GENIE: SIMPLIFYING THE RESTAURANT FINDING PROCESS

LAT35N Team91

Blaine Davenport, Rohit Pegallapati, Anthony Song, Shinji Ueno, Lingqi Zhang

Georgia Institute of Technology

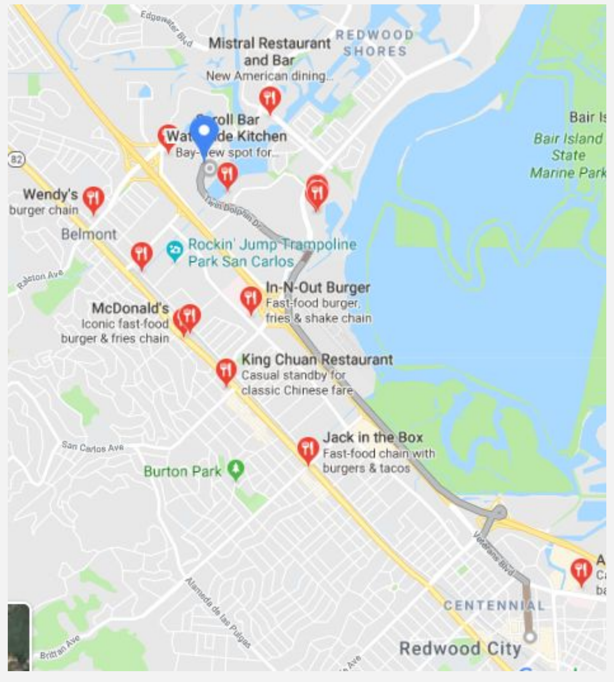
{bdavenport7, rpegallapati3, asong45, sueno9, lzhang624}@gatech.edu

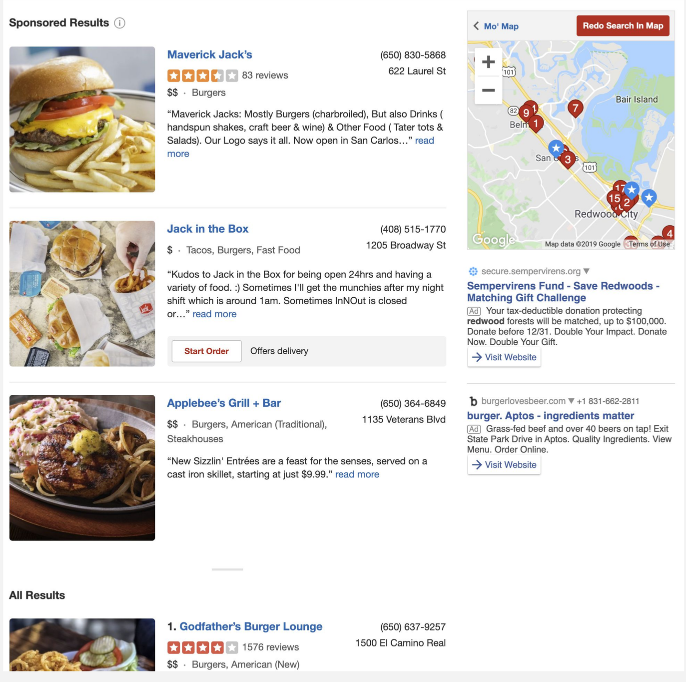
Introduction

There are more than **1 million** restaurants in the United States, and **90%** of restaurant guests research a restaurant online before dining. With the vast amount of restaurant review data available through sites like Yelp, Google, and TripAdvisor, it is often tedious and time-consuming to navigate through and find the perfect restaurant. To add difficulty to the problem, there are no commonly used applications that allow the user to search for restaurants along a route. **GENIE** combines maps, user interaction, machine learning, and visualizations, to help the user quickly and efficiently find the best restaurant along a route according to their preferences.

