Solving the Hyderabadi Word Soup

Dinis Fernandes nº20221848
Dinis Gaspar nº20221869
Inês Santos nº20221916
Luis Dávila nº20221949
Sara Ferrer nº20221947

1 Introduction

Hyderabad Tourism Board contracted our team to conduct an analysis of the city's restaurant landscape. Our objective is to try and understand the meaning hiding in restaurant reviews, so that we are able to provide more meaningful recommendations to the organizations. By analyzing these reviews, we will develop models that are able to classify restaurants by cuisine type. The previous results will then help us to see the distribution of cuisine types across the city and to point out potential issues.

As we performed a Text Mining process, our analysis was led by the CRISP-DM (Cross-industry standard process for data mining) methodology. This process ensured coherency between the analysis as well as a good step-by-step approach, guaranteeing an efficient workflow. To analyse the landscape of the restaurants of Hyderabad we used two datasets, one with detailed information about the restaurants, and another one with the zomato reviews regarding the restaurants in the first dataset. To carry out a complete understanding of the restaurants of the city we divided our work in four different notebooks, with their respective data understanding, analysis, cleaning and preprocessing, models and results. Those four notebooks comprise the following tasks:

Sentiment Analysis : Where we extracted and deeply analyzed the sentiment of the zomato reviews, and pinpointed worrisome areas as well as satisfactory sectors.

Multilabel Classification : Where we classified the restaurants by cuisine type, only based on their reviews, looking for patterns which relate customer opinions to restaurant cuisine type.

Co-occurrence Analysis and Clustering: Where we identified clusters of cuisine types seeing the dishes that are usually mentioned together.

Topic Modelling : Where we point out the main and specific topics of the provided reviews and identify customer sentiments.

With this analysis we aim to provide meaningful insights about the restaurants in the Hyderabad area, as well as identify key factors driving some restaurants' success or failure. We also aim to provide a guide to the tourism board on areas of improvement in the city.

2 Literature Review

After some research on papers and methods that have been previously used to solve problems related to the project requirements, we found 5 papers that helped us. In our project paper [1] motivated this part of the project as this analysis is also important for food inspection and overall safety, as most complaints are never formal, but instead made on social media. More specifically for sentiment analysis [2] highlighted the importance of implementing good preprocessing and NLP techniques, like word embeddings for sentiment prediction. [3] Showed us how using pre-trained models we can identify (using NER) the dishes in a corpus. [4]Showed us a baseline to compare to when performing cuisine classification based on text data (the paper classifies recipes, but still has some interesting insights) [5] Showed us the effectiveness of topic modelling techniques like LSA, LDA and BERTopic to extract valuable insights about customers.

3 Sentiment Analysis

3.1 Data Understanding

We started by checking the data and its format. For this analysis, we kept only the columns Restaurant, raw_review, and Rating as these were the only important ones to conduct sentiment analysis. Then we checked that both raw_review and Ratings had missing values. As we are using raw_review to predict Rating it did not make sense to input missing reviews or ratings, the percentage of missing values was very low for each variable, so we decided to drop them. Then we started analyzing the Rating data type issue by checking its unique values and quickly understood that there was a review that had a rating that was a string, 'Like'. As this only happened in one row, we dropped it. We then plotted the Ratings distribution, to check if there was a significant difference between ratings and concluded that there were more reviews with rating 5, which is already a good indicator of the restaurant landscape.

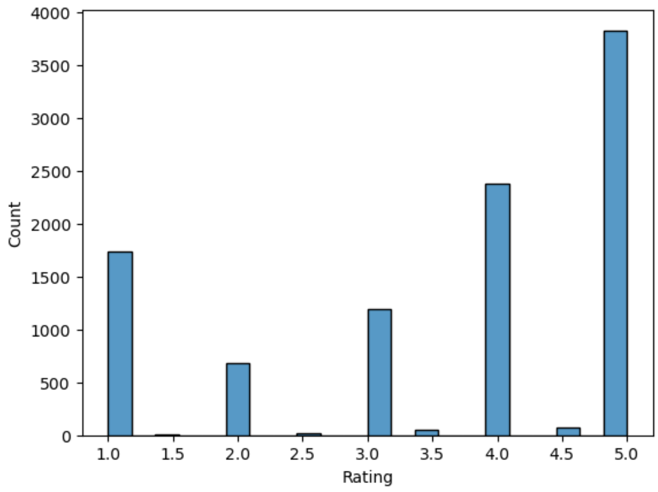

Figure 1: Rating distribution

We then created word frequency based visualizations to obtain a better understanding of the vocabulary of our corpus. We also generated word clouds per rating, with and without stop words so that we could better analyze which words were connected to which rating, these word clouds can be found in the folder ”word_clouds/ratings”. Only the word clouds without stop words gave more interesting findings:

