Customer Complaint Type Modeling

Eyasu Woldu *eyasuwoldu@temple.edu*
Computer Science
Temple University
Philadelphia, PA, 19140, USA

Natnail Belete *tug62764@temple.edu*
Data Science w/ concentration in Computer Analytics
Temple University
Philadelphia, PA, 19140, USA

Abstract

This paper addresses the problem of finding meaningful feedback from customer reviews for businesses. We used a Latent Dirichlet Allocation (LDA) model, and a KNN Text Classification Algorithm to classify different types of complaints after which we applied our model to 500 businesses to do an exploratory data analysis.

1. Introduction

Topic modeling in Natural Language Processing has only been around over the past 20 years. The advent of Latent Dirichlet Allocation (LDA), a generative probabilistic model for a large collection of text data, in the early 2000s and the rise of computational processing power has seen an increase in the applications of Topic Modeling and LDA. For example, LDA has been used to , build a New York Times recommendation system, analyze the progression of Science Journals, as well as in classifying different social groups.

One area that we applied LDA to is customer reviews; for the purpose of giving business insights into unsatisfied customers. Small businesses can easily evaluate their performance and customer satisfaction from a handful of reviews. Big businesses however, with hundreds if not thousands of reviews might not have the capacity to go over each of the reviews to assess the problems that the reviewer encountered.

When we first started, we had a large data-set from Yelp containing millions of reviews and businesses. We wanted to see what reviewers were saying (topic) about these businesses, in particular reviews with low star ratings. Our initial idea was to classify these reviews by using only a set of custom keywords that we believe might be associated with a topic. However, We realized an approach with pre-set topics and variables is not a wise approach considering a lack of analysis on the data. Research in the field of text analysis lead us to discover forms of Natural Language Processing and Topic Modeling using LDA models.

After some learning and experimentations, we realized that by building a model that can classify different types of customer complaints, we can perhaps create some type of a dashboard that can be utilized by the businesses themselves to improve on areas that customers frequently complain about.

2. Approach

In this section, we describe the dataset that we used, as well as go in detail about our LDA, and KNN models. In addition, we also discuss about the various challenges, limitations and decisions that we had to make.

2.1 Data

The data we used was from the 2019 Round 13 Yelp Dataset Challenge. The dataset was divided into several JSON formatted files. Of these, we used 'business.json' and 'review.json'. review.json was about 4.7 GB and contained about 10 columns/attributes and 6 million rows, with each row representing an individual review. Some of these attributes were 'business_id',a unique identifier for every business, 'star-rating' which is the user rating out of five as well as 'text' which contained the review in a string format. business.json had about 10 attributes containing info such as business type, closure status, number of reviews etc. It also contained about 1 million row with each row being a unique business identified by its business_id.

Because we're interested in customer complaints, it makes the most sense to focus on reviews with a low star rating. We initially decided this would be reviews with ratings of three or less, but because these ratings often had mentions of some positive experience, (which might skew our model), we decided to focus on reviews with two or less stars, of which there were about 1.5 million reviews across 150,830 businesses.

Obviously, using all 1.5 million reviews to build our models is impractical as it will require immense computational resources which we don't have, thus we had to figure out an efficient way of dividing our data. One idea was to get the first 20,000 - 100,000 reviews as our training data. However, because the reviews were sorted randomly, this might affect our model since we'll essentially get a few reviews per business. In addition, this might prove challenging when conducting EDA because businesses might not have enough reviews captured to make some conclusions about the problems facing them. To mitigate this issue, we decided to get the top 100 businesses, ranked by the total number of reviews in each business. This way, we can get a variety of businesses while ensuring that we get on average about 200 reviews per business or a median value of about 147 reviews per business. A higher number of n businesses required higher number of reviews to be feed to the LDA model, which increased training time.

It should be noted that a small subset of the reviews are written in the English language.

2.2 NLP Pre-Processing, LDA Model and Topic Building

Pre-processing the reviews in an NLP style, and designing the LDA model proved to be both the most time-consuming and challenging aspect as it was the backbone of this project. Most of the time here was spent on researching NLP pre-processing techniques such as stemming, lemmatization, n-gram models, as well as reading other research papers that utilized LDA models. We also did a lot of experimentations with different parameters for the LDA model.

