**Team 2 - Aspect Analyzers**

**Comparative study of LDA, NMF and BERTopic to perform Topic Modeling and Aspect Based Sentiment Analysis on Yelp reviews**

**Aparna Gopalakrishnan (**[**apgopala@umich.edu**](mailto:apgopala@umich.edu)**)**

**Mexi Liang (**[**mexil@umich.edu**](mailto:mexil@umich.edu)**)**

**Xue Xu (**[**xuxue@umich.edu**](mailto:xuxue@umich.edu)**)**

*Project Advisor:* **Prof. Rachael Tatman**

**Introduction**

Sentiment Analysis is a technique to understand sentiment or emotions associated with text data. It is widely used to understand customer emotions in product reviews, survey and feedback data, social media posts, etc. Sentiment Analysis is a complex subject given the nuances of how varied humans write and also due to the short-form nature of some of these texts. Additionally, how does one deduct sentiment from a text like, ‘The atmosphere was splendid, but the food was terrible’? Would we say sentiment is ‘positive,’ ‘negative’, or ‘neutral’? The approach to this is Aspect Based Sentiment Analysis. For the aspect ‘atmosphere’, the sentiment is positive, and for the aspect ‘food’, the sentiment is ‘negative’. So, essentially, ABSA helps us understand not just whether people like or dislike something but also why they feel that way by looking at different aspects individually.

We intend to study Yelp reviews through Aspect-Based Sentiment Analysis (ABSA) to

discover the aspects or dimensions and their sentiments in the review text. With respect to restaurant reviews, aspects mean dimensions like 'food quality', 'ambiance' etc.. The goal of our project is twofold: For customers searching for a specific restaurant we want to provide a list of top aspects along with their sentiment score for that restaurant. For business owners, we want to provide a comprehensive solution for improving aspects of reviews written for their restaurants, and this would be accomplished by finding similar restaurants which have high ratings for these aspects and providing most useful reviews associated with these aspects.

**Literature Survey**

Topic modeling in the field of NLP has been studied widely over the years. In the paper, Egger et al., 2022 shares the importance of applying different topic modeling algorithms on the same dataset to validate the results and understand each algorithm's nuances. This paper discusses the complexities of modeling unstructured text data and how algorithms like LDA, NMF, and BERTopic can yield varying results and also provides techniques to evaluate topic modeling results on Twitter data. Our project is similar to this paper in that we are going to use LDA, NMF and LLM-based approaches and compare the results on Yelp reviews data but different in the sense that we are aiming to do Aspect based sentiment analysis as one part of our project.

Sentimental analysis is useful to understand overall sentiment of text content but it doesn’t tell about sentiment of a text with respect to a specific aspect. Anoop, V et al., 2018 made us realize the importance of considering aspect-specific sentiments in text data analysis, providing deeper insights into customer preferences and experiences. This could be crucial for improving the restaurant's business service and customer experience in our projects. Their method involves preprocessing the reviews, applying topic modeling, mapping topics to product aspects, and calculating aspect-level sentiment polarity. We take inspiration from this approach and use it as a guide for our projects while specific models and methods can be explored and refined during practical implementation. We also explore the SetfitABSA model and pre-train our model as demonstrated in the article by Chiusano, F.

**Data Preprocessing and Cleaning**

The dataset is made available by Yelp on their website: <https://www.yelp.com/dataset>. The main datasets that we are using are reviews and business. Reviews dataset contains review data for all of the merchants while business dataset contains descriptive data for the merchants. The original data is well formatted and clean already so we didn’t need further data cleaning. However, the datasets are in JSON format and the reviewes dataset itself is 5.34GB with close to 7 million records. We converted the dataset format from JSON to csv for better performance.

The review dataset includes customer reviews for all types of businesses but our project focus is restaurant reviews so we had to filter out non-restaurant reviews. For our training purpose, we decided to go with restaurant reviews for California given the smaller data size and diversity of this state.

The approach we took to get rid of non-restaurant reviews is to gather all the unique values in *categories* column from the business dataset and then manually list out all the food related labels. This list of labels is used to determine if a business is a restaurant. Please refer to ca\_restaurants\_used\_labels.csv which is in the following [Google Drive location](https://drive.google.com/drive/u/0/folders/1qXAxWMqGp2nNc_dIkvIRgi7S9OUIU9Mt) for details.

**Methodology**

**1. Comparative study of LDA, NMF and LLM based approaches to perform ABSA on Yelp reviews.**

**1.a. LDA and NMF Topic Modeling:**

For LDA and NMF based approaches: We plan to conduct a preliminary exploration of aspects in reviews using LDA and NMF topic modelings. This involves identifying dimensions within reviews and visually representing the vocabulary associated with each dimension. The analysis will apply to all California dataset we captured.

We adopt NLTK and Gensim library helping on data processing, including the following steps: 1) Lower Case, 2) Tokenization, 3) Remove Stop Words, 4) Removing Emojis, URL, Emails, and Special Chars, 5) Stemming, 6) Replace cuisine-related entities (sushi, roll, ) by a general word 'food'. For Step 4, we did a data exploration regarding Emojis. Although Emojis convey emotions, we chose to filter out emojis. Because in our analysis only three Emojis catched: !!, ®, ™. The last step is essentially a post-processing step to make the LDA and NMF topic model result more aspect oriented instead of cuisine oriented.

We will find the optimal parameters for the two models, and extract 10 top words for each of the topics. By manual inspection and the help of chatGPT, we will be able to give the topic names for topics extracted from the two models.

**1.b. LLM based Approaches:**

1. Using pre-trained models:

For the LLM based models, we plan to use two approaches: The first one relies on the BERTopic model to extract aspect-related topics and the second one focuses on extracting aspects and sentiment associated with each aspect using the pre-trained Deberta model

BERTopic model:

For the first phase, we converted the review text to lowercase and then removed stop words. The results obtained were very cuisine centric. The model was able to cleanly identify cuisines from the reviews but given that we were looking for aspect specific topics. So we finetuned the approach and the modified approach is detailed below:

1. Firstly, we converted the text reviews into lower case
2. Then, for the aspects we were interested in, we used a list of initial aspects and used ChatGPT to extract 10 synonyms for each of the aspect
3. We only extracted sentences with these aspects or synonyms of the aspects for the inference in Bertopic model
4. Lastly, we used seed words to ensure aspect specific words are weighted higher
5. Finally, for tuning the BERTopic model we wanted to reduce the number of irrelevant topics by setting the minimum topic size to 100.

This resulted in getting topics that were more aspect-oriented than our initial approach.

Pre-trained ABSA using Deberta Model:

For the pre-trained Deberta model, we used the ‘yangheng/deberta-v3-large-absa-v1.1’ model from Hugging Face. The detailed approach is given below:

1. We used the review text for the specified aspect, extracted sentiment for each aspect
2. For each aspect along with the sentiment, the model also provides a confidence score
3. We will extract sentiment for a set of aspects that we obtained from a manually annotated data set. This manually annotated data set is used later for Setfit ABSA. Results from the pre-trained Deberta contain a sentiment label and a confidence score for every aspect we input.
4. If the score is too low, we don’t want to consider it in our analysis. For example, if the score is 0.5 and sentiment is ‘Negative’ we don’t want to consider the sentiment as Negative since the score is only 0.5. So we further applied a condition to only keep sentiment as is if the score is greater than or equal to 0.7.
5. Additionally, the sentiment scores are obtained per review per aspect which we then aggregated at a restaurant level.
6. There were cases when nothing was mentioned about a particular aspect for a restaurant in which case the sentiment would be “no mention” or when an aspect is rated equal # of times positively and negatively we marked the sentiment as “neutral”. Thus we ended up with 4 sentiments per aspect - positive, negative, no mention and neutral.

This is what was used for further evaluation and analysis.

B. Pre-training our own model

The second approach we used is the SetFitABSA model wherein we pre-trained our model with manually annotated dataset. We randomly selected 40 samples from the California restaurant reviews dataset and then manually extracted aspects and sentiments from the reviews to use as our training dataset. There are 19 aspects that we extracted from the samples, such as food, service, seating, price, and etc .Sentiments that we have identified include positive, negative, neutral and mixed. Each of the aspects extracted from each sentence becomes an individual training record. 102 training records were prepared from these 40 sentences. We splitted this dataset into 2 parts. The first part is the training data for the model, which contains 50 rows of records. The second part is the evaluation data for the model, which contains 52 rows of records.

