA, behavioral analytics, and machine learning techniques will shape the next phase of sentiment-driven business intelligence.