The two figures below display typical output of search results using today's most popular search tools (Google Maps and Yelp). The left figure shows route information and several restaurants nearby. However, it is not easy to find out the detail of each restaurant and their corresponding ratings. The right figure includes detailed information for each restaurant but also contains "sponsored" results, and the rating of each restaurant basically reflects the summary of different types of users' ratings, and the search results only show restaurants near one location.

The purpose of this project is to use machine learning algorithms with summary comparison visualizations to enable the user to quickly select the best restaurant for their purpose (e.g., meet up friends, travel) and preferences (e.g., history of their restaurant ratings.)



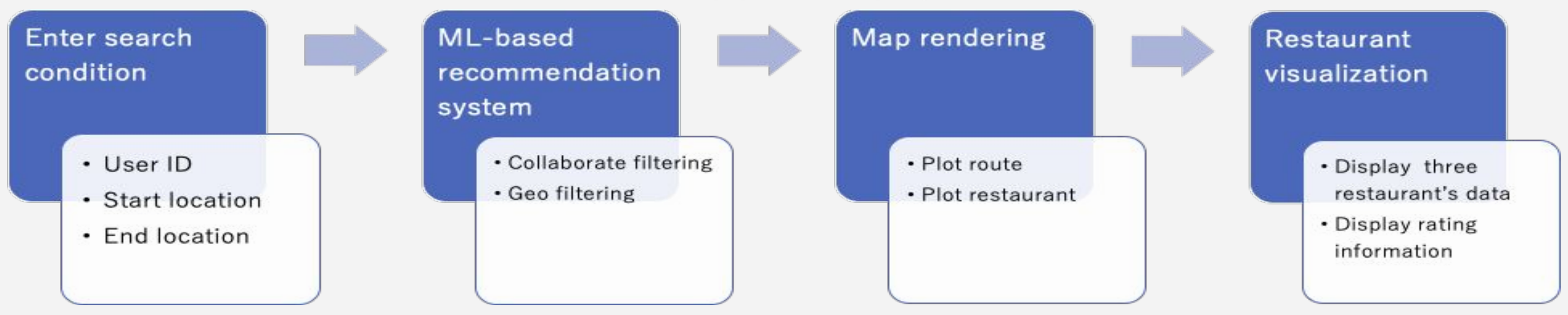


Methods

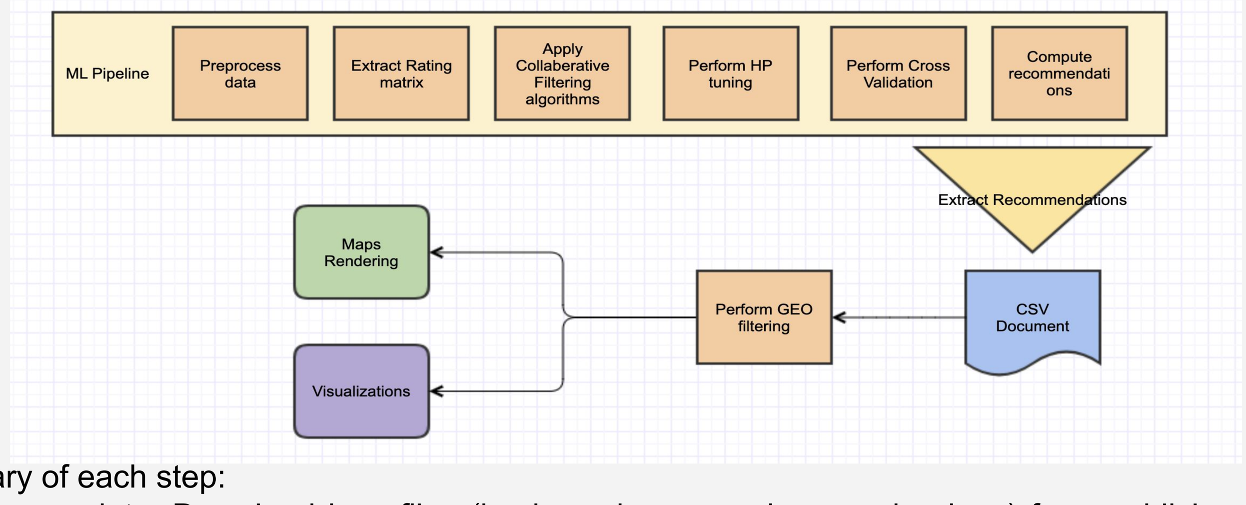
Our project expands on current state of the art with these **3** innovations:

- Recommends and compares restaurants along a route rather than around a single location
- Implements Machine-Learning based recommendation system to provide customized result for each user, not just a star rating based recommendation
- Displays useful summary comparison visualizations for easy judgement

Overall pipeline of project



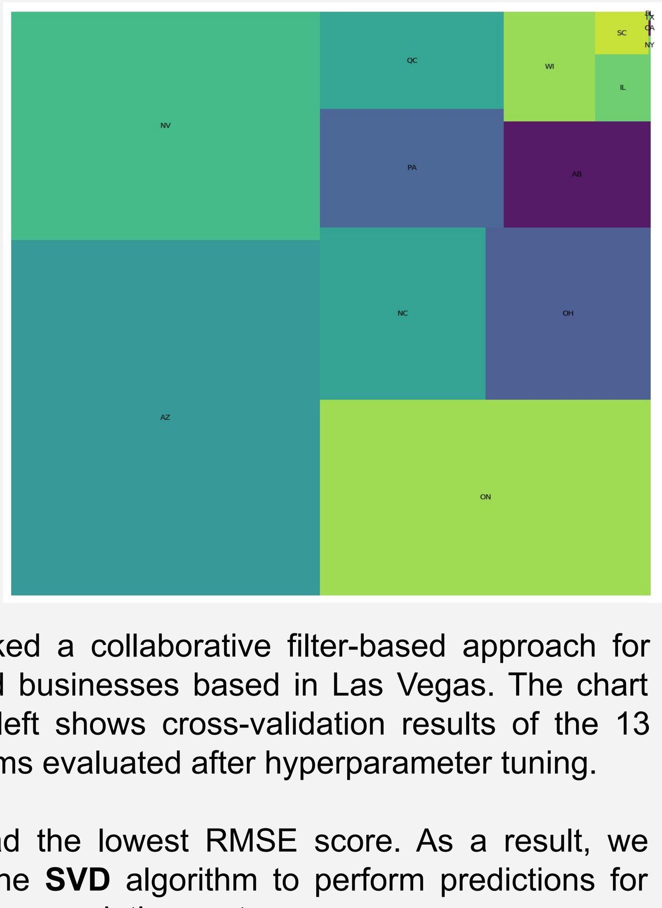
The key element of our project is a ML-based recommendation system. Detailed pipeline is shown below:

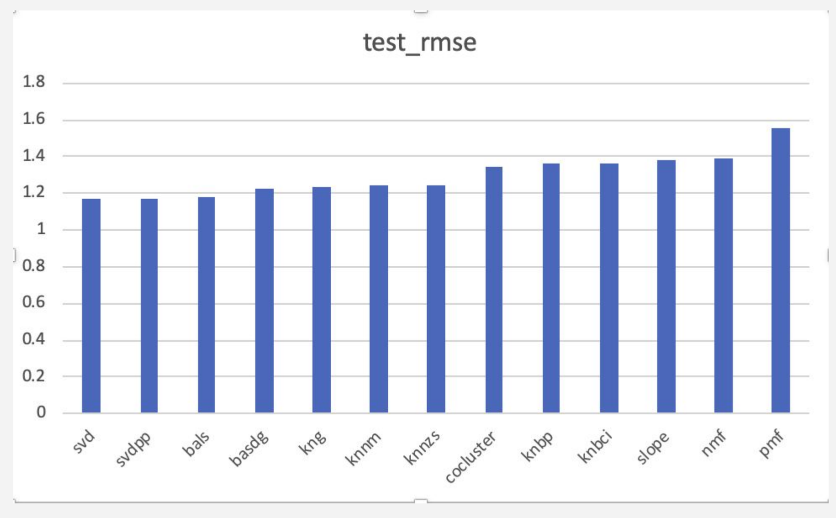


Summary of each step:

- Process data: Download json files (business.json, user.json, review.json) from publicly available Yelp dataset (8.7 GB uncompressed), filter the data to include only restaurants and food businesses in Las Vegas
- Extract rating matrix: Modify data for collaborative filtering to contain three parameters (user_id, business_id, star_rating), then split into training and test data
- Collaborative filtering algorithm: Apply 13 different algorithms by using Python scikit library Surprise
  - Neighborhood based algorithms: KNNBasic, KNNWithMeans, KNNWithZScore, KNNBaseline
  - Matrix Factorization based algorithms: Singular Value Decomposition(SVD), SVD++, PMF, NMF, SlopeOne, CoClustering, BaseLine Alternating, Least Squares, Baseline Stochastic Gradient Descent
- Hyperparameter tuning: Perform hyperparameter tuning for each of the 13 algorithms using GridSearchCV, then set optimal hyperparameters for each algorithm
- Cross validation: 5 Fold Cross validation for data and retrieving metric RMSE to pick best algorithm
- Compute recommendation: Compute recommendation by using test data and best algorithm
- Geo filtering: Filter a list of recommended restaurants by using start and end location. Any restaurant outside of "bone-shaped contour" area are filtered out

Our project originally planned to cover entire US based restaurants. However, exploratory data analysis on the Yelp dataset showed that the Yelp provided dataset was limited and did not contain all of the data. The treemap on the right shows that AZ and NV have more businesses compared to states like CA or NY. Also our focus is food-related businesses. By considering these limitations, we found that Las Vegas had the most data for food-related businesses. Thus, we decided to focus on the **Las Vegas** area to create customized restaurant recommendations.





We picked a collaborative filter-based approach for the food businesses based in Las Vegas. The chart on the left shows cross-validation results of the 13 algorithms evaluated after hyperparameter tuning.

