11/13/2016

Yideas!

Advanced Analytics Toolkit for Yelp

Abbinayaa Subramanian
Ananya Roy
Kimisha Mody
Vansh Khurana

TABLE OF CONTENTS

PROBLEM DEFINITION

Yideas! is an advanced analytics toolkit for Yelp that leverages machine learning and data mining techniques to provide useful insights to users and businesses.
Given the amount of information that Yelp reviews provide, we believe there is a strong need for a single integrated personalized tool that can present relevant insights to users.

These insights and main features of Yideas include:

1. Normalized rating for reviews based on underlying sentiment analysis
2. Finding cultural trends across states in US
3. Content based personalized recommendations
4. Personalized integrated user dashboard

SURVEY / LITERATURE REVIEW

Our survey was broken down into the following main focus areas:

1. Finding Cultural Trends

   Several approaches were discussed to break down text into its sub-topics. Schomberg, Hayes and Anton-Culver discussed the correlation between a keyword-rubric pair to identify cultural trends in their paper (John P. Schomberg, 2016). A few other papers (James Huang, 2014) (Julian McAuley) describe an approach to detect hidden sub-topics using a Machine Learning Technique called Latent Dirichlet Allocation (LDA).

2. Sentiment Analysis of reviews

   Different approaches (Agrawal) (Bo Pang, 2015; Linshi, 2016) (Mingming Fan, 2016) (Luca) (Hanhoon Kang, 2012) to solve the rating inference problem have been discussed to determine the user's rating on a multipoint scale based on textual information. Classification models such as SVM, methods for opinion mining and sentiment analysis based on WordNet or Parts of Speech tagging can be incorporated in our approach to calculate normalized rating for yelp reviews.

3. Personalized Recommendation

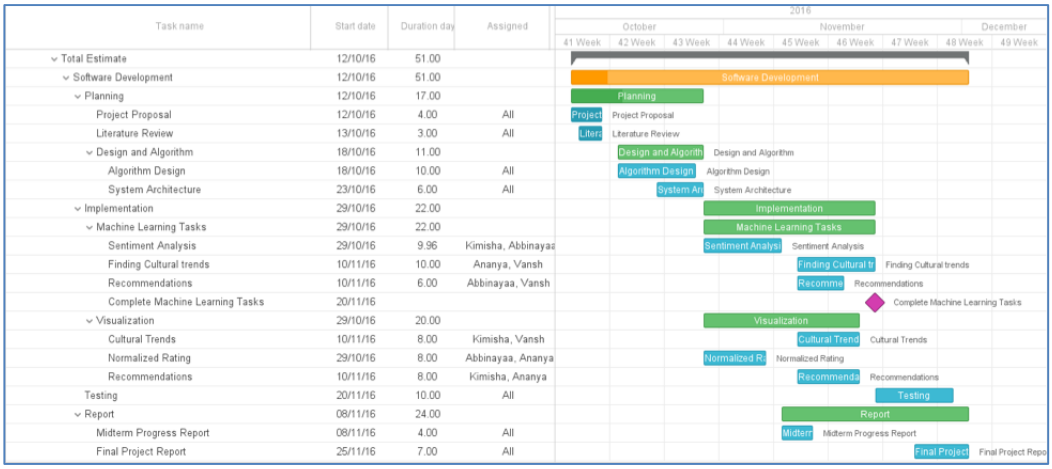
   The paper (Zhang) (Guokun Lai, 2016) discusses an approach to making personalized recommendations of Yelp products based on textual reviews (content-based collaborative-filtering) in addition to product features (item-based collaborative filtering). This approach can be leveraged in making quality recommendations based on a user's personal preferences.

4. Visualization

The paper (Karthik Ramesh, 2016) discusses methods for analysis and visualization of eating trends across various cities. This is relevant to our project as we aim to find cultural trends in products across regions and use them to provide insightful visualizations as well.

We also reviewed other papers on topics such as Virality Prediction (Lilian Weng) to understand how the Yelp network structure of friends can be used to identify diffusion of information and predict star users.

TIMELINE



Task name	Start date	Duration day	Assigned
Total Estimate	12/10/16	51.00	
Software Development	12/10/16	51.00	
Planning	12/10/16	17.00	
Project Proposal	12/10/16	4.00	All
Literature Review	13/10/16	3.00	All
Design and Algorithm	18/10/16	11.00	
Algorithm Design	18/10/16	10.00	All
System Architecture	23/10/16	6.00	All
Implementation	29/10/16	22.00	
Machine Learning Tasks	29/10/16	22.00	
Sentiment Analysis	29/10/16	9.96	Kimisha, Abbinayaa
Finding Cultural trends	10/11/16	10.00	Ananya, Vansh
Recommendations	10/11/16	6.00	Abbinayaa, Vansh
Complete Machine Learning Tasks	20/11/16		
Visualization	29/10/16	20.00	
Cultural Trends	10/11/16	8.00	Kimisha, Vansh
Normalized Rating	29/10/16	8.00	Abbinayaa, Ananya
Recommendations	10/11/16	8.00	Kimisha, Ananya
Testing	20/11/16	10.00	All
Report	08/11/16	24.00	
Midterm Progress Report	08/11/16	4.00	All
Final Project Report	25/11/16	7.00	All

Figure 1 Timeline

INNOVATION

Yelp's average rating for different businesses and products are based on individual ratings and the number of reviews. This could be misleading as different people have different baselines and scales. Hence, we have provided a single holistic score for each product based on all product reviews.