To begin with the training, we must use a new AbsaModel, and the sentence transformer model used for aspect filtering, the sentence transformer used for polarity classification and spaCy model used is "sentence-transformers/all-MiniLM-L6-v2", "sentence-transformers/all-mpnet-base-v2" and "en\_core\_web\_sm" respectively. We then feed our training dataset to AbsaTrainer from the setfit[absa] library and use the evaluation method provided by the same library to inspect accuracy of the aspect and accuracy of the polarity. We got 96% accuray on aspect and 65% accuracy on polarity after running evaluation.

**Part 2:** To improve the low scoring aspect for a restaurant, we will rely on finding similar restaurants that are rated highly for the same aspect and find the most useful reviews with the corresponding aspect in it.

Our methodology encompasses several critical steps:

### **Step 1: Restaurant Selection Based on Similarity**

To identify similar restaurants, we first filter potential candidates based on the same price level. Next, we compute the cosine similarity between restaurants based on their cuisine type and dining style. Cosine similarity measures similarity between multidimensional vectors where each dimension corresponds to a category attribute.

### **Step 2: Similarity Threshold Setting**

A similarity threshold is set. Restaurants with a similarity score above this threshold are considered. In this study, we establish a threshold value of 0.7, which can be adjusted based on empirical analysis.

### **Step 3: Filtering and Ranking Based on Aspect Ratings**

From the list of similar restaurants, we proceed to filter out reviews with negative labels for the aspect restaurant owners care about. Our focus is solely on positive reviews to understand better practices that lead to higher satisfaction in that aspect. Subsequently, the positive reviews are ranked according to their confidence scores.

### **Step 4: Reviews Extraction**

The top ten reviews, based on the highest confidence scores, are selected and presented to users. These reviews might provide insights concerning the specified aspect. By analyzing these reviews, they can derive actionable feedback for improving the same aspect in the target restaurant.

**Evaluation**

**Manual Evaluation of Deberta Model results**

We manually annotated the data set to measure the performance of the model. We calculated the proxy accuracy (we call it proxy since we only labeled a sample of the reviews) and we got the accuracy value of 80.0%. The confusion matrix is attached in Appendix as figure ‘Appendix 1.A’.

**Manual Evaluation of SetFit ABSA Model results**

Similar to what we did for the Deberta model, we did a similar manual evaluation for SetFit ABSA Model results as well. For SetFit ABSA model there were two types of evaluation that was performed:

1. Evaluation on if aspects were extracted when they were present and we found the accuracy to be 50%. The confusion matrix (figure Appendix 1.B) and details are captured in Appendix.
2. Evaluation on if Aspects extracted were correct and we found the accuracy to be 64%. The confusion matrix (figure Appendix 1.C) and details are captured in Appendix.

**Evaluation of Topic Modeling Results**

We evaluated the Topic modeling results using two different approaches. The first one is relying on the coherence score measure to find the appropriate # of topics to be extracted and the second one is manual evaluation of the results by inspecting the top keywords extracted in each Topic and using best judgment to see if the Topics makes intuitive sense.

**Evaluation of Restaurant Recommender**

ReviewsRecommaner file provides a method: *get\_reviews\_recommender*(*res\_id, asp*

),which enables users (restaurant owners) to get a review list based on restaurants id and aspect name they care about. We can evaluate by samples in [this file](https://github.com/mexiliang/SIADS699/blob/main/2_Model/3_Recommender_Feature/ReviewsRecommander.ipynb).

**Results and Visualizations**

**Results from LDA and NMF Topic Modeling**

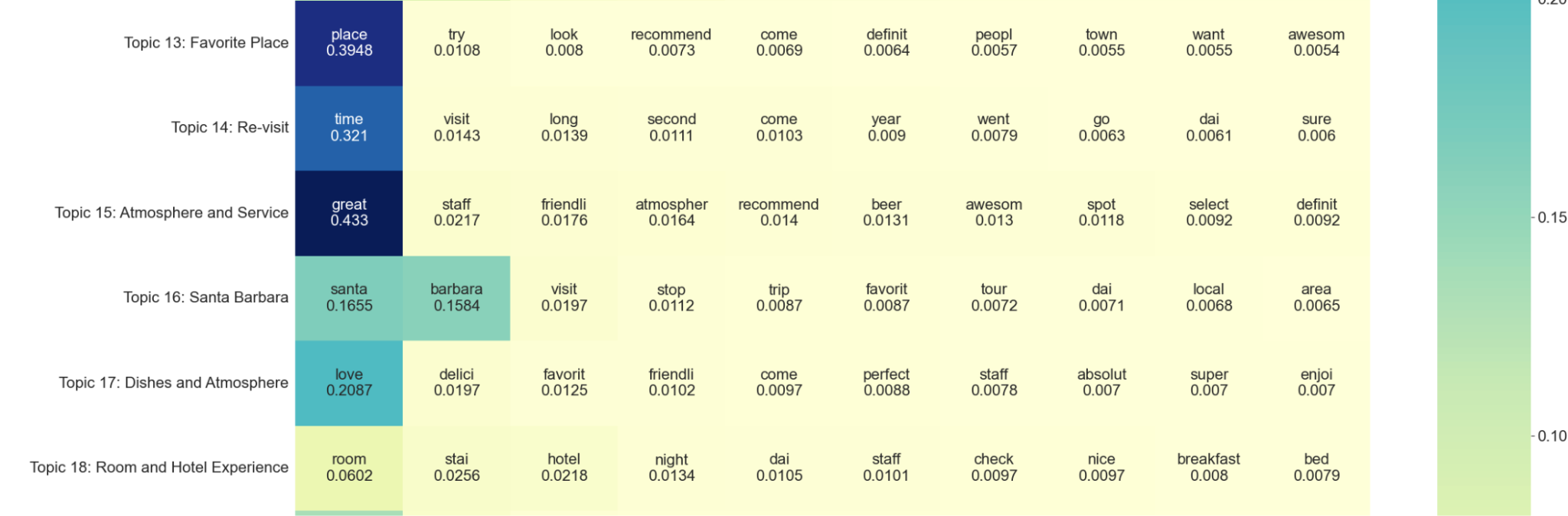
By plotting the coherence scores, we identified several potential numbers of topics. Combining human judgment and domain-specific knowledge into our evaluation process, we determined that the number of topics for the LDA model is 29, the NMF model is 22 topics. Fig 1 and Fig 2 show the results of the LDA and NMF model, displaying the topics identified by each model along with their top associated words. Y-axis is the topic name where we could extract aspects from. Each cell represents a word with its weight contributing to the corresponding topic. Full figures could be retrieved in [Github](https://github.com/mexiliang/SIADS699/blob/main/model%20training/TopicModel.ipynb).

In the LDA model, a couple of compelling topics emerged. Notably, Topic 11 and Topic 14 stand out, pointing to the prevalence of gluten-free and vegetarian options in Californian restaurants—a trend not captured in our initial training dataset. Moreover, Topic 28 introduces a topic of cleanliness and COVID-19, reflecting the increased focus on sanitary measures and precautions since the pandemic's onset, which suggests a novel aspect for consideration in future data training.

Upon inspecting Fig 2, we observe a substantial overlap between topics derived from the NMF and LDA models. While the NMF model misses some topics(Good for Group, Gluten-free), it distinguishes itself with some unique topics - Topic 14 'Re-visit', which talking about frequency of visiting, is not found within the LDA topic range. In the forthcoming section, we will explore the degree of similarity of three topic models LDA NMF and BERTopic model, providing further insight into the potential overlaps.