Rating 1: Negative adjectives, such as: ‘worst’, ‘bad’ and ‘pathetic’ are very frequent showing customers dissatisfaction. Some concerns on customer service quality (‘service’, ‘served’, ‘staff’, ‘manager’) and on waiting times (‘time’, ‘table’) are suggested. The words ‘food’, ‘experience’, ‘quality’ and ‘money’ may express negative thoughts regarding the price - quality ratio. Some dishes appear frequently, ‘chicken’, ‘biryani’ and ‘veg’, possibly due to poor preparation.

Rating 2: The adjectives, ‘average’, ‘better’, ‘disappointed’, ’overall’ and ‘ok’, still show dissatisfaction but not as negative as before. Concerns on customer service and waiting times persist (‘service’, ‘served’, ‘staff’, ‘time’, ‘table’). The concerns on price - quality ratio remain, along with concerns on restaurant atmosphere (‘ambience’), food quantity (‘quantity’) and with the menu (‘menu’).

Rating 3: More positive adjectives began to appear (‘better’, ‘great’, ‘nice’) and show more satisfied clients. The words ‘service’ and ‘staff’ still appear, but now they may either be referenced in a positive or in a negative way. The dishes mentioned remain the same, likely due to their popularity. Numbers like: ‘3’, ‘3.5’ and ’5’ also appear indicating that clients are probably rating restaurant’s elements in the reviews, with ‘5’ being the most frequent one.

Rating 4: The high frequency of positive adjectives (‘best’, ‘amazing’, ‘like’) show customer satisfaction. The words ‘staff’, ‘service’ and ‘time’ still appear but in this case are probably associated with positive opinions. The same popular dishes are mentioned, supporting previous observations. The words ‘taste’ and ‘try’ may indicate very well-prepared dishes and the word ‘friend’ may mean that clients are suggesting the restaurant to others.

Rating 5: With this rating, adjectives like: ‘awesome’, ‘love’ and ‘delicious’ are very frequent, showing very satisfied customers. Customers are very satisfied with their food as the word is usually linked to words like ‘delicious’ and ‘best’. The same popular dishes are mentioned but now in a positive way. Mentions of ‘family’ and ‘friend’ together with ‘recommend’ indicate that clients are suggesting the restaurant to others. The word ‘friendly’ and the expression ‘good service’ are usually associated with ‘staff’ which shows positive opinion on the staff.

Still without stop words, we made a bar plot to check the most frequent words and concluded that one of the most frequent expressions are ’main course’ and ’must try’ which shows that the main course is often mentioned in reviews, either in a positive or negative manner, and ’must try’ suggests satisfaction from the customers as they may be suggesting to others to try that restaurant.

3.2 Data Preparation

The data was prepared according to what model was being used. We used two models, Vader and TextBlob, which we specifically pre-processed data, and another two pretrained models that as they are more complex we did not apply any preprocessing. For Vader, we retained some elements such as emojis, hashtags, punctuations and stop words, as these elements contribute to sentiment analysis. We only converted diacritics, which means that we removed accents and special marks. For TextBlob, the preparation was slightly different. We removed emojis, the hashtag symbol itself though the words where kept, and as before, we removed accents and special marks.

3.3 Modelling

To really have a good analysis we decided to test four different models.

We decided to start with Vader, which we decided to get the compound score for each review. The median score of all of our reviews is 0.765 which is a very positive insight regarding our restaurants, as this means that most of our reviews are positive ones. However it is also notable that our minimum value is -0.994 which means that there are reviews overwhelmingly negative about the restaurants.

We then determined to repeat the process for TextBlob, but instead of using the compound score, we used the equivalent in TextBlob, that is the polarity. TextBlob overall assumes the polarity of the reviews more negatively, as the median in this case is 0.275, and the most negative score given is actually -1.0.

After it we tested the pre-trained models, using transfer learning. Our models were RoBERTa and DistilBERT [7].