We are using the normalized ratings from sentiment analysis to identify cultural trends from review text across locations. Yelp's cultural trend analysis tool, Yolo monitors trends from the past 10 years by analyzing how often certain words have been used in different regions. However, this approach is limited as it only compares how the popularity and demand for a certain product has changed over time but does not portray cultural trends of products across different locations.

For personalized recommendations, we have applied latent dirichlet allocation on review text to extract latent factors like personal preferences and product properties to build a meaningful content based recommendation system instead of a collaborative filtering based model.

Thus, combining all these features into an integrated system we feel that our approach goes beyond the state of the art solutions.

SYSTEM ARCHITECTURE

Below we present the high-level system architecture for Yideas:

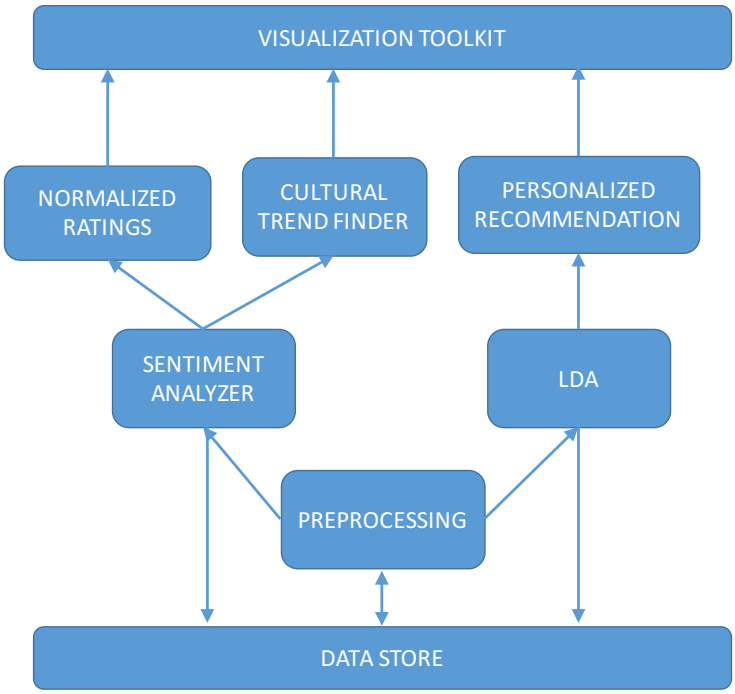


Figure 2: System Architecture

DATA STORAGE

Data Setup: We are using the academic dataset from the yelp dataset challenge (JSON files) and MongoDB for data storage. The data is around 4.4 GB with over 25 million review records. Much of the used data was qualitative (textual reviews).