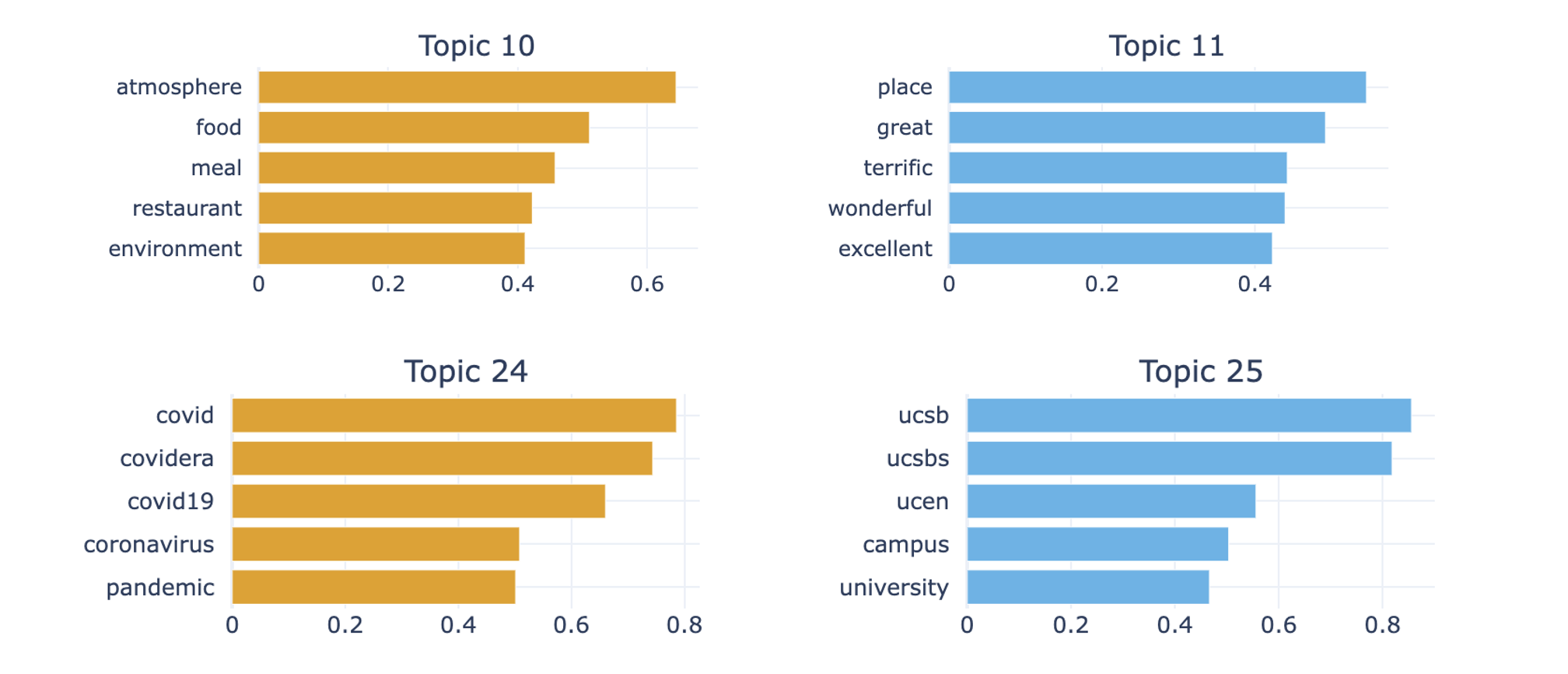
*Fig 1. Top words Heatmap of LDA model*



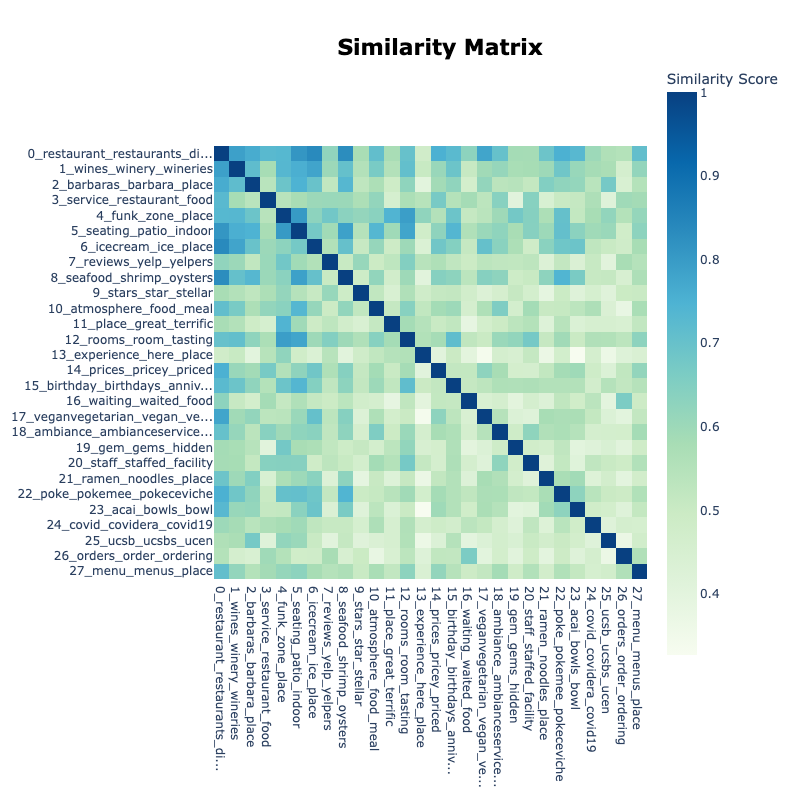
*Fig 2. Top words Heatmap of NMF model*

**Results from BERTopic Topic Modeling**

For the BERTopic model, we used the same number of topics as LDA which is 29 and set the minimum topic size to be 100 to ensure we are getting relevant topics. Fig 3 shows partial results obtained from BERTopic (the detailed visualization can be accessed via this GitHub [link](https://github.com/mexiliang/SIADS699/blob/main/4_Result_files/1_Topic_Modeling/3_TopicModel_BERTopic_only_aspect_based_sentences.html)). The topics are aspect related and some cuisine related and some are aspects which we didn’t input for inferencing like COVID related and University related.



*Fig 3. Top words per Topic from BERTopic model*



*Fig 4. Similarity matrix of the topics from BERTopic model*

Fig 4. below shows the similarity matrix returned by the BERTopic model. It uses cosine similarity using topic embeddings and generates a heatmap to show similarity between the topics matrix between topic embeddings, a heatmap is created showing the similarity between topics. Looking at the similarity matrix, we see topics with high similarity scores with other topics like vegan/vegetarians with ice cream, atmosphere with seating/patio, birthday with seating/patio etc suggesting that people talk about these topics together much more than other topics.