Starting with RoBERTa we used the model already fine tuned for restaurant reviews data [8] [9].This model had an output of two scores, with negative and positive labels, however as they were complementary (summed up to one) we could only use one as our score. With this model we understand that the median score is also very positive, 0.99 however, we would like to note that on contrary to the previous two models, the scale is not from -1 to 1 but from 0 to 1, so the direct comparison is harder.

We then tested DistilBERT that was fine tuned for tweets analysis [10] , [11] as the language in tweets is very wide ranging, informal and mainly people still give their opinions, it still makes sense to use it in our analysis . This model outputted three scores: negative, neutral and positive and therefore are not complementary like in RoBERTa. The most dominant score is the positive with a median of 0.854, followed by the neutral one with a median of 0.0578 and the negative being the last one with a median of 0.0386. As in RoBERTa, the scale is from 0 to 1, therefore more difficult to compare with the other models.

3.4 Outlier Analysis

During our analysis, we encountered observations that we considered outliers, as their rating was high/low but our models had low/high scores. For both situations, stopwords were removed keeping only 'no' and 'not' and we also decided to keep the punctuation marks as these help in predicting the sentiment in a sentence. We then made a wordcloud with bigrams and a bar plot with the most frequent words, for the data we preprocessed and for invalid words in the reviews, for both situations. We concluded that high ratings with low sentiment scores may occur due to clients being overall satisfied but some negative aspects are mentioned and the models are probably prioritizing this over the positive sentiment. For low ratings with high sentiment scores, we concluded that customers are also satisfied but very dissatisfied with specific aspects and in this case the models are prioritizing the positive sentiment. For both situations, it may be due to customers mistakenly selecting the ratings or that the customer gives high/low ratings out-of-habit.

3.5 Evaluation

To evaluate our models we used the Pearson and Spearman correlation of the scores outputted from the model and the rating, obviously we aim for the highest correlation possible.

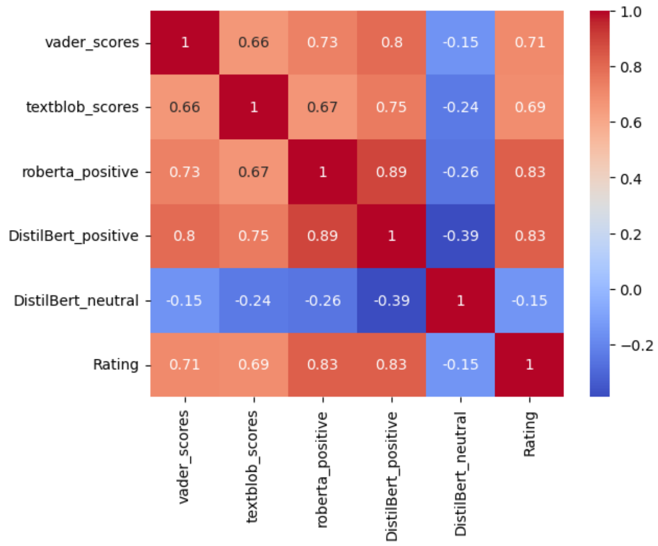

Figure 2: HeatMap of Models with Rating

As we can observe the highest correlated models with the target are the pre-trained ones, with the same correlation, it's also notable that the distilbert neutral score, is in fact a low negative one. To further analyse our models we decided to plot the average scores against the rating.