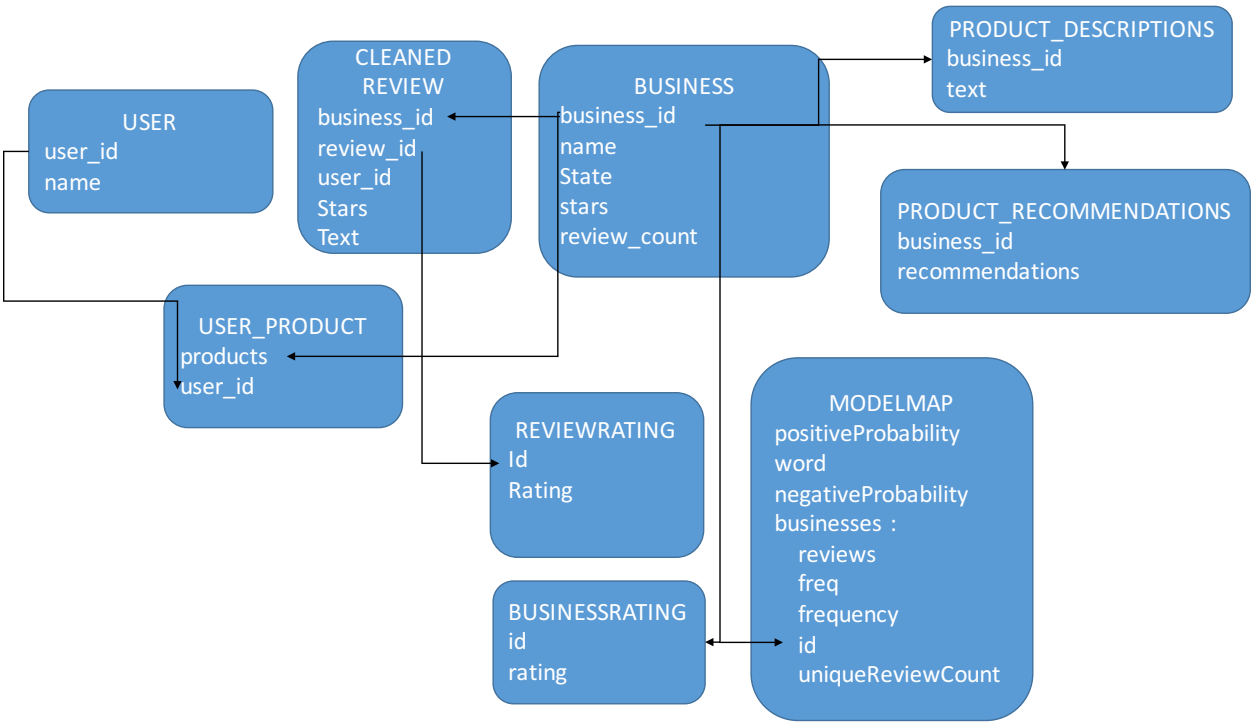


Figure 3 Data Storage - MongoDB schema

DATA CLEANING

The review data was cleaned by removing stop words, punctuations, control characters, digits and stemming. We also used n-gram fingerprinting to merge similar/abbreviated words. The cleaned data was dumped into MongoDB as a new collection called cleanedReview to serve as input for LDA and sentiment analysis.

IMPLEMENTATION

We have presented the visualization in the form of a user web interface where a user can log in, view personalized recommendations based on his preferences, get a succinct analytical view of reviews and search for cultural trends.

This is done using a dashboard built on the Bootstrap framework that gives a consistent look and feel to all individual components.

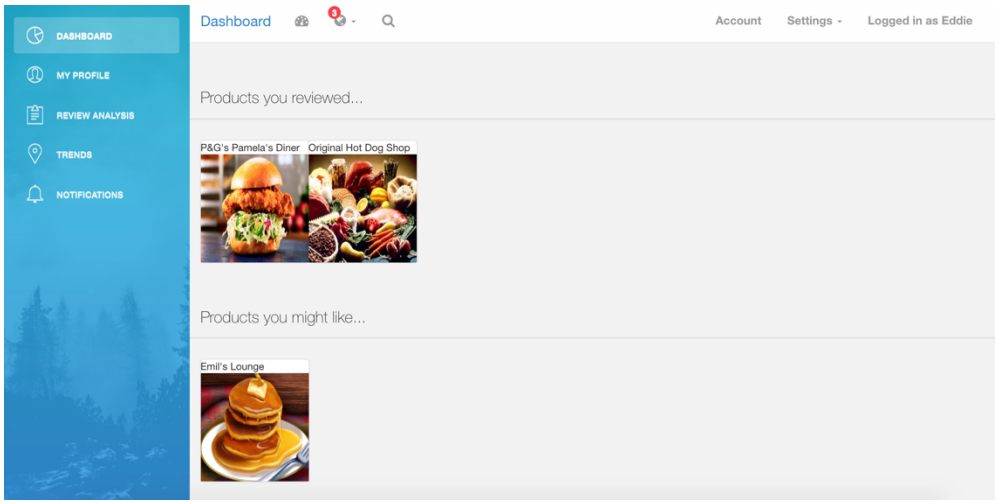


Figure 4 Personalized Dashboard

NORMALIZED RATINGS

Sentiment Analysis is core to our project as we want to derive sentiment from a user's review by predicting latent factors that trigger user's decision and anticipate how a user will rate another product.
We believe that ratings don't always justify text, hence we have come up with a rating prediction based on the underlying sentiment in the review. We are using a Bayesian approach to treat Sentiment Analysis as a classification task. Positive sentiment reviews are expected to be rated higher and negative sentiment reviews map to lower rating.
The actual assignment is based on a probabilistic function:

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_{i \in positions} P(x_i \mid c_j)$$

Figure 5: Text classification using Naive Bayes (Jurafsky, 2016)

Where x = word, c = class
We assign a hybrid rating to products based on star rating and usefulness of review text by calculating the probability of every review being negative or positive, based on individual word probabilities. Comparing this conditional probability with the threshold (<2.5 negative, > 2.5 positive), we classify each review as positive or negative.
Next we treat the reviews as a mixture of Gaussians and assign the review rating based on the amount of standard deviation from mean. Finally, we use the individual review scores to calculate cumulative product scores. Taking into account multiple factors from the review, this will be a more accurate predictor of the overall rating of a product.