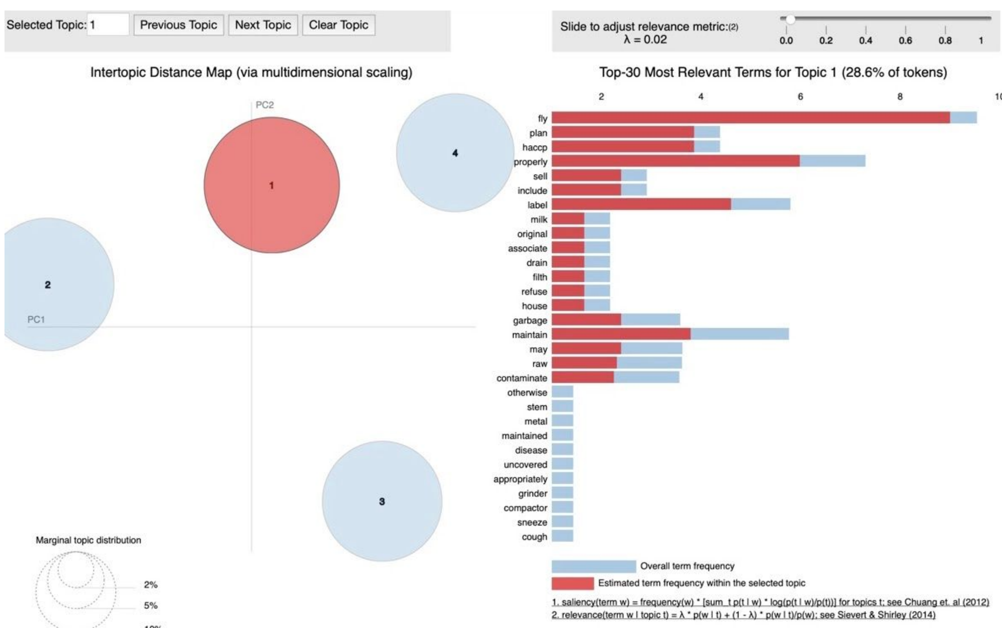
Because LDA is an unsupervised learning algorithm, it is almost impossible to determine how many topics the user should enter for the model to generate. Based on our research however, coherence score was a popular method to estimate the ideal number of topics to generate. Computing the coherence score for our data, the ideal number of topics to generate was between 20 and 30, with the coherence value hovering between 0.58 and 0.59. These values were not always consistent but the highest scores tended to be within a range

Training on 30,000 documents/reviews while doing 500 passes and a chunksize of 5000, took about an hour while using the GenSim MultiCore support on a hexa-core Intel CPU. Doing 500 passes across the 30,000 documents yielded topics that were more coherent and understandable as they'll be discussed in more detail in the Results section.

2.3 KNN Text Classification using Keywords

K Nearest Neighbor is the supervised learning portion of the project that results in a classification of review into four most commonly found complaints, Sanitation/Food Quality, Customer Service, High Price and Long Wait Time. These topics were generated after extensive analysis of text using NLP.

The NLP analysis resulted in a list of frequently occurring keywords that were open to Interpretation. Further analysis lead to a break through! We were able to block out unnecessary noise and cluster similar related keywords together resulted in what is seen below.