**Comparison of results between LDA, NMF and BERTopic**

Table 1. Aspects identified by BERTopic, LDA and NMF model with Keywords

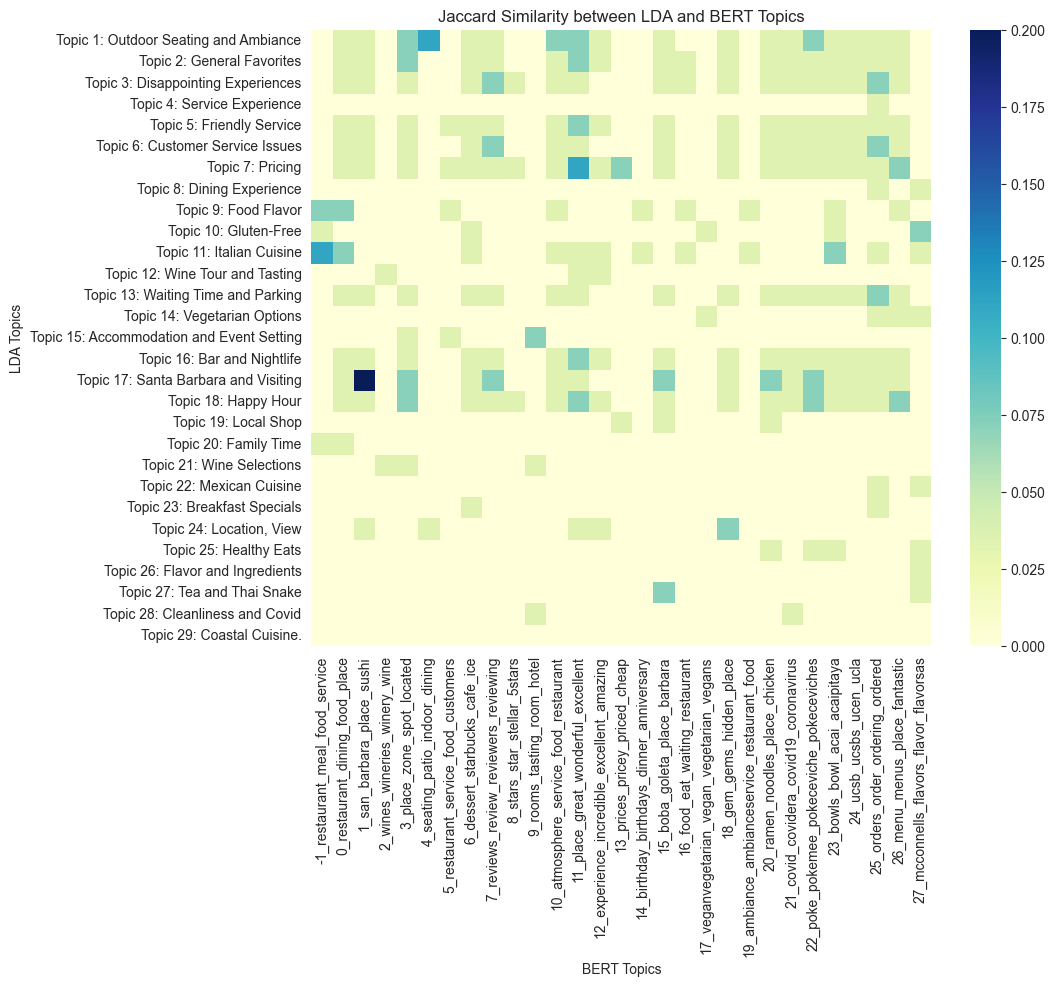
| **No.** | **Aspect/Topic** | **LDA Keywords** | **BERTopic Keywords** | **NMF Keywords** |
| --- | --- | --- | --- | --- |
| 1 | Seating | tabl, enjoi, good, area, lot, like, outsid, littl, sit, patio, space, dog, nice, friendli, great, love, place, outdoor, seat, insid | dining, sit, restaurant, seating, patio, outdoor, indoor, beach, picnic, sitting | tabl, ask, come, server, minut, waiter, check, water, took, waitress, seat, said, came, brought, arriv, reserv, told, hostess, final, sit |
| 2 | Ambiance | great, servic, friendli, place, staff, recommend, delici, love, amaz, good, definit, nice, super, highli, excel, come, time, awesom, atmospher, best | pleasant, superb, delicious, ambianceservice, restaurant, dining, meal, food, terrific, ambiance | great, staff, friendli, atmospher, recommend, beer, awesom, spot, select, definit, view, fun, price, fantast, experi, lunch, enjoi, locat, highli, excel |

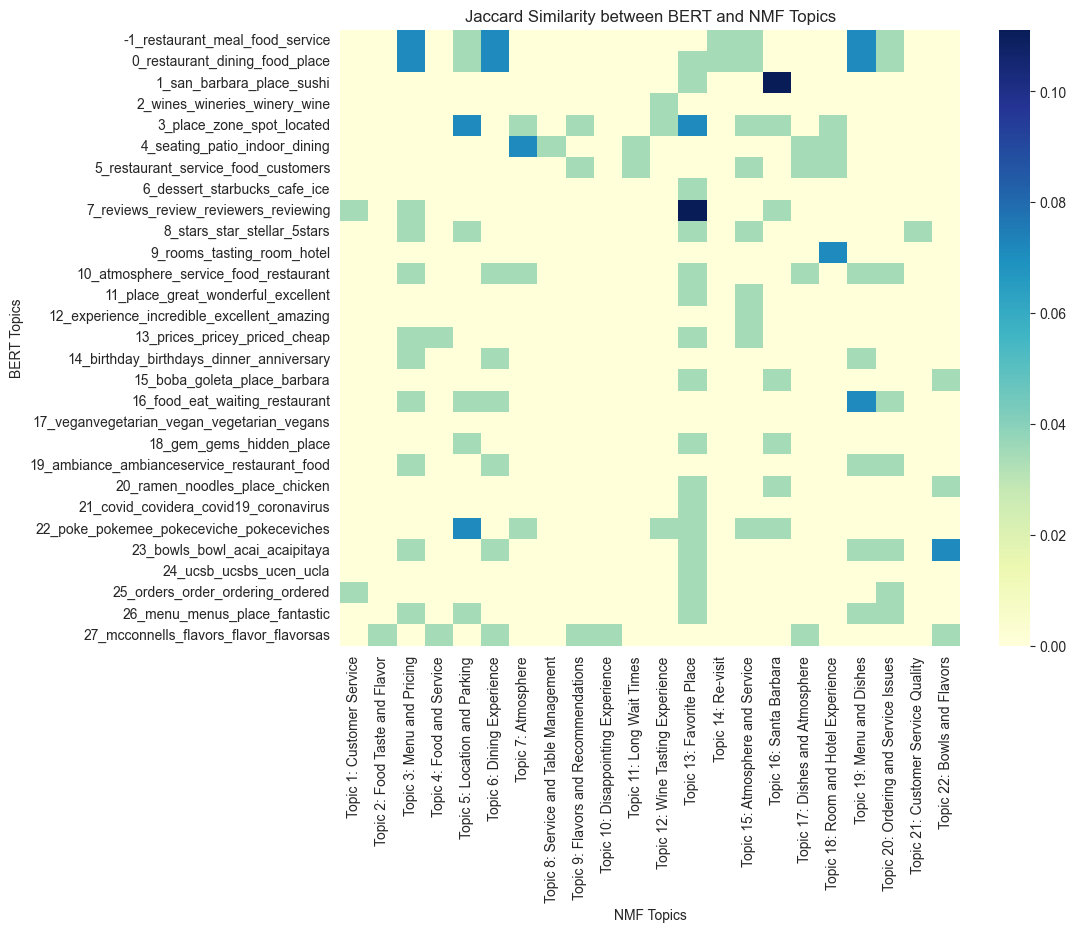
Table 1(full table is attached in Appendix) displays the aspects extracted from three topic models with their keywords. The three models focus heavily on the dining experience, identifying various aspects from the reviews such as “Setting”, “Service”, and “Ambiance” in Table 1. However, each model displays different granularity and focus. For example: LDA model seems to capture broader themes, combining various elements like food, service, and ambiance into one topic. It incorporates some unique and modern concerns like COVID-19 precautions very well. Although LDA includes words "gluten," "vegan," in some topics, the topic is kind of mixed; NMF provides more detailed insights into dining experience but with some degrees of overlap between topics. We observed that NMF mixes attributes with actions more frequently, as seen in topics like "want, said, ask, customer, manager," which are more reflective of customer service interactions; BERTopic could understand syntax well with the benefit of transformers, is able to extract specific cuisines or offerings (e.g., vegan, vegetarian), and geographical references and dining settings. Noticeably, it extracted some university oriented and Covid oriented topics.

