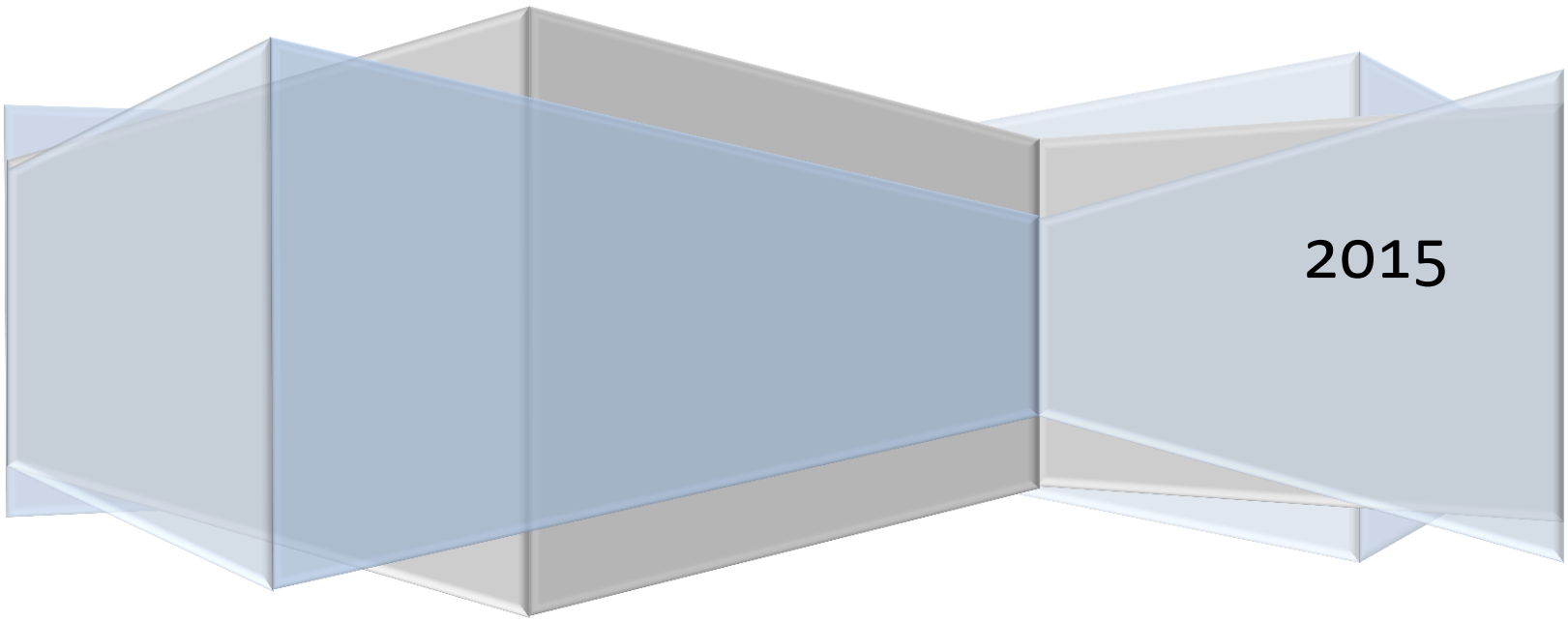
CSE 5334 Course Project

Yelp data - Project Proposal

Puneeth Umesh Bharadwaj

Shreekanth Kuppur Shreedhar



Proposal:

We see that most of the reviewers visit and provide reviews to the restaurants that are in their vicinity as they have knowledge of restaurants in and around there are or hometown. We provide user with suggestions in other locations (cities, neighborhoods) based on his/her reviews, ratings, check-in and cuisines. This would help a user when he is out of town or in other location and wants to know the best restaurant of his favorite cuisine in that locality.

Project Title

Restaurant Predictor.

Base of Interest and Significance

The user may want to explore different areas, cities or neighborhoods, providing certain specific cuisines. Our suggestion model pin-points exact locations based on above mentioned parameters. The significance is on the psychology of the user. He can provide his new location and based on his previous reviews, ratings, etc. he can get the suggestion of the restaurant best suited for him. We would provide the top 5 restaurants best suited with his liking.

Data Mining Techniques

We retrieve all the information about the user like his check-ins, reviews weather good or bad, ratings provided by him. Then we take the present location i.e. the location for which he wants restaurants of his choice to be displayed and apply a K Nearest Neighbor classification technique. The classification technique is applied to get the top restaurants in the new location which are closest to his liking.

Final Deliverables

Final deliverable includes a website where given the name of the user we retrieve all his reviews, ratings, check-ins so we know his liking and interest. We take the new location and mine the available data by applying K Nearest Neighbor classification to display top five restaurants in his give locality. A map of his region is displayed with pins on top five restaurants. The demo would include the display of the top 5 restaurants based on the previous reviews of the user.

Challenge areas

The challenge areas in this project would be the consideration of features while applying classification. Also the order of preference of features while finding the K Nearest Neighbors. We also need to care of the execution time as the data set being huge the algorithm should be efficient, as user should not wait to get the results.

Addressing Challenges

As it is a location based classification we would first give preference to his check-ins feature later his reviews and ratings while applying classification. To work on the execution time we would use efficient data structures like dictionaries where we can achieve O(1) operations.

Efficiency evaluation

Efficiency can be evaluated by using the data of a user who has visited two locations and we have his check-ins, reviews and ratings. We can use the features of one location to predict and verify if we get the same restaurants he has visited in other location.

Partition and Coordination of Tasks

We would partition the tasks such that the load is equalized on both of us. One would work on getting the features of the users, and other on the features of different location. We would work on developing the algorithm for K Nearest Neighbor classification and also the website development.