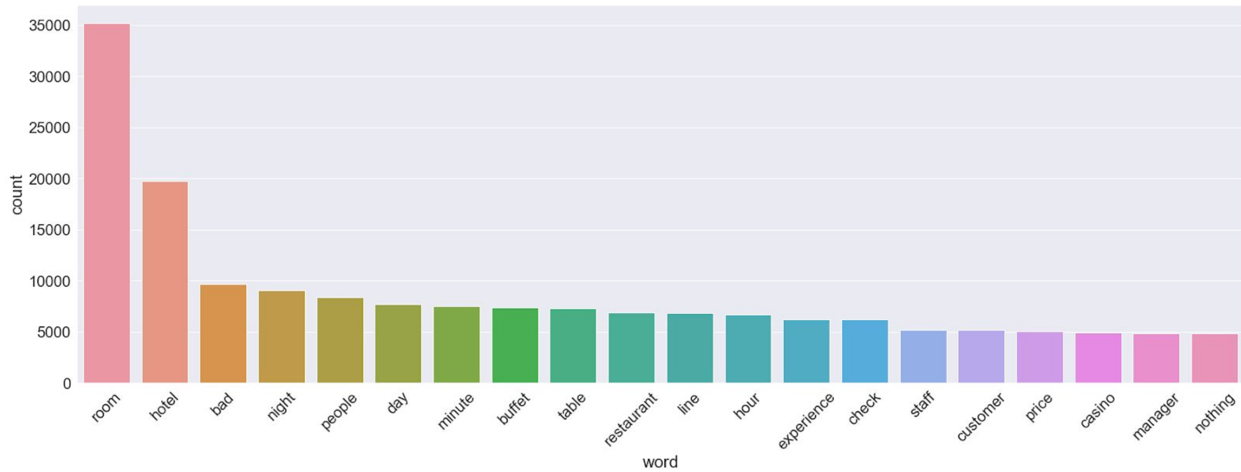
Taking these findings, we built a keyword bank where KNN algorithm can use to train its model. The accuracy of the algorithm was further increased using synset, a set of synonyms s and closely related words, on each keyword. Therefore, even if the exact keyword was not found in the text, KNN found the nearest match and is able to predict the correct outcome.

To take our project a step further, we decided to connect the KNN algorithm to a Yelp API. Yelp has a convenient API that allows a user to find a business/restaurant using name and location. Based on that information, Yelp then returns general business details, number rating and most importantly, 3 review excerpts for a business based on yelp default order. Our Algorithm takes these reviews, checks if any of the rating is below a threshold (usually 2) then makes a prediction as far as what the customer is reporting to be the problem.

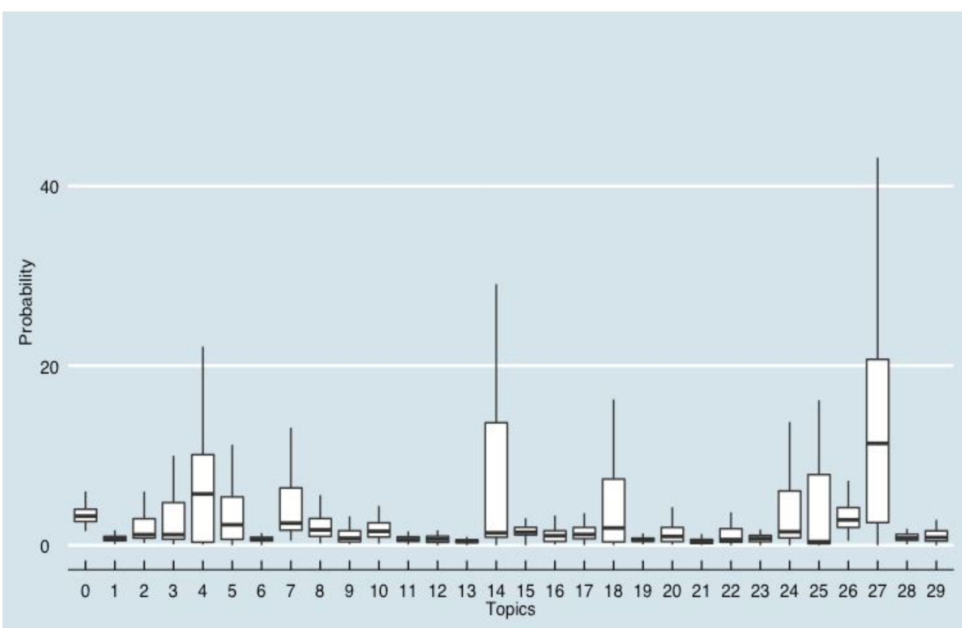
```
Input Name:  Diamond Pizza
Input Location:  1700 N Diamond St Philadelphia, PA 19121

A reviewer chosen by Yelp gave a bad Review

The delivery and customer service is terrible. It took over 3 hours for it to arrive and we only live about 20ish min walking d
istance. We called several...

                                        text  code            answer
0  The delivery and customer service is terrible....     1  customer Service
```

3. Results

The following bar graph is a frequency chart for the top 30 words in the lemmatized text following pre-processing. This is right before running our large bag of words through the LDA model. As we can see a huge section of this list is the words hotel and restaurant. Casino even makes it to the list! This however doesn't come as a surprise considering that most of the businesses in this dataset are in located Las Vegas.