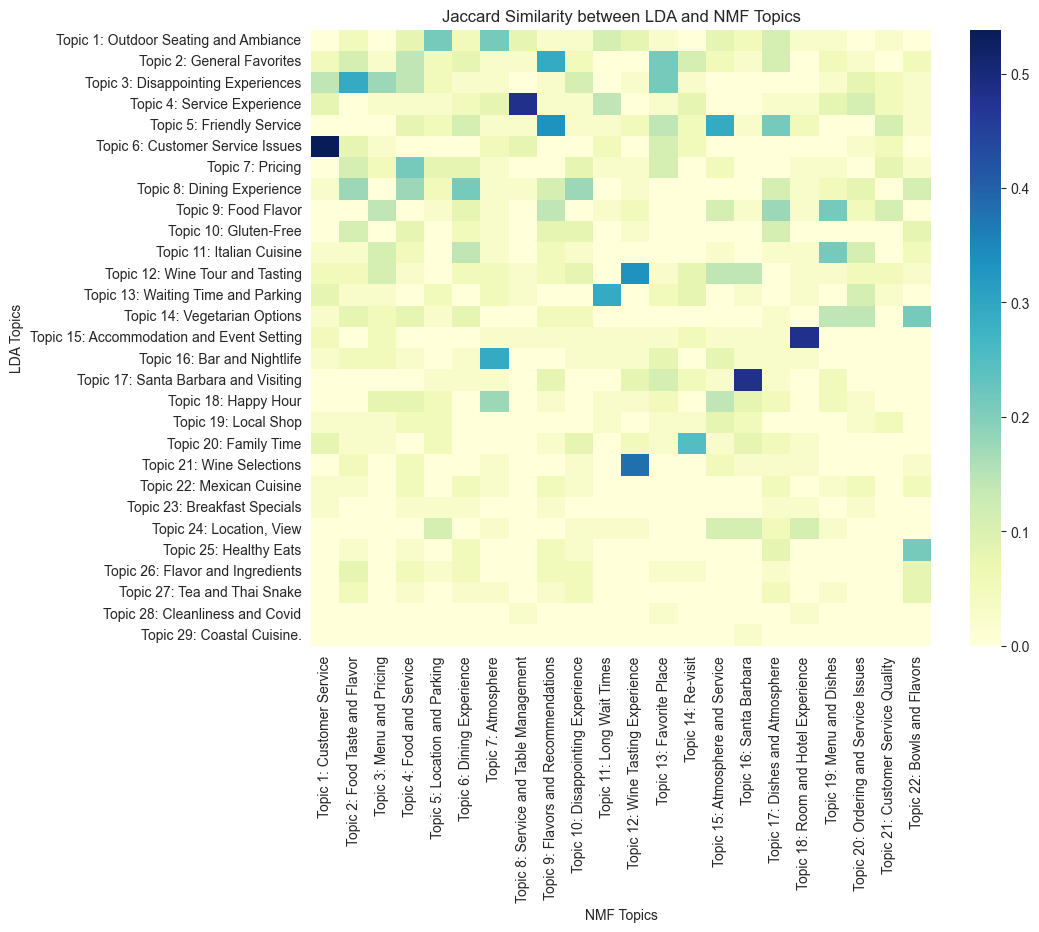
Re-occurrence words in different topics may be caused by topic overlap, but all three models frequently mention basic dining aspects like "food", "place", "flavor", and "service"(shown in Table 1). Instead of topic cross, it also makes sense that universal focus on these words is a fundamental aspect central to many discussions in restaurant reviews. And that is the main reason we did not use tf-idf in topic models, we do not want some core aspects to lose their weights by tf-idf method. NMF tends to have overlap in terms like "flavor," "sauc," and "dish" across topics related to food quality and preparation, indicating perhaps need to adjust parameters to distinct separation of topics.

To analyze the topic similarity across the three models, the Jaccard similarity(Fig 5) measure is employed. A high Jaccard score indicates a larger intersection relative to the union of the words in the topic sets, indicating greater similarity. The highest similarity score between LDA and NMF is around 0.5, which means they share almost 50% words in similar topics. While BERTopic only has 0.2 with other models. This may be caused by different preprocessing of text.

From Fig 5, cells regarding Santa Barbara, Customer Issues, and Service experience for all three Figures are darker, that means these kinds of topics are covered by three models. This reveals consistency in topic extraction. And we can also pinpoint the unique topics that each model, for example, Topic 29 Coastal Cuisine for LDA model and Topic 14\_birthday\_birthdays\_dinner\_anniversary for BERTopic model barely share common keywords with other models. Which could help in finding more unique topics by cross referencing.



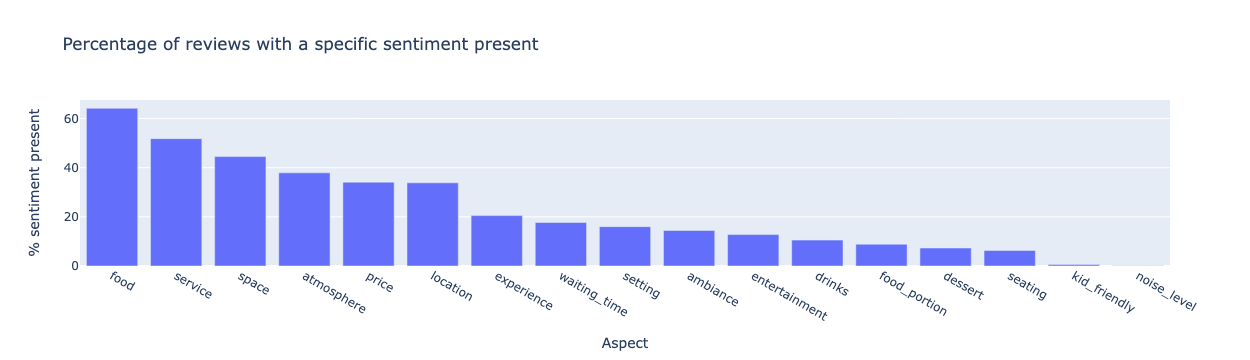




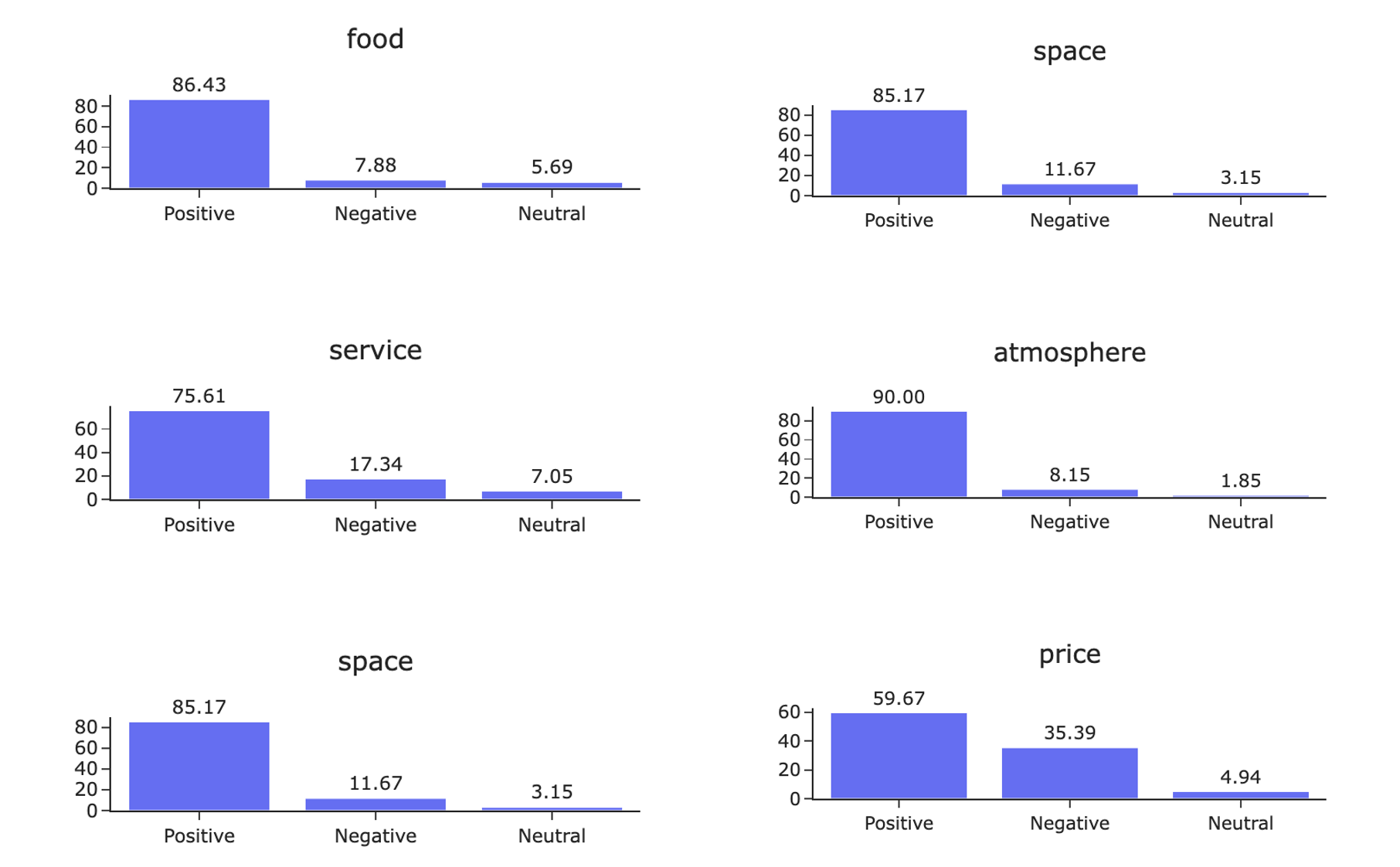
*Fig 5. Jaccard Similarity between topic models*

**Results from ABSA using pre-trained Deberta model**

For the results from the Deberta Model, given that we are asking the model to infer sentiment for a set of aspects, we first need to evaluate the top aspects for which sentiments were inferred by the model. Figure 6 below shows in descending order the aspects with the percentage of reviews with any sentiment retrieved by the model and we see our 6 top ones are: food, service, space, atmosphere, price and location.