We have used bubble charts to display the sentiment associated with words and reviews. The user can search for a restaurant and view all words associated with respective reviews as well as the sentiment of each word. The size of the bubbles corresponds to frequency of

the word and the color indicates negative/positive sentiment. On clicking each word, reviews associated with the word and their normalized ratings are displayed.
Hence this visualization provides a very intuitive way for users to derive insights on products.

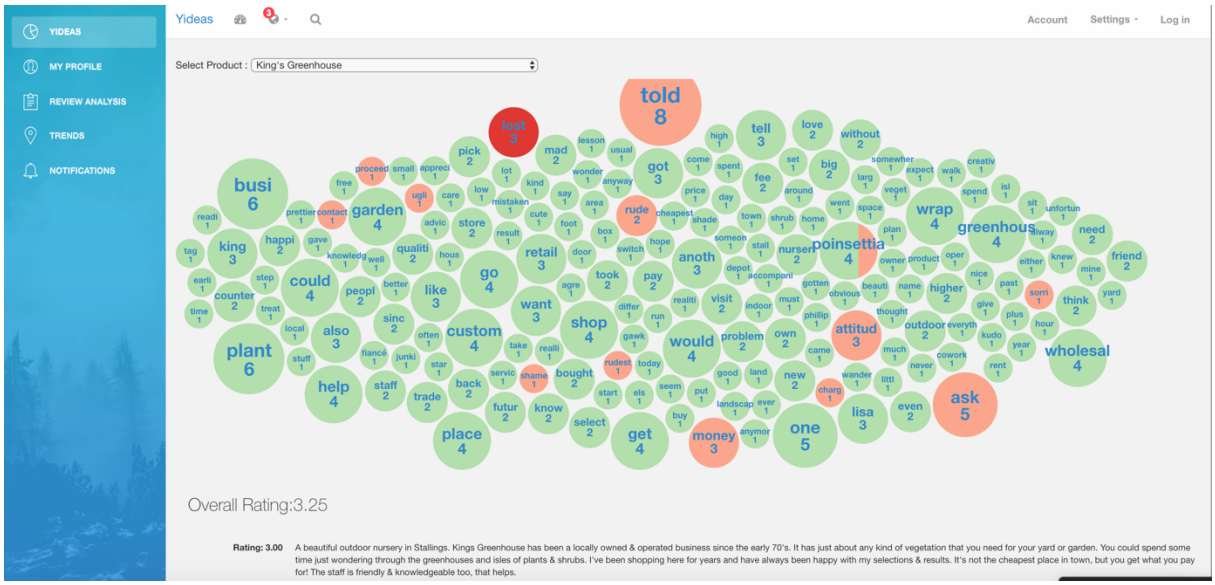


Figure 6 Sentiment Analysis Bubble Chart

PERSONALISED CONTENT BASED RECOMMENDATIONS

We have leveraged an algorithm called **Latent Dirichlet allocation (LDA)** for providing content-based personalized recommendations to users.
Latent Dirichlet allocation is a statistical topic model that generates topics from a set of documents based on word frequency and is widely used for discovering abstract topics from text.

We have leveraged the gensim Python module for this task. This module allows LDA model estimation from a training corpus and inference of topic distribution on unseen documents. The output is a list of topics with their probabilities in the training corpus.

For this task, the input is the training set, which is a subset of the original data from Yelp. The LDA model so trained will be run against the remaining subset using cross validation techniques. The final output will be the test document set with topic probabilities in each document saved into a new collection called product descriptions with a list of products and corresponding topics associated with each.

The following algorithm has been implemented for personalized recommendations for a user:

1. Find topics related to each review using LDA. We find the optimal number of topics by computing perplexity using 10-fold cross validation

   Evaluating log-likelihood:

   $$\mathcal{L}(\boldsymbol{w}) = \log p(\boldsymbol{w}|\boldsymbol{\Phi}, \alpha) = \sum_{d} \log p(\boldsymbol{w}_d|\boldsymbol{\Phi}, \alpha).$$

   Likelihood of unseen documents can be used to compare models; higher likelihood implying a better model. Performing cross-validation to find perplexity using different number of topics, we can obtain optimal number of topics.
   Perplexity of each test set is computed as below:

   $$\text{perplexity(test set } \boldsymbol{w}) = \exp\left\{-\frac{\mathcal{L}(\boldsymbol{w})}{\text{count of tokens}}\right\}$$

   where w is the number of unseen documents and the model is described by the topic matrix and hyper parameter.
   After the cross validation, we found optimal number of topics to be 200.

2. Find aggregated topics (or description) per product- these are the latent topics that describe the product
3. To make recommendations to users, find all products that he has reviewed and that have been rated highly above a certain threshold
4. For each of the products we index the topics appearing in the product descriptions and compute TF-IDF matrix.
5. Use cosine similarity to find all products similar to these products and recommend them to user. This similarity is based on the topics appearing in the reviewed document and the TF-IDF matrix we computed in step 4.