Based on the best coherence score, our LDA model generated 30 topics. **The full list of the topics and the keywords associated with them can be seen on the "LDA Generated Topics" file.** Here we do some statistics about the average distribution of the topics across the 100 businesses. This was achieved by first finding the average probability of each topic for each business.




```
Topic 3: Hotel and cleanliness

Topic 4: Entree

Topic 5: Steak/meal price

Topic 7: Hotel car service

Topic 14: Hotel front-desk

Tpic 18: Fried food

Topic 24: Casino/Strip

Topic 25: Hotel Time?

Topic 26: Long line

Topic 27: Table waiting
```

Next to the boxplot chart, we listed our interpretation for some of the topics that had high probability rates in according to the boxplot chart. The interpretation of topics in LDA can be a subjective matter, but a good model can reduce the subjectivity. Note: probability here is shown as out of 100 (multiplied by 100).

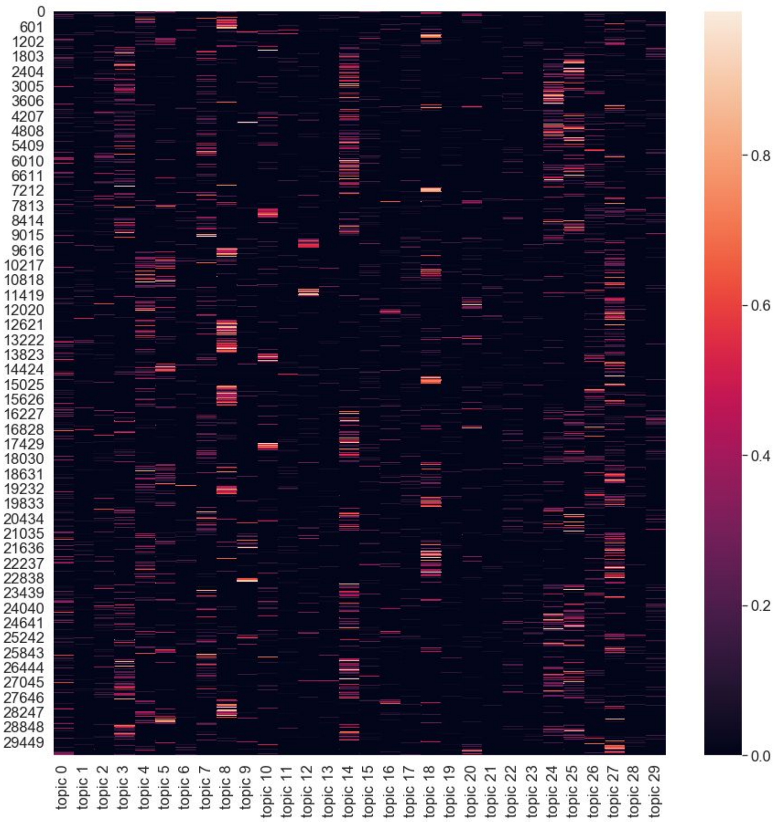
The LDA model was able to classify complex long sentences correctly. For example the following review was given a very high probability of 0.98 for being Topic 27 which in general deals with waiting. A brief gimps at the review hints that the customer is waiting for the servers.