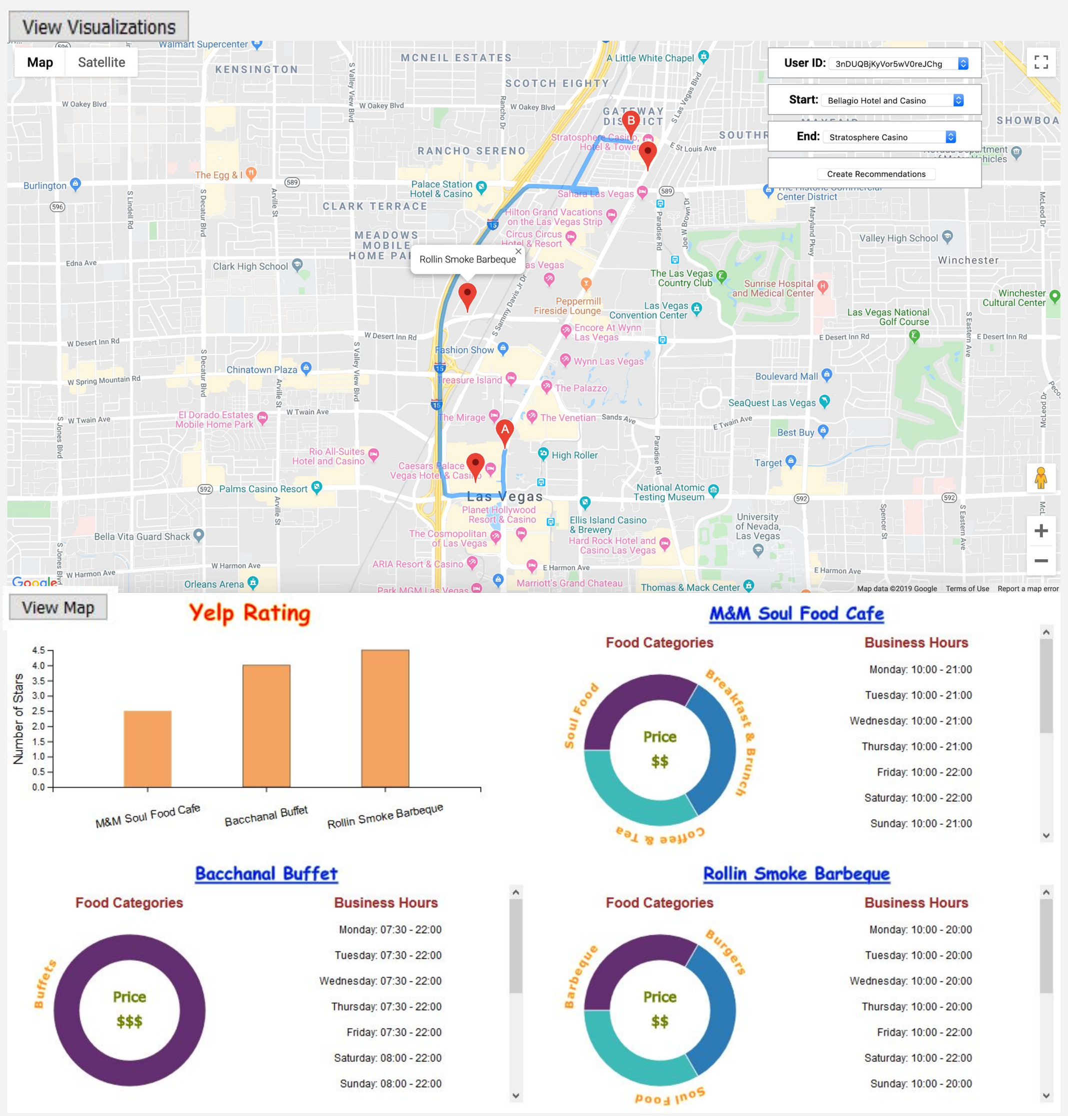
SVD had the lowest RMSE score. As a result, we chose the **SVD** algorithm to perform predictions for our recommendation system.

Results

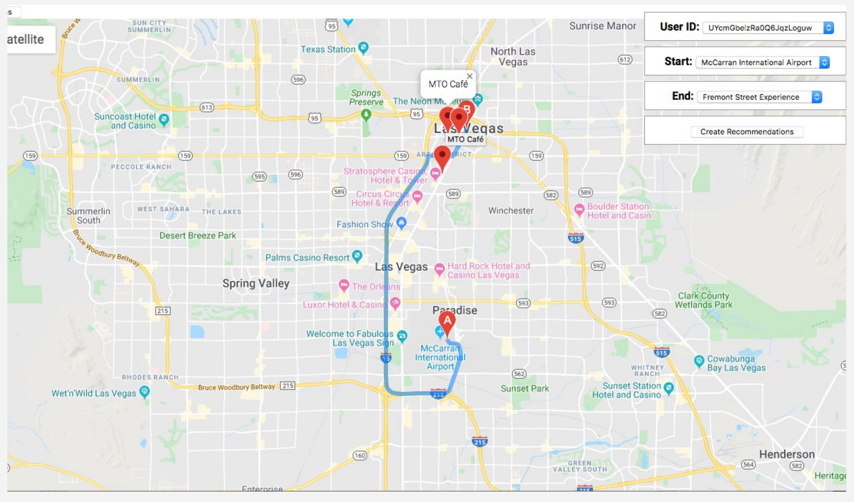
Output of the recommendation system: the upper half of the screen shows the route from start point (Bellagio Hotel and Casino) to end point (Stratosphere Casino) with three recommended restaurants for selected user along the route. Upon mouseover, the name of the restaurant will appear above the marker for each recommended restaurant.