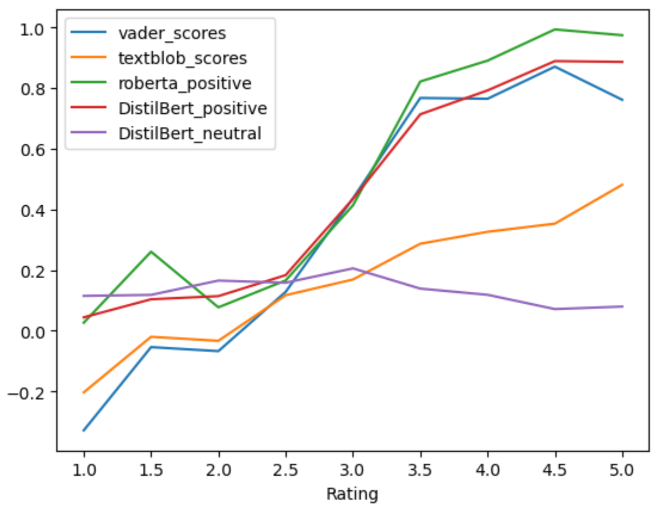

Figure 3: Progression of Models with Rating

It was expected that all models always increased their scores with the rating, however that does not happen. In fact, only Distilbert and textblob always increased their scores. As the correlation of Distilbert and the target is higher we decided to choose this model, as the appropriate one for sentiment analysis, in this context. After choosing this model we decided to just calculate the median score again that was of 0.854, as other analyses, for example word clouds, were already computed before. We can see that according to distilbert the overall predicted sentiment of the reviews is positive, which is in line with the actual ratings, as seen in the beginning. We should also inform the Hyderabad Tourism Board that most of the restaurants are secure and good, but there are some serious concerns regarding the hygiene of the restaurants and food poisoning.

4 Cuisine Classification (Multilabel Classification)

4.1 Data Understanding

As with any other classification problem, before proceeding with modelling we must first get an understanding of our data. Firstly, we checked some information on the structure of our reviews and concluded that,on average, reviews are approximately 4 sentences and 281 characters long. It's also possible to conclude that there are some reviews which are 1 character long and some which are up to 5247 characters long, but are still actual reviews. We can also see that we have reviews with up to 66 sentences. We'll now plot the frequency of the different cuisine types.

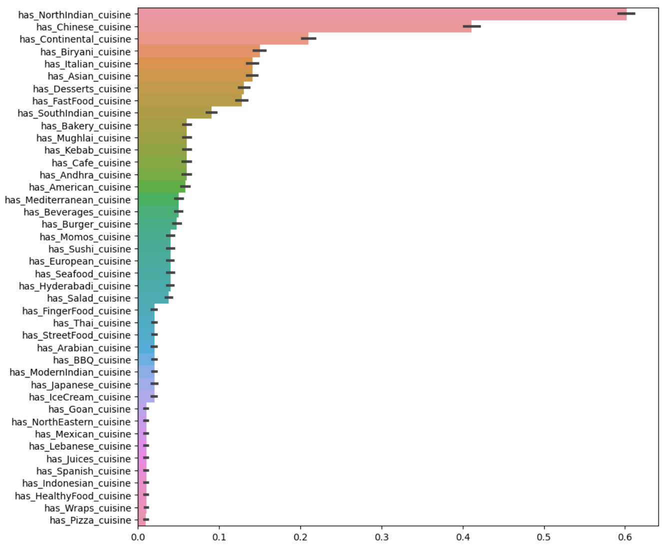

Figure 4: Frequency of cuisine types

There are some take-aways from the plot above:

- As expected, due to the region our data originates from, we can see that asian cuisines are much more common than western ones. With the most commonly reviewed cuisine type being North Indian, the second most common being Chinese, Continental in third and Biryani in fourth.

- Interestingly, Italian cuisine is quite high on the list, this may be due to it being more internationally recognizable than other western cuisines.

- We can also see that latin cuisines (excluding Italian) are very seldom reviewed. The least reviewed cuisine is Pizza, which makes some sense since it can actually be seen as a subdivision of Italian cuisine and, thus, it may simply be more specific than other cuisines present on the list.

- The healthy food cuisine also contains very few reviews, this could be due to the people in this particular region being less conscious/worried about healthy eating habits, compared to western cultures.

- The distribution is, as expected, unbalanced, as such we'll use techniques like stratification, class weights and evaluating with F1-score to attempt to mitigate this unbalance.

We then produced wordclouds for each cuisine type. As we have 42 unique cuisines, we will obviously not display the wordclouds in this report, they can be found in the "word_clouds/cuisines" folder of the delivery. We will however draw some conclusions from analysing them:

- Words like "good", "food", "place" and "service" are quite common across all cuisine types

- The word "chicken" also appears frequently across a few cuisines, which makes sense as most of the cuisines present in our data have some type of chicken dish

- Bakery, Beverages, Cafe and Dessert cuisine types have a lot of common terms across them, such as "cake", "donut" and "coffee", this makes sense as these 4 cuisine types are obviously closely related to each other.

- Pizza cuisine has a very distinct vocabulary with terms like "pizza", "dominos", "order" and "delivery", this makes sense as pizza restaurants traditionally do a lot of their business through deliveries.

- The Ice Cream cuisine also contains a unique set of terms, such as "ice cream", "chocolate", "nut" and "cream", these are mostly unique to this cuisine.

- The remaining cuisines have some differentiating factors but are not as distinguishable as the ones described above.