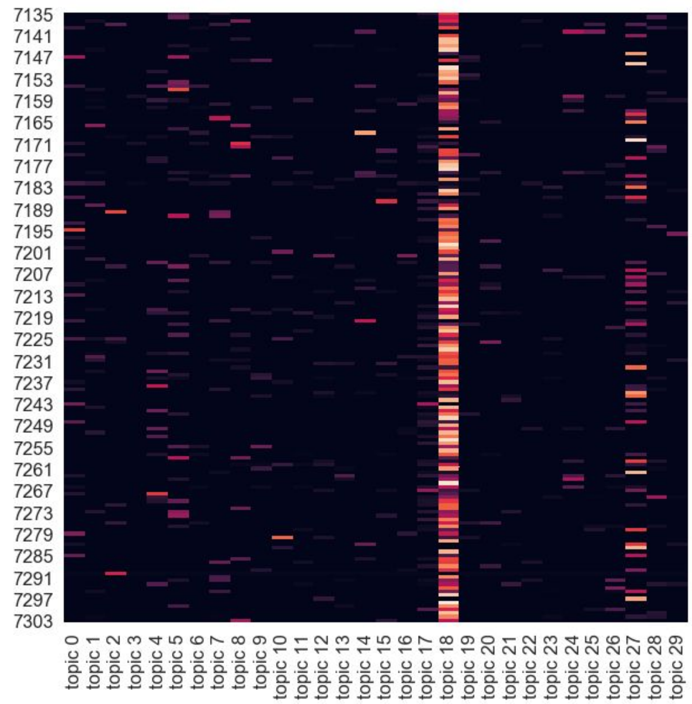
> I used to rave to people how great this place is. Don't get me wrong, the food is still delicious; however, last weekend I came by and have decided that I will not go to this location again. We got into the restaurant the same time as a large party of 7. We stood behind them as they were getting greeted by the employees. Looked like it was a server that was setting up their table. He cleaned off the table for them and put together tables to form a big one. At this point, we were still standing behind the large party. About 6 minutes later, the server sat the party. We then were able to move up closer to the "host stand" and waited for our turn to get seated. Nobody greeted us, nobody looked at us. The server that sat the large party kept walking pass us back and forth as he was getting the large party water and asking if they had questions on the menu (keep in mind that the large party was seated right next to the host stand, so we were for sure within sight). I kept trying to make eye contact with someone, but nobody seemed like they were paying attention to the host stand. After about another 7 minutes, someone finally saw us and said "someone will be right with you," which still took another minute for a server to come greet us and seat us down. I told the server in a polite way that we were standing there waiting for awhile, not only did he not apologize, he just said okay and walked us to a table. It seemed to me that the workers didn't really carry any customer service whatsoever. As we were eating, we also noticed the several parties that were coming in had to wait quite awhile to get someone's attention. Even though the food was really good, I will not be returning to this specific location due to the lack of customer service.

The LDA model was also able to correctly classify reviews with multiple types of complaints. The following review had a probability of 0.42 for being topic 8 () and 0.36 for being topic 26 and .08 for topic 4 ()

> Gosh...where to start? Just kidding! I've been waiting in line for over an hour just to pay! They pick out larger parties in the pay line to cut in front of parties of two...then you have to wait in another long ass line just to be seated to eat. 1 1/2 hr later.....OMG, We made it to the front of the "eat" line!!!....more to come! Biggest disappointment ever! Food was mediocre. The crab legs were thin and the meats were dry. The only thing that was worth eating were the fresh oysters. I had to stuff myself with shrimp cocktails just to make it worth my money. When exiting the Bacchanal, I was tempted to yell " IT'S NOT WORTH IT!" at everyone waiting in the long ass line! Service was good though...

Plotting all 30,000 reviews vs the 30 topics (left figure) with the rows sorted by business_id it becomes apparent why some of the topics such as topic 13 and topic 21 barley show up in comparison to the other topics. A tight and bright cluster of horizontal lines here represents individual businesses whose topic distribution across the majority of its reviews is very similar. For example, topic 18 represents Burgers and anything fried. Looking around row 7212, column 18, we find a very bright cluster. A closer inspection is provided on the figure to the right. The name of the restaurant in question is Burger Bar located in Las Vegas. When glossley looking over the reviews, a vast majority of the reviews complained about the burger as shown in the figure below the two charts. The average probability for this restaurant concerning topic 18 (Burgers) is 0.5 out of 1, which is twice as high as the upper quartile for the distribution of topic 18.





