

Khanh Ngo  
Thomas Mathew

## **Cuisine Palate Expertise Clustering**

### **Project Objective Overview:**

The objective of the Cuisine Palate Expertise Clustering project is to group a collection of people into experts for each different type of cuisine. Each human palate varies between person to person so there are instances where a rating at a restaurant could be unreliable. The motivation for this project is to determine whether a rating made by a person is trustworthy. To measure a person's trust or palate, we will determine an expertise weight that correlates to their related attributes. By using this type of clustering, users can feel more confident about a restaurant's rating made by others.

### **Data Mining Tasks:**

Our general algorithm for this project can be broken down into three steps. First we need to determine the level of expertise a user has with a particular type of cuisine. A formula will be developed in order to accurately weigh and calculate each user's affinity for a type of cuisine. Next we will create a vector which will contain the expertise level of each user where each dimension is type of cuisine. Finally we will cluster similar vectors to determine which users are experts for each cuisine.

### **Plan and Implementation:**

By the end of the semester, we plan to deliver a formula will determine the expertise weight, a program that can rate a person's restaurant rating through their expertise level, and a demo of our project. Our demo will be a functioning website that will use the Yelp datasets to display whether or not a user's rating is reliable. The website will contain visual representation of the theory behind our clustering, methods, and results.

### **Challenges:**

One of challenges we must overcome from this project is to formulate a formula that will represent the expertise weights of cuisines as accurately as possible. Another challenge is choosing a threshold in our vector space that will represent each partition space. To address these problems, we will seek guidance from the professor to ensure that our weight formula and threshold are properly constructed.

**Evaluation Methods:**

The method we will use to evaluate the efficacy of our formula is to check the consistency of the weight formula. We can check the divergence between old and new weights by testing with new similar attributes. This will ensure that nearest neighbors falls into the correct partition. The next method to evaluate our K-means value is to use the rule of thumb. The rule of thumb states that K is similar to the square root of N (the number of data points) divided by 2. By using the K value from the rule of thumb, we can check the new partition with our original K value partition to determine if they resemble one another.

**Partitioning:**

The tasks for this project will not be partitioned into parts because both members will be present when working on this assignment. This will include, developing, coding and presenting the project together.