TFIDF

For a term i in document j:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j
df_i = number of documents containing i
N = total number of documents

Figure 7 Computing TF-IDF Matrix

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Figure 8 Cosine Similarity

For visualization, we have created a personalized dashboard, where a user can login and view the list of products recommended based on already reviewed products.

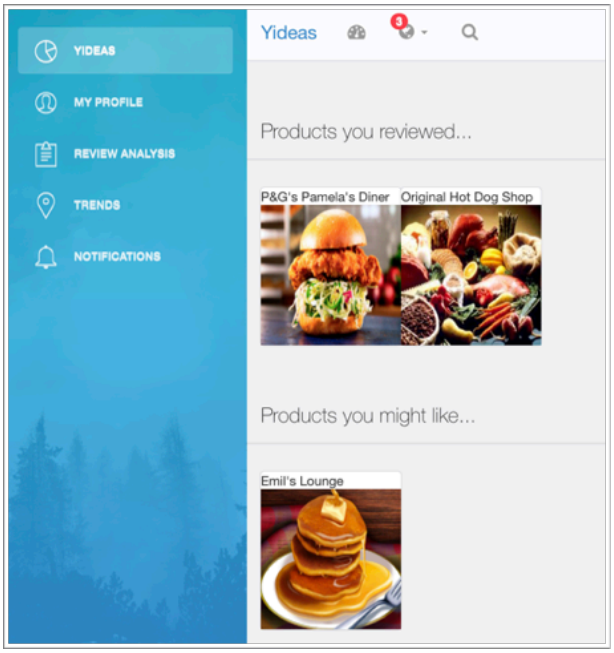


Figure 9 Personalized Recommendations

CULTURAL TRENDS

We have leveraged the sentiment analysis algorithm to find cultural trends. Through the sentiment analysis algorithm, we already have the normalized score for each business by taking the average over all review ratings calculated for that business.

For example, to find cultural trends for Indian businesses in the USA, we search for "Indian" using the interactive visualization GUI:

1. Our algorithm uses business location information from the dataset to filter all reviews that have the keyword Indian in them.
2. The normalized score for each business is used to determine the cumulative rating for all businesses within states containing topic "Indian". This, in effect, gives a measure of the popularity of Indian businesses across different states

3. We then use heat maps to effectively visualize the cultural trend across states.
On hovering over the state, users can see the sentiment value of the search term in each state and they can zoom into a particular state to see the top 5 businesses for that trend.

