Enhancing Restaurant Business Strategies with Sentiment Analysis of Yelp Reviews

Duc Dat Hoang
Department of Computer Science
University of Exeter, Exeter, UK

Supervision:
Dr. Federico Botta
Supervisor affiliation

***Abstract*—Since 2023, the restaurant industry has faced significant challenges, including financial strains, changing in customer's behaviour, and a shortage of labor. These factors have contributed to the ongoing struggle of many restaurant businesses. This study proposes a novel approach to harnessing Yelp reviews for strategic business enhancement through sentiment analysis employing a Random Forest algorithm. By extracting and analyzing the polarity scores of individual words within the reviews, the model identifies key factors contributing to customer satisfaction and dissatisfaction. This methodological approach allows for a nuanced understanding of customer sentiment, revealing the inherent advantages and drawbacks specific to each restaurant. The polarity scores, indicative of the significance of each word, enable the pinpointing of areas for improvement and potential competitive edges. The outcomes of this analysis provide restaurant owners and managers with actionable insights to formulate strategies that enhance positive aspects and address negative factors, aiming to elevate customer experiences and improve business performance.**

I. Section

Yelp has emerged as a pivotal platform for patrons to express their dining experiences, offering a rich dataset for extracting actionable insights into restaurant features. While previous studies have implemented Support Vector Machines (SVM) for sentiment analysis [1], this research adopts a Random Forest approach to delve deeper into the nuances of customer feedback. Motivated by a comparative study by Sahu and Selot (2022) [2], which highlighted the superior performance of Random Forest over SVM in analyzing sentiment on tourism reviews, this study aims to uncover the multifaceted aspects influencing restaurant patronage. In this study, I embarked on innovative analyses, uncovering intriguing insights from Yelp reviews. I observed that the general sentiment polarity leaned towards service aspects, suggesting that customers may prioritize service when choosing their dining experiences. Additionally, my deep dive provided valuable insights not readily apparent on Yelp's interface. While Yelp primarily presents an aggregated rating for businesses, my analysis delved into the finer details, evaluating sentiments at a more nuanced word-level, thus offering a multidimensional perspective on customer feedback.

Various methodologies have been employed to decode the sentiments embedded in textual content or entire documents. Among the prevalent machine learning techniques in the domain of natural language processing (NLP) are Naive Bayes (NB), Support Vector Machine (SVM), Random Forest [3] and unsupervised learning approaches (Sana and et al. 2021) [4]. Nowsaday, with the surge in applications of neural network-based methods, there are following study by Li, Yand and Yen (2021) who introduced approach based on neural network algorithms that claimed to improve sentiment classification of restaurant reviews.

The structure of this paper is as follows: Section I provides an introductory overview of Yelp and Natural Language Processing (NLP), setting the stage for my analysis of Yelp reviews through natural language techniques. Section II details the dataset utilized and the methodologies employed for extracting insights from Yelp reviews. Section III outlines the outcomes derived from my analytical models. Section IV delves into a discussion on the findings across various cuisine types. Finally, Section V wraps up the study, offering conclusions and potential extensions of this research into other domains.

II. Data and Methods

A. Data description

The data for this study was sourced from the Yelp Open Dataset, which is comprised of five segments, offering 150,346 entries of basic business details (like hours, address, and ambiance), 6.9 million customer reviews, totaling used data approximately 5.11GB.

My analysis was narrowed to restaurant reviews, utilizing both the customer feedback and business attribute datasets, both formatted in JSON. Post-filtering to isolate restaurant-specific data and also reduced the number of restaurants to lower the execution's time, I amassed 2,782,263 reviews from 43,819 distinct eateries.

The review dataset features include business ID, full address, price, category, and business type, among others. For my purposes, the model focused on business ID, the textual content of reviews, and the ratings. The business ID was crucial for data processing, while the text of the reviews served as the primary dataset for my sentiment analysis, with the ratings aiding in distinguishing between positive and negative sentiments.

B. Data Cleaning

I combined the business and reviews datasets using the "business id" as a key attribute. Subsequently, I tokenized the text in each review, eliminating all punctuation to create a "bag of words" for each entry. Following this, I applied stemming, lemmatization, and stop word removal to each word collection, utilizing Python's NLTK package's built-in functions.

The combined dataset was then randomly split into training and testing sets in an 8:2 ratio. For the purpose of this analysis, reviews with a rating of 4 or higher were categorized as positive, while those with lower ratings were deemed negative. This categorization was used in most of the article that I have reviewed [2].