****

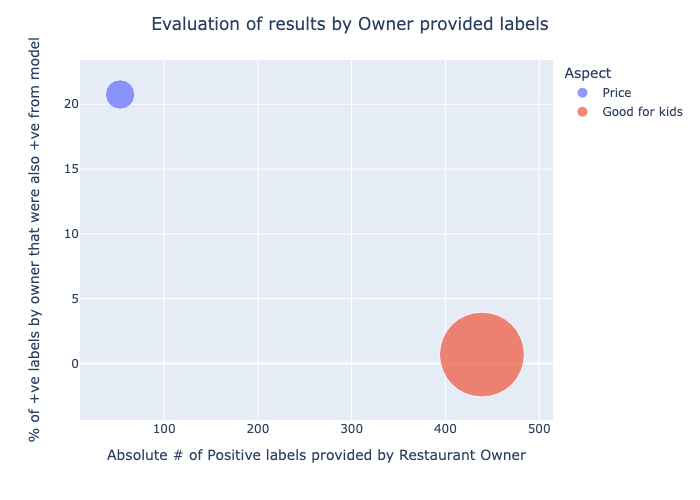
*Fig 6. Percentage of reviews with a sentiment present for a given aspect*

****

*Fig 7. Distribution of sentiment per Top 6 aspects*

With all these insights, we then looked at the distribution of sentiment for each of the top 6 aspects and what we found were: when people review about ‘price’, 35% of the time it is with negative sentiment. Price is the aspect that when people talk, they talk with a negative sentiment than aspects like service or food. When people talk about atmosphere or food it is predominantly to share a positive sentiment.

With the yelp data, there were labels that were either restaurant owner provided or provided by users (we didn’t have a way to distinguish between the two). We used the aspects obtained from the model to validate what the restaurant owners provided about them. In Figure 8, we look at the two aspects ‘price’ and ‘good for kids’ (the only two ones for which we could do the comparison), we see that for the aspect ‘price’, which has very few positive labels (sample size less) but has been provided positive labels in 20% of cases by the model. Interestingly, for “kid friendliness”, we see that when restaurant owners say a restaurant is kid friendly we see very small % of cases the model returning or in other words people reviewsing that the restaurant is kid friendly. It can be due to people not thinking that the restaurant was kid friendly or due to people not mentioning it in their review.

****

*Fig 8. Distribution of sentiment per Top 6 aspects*

**Results from SetFit ABSA**

Based on the evaluation results, we observed a 50.5% of accuracy of the SetFit ABSA model extracting aspects when they are present and 64.3% of accuracy on correctly predicting the sentiment of an aspect. Interestingly, this model can only understand exact word matches, meaning that it cannot understand a burrito is a type of food and if people talk about a delicious burrito, it cannot extract food as an aspect and positive as sentiment from this review. While inspecting the model predictions, we have only seen positive sentiment mapped to reviews, the other three sentiments negative, mixed and neutral that we identified during model training were not present in the model predictions.

**Comparison of results between SetFit ABSA and Deberta model**

Given that both SetFit ABSA and Deberta models are both ABSA models, we compared the sentiment extracted for reviews whenever there was a match. Firstly, SetFit ABSA only gave sentiments in 51% of the times when the Deberta model produced results. Secondly, when we could find a match of results between SetFit ABSA and Deberta model in 80.3% of cases both the results matched.

**Results from Restaurant Similarity**

In order to prepare potential improvements for certain types of restaurants, we created a review recommendation to allow users to explore similar restaurants with positive reviews. This recommendation requires a business id of a restaurant, and an aspect that a user is interested in reading. To achieve this goal, firstly we filtered similar restaurants by price range and restaurant category description, so that restaurants sharing similar characteristics can be easier grouped together. And then we sorted the results by aspect confidence score, and returned top ten reviews from similar restaurants.

**Discussion**

Our initial hypothesis before implementing the project was that we would be able to extract aspect related topics using approaches like LDA, NMF and BERTopic fairly easily. We also had high expectations of the results generated by large language models. Contrary to our beliefs, we found few surprising results like COVID and vegan/vegetarian related topics evolving as part of the analysis. For LDA and NMF, they are able to generate a variety of engaging topics under different topic numbers. From BERTopic, to extract aspect based topics we had to tune the model as well as pass modified reviews (which was different from originally planned). Additionally for the ABSA models, from the Deberta model we were able to see that the aspect with the most negative sentiment mentioned is the price and the aspects with the most positive sentiment shared are atmosphere and food.

There were few limitations we faced as part of this project. In our study, the results of topic models are pretty sensitive to parameters. For example, the quality of the topics generated highly depends on the choice of the topic number. Another limitation is that topics are too broad compared to the fine-grained aspect we expected, and aspect extraction relies on human inspection and domain knowledge. For the SetFit ABSA model, given that we are training it on yelp reviews data set, we expected high accuracy. But the model only correctly identified presence of an aspect 50.5% of the times and when it extracted the absence the accuracy was only around 64%.

**Ethical considerations:** We expect a few ethical challenges from our dataset such as 1) dishonest reviews submitted by users and accounts that are created by the business itself for review only purposes in order to boost their appearance on Yelp. 2) Inaccurate reviews by paid Yelp users to make compliment on the restaurants that aren’t necessarily good. Considering this should be a minority of the entire dataset, we shouldn’t need special treatment for this but just keep this in mind while performing our analysis. Also, an additional ethical aspect that our project advisor flagged to us was the sentiment biases introduced by LLMs as mentioned by Kiritchenko et al., 2018 We found this first hand while doing manual evaluation of our results of LLM wherein the model was hallucinating an aspect to be present and providing a sentiment. We hope that by manually evaluating the results and by fine tuning the input to the model we have to a certain degree treated for these biases but it is possible that there are biases that we have not accounted for.

**Conclusion**

We did a comparative study and performed Topic modeling and Aspect based sentiment analysis on Yelp reviews. For the Topic Modeling part of the project we used LDA, NMF and BERTopic to identify aspect based topics and then used ABSA models like SetFit ABSA and a pre-trained ABSA model using Deberta for sentiment analysis. Our learnings are summarized below: firstly, in topic modeling the three models focus heavily on the dining experience, identifying various aspects such as “Setting”, “Service”, and “Ambiance” and the three models showed consistency in topic extraction. Secondly, with fine-tuning Deberta model performance improved significantly and we saw users providing positive sentiment about aspects like atmosphere and food. Thirdly, the pre-trained SetFit ABSA model even with training on our own yelp reviews data set performed poorly in terms of accuracy which was unexpected. Fourthly, regarding a certain aspect, a review recommendation is provided for users to explore similar restaurants with positive reviews. Finally, as our project advisor shared with us in meetings, it is very pertinent to manually evaluate results in case of text data projects. We were able to make changes to our projects and get better results due to the manual evaluation we did as part of the fine tuning. For the future iterations of this work, we would like to apply more NLP techniques like stemming and lemmatization to our data, applying different embeddings and other LLMs and also exploring data from a different city to compare and contrast the results.