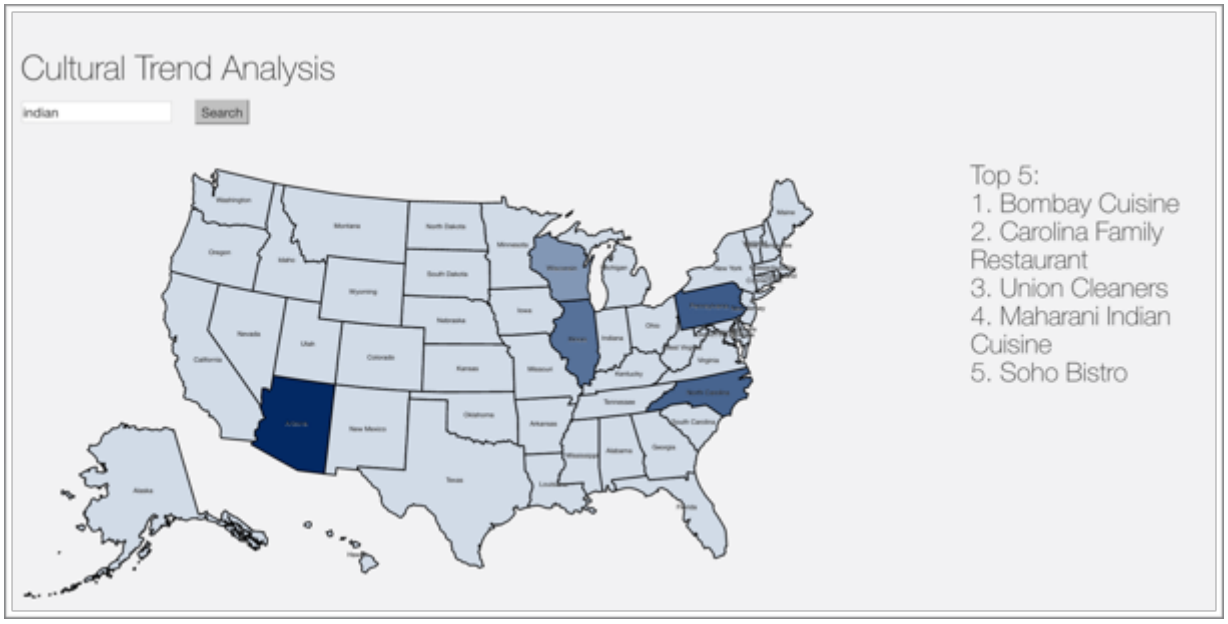


Figure 10 Cultural Trends Analysis Heatmap

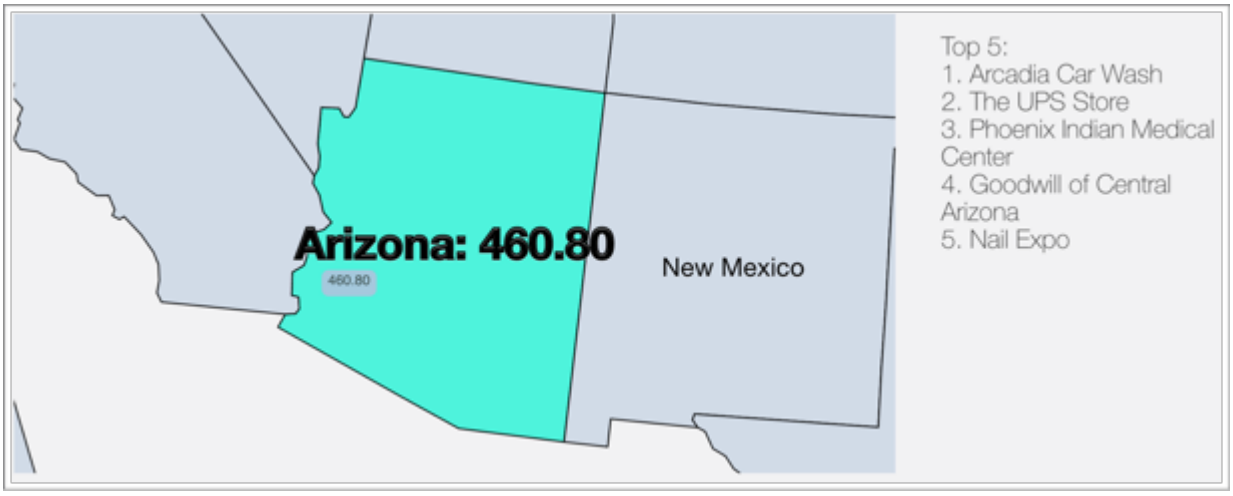


Figure 11 Cultural Trends Analysis Heatmap zoomed in

EXPERIMENTS AND EVALUATION

USER SURVEY

We have conducted user surveys to gauge the value addition of our product in terms of the features and user experience.
The survey mainly included validating accuracy of content-based ratings and how they compare against actual yelp ratings. Users searched for products, reviewed sentiment and normalized scores associated with its reviews and obtained personalized product recommendations. They also assessed the accuracy of the review sentiments and personalized recommendations, and provided feedback.

Our survey had the following questions:

1. How intuitive was the user experience? (Rate on a scale of 1 to 5)
2. How responsive was the system? (Rate on a scale of 1 to 5)
3. Do the results of personalized recommendations make sense to you?
4. Do the results of cultural trends you searched for make sense to you?
5. Do you agree with our review and business ratings?

We also tested the user experience and received feedback as part of our user surveys. Around 15 people out of 20 have given very positive, rating their experience between 4 – 5.
Out of the remaining, we have 4 people rating our product with 3 starts and 1 with 2 stars.
The feedback provided by people with lower ratings mostly corresponds to the lack of complete dataset, leading to 0 scores for most of the states in the Heatmap and slowness in data load, as we are running the server on localhost.
Overall, users have enjoyed the experience and the feedback indicates that they are in favor of the interactive visualizations and feel the application is intuitive and easy to use.

Some of the constructive feedbacks we got are as below:

- Cultural Trend Map shows a null score for most of the states:
  This is due to the fact that we have a limited yelp dataset which contains information about businesses only from 7 states in US. Given the complete dataset, the heat map should produce useful insights about trends across US.
- Recommended products are similar to already reviewed items but are not matching user requirements:
  Currently we have implemented content based recommendations considering all the topics present in the products users have reviewed, which is mainly based on product features rather than user preference. As an improvement to our current approach, we could use a combination of content based and collaborative filtering to take into account both product features and user preferences

HUMAN CURATION

We also took a subset of the test set and manually curated the output to verify that the normalized scores are in line with the expectation. We extracted user and reviews data from yelp to our database to validate whether the personalized recommendations generated match our interests. Detailed results have been discussed later.

For cultural trends discovery, we evaluated results using well-known trends. For instance, we know that certain areas are known to have more of a certain ethnic group than another and when we search for a trend for that particular group of people, the trend should reflect the demographics. In the example shown above in Figure 10, when we search for the trend 'Indian', we get the highest trend rating from Arizona, closely followed by Pennsylvania and North Carolina. The demographics of these states do show a greater percentage of Indians than the states for which the trend rating is low like Wisconsin and Illinois. The top 5 businesses are also mostly Indian restaurants or businesses owned by Indian owners.