C. Model Tuning

In my analysis, I utilized an exhaustive search over specified parameter values algorithm to fine-tune the parameters of my Random Forest model, aiming to elevate its predictive accuracy. I implemented cross-validation with 5 folds to ensure a rigorous evaluation framework, which helps prevent overfitting and offers a dependable measure of the model's performance on new data. Accuracy was chosen as the key metric for comparison during the exhaustive search, with the goal of finding the parameter configuration that yields the highest accuracy. This meticulous parameter optimization was carried out for two models adopting distinct feature selection techniques: the bag-of-words and tf-idf. Through the exhaustive search, I methodically investigated all possible parameter combinations, enabling us to pinpoint the most effective settings for each method of feature extraction, thus optimizing the model's capacity to analyze sentiments in restaurant reviews accurately.

D. Methods

The random forest algorithm uses "bootstrap aggregating" and "random subspace process" methods to generate a series of decision trees that are categorized using a set of decision trees (Breiman, 2001) [5]. The output categories in a random forest includes several decision tree classifiers determined using the decision tree classification results mode. The random forest algorithm uses two random selection methods to construct a single decision tree, the first is a random selection of training samples, and the second is a random selection of sample feature attributes. After all the decision trees have been created, the final classification result is determined using the equal weight voting method. This classifier has become popular in data processing due to the accuracy of its classifications. Correlation between trees is minimized by random selection of features, which increases the ability and efficiency of forecasting.

For feature selection within the Random Forest model, I utilized two approaches: the "bag-of-words," which accounts for the occurrence rates of different words in each review, and "tf-idf," which stands for term frequency-inverse document frequency, a statistical measure used to evaluate the importance of a word to a document in a collection or corpus. Reviews were labeled as "positive" or "negative" based on their rating values to facilitate this analysis. After training the model and metrics evaluation, A comparison of the models would be made and the model with best accuracy would be selected for implementing the next step of calculating overall polarity score of each words.

In my model, each word is considered an individual feature, which could lead to a high-dimensional, sparse feature matrix. To address this, I utilized Python dictionaries for each review to store only the words that occur, reducing the prevalence of zero entries and thereby reserving significant computational resources.

To identify the words that reflect customer concerns or highlight distinctive attributes of each restaurant, I excluded words which are simple sentiment descriptors like "good", "amazing", or "terrible". I calculated a 'polarity score' for each word in a review by multiplying its overall feature importance with its frequency across all reviews and the number of time it appear in the review, while the sentiment will decide whether the value is positive or negative. Subsequently, summing all the 'polarity score', I have achieved the 'Overall polarity score' for each word.

$$Overall\_polarity\_score(w) = \sum_{i}^{n\_reviews} w\_importance(w) \tag{1}$$

$$* count(w, r[i]) * frequency(m) \tag{2}$$

$$* predicted\_sentiment(r[i]) \tag{3}$$

Additionally, by extracting the most significant positive and negative words across all reviews using 'Overall polarity score', I gain insights into the attributes most valued by customers and discern broader trends in preferences for dining experiences globally.

III. Results

Utilizing the bag-of-words model and after fine-tuning, the Random Forest classifier achieved optimal accuracy with default parameter settings. On the test dataset, the classifier's accuracy reached 87.861% when configured with 100 trees. Conversely, when employing the tf-idf feature selection method, adjusting the minimum number of samples required to split an internal node to 10, while maintaining the same number of trees, resulted in a slightly lower accuracy of 87.943%.

Contrary to expectations, the taste of the food was not the foremost positive aspect highlighted in reviews. Instead, the quality of service stood out, with words like 'friendly' and 'time' ranking highly. This indicates a trend where customers prioritize service over the flavor of the dishes, with a particular preference for freshness in their culinary choices (Figure 1).