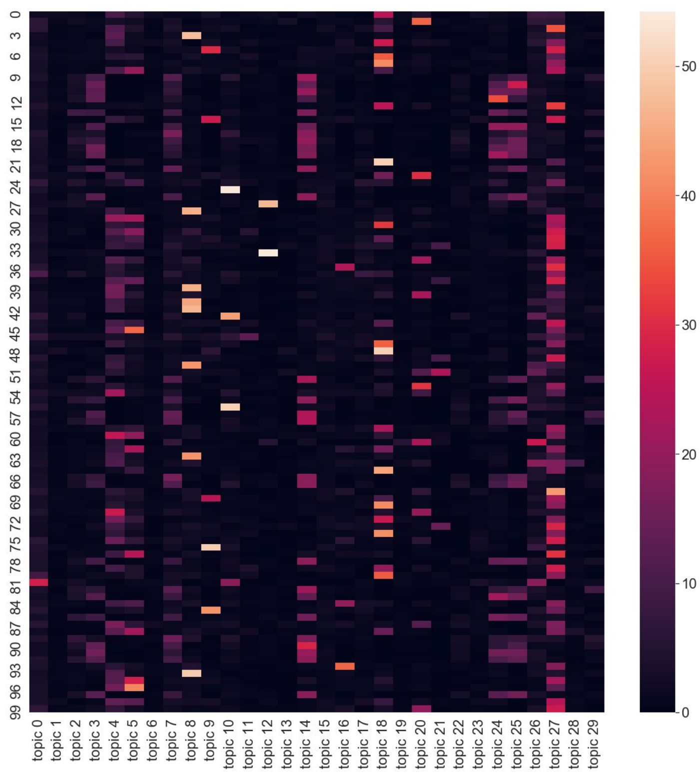
Topic distribution for Burger Bar

30,000 reviews vs their topic distribution

```
22 .  For a celebrity chef's burger restaurant it was very disappointing! Had sliders and they were poor to fair. Fi
rst order arrived cold and sent it back - refire was cold too! Tasted pre made and industrial. A disappointment and
I just had a simpler from the happy our menu. Glad I was out only 10 bucks for my sliders and a beer. Again, disappo
inting !

23 .  Completely disappointed, under seasoned and overpriced...the Rossini did not even resemble the picture. The fr
ies were stale and dry...all of the batter fried food were terrible...it has chunks of batter that is as hard as roc
ks.

24 .  I came here not too long ago based on the Yelp reviews.  The wait was long (1 hour+), the service was terrible
and the fries were cold.  The burger was just okay.
```

Review # 24 also goes along with Topic 27 (Table wait) which appears to be the second most popular topic/complaint type for this restaurant

Finally, for this part of the analysis, we calculated the average topic distribution for every restaurant and plotted them. Rows are individual restaurants. As we can see from the chart on the left, a huge portion of complaints seem to be clustered our topic 27 which relates with waiting for a long time at a restaurant. On the right side, we display the businesses with the highest average for some of these complaints.



Resturant Name	Topic 27 (table wait time) Avrg Probability Distribution
Gen Korean BBQ House	43.16906
Carson Kitchen	34.76907
Bachi Burger	32.09370
Echo & Rig	31.09339
Sushi House Goyemon	30.55685
La Santisima	29.10222

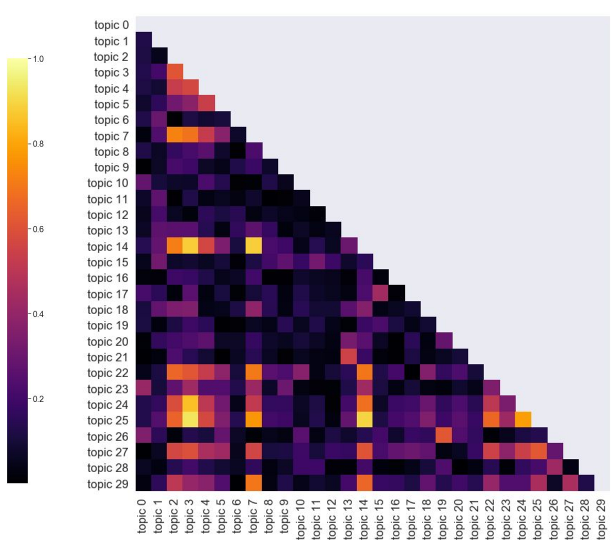
Resturant Name	Topic 18 (Fried stuff), Avrg Probability Distribution
Shake Shack	50.62230
Burger Bar	50.45225
Gordon Ramsay BurGR	44.36382
Hash House A Go Go	41.70004
Hash House A Go Go	41.46355
Rehab Burger Therapy	41.27652