RESULTS AND OBSERVATIONS

LDA

To evaluate relevance of topics extracted by LDA, we manually compare the LDA output for the following 2 example reviews

Review 1 (Rated 5 stars)

Overall, its a good alternative to Subway, which is down the road.", "type" : "review", "business_id" : "5UmKMjUEUNdYWqANhGckJw" }
{ "votes" : { "funny" : 1, "useful" : 1, "cool" : 0 }, "user_id" : "auESFwWvW42h6alXgFxAXQ", "review_id" : "fFSoGV46YxuwBr3fHNuZig", "stars" : 5,
"date" : "2015-10-31", "text" : "Yes this place is a little out dated and not opened on the weekend. But other than that the staff is always
pleasant and fast to make your order. Which is always spot on fresh veggies on their hoggies and other food. They also have daily specials and
ice cream which is really good. I had a banana split they piled the toppings on. They win pennysaver awards ever years i see why.", "type" : "
review", "business_id" : "5UmKMjUEUNdYWqANhGckJw" }

Review 2 (Rated 1 star)

{ "votes" : { "funny" : 0, "useful" : 0, "cool" : 0 }, "user_id" : "jBoH6qKGO7wdYyg_YjBcQA", "review_id" : "V-bqYx62zpxfH2oFkzXPzw", "stars" : 1,
"date" : "2016-04-10", "text" : "Normally, I do not do reviews of an establishment unless the rating is exceptionally great or exceptionally
bad. If I had not felt sucker punched and mugged after I left Mr. Hoagie, I would not be writing this bad review. On my first and last trip
there, I paid $24.59 for two whole hoagies. \n\n1) The Italian (ordered because the woman (she) working there told me it was their most
popular hoagie) I ordered with lettuce and onions (no tomato, italian dressing which all come free with the hoagie). I had to pay $1 to add
cheese to the hoagie. Because I was traveling and would not be eating the hoagie right away, I asked for the italian on the side. She told
me there was a charge for the dressing on the side. I passed on that on principle. Are you kidding me? I just saved her money by declining
the tomato. I was not going to pay for the dressing on the side. At this point, I am certain something is not right and this place appears
to me to be in financial trouble. That pales in comparison to what happened next. I opened the hoagie and there were three microscopically
thin slices of meat on this hoagie. Ham, salami, and bologna. Bologna!?!?!?!? Who in the hell puts disgusting bologna on an Italian
hoagie? That disqualifies that hoagie as an Italian. The red flags indicating this place is in financial trouble are blinding my view. One
area of this hoagie only had one slice of bologna, no ham and no salami.\n\n2) The Steak hoagie came with nothing on it but lettuce, tomato,
onions, and dressing. I ordered it without lettuce, tomato, onions, or dressing. I asked for mayo on the side. I had to pay for that so I
declined. Are you kidding me? I just saved them money by passing on the lettuce, tomato, fresh onions, and dressing! I paid $3 to add
sautéed mushrooms, onions, and cheese. Here we go again. Canned mushrooms that were not even cooked through (were still cold) and onions
that were still crunchy/undercooked. News flash: you need to serve fresh mushrooms that have been sautéed (canned mushrooms are another sign
this place is in financial trouble). I literally had to open the hoagie to find the steak because I could not taste or see it. The hoagie
was 50% mushrooms, 30%onions, 10% cheese and 10% steak.\n\nI could have personally made 10 whole hoagies with substantial quality ingredients
and toppings for the $24.59 I paid for these two whole hoagies. I intentionally patronize non-chain restaurants to support small businesses.
In this case, Never. Again. (The hoagie bun was better than average.) Wake up or you are going to lose your business.", "type" : "review", "
business_id" : "5UmKMjUEUNdYWqANhGckJw" }

For review 1, 3 topics were derived from the training dataset:

Topic 1: "always fast"
Topic 2: “food pennysave”
Topic 3: “good food”

The probabilities for each topic within a text are represented as json:


```
{
  "topicProbs":[
    {
      "prob":0.9809295463826615,
      "topic":" food pennysave"
    }
  ],
  "review_id":"fFSoGV46Yxuwbr3fHNuZig",
  "stars":5
}
```


PERSONALIZED RECOMMENDATIONS

In addition to checking algorithm output, we also verified the recommendation end results.

For example, based on Eddie’s restaurant reviews, the following restaurants were recommended to him. We verified the relevance of the recommendations by looking up the latent topics associated with each recommended restaurant and doing an eyeball check on those of the reviews the user wrote.