With all of this in mind, we believe that our models should be able to find some patterns, but will struggle to have high scores.

4.2 Data Preparation

We started by checking for missing values, we found 45 (0.45% of our data) reviews without an actual review, these will be dropped. Then, the 1 character reviews are dropped, as they contain only random characters or emojis, this results in the removal of 19 (0.19%) rows, leading to a total of 0.64% dropped rows.

4.3 Modelling

We will now begin creating models. We will test 4 feature extraction methods: Doc2Vec, TF-IDF, BagOfWords and BERT tokenizer [12]. For TF-IDF and BagOfWords, we will test full preprocessing (i.e. stopword, punctuation, html tag, etc. removal) with and without lemmatization, we performed some testing and including stopwords or punctuation did not improve results, this testing is not included in the report or notebooks. We will test ClassifierChain and OneVsRest classifiers with RandomForestClassifiers and LogisticRegressions. For the ClassifierChains we will always use order descending, which means the first model predicts the most common class and the last model predicts the least common class, this is based on the assumption that minority classes will benefit more from taking as input in their corresponding models the previous prediction and the relationships between them, and we will pass the prediction probabilities rather than the actual predictions between models, these methods provide better results from some testing. We also developed a function to perform crossvalidation as a way to make our results more robust. As we tested a wide variety of combinations, we will only analyze the test set predictions on our preferred model.

4.4 Evaluation

Our final model uses BagOfWords vectorization and a LogisticRegression ClassifierChain without lemmatization. We'll now analyze its performance. Looking at the test set classification report, we can see that our model achieves an average F1-score of 0.43 and a weighted average F1-score of 0.56. Let's now dive a little deeper and highlight some of the biggest takeaways:

- The model, as expected, captures the more common classes quite well, with the majority class (NorthIndian) having an F1-score of 0.8, which is comfortably the highest and the second and third (Chinese and Continental, respectively) most common cuisines having F1-scores above 0.6.

- It is also worth noting that some of the least common classes (such as Wraps and NorthEastern, for example) also achieve decent F1-scores above 0.55.

- As expected from the wordcloud analysis, despite not being very common the Ice Cream class achieves a satisfactory F1-score of 0.57.

- On the other hand, the Pizza class, which despite being rare had a very distinct vocabulary, was not picked up well by our model, achieving an extremely poor F1-score of 0.16.

- The Bakery and Café cuisines described as having similar vocabularies which were quite unique compared to the rest of the cuisines, also achieved decent F1-scores, around 0.55.

Overall, as expected, our model was able to pick up some patterns describing cuisines in the reviews, but the results are not very good, this could be since reviews in some cases may be too short or don't contain enough information to allow for the inference of cuisine type.

5 Dishes Recognition Co-occurrence analysis and clustering

5.1 Data Understanding

We start by joining the reviews with their respective restaurant to have their cuisines types, then was checked if the reviews had some content, dropping the ones with missing values. Afterwards a sentence length analysis was performed and concluded that most reviews have less than 500 characters, and they tend to have less than 5 sentences in each review.

5.2 Data Preparation

To analyses the dishes in a meaningful way it was needed to extract the dishes that were mentioned in the reviews, to do so we used a pre-trained RoBERTa model fine-tuned to find dishes in text using name entity recognition ('roberta-base-food-ner' [6]), we process that output cleaning the text and removing numbers that the model captures, obtaining only a list of dishes. After the dishes were obtained, we saw that 4955 of the 9955 not nulls reviews didn't mention any dishes, and there were 9001 unique dishes mentioned, then we saw that some words were misspelled so we join every words that have a Levenshtein distance of 1 and obtain that in reality we only had 8208 unique dishes (here dishes consider everything that could be drink or eat, like chicken, pizza, coffee, etc).

5.3 Modelling

5.3.1 Co-Occurrence Analysis

With the dishes mentioned in each review, we build a co-occurrence matrix to identify relationships between dishes appearing in the same context. Creating a network graph to visualize these relationships, showing only the most dishes mentioned together.