**Statement of work**

See below in detail the work each of the team members did for this project. We summarized the role as primary when the team member was the main contributor for that module of the project.

|  | Aparna | Mexi | Xue |
| --- | --- | --- | --- |
| Data Preprocessing |  | Primary |  |
| SetFit ABSA model  *(Training, Inference and Manual evaluation)* |  | Primary |  |
| Deberta ABSA model  *(Inference and Manual evaluation)* | Primary |  |  |
| LDA Topic model |  |  | Primary |
| NMF Topic model |  |  | Primary |
| BERTopic model | Primary |  |  |
| Similar Restaurant Recommender |  |  | Primary |
| Team meetings coordination and planning | Primary | Primary | Primary |
| Report | Primary | Primary | Primary |

**References**

Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2VEC, and BERTopic to demystify Twitter posts. Frontiers in Sociology, 7. <https://doi.org/10.3389/fsoc.2022.886498>

Chiusano, F. (2022, August 19). Quick intro to Aspect-Based Sentiment Analysis - NLPlanet - Medium. Medium. <https://medium.com/nlplanet/quick-intro-to-aspect-based-sentiment-analysis-c8888a09eda7>

Anoop, V. S., & Asharaf, S. (2018). Aspect-Oriented Sentiment Analysis: A Topic Modeling-Powered approach. Journal of Intelligent Systems, 29(1), 1166–1178. <https://doi.org/10.1515/jisys-2018-0299>

Svetlana Kiritchenko and Saif M. Mohammad (2018). Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. Proceedings of the 7th Joint Conference on Lexical and Computational Semantics (\*SEM). <https://aclanthology.org/S18-2005/>

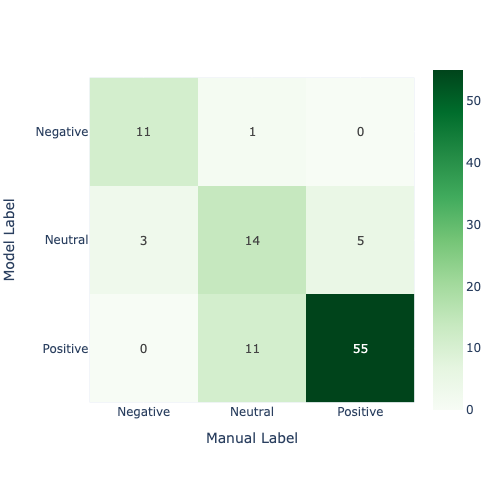
# Appendix

Based on Yelp documentation <https://www.yelp.com/dataset/documentation/main>

| **column name** | **description** | **data type** |
| --- | --- | --- |
| review\_id | An unique identifier of a review (22 characters) | string |
| user\_id | An unique identifier of an user (22 characters) | string |
| business\_id | An unique identifier of a business (22 characters) | string |
| stars | Star rating of a business | int |
| useful | The number of useful votes received | int |
| funny | The number of funny votes received | int |
| cool | The number of cool votes received | int |
| text | Actual review | string |
| date | Date the review received | date format YYYY-MM-DD |
| name | Name of the business | string |
| address | Address of the business | string |
| city | City of the business is located | string |
| state | State of the business is located | string |
| postal\_code | Postal code of the business is located | string |
| latitude | Latitude of the business | float |
| longitude | Longitude of the business | float |
| review\_count | Number of reviews | int |
| is\_open | Status of the business, 0 or 1 for closed or open | int |
| attributes | Business attributes to values | dict |
| categories | An array of strings of business categories | list |
| hours | Key day to value hours, hours are 24hr format | dict |
| is\_restaurant | Flag to identify if a business is a restaurant, Y or N for restaurant or non-restaurant | string |

**Manual Evaluation of Deberta Model results**

*Fig. Appendix 1.A Confusion matrix Manual label and Model label Deberta model - aspect extracted are correct?*

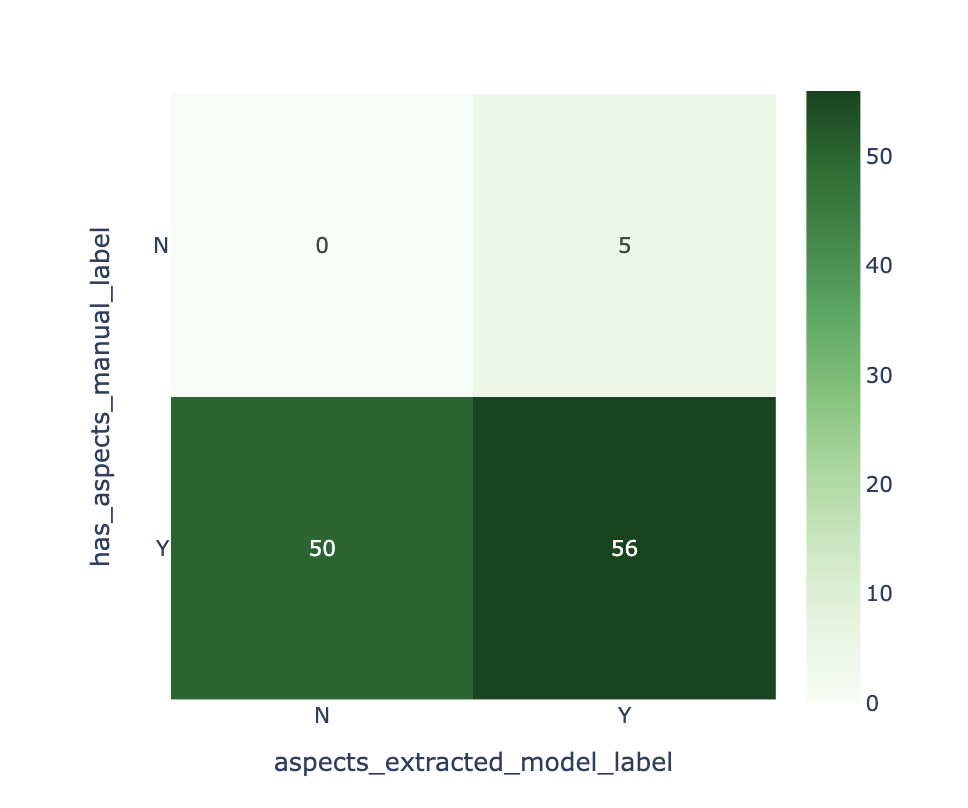


**Manual Evaluation of SetFit ABSA Model results**

In order to evaluate the performance of SetFit ABSA model, we randomly selected 2000 records from the California restaurant reviews dataset and dropped any records that were previously used in our model training. Then we inferred the random samples with our trained SetFitAbasa model. There were only 886 records from this dataset that contain one or more aspects. We then arbitrarily selected 100 records from the 2000-record dataset to perform manual validation to understand whether there are actual aspects and sentiment embedded in the review text. We used has\_aspects\_manual\_label to describe if there is actually an aspect embedded in the review and aspects\_extracted\_model\_label to describe if the model was able to extract the aspect. The aspect column is to capture the aspect, and model\_label and manual\_label is to narrate the sentiment predicted by the model and the sentiment manually labeled by us correspondingly. The last column is\_actual\_restaurant represents whether this business is a restaurant. During our data cleaning process, we used a list of food related labels to filter out non-restaurant. Therefore, a business that has a food related category would be considered as a restaurant even though it might not be a true restaurant.

1. Evaluation on if aspects were extracted when they were present

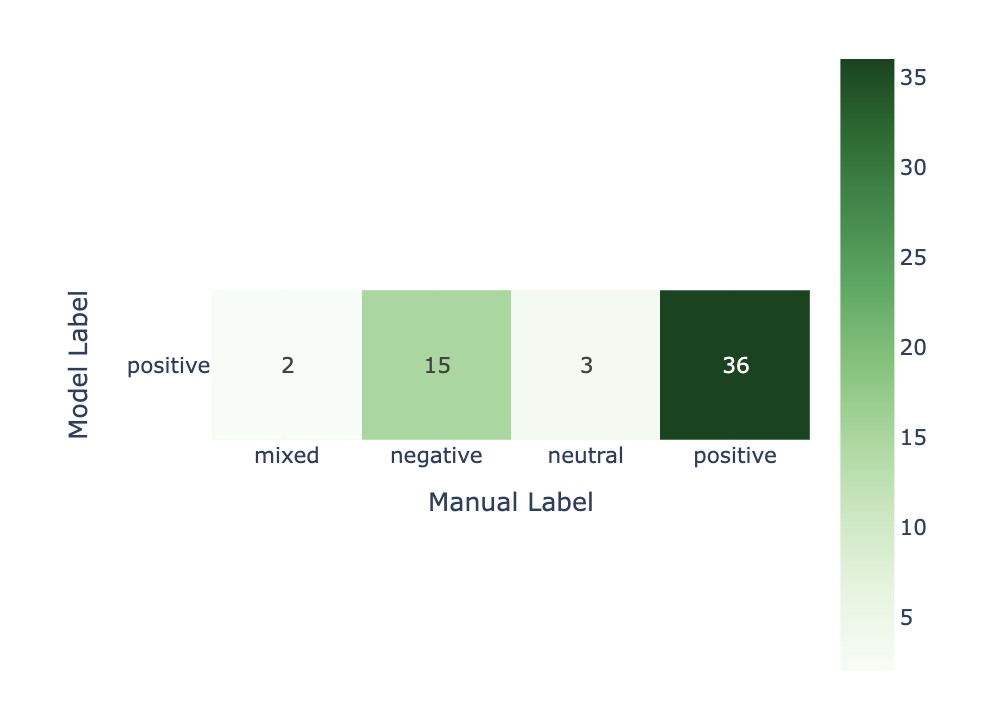
SetFit ABSA sometimes doesn’t return even when there is a relevant aspect present. So we did first the evaluation to see aspects were extracted when they were present and plotted the confusion matrix as below. We calculated the accuracy to find that it is just 50% meaning this is a big area of improvement for the SetFit ABSA model and any future iterations we do on this model