Business Name	Topic 14 (Front Desk), Avrg Probability Distribution
The Palazzo Las Vegas	29.07328
The Venetian Las Vegas	23.64331
Vdara Hotel & Spa at ARIA Las Vegas	23.60088
ARIA Resort & Casino	22.70374
The Cosmopolitan of Las Vegas	21.74750
Bellagio Hotel	21.33418

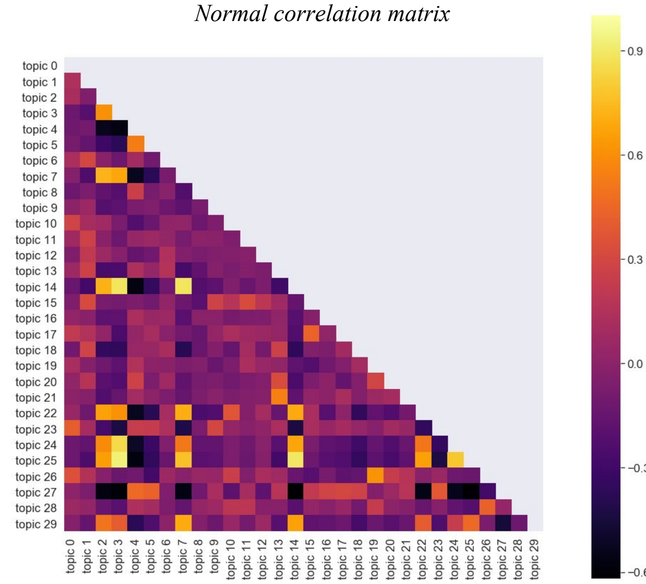
Businesses with the highest complaints by topic

Examination of the relationship between the topics shows some correlation. On the chart to the left, we took the absolute value of the correlation matrix for better visualization to see topics that correlate. There seems to be very strong (positive) correlations concerning topics 3 and topic 25 (3: Hotel and cleanliness, 25: Time). One hypothesis for this is that it has to do with the rooms not being cleaned up on time. Looking at chart on the right, we see a normal correlation matrix with both positive and negative values.

Correlation matrix with abs value



Normal correlation matrix



4. Conclusion

The results are promising, the LDA model performed much better than anticipated and was able to generalize different types of customer complaints across a wide array of business types. An LDA model with a larger set of reviews, trained on a high performance machine, and optimized parameters will without a doubt be able to perform better. In the future, with a larger dataset and the assistance of the cloud, we plan on expanding what we started here by getting into time-series analysis for business complaints as well as social modeling of the hidden nodes that interconnect the user with a business. Perhaps the lessons we've learned here might be a peek to an untapped market. We predict that within the next 10 years that we will see similar features starting to arise were businesses will leverage AI and Machine Learning to gain hidden insights into their performance.

5. Acknowledgments

We would like to thank Dr. VG Vinod Vydiswaran and his team behind 'Applied Text Mining' course from University of Michigan for providing us an understanding of the fundamentals of NLP that helped us complete our project. In addition we would like to thank Feyzullah Kalyoncu, Engin Zeydan, Ibrahim Onuralp Yigit, and Ahmet Yildirim, whose paper "A Customer Complaint Analysis Tool for Mobile Network Operators" proved to be an invaluable resource to our efforts to build our LDA model and finally, we would like to thank Summit Du provided a great resource on KNN Text classification thank you for explanation and code.