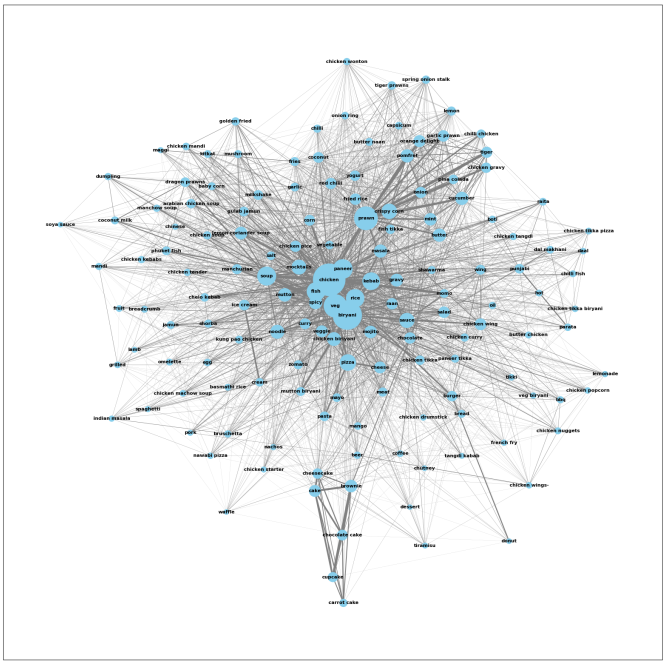

Figure 5: 150 most frequent dishes mentioned together

5.3.2 Clustering

We then perform some clustering algorithms (K-means, Hierarchical, HDBScan and Optics) on top of a bag of word representation of the dishes mentioned to see if they could capture some relations and as end result capture some clusters of cuisines.

5.4 Evaluation

The clusters models didn't achieve any meaningful results, not being able to distinguish dishes nor cuisines, giving almost the same dishes for all cuisines or creating many clusters with only a few observations. Because of that we will use the co-occurrence graph to identify clusters of cuisines. So, after seeing the wordclouds of the most frequent words that appear in each of the cuisines and seeing the co-occurrence graph of the dishes we can conclude the following:

- The group of dishes formed by cheesecake, cake, carrot, cake, cupcake, chocolate cake and brownie can be associated to a group of similar cuisines: Bakery, Beverages, Cafe and Dessert.

- The pair ice cream and cream clearly belongs to the Ice Cream cuisine.

- The group of donuts, burger, chicken wings, wings, bbq, French fries clearly relate to American, BBQ and Burger cuisines (which are all American based cuisines).

- There is a cluster that follows only the combination of burger, coffee and donuts that relates to the Beverage and the Cafe cuisines.

- The group of FastFood dishes, which involves the following dishes: chicken, donuts, burger, French fries, pizza, cake, cheese and wing is related to the Fast Food cuisine as well as the Pizza and Burger cuisines, which are also fast food.

- The Italian cuisine from a distinct cluster containing pizza, pasta and cheese.

- The combination of fish, prawn and pomfret clearly indicates reviews about the Seafood and Goan cuisines.

- The biggest group is the one formed by the combination of chicken, biryani, rice, spicy, curry, fish, paneer, mutton, veg, kebab, momos and soup, this group relates to: Andhra, Arabian, Asian, Biryani, Chinese, Indonesian, Modern India, Mughlai, Kebab, Momos, Northeastern, North Indian, SouthIndian, Hyderabadi and Thai cuisines. We could only consider the relations that correspond to each of the cuisines, but we are keeping all of those as a group since they share similar foods between them and thus their dishes end up clustered together.

- The rest of the cuisines are not easily identifiable using the co-occurrence of the dishes in the reviews, but we can clearly state that the cuisines respect their themes and the reviews mention them (with a positive or negative connotation).

6 Topic Modelling

6.1 Data Understanding

To perform Topic Modelling we only kept the reviews column. Before proceeding with the Topic Modelling itself we started by checking the missing values, as mentioned early there were found 45 missing values, which ended up being dropped. Besides the missing values, there were found two reviews with gibberish that were affecting topic modelling, which ended up being dropped.

6.2 Data Preparation

This process involved multiple stages of text cleaning, tokenization and vectorization to transform raw text into structured format suitable for topic modelling.

During the whole analysis of the reviews we noticed that there were many reviews that contained gibberish and were affecting topic modelling, so to deal with them we created a function that tried to remove some of that gibberish by transforming for example "aaaamazing" to "amazing".

Following this, the main preprocessing pipeline was applied. The pipeline processed the text by performing lemmatization and removing some stopwords that are often too common and do not provide meaningful information. In parallel, another preprocessing pipeline was applied to generate tokenized reviews. A third pipeline was used to process text for Doc2Vec, in this case is neither applied lemmatization, removal of stop words or lower case transformation because Doc2Vec requires the original form of the text.

