**WOM Restaurant Recommendation System**  
Team Revengers

Arun Ramakrishnan, Juily Vasandani, Maharshi Dutta, Samir Husain, Yizhu Liao

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

[ramakr17@purdue.edu](mailto:ramakr17@purdue.edu); [jvasanda@purdue.edu](mailto:jvasanda@purdue.edu); [dutta33@purdue.edu](mailto:dutta33@purdue.edu);

[husain2@purdue.edu](mailto:husain2@purdue.edu); [liao71@purdue.edu](mailto:liao71@purdue.edu)

[Video](https://youtu.be/M0jvTXESfaQ) | [Onedrive](https://purdue0-my.sharepoint.com/:f:/g/personal/husain2_purdue_edu/EjA_Crms8_FHkY7hEo7u2EoBpyiW1iGcyhLU_jX9wRnJrg?e=gcGgHF) | [Github](https://github.com/samirhusain26/WoM-restaurant-recommendation)

## Abstract

By digitizing Word of Mouth (WOM) Advertising, our team has restructured the logic behind recommendation systems, revolutionizing their algorithmic process. Current recommendation systems direct users to options in the same category or ones with similar features or attributes, developing a gap between a consumer’s expectations and what the system can offer. Our algorithm instead maps a user’s history to recommend restaurants that matches them to restaurants that have received similar reviews, shifting the focus away from restaurants and towards personal experience. This was done by combining descriptive and prescriptive analytical techniques and integrating it within the R Shiny environment.

*Keywords:* natural language processing, cosine similarity, document-term matrix, R Shiny, recommendation systems, word-of-mouth advertising

## Business Problem

We believe that current restaurant recommendation systems do not satisfy consumer’s expectations, because their algorithm is highly geared towards restaurants, rather than being customer-focused, directing users towards options with similar features and categories, or sponsored results. Clients expecting personalized recommendations are given options that aren’t unique or exciting, which can lead towards low customer satisfaction and a reduced customer base, subsequently impacting restaurants who lose out on potential profits. Knowing that users trust personal recommendations – usually obtained through verbal communications aka word-of-mouth – our solution depended on our ability to translate this tradition of sharing information into a digital service and automate it. By doing this, our recommendation system will ease the decision-making process for clients and reduce profit loss for restaurant owners.

## Analytics Problem

With the main objective of our project being the creation of a recommendation system to enable hyper-personalized restaurant recommendations, the underlying analytical problem for this objective would be to identify similarities between restaurants that weren’t simply based on what cuisine the restaurant had or whether or not it served vegan food or had a patio. Instead, to incorporate that crucial word of mouth dimension, customer reviews became the backbone of our system. Our app had to be able to read a user’s review history as an input, calculate its similarity to reviews from other restaurants, and output a list of unique recommendations for the user to explore.

In its current functionality, our user is assumed to have a Yelp account, have reviewed several restaurants and be in the Madison, Wisconsin area. This was done to ensure ease of computational processing and usability of the Yelp dataset from Kaggle. As Yelp is the largest crowd-sourced review forum, it was necessary to source our project data from their resources, with the goal being to develop a beneficial add-on feature for Yelp users to explore restaurants outside of what the in-house Yelp recommender could offer.

## Data

In order to effectively build a text mining system, we needed to find a dataset large enough for us to process and tokenize individual reviews per restaurant. The most appropriate dataset for our project was the Yelp Review Dataset[[1]](#footnote-2) on Kaggle.

To perform pre-processing, we used the tidyText and SnowballC packages to clean[[2]](#footnote-3), tokenize, and remove stop words from the dataset. The resulting list of useful words was further converted into a matrix with rows representing each restaurant and the columns representing a unique word. The values of each cell represented the frequency with which each word appeared in the cumulative list of reviews for that given restaurant. This frequency is the crux of the calculation for a similarity score between restaurants.

## Methodology

The underlying approaches to our recommendation system lie in the fields of descriptive and prescriptive analytics – by utilizing text mining techniques and building the recommendation algorithm. The broader idea for this app came from a Surprise Me kernel[[3]](#footnote-4) built on the Yelp dataset which uses Euclidean distance between two points in a matrix to calculate the similarity score. We built on this by using cosine of the angle between two vectors so that we get a more relevant set of recommendations.

Although the kernel owner used Python to develop their model and algorithm, we found that the variety of packages offered in R and the flexibility of R Shiny as a hosting platform made it more feasible to develop and improve upon this concept.

## Backend Process

The process starts by obtaining a user’s history on Yelp (focusing on restaurants they have reviewed) and assigns all other restaurants a rank on the basis on the similarity score they receive from the algorithm. A weighted average of all ranks is calculated to get a consistently higher ranked restaurant to the top. The top 10 restaurants from this final ranking are shown as recommendations.

## Functionality

The app consists of a login page where a user enters their username. They are then taken to the profile page where they see their details and a list of restaurants in Madison, WI that they have reviewed. Clicking on the ‘Recommend’ button us shows a pop-up box with a list of 10 restaurant recommendations that are built using a cumulative ranking system and similarity scores from the cosine similarity matrix. In order to work with the given dataset, we have built usernames for all users using their name and Row number. Also, a map popup is shown with the restaurant locations. These popups were created using modal dialogs and leaflet package for rendering the interactive map.

## Conclusions

WOM Restaurant Recommendation System provides personalized recommendations to customers based on similarities between Yelp reviews for restaurants to increase customer satisfaction. This process takes advantage of Word of Mouth marketing strategies in order to revolutionize the way recommendation systems work, shifting focus from restaurant features to customer experiences.

We have also built this app with the goal of scaling the system to be applicable to other institutions. By doing so, we hope to change the logic behind recommendation systems, providing a service that exceeds customer’s expectations and can empower all business to increase profits.

References

1. Yelp, Inc. (2019, February 5). Yelp Dataset. Retrieved from <https://www.kaggle.com/yelp-dataset/yelp-dataset>.
2. fahd09. (2019, January 4). Yelp Dataset: SurpriseMe Recommendation System. Retrieved from <https://www.kaggle.com/fahd09/yelp-dataset-surpriseme-recommendation-system>.
3. liu1498. (2018, October 12). Movie-Recommender-System-ShinyApp. Retrieved from <https://github.com/liu1498/Movie-Recommender-System-ShinyApp>.
4. Natural Language Processing with R. (n.d.). Retrieved from <http://rpubs.com/LuizFelipeBrito/NLP_Text_Mining_001>.

1. Yelp, Inc. (2019, February 5). Yelp Dataset. Retrieved from <https://www.kaggle.com/yelp-dataset/yelp-dataset>. [↑](#footnote-ref-2)
2. Natural Language Processing with R. (n.d.). Retrieved from <http://rpubs.com/LuizFelipeBrito/NLP_Text_Mining_001>. [↑](#footnote-ref-3)
3. fahd09. (2019, January 4). Yelp Dataset: SurpriseMe Recommendation System. Retrieved from <https://www.kaggle.com/fahd09/yelp-dataset-surpriseme-recommendation-system>. [↑](#footnote-ref-4)