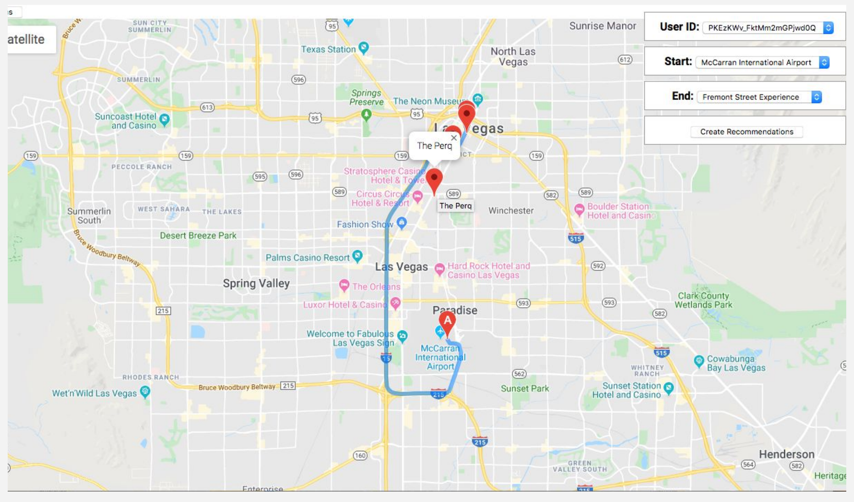
The bottom half of the screen shows information about the three recommended restaurants, including comparison of Yelp ratings and useful information for each restaurant.

The visualization part extracts the restaurant information such as business name, business hours, food categories, latitude and longitude from Yelp Fusion API based on the Business ID's passed from the recommendation system. The map and directions used in this project is implemented using Google's Maps and Directions API.



This customized recommendation system returns a different set of restaurants based on UserID. The two figures below show different recommendations given the same start and end locations with a different UserID.





The visualizations also have interactive features, including detailed information for the rating and hyperlink to the restaurant's Yelp page, which will open in a new window.

Conclusions

For the system performance evaluation, we used **RMSE** as the metric for testing the algorithms and confirmed that the **SVD** algorithm performed the best. We executed user testing for the visual part with the questions below:

- Is the product useful? (range of 5)
- Does the product meet the needs of the user? (range of 5)
- Is the product easy and intuitive to use? (range of 5)
- Does the user like the way the product looks and feels? (range of 5)
- Does the user prefer this product over similar products? (range of 5)
- What could be improved? (Comments)

The current recommendation system is based on collaborative filtering, which inherently suffers from the cold start problem, in that the system cannot draw any inferences about the users or items that have not interacted with the system. This being the fundamental limitation, it manifests into no recommendations for users along a particular route if there is a lack of any data about restaurants or users along that route. Nevertheless, for routes along which data is available for restaurants and users, the system was able to output recommendations. There are ways in which this problem could be overcome by collecting metadata about new users while they register into the system, and by adding the feature of content-based filtering in order to make the recommendation system an ensemble learner. Our existing implementation is focused on the Las Vegas area, and because of the cold start problem, the users are selected from the test set. Based on that limitation, the system works as planned. The system returns three restaurants along the route, and provides a different set of recommendations based on user ID.