*Fig. Appendix 1.B Confusion matrix Manual label and Model label SetFit ABSA model - if aspects are extracted*

1. Evaluation of if Aspects extracted were correct

Next, we evaluated when the aspects were extracted if that was correct and plotted the confusion matrix which is shown in figure 3 below and calculated the accuracy and it was 64.3% which is not very good but still better than coin toss. Another limitation from the evaluation that arose was how the model label were always positive even though during manual evaluation we found other sentiments like negative, neutral etc.. for reviews.

****

*Fig. Appendix 1.C Confusion matrix Manual label and Model label SetFit ABSA model - if aspect extracted are correct*

Table 1. Aspects identified by BERTopic, LDA and NMF model with Keywords

| **No.** | **Aspect/Topic** | **LDA Keywords** | **BERTopic Keywords** | **NMF Keywords** |
| --- | --- | --- | --- | --- |
| 1 | Seating | tabl, enjoi, good, area, lot, like, outsid, littl, sit, patio, space, dog, nice, friendli, great, love, place, outdoor, seat, insid | dining, sit, restaurant, seating, patio, outdoor, indoor, beach, picnic, sitting | tabl, ask, come, server, minut, waiter, check, water, took, waitress, seat, said, came, brought, arriv, reserv, told, hostess, final, sit |
| 2 | Ambiance | great, servic, friendli, place, staff, recommend, delici, love, amaz, good, definit, nice, super, highli, excel, come, time, awesom, atmospher, best | pleasant, superb, delicious, ambianceservice, restaurant, dining, meal, food, terrific, ambiance | great, staff, friendli, atmospher, recommend, beer, awesom, spot, select, definit, view, fun, price, fantast, experi, lunch, enjoi, locat, highli, excel |
| 3 | Experience | place, love, best, good, amaz, try, time, come, eat, like, definit, delici, favorit, town, thing, great, want, tri, got, know | excellent, great, pleasure, experience, incredible, here, wonderful, amazing, impressions, enjoyed | came, nice, definit, pretti, delici, sauc, dish, recommend, flavor, seat, meal, try, friend, bit, wasn, dinner, overal, cook, tasti, plate |
| 4 | Waiting Time | order, wait, line, time, park, long, open, close, place, minut, door, busi, hour, lot, walk, peopl, drive, pick, worth, readi | dinner, prepared, waiting, busy, orders, waited, sandwich, food, restaurant, eat | wait, minut, long, seat, line, hour, tabl, staff, peopl, worth, min, busi, arriv, final, outsid, sit, walk, told, hostess, reserv |
| 5 | Service | tabl, time, ask, server, servic, wait, waitress, arriv, order, waiter, minut, got, come, came, restaur, check, said, took, drink, seat | service, meal, dinner, customer, delicious, quality, dessert, food, restaurant, lunch | want, said, ask, custom, manag, know, told, look, went, work, peopl, call, busi, owner, rude, order, go, dai, review, wai |
| 6 | Re-visit |  |  | time, visit, long, second, come, year, went, go, dai, sure, busi, lunch, week, try, tri, coupl, twice, work, usual, took |
| 7 | Quality |  | dessert, lunch, delicious, quality, restaurant, dinner, meal, food, service, customer |  |
| 8 | Wine Selection | tour, wine, tast, great, dai, recommend, wineri, experi, time, fun, knowledg, trip, stop, group, wai, pick, guid, friend, highli, took | wines, brewery, tasting, delicious, winery, taste, wineries, tastings, wine, beers | wine, tast, tour, wineri, bottl, glass, experi, enjoi, knowledg, red, dai, recommend, stop, pour, white, fun, area, visit, nice, flight |
| 9 | Food Flavor | flavor, onion, top, crust, slice, butter, pickl, turkei, like, caramel, pie, bun, tomato, avocado, ingredi, ring, salt, peanut, try, dough | sauce, flavor, order, good, cook, delicious, perfect, perfectly, sweet, crispy | like, tast, look, feel, flavor, sauc, think, thing, try, know, felt, sweet, pretti, better, wasn, bad, littl, peopl, wai, bit |
| 10 | Gluten-Free Options | cake, free, sweet, gluten, pastri, cupcak, flavor, cooki, latt, delici, vanilla, bakeri, almond, tast, dessert, drink, like, milk, love, good | restaurants, veggie, vegetarians, vegan, veganvegetarian, vegetarian, glutenfree, vegetarianvegan, gluten, vegans |  |
| 11 | Pricing | good, price, servic, place, pretti, great, reason, portion, star, qualiti, littl, bit, nice, small, worth, decent, like, restaur, size, high | pricey, price, expensive, cheap, prices, cost, pricing, overpriced, priced, decent | good, pretti, price, nice, sauc, overal, bit, tasti, flavor, beer, thing, tri, portion, littl, try, tast, select, wasn, come, bad |
| 12 | Vegetarian Options | dish, rice, soup, spici, order, vegetarian, noodl, restaur, beef, menu, like, flavor, good, chines, option, broth, tuna, veggi, sauc, hot | restaurants, veggie, vegetarians, vegan, veganvegetarian, vegetarian, glutenfree, vegetarianvegan, gluten, vegans |  |
| 13 | Nightlife | drink, bar, night, place, bartend, music, great, fun, good, like, friend, cocktail, peopl, plai, live, cool, late, crowd, vibe, atmospher |  | drink, bar, beer, bartend, peopl, night, happi, tabl, hour, seat, cocktail, sit, friend, enjoi, nice, want, fun, area, patio, music |
| 14 | Events and Setting |  | bday, dinner, celebrate, anniversary, birthdays, birthday, celebration, wedding, guests, celebrated | love, delici, favorit, friendli, come, perfect, staff, absolut, super, enjoi, husband, flavor, sweet, wonder, atmospher, famili, yummi, beauti, feel, patio |
| 15 | Good for Group | year, kid, time, dai, famili, wife, went, ago, visit, old, week, daughter, month, son, husband, lunch, coupl, mom, look, forward | year, kid, time, day, family, wife, went, ago, visit, old |  |
| 16 | Healthy Options | bowl, healthi, juic, scoop, fruit, yogurt, smoothi, strawberri, flavor, poke, option, banana, churro, top, love, size, kale, green, delici, quinoa |  |  |
| 17 | Location and View | street, locat, view, beach, state, walk, right, ocean, park, beauti, great, coast, awai, restaur, hidden, enjoi, montecito, block, downtown, boathous | jewel, location, inside, place, home, hidden, street, gem, gems, located | eat, littl, locat, lot, park, nice, breakfast, street, friendli, spot, area, lunch, seat, menu, right, look, small, try, outsid, shop |
| 18 | Happy Hour | beer, happi, hour, great, select, good, tap, bar, drink, menu, breweri, brew, local, tapa, spot, game, place, craft, enjoi, favorit |  |  |
| 19 | Cleanliness and Covid Precautions | clean, covid, croissant, bathroom, dirti, water, smell, cup, hand, restroom, mask, us, wear, safe, wash, paper, social, floor, station, molli | seating, covid19, place, coronavirus, precautions, covid, virus, covidera, pandemic, quarantine |  |
| 20 | Dessert |  | creamy, dessert, cream, smoothies, cafe, smoothie, place, ice, donuts, starbucks |  |
| 21 | Drinks |  |  |  |
| 22 | Kid Friendly | year, kid, time, dai, famili, wife, went, ago, visit, old, week, daughter, month, son, husband, lunch, coupl, mom, look, forward | year, kid, time, day, family, wife, went, ago, visit, old |  |