After cleaning and preprocessing the text, we proceed to vectorization.The first method was BOW (bag-of-words), where each word in the vocabulary was represented as a feature and the count of each word in each review is recorded. The second method of vectorization was TF-IDF, where we applied TfidfVectorizer in order to assign higher weights to words that are less frequent across the entire corpus but are more frequent in a particular document. Finally Doc2Vec was used to generate document level embeddings, where each review was represented as a fixed-length vector. The reviews were converted into TaggedDocument objects and then a Doc2Vec model was trained using the processed reviews. To finalize we converted the resulting vectors into NumPy arrays for compatibility with the topic modelling algorithms and for more efficiency.

6.3 Modelling

This part is divided into 3 main Topic Modelling algorithms (LSA,LDA and BERTopic), in all of them to arrive at the optimal number of topics we calculated the topic coherence for each number of topics within a range. We decided to use Coherence as the measure because it is the one that provides the most reliable, interpretable and meaningful topics.

The first model applied was the LSA using sklearn implementation of TruncatedSVD, LSA is a technique for dimensionality reduction and topic extraction, and TruncatedSVD is the method behind it. The optimal number of topics was 10 which corresponds to prior analysis where using larger topics we noticed through singular_values_ attribute that most of the variance is explained by the first components and there was a large drop off after the first few topics.

The second model applied was Gensims's LSI, an implementation of LSA. Which gave very similar results since both methods rely on the same fundamental technique. Through the attribute show_topics we were able to analize some customer sentiment and with projection.s we concluded that by far the most prominent topics were related to service and food quality.

The third model applied was the sklearn implementation of LDA which performs topic modelling by learning the topic distribution for a set of documents. During the implementation of this model we noticed that there were some difficulties, like the lack of interaction in terms of direct access to topics and model parameters and it didn't provide native tools for evaluating topic coherence.

The fourth model applied was LDA with Gensim, a model specifically optimized for large-scale text and batch learning, making it more efficient in terms of memory usage and speed. When calculating the number of optimal topics it gave 20 regarding perplexity and 5 regarding coherence, for us to not have either a broad or a specific topic modelling we met in the middle and assigned ten topics. With ten topics we arrived at a coherence of 0.4 which in real practice for restaurant reviews is a good score close to the ideal. Through the attribute show_topics we were able to identify positive and negative customer sentiments related to food and service, with the attribute get_coherence_per_topic we concluded that the most clear and specific topic was related to the vegetarian and non vegetarian option of biryani (famous dish of the Hyderabadi Cuisine).

The fifth and final model is BERTopic, a more advanced model that can capture finer details like customer sentiment and food trends. It is better because it uses transformer-based embeddings to understand the meaning behind words more accurately. Additionally, BERTopic automatically adjusts the number of topics and generates more coherent topics, making it a better fit for analyzing restaurant reviews compared to LSA and LDA. Based on coherence, we arrived at 85 topics with a coherence of 0.46 which can almost be considered a high coherence since we have reviews that are very difficult to interpret. With this model we arrived at many emergent topics like specific cuisines (ex. American, Chinese, Italian), food trends (ex. multiple types of pastry), ordering problems (ex. refunds), specific complaints (ex. salty food), recommendations for dinner with friends and much more.

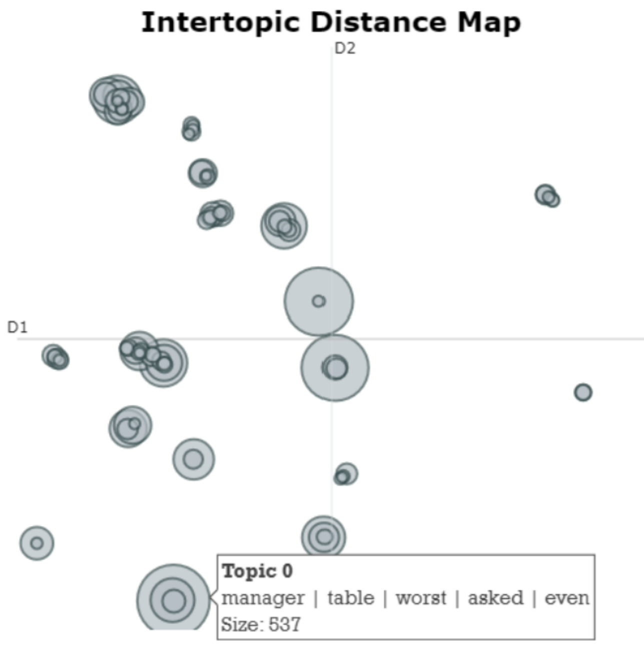

Figure 6: InterTopic Distance Map

Through the Intertopic Distance Map we can conclude that there are many topics similar to each other but with some interesting nuances, and the fact that they are mainly focused on the left side could point that there are more general topics in the dataset.

