**YELP CHALLENGE: Predicting Health Score for Businesses**

**Stakeholders**: Yelp, Government

**What you plan to do?**

Evaluating restaurants is important to keep a check on how a restaurant is performing and what improvements might be needed. One of the standard measure used by US government to evaluate restaurants is Health Score. This project aims to predict health score of restaurants based on available features in yelp and US government health data. The extended goal (if time permits) is to design a solution to allocate limited government officials for health inspection of restaurants.

**Why it is important?**

1. For government -

* Prediction of a standard health score for each restaurant (based on available feature data) can help the government to decide the risk of a restaurant. It will ensure public health and safety.
* This can help them to allocate limited official resources for restaurant inspection efficiently. This project can provide an efficient strategy based on priority of inspection and distance for reaching out to inspect a restaurant at the much-needed time.

2. For yelp –

* They can collaborate with businesses to provide recommendations to improve their business and user needs. This can provide mutual benefit to both yelp and businesses.
* They can collaborate with government to provide health score and health analysis.

**What algorithms do you plan to employ?**

1. Supervised Machine learning: To predict health score (output class) based on input features (such as ratings, reviews, violations, and others). The data for features would be taken from the yelp dataset and online open source data provided by government health agencies.

2. Find similar items and users, recommendation systems: To fill missing entries in yelp data. Multiple restaurants can be compared to find similarity in restaurants. Multiple features can be compared to derive similarity in features.

3. Clustering techniques (such as k-means): [Extended goal if time permits] To cluster restaurants based on area to find which areas have good restaurants and which ones have poor restaurants. This can be useful for the government agencies or users when they want to see areas performing good vs bad.

**How you plan to evaluate?**

Precision and Recall: To create a good Health Score standard.

Root Mean Squared Error: Improving prediction model using validation and test data.

**Team**

Ajay Anand ([ajayanan@usc.edu](mailto:ajayanan@usc.edu)), Devershi Purohit ([dupurohi@usc.edu](mailto:dupurohi@usc.edu)), Rajdeep Kaur ([kaurr@usc.edu](mailto:kaurr@usc.edu)), Rimsha Goomer ([goomer@usc.edu](mailto:goomer@usc.edu))