References

1. Bejarano, Andres, et al. "Measuring User's Influence in the Yelp Recommender System." *PSU Research Review*, vol. 1, no. 2, 2017, pp. 91–104., doi:10.1108/prr-02-2017-0016.
2. Bresciani, Sabrina, and Martin J. Eppler. "The Pitfalls of Visual Representations: A Review and Classification of Common Errors Made While Designing and Interpreting Visualizations." *SAGE Open*, vol. 5, no. 4, 2015, doi:10.1177/2158244015611451.
3. Danone, Yaakov, et al. "Visualizing Reviews Summaries as a Tool for Restaurants Recommendation." *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval - IUI '18*, 2018, doi:10.1145/3172944.3172947.
4. Doan, Tri, and Jugal Kalita. "Sentiment Analysis of Restaurant Reviews on Yelp with Incremental Learning." *University of Colorado, Colorado Springs*, 2016, cs.uccs.edu/~jkalita/papers/2016/DoanTriICMLA2016.pdf.
5. Hang, Lei, et al. "Design and Implementation of an Optimal Travel Route Recommender System on Big Data for Tourists in Jeju." *Processes*, vol. 6, no. 8, 2018, p. 133., doi:https://www.researchgate.net/publication/327090246_Design_and_Implementation_of_an_Optimal_Travel_Route_Recommender_System_on_Big_Data_for_Tourists_in_Jeju.
6. Keck, Mandy, and Dietrich Kammer. "Exploring Visualization Challenges for Interactive Recommender Systems." *CEUR Workshop Proceedings*, May 2018, ceur-ws.org/Vol-2108/paper3.pdf.
7. Kumar, Sudhanshu, et al. "Movie Recommendation System Using Sentiment Analysis from Microblogging Data." *ArXiv*, 27 Nov. 2018, arxiv.org/pdf/1811.10804.pdf.
8. Li, S., and J. Gong. "Mashup: A New Way of Providing Web Mapping/GIS Services." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVII, no. B4, 2008, pp. 639–648.
9. Pazzani, Michael J., and Daniel Billsus. "Content-Based Recommendation Systems." *The Adaptive Web Lecture Notes in Computer Science*, pp. 325–341., http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.130.8327&rep=rep1&type=pdf
10. Perez, Luis. Predicting Yelp Star Reviews Based on Network Structure with Deep Learning. 2017. arxiv.org/pdf/1712.04350.pdf.
11. Saggi, Mandeep Kaur, and Sushma Jain. "A Survey towards an Integration of Big Data Analytics to Big Insights for Value-Creation." *Information Processing & Management*, vol. 54, no. 5, 2018, pp. 758–790., doi:10.1016/j.ipm.2018.01.010.
12. Sawant, Sumedh, and Gina Pai. "Yelp Food Recommendation Systezhum." *Stanford University*, 2013, cs229.stanford.edu/proj2013/SawantPai-YelpFoodRecommendationSystem.pdf.
13. Trattner, Christoph, and Elsweiler, David. "Food Recommender Systems: Important Contributions, Challenges and Future Research Directions." *arXiv:1711.02760v2 [cs.IR]* 10 Nov 2017
14. Tsai, Chieh-Yuan, and Shang-Hsuan Chung. "A Personalized Route Recommendation Service for Theme Parks Using RFID Information and Tourist Behavior." *Decision Support Systems*, vol. 52, no. 2, 2012, pp. 514–527., https://lms.ctl.cyut.edu.tw/sysdata/22/27122/doc/983f1e1f02a96e69/attach/1520796.pdf.
15. Vartak, Manasi, Huang, Silu, Siddiqui, Tarique, Madden, Samuel, and Parameswaran, Aditya. "Towards Visualization Recommendation Systems." *SIGMOD Record,* Vol. 45, No. 4. December 2016
16. Wang, Ji, et al. "Clustered Layout Word Cloud for User Generated Review." *Yelp*, 2013, www.yelp.com/html/pdf/YelpDatasetChallengeWinner_WordCloud.pdf.
17. Wang, Yong, et al. "Towards Easy Comparison of Local Businesses Using Online Reviews." *Computer Graphics Forum*, vol. 37, no. 3, 2018, pp. 63–74., doi:10.1111/cgf.13401

Genie icon made by Freepik from www.flaticon.com