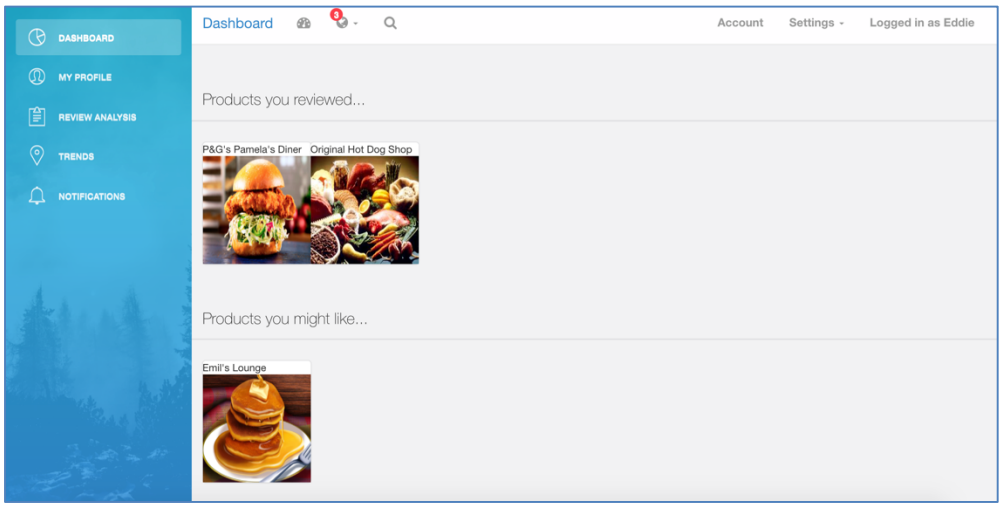


Figure 12 Personalized Recommendations

In 85% of cases, the recommendations generated matched expectations with a few anomalies. There were, however, certain cases where topics generated did not exactly

correlate to the main idea of the review. Thus, there is scope of improvement in this area by refining the algorithm.

SENTIMENT ANALYSIS – CROSS VALIDATION

We evaluated sentiment analysis by comparing the label assigned by the sentiment analysis text classifier against the expected label of a review based on its Yelp Rating. The expected label is positive if Yelp Rating >= 2.5 and negative for Yelp Rating <2.5. After defining our metrics, we used 10-fold cross validation over the 26 million reviews and we observed that we have a normalized accuracy of 80.9%. Below we present the confusion matrix:

	Actual Positive	Actual Negative
Predicted Positive	20254520	4963920
Predicted Negative	780	780780

CONCLUSION

Yideas helps address some of the problems that users and businesses on Yelp are facing today. Using sentiment analysis of reviews, we are able to provide users a holistic rating to help evaluate a business quickly and eliminate bias in the rating scales of different users. We also use content based filtering and topic matching to provide better recommendations compared to collaborative filtering models today. Finally, our cultural trend analysis helps provide businesses and users with useful insights about new growth avenues and top recommendations for a particular trend by state.

Effective visualization strategies such as heat-maps and bubble charts give users easy access to important information like sentiment related to a particular product and cultural trend across locations, which can help drive strategies for businesses as well as help people explore their communities better.

FUTURE WORK

Due to the large amount of data, the algorithms for LDA and Sentiment Analysis are computationally expensive and take a long time to run. In order to make the process more efficient and scalable, we could implement map reduce to parallelize the process.

WORK DISTRIBUTION

All team members contributed similar amount of effort. (Refer to Timeline for work distribution).

BIBLIOGRAPHY

Agrawal, S. (n.d.). Feature based Star Rating of Reviews: A Knowledge-Based Approach for Document Sentiment Classification.

Bo Pang, L. L. (2015). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales.

Guokun Lai, M. Z. (2016). Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis.

Hanhoon Kang, S. J. (2012). *Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews.* Expert Systems with Applications 39 6000-6010.

James Huang, S. R. (2014). Improving Restaurants by Extracting Subtopics from Yelp Reviews. *SOCIAL MEDIA EXPO 2014*.

John P. Schomberg, O. L.-C. (2016). Supplementing Public Health Inspection via Social Media. *PLoS ONE 11*.

Julian McAuley, J. L. (n.d.). Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text.

Jurafsky, D. (2016). *Text Classification and Naïve Bayes*. Retrieved from https://web.stanford.edu/class/cs124/lec/naivebayes.pdf

Karthik Ramesh, P. C. (2016). Yelp Unbounded: Visual Analytics.

Lilian Weng, F. M.-Y. (n.d.). Virality Prediction and Community Structure in Social Networks. *arXiv:1306.0158v2 [cs.SI]*.

Linshi, J. (2016). Personalizing Yelp Star Ratings: a Semantic Topic Modeling Approach.

Luca, M. (n.d.). Reviews, Reputation, and Revenue: The Case of Yelp.com.

Mingming Fan, M. K. (2016). Predicting a Business' Star in Yelp from Its Reviews' Text Alone.

Zhang, Y. (n.d.). Incorporating Phrase-level Sentiment Analysis on Textual Reviews for Personalized Recommendation.