The analysis of the top negative words revealed 'manage' as a major issue, significantly impacting customer satisfaction. Additionally, slow and rude service emerged as critical areas for improvement in the hospitality sector. The context of 'money' remains common when it come to customer's negative reviews (Figure 2)

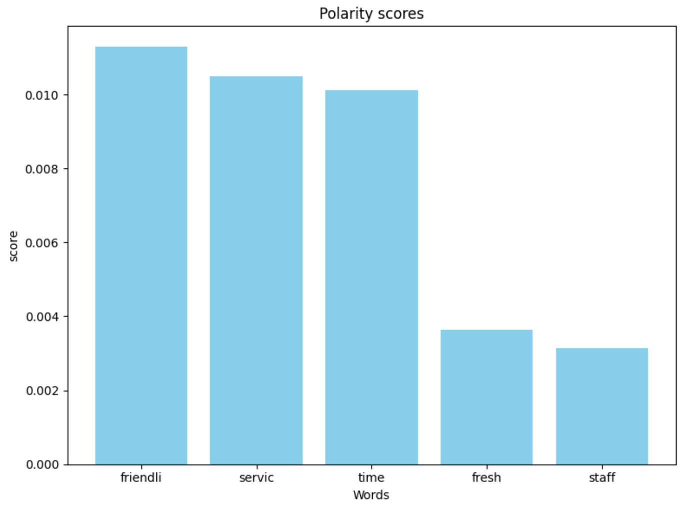

Fig. 1. Top 5 Positive Words of Cuenelli's Peruvian Restaurant

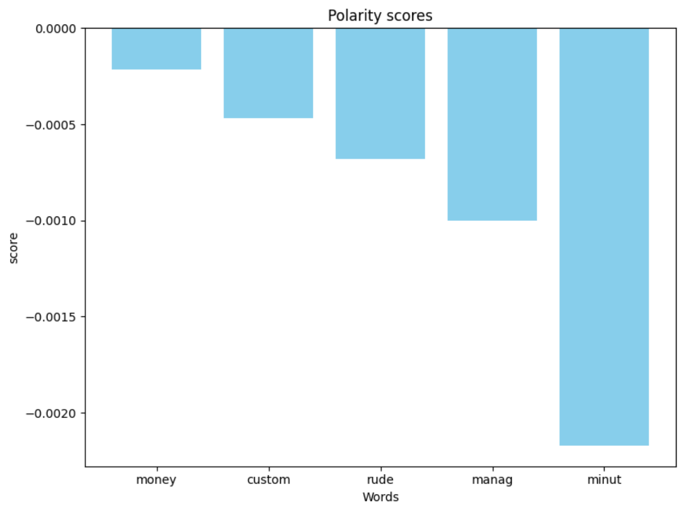

Fig. 2. Top 5 Negative Words of Cuenelli's Peruvian Restaurant

IV. Discussions

Since our analysis may help to extract specific features from any set of reviews, restaurant owners can make good use of it for essential information once they receive a certain amount of Yelp reviews. From those reviews, we can understand why customers love or dislike their restaurants, maybe great reviews primarily due to fresh food, or perhaps unsatisfied reviews caused by too high prices. From the point of customers, a more user-satisfying recommendation and a more considerate appraisal of the restaurants can be expected if Yelp includes more features in its overall evaluation of each restaurant using the technique we have developed. Although the performance of our model is decent, there are still a lot of spaces for improvement. One of the suggestions for future work is to try other classifiers like boosting or neural networks to check whether they may outperform the Random Forest model.

V. Conclusion

In this study, I introduced a innovative approach to uncover various features of restaurant, leveraging a highly precise Random Forest model to compute word scores and assess sentiment polarity. The pivotal attributes identified through my methodology can aid customers in selecting their preferred cuisine and offer restaurants insights into their strengths and improvement areas. Moreover, this approach is versatile and can be applied to other forms of feedback analysis, such as movie reviews or social media commentary. For example, implementing this sentiment analysis technique on reviews from platforms like IMDB or Rotten Tomatoes could pinpoint movies that excel in depicting 'love'. Looking ahead, I aim to apply this strategy to collate and interpret public opinions, extract meaningful insights, and generate actionable recommendations based on these insights.

References

[1] B. Yu, J. Zhou, Y. Zhang, and Y. Cao, "Identifying restaurant features via sentiment analysis on yelp reviews," *arXiv preprint arXiv:1709.08698*, 2017.

[2] S. SELOT and M. K. SAHU, "Support vector machine and random forest machine learning algorithms for sentiment analysis on tourism reviews: A performance analysis," 2021.

[3] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197–227, 2016.

[4] T. A. Rana, K. Shahzadi, T. Rana, A. Arshad, and M. Tubishat, "An unsupervised approach for sentiment analysis on social media short text classification in roman urdu," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 2, pp. 1–16, 2021.

[5] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.