6.4 Evaluation

With this Topic Modelling project, we concluded that algorithms like LSA and LDA excel at identifying broad topics such as service quality, food taste, ambiance and even customer sentiment. However, they often missed to capture finer details or aspects of customers feedback. With BERTopic we were able to identify those finer details that traditional models overlook. This included topics such as trendy dishes, ordering problems, quantity problems and more personalized customer preferences. The ability of BERTopic to extract more insightful topics makes it particularly useful for Topic Modeling.

7 Conclusion

In this project, we understood the value of performing text mining tasks to gain insights into the restaurant industry, reaching interesting conclusions for the different tasks. We faced some challenges with the very informal text, which contained a lot of typos in the review. Ideally, we would have liked to be able to perform some deeper searches in the modelling stages, but due to time and computational capacity constraints this was not possible. We believe that the insights gathered bring genuine value to a governing authority, such as the Hyderabad Tourism Board.

References

[1] F. J. V. F. Monteiro, Predicting Restaurant Inspection Scores Based on Yelp Data and Sanitary Inspection Reports, M.S. thesis, Telecommunications and Computer Engineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal, supervised by B. E. G. Martins and M. J. C. G. da Silva, Nov. 2018 https://fenix.tecnico.ulisboa.pt/downloadFile/563345090416349/76365_Predicting_Retaurant_Inspection_Scores_Based_on_Yelp_Data_and_Sanitary_Inpsection_Reports.pdf..

[2] Guda, Bhanu Prakash Reddy Srivastava, Mashrin Karkhanis, Deep. (2022). Sentiment Analysis: Predicting Yelp Scores. 10.48550/arXiv.2201.07999. https://www.researchgate.net/publication/357987621_Sentiment_Analysis_Predicting_Yelp_Scores.

[3] M. Goel et al., "Deep Learning Based Named Entity Recognition Models for Recipes," in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, pp. 4542–4554. [Online]. https://aclanthology.org/2024.lrec-main.406.pdf

[4] T. Sharma, U. Upadhyay and G. Bagler, "Classification of Cuisines from Sequentially Structured Recipes," 2020 IEEE 36th International Conference on Data Engineering Workshops (ICDEW), Dallas, TX, USA, 2020, pp. 105-108, doi: 10.1109/ICDEW49219.2020.00008.

[5] cBaby, Anusuya. (2023). Exploring the Power of Topic Modeling Techniques in Analyzing Customer Reviews: A Comparative Analysis. 10.48550/arXiv.2308.11520

[6] "RoBERTa-Base Food NER," Hugging Face, 2024. [Online]. Available: https://huggingface.co/carolanderson/roberta-base-food-ner.

[7] A. Joshy and S. Sundar, "Analyzing the Performance of Sentiment Analysis using BERT, DistilBERT, and RoBERTa," 2022 IEEE International Power and Renewable Energy Conference (IPRECON), Kollam, India, 2022, pp. 1-6, doi:10.1109/IPRECON55716.2022.10059542.

[8] Hugging Face. "cardiffnlp/twitter-roberta-base-sentiment." [Online]. Available: https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment.

[9] Hugging Face. "pachequinho/sentiment_roberta_restaurant_10." [Online]. Available:https://huggingface.co/pachequinho/sentiment_roberta_restaurant_10.

[10] Hugging Face. "austinmw/distilbert-base-uncased-finetuned-tweets-sentiment." [Online]. Available:https://huggingface.co/austinmw/distilbert-base-uncased-finetuned-tweets-sentiment.

[11] Hugging Face. "distilbert/distilbert-base-uncased." [Online]. Available:https://huggingface.co/distilbert/distilbert-base-uncased.

[12] "BERT" Hugging Face, 2024. [Online]. Available:https://huggingface.co/transformers/v3.0.2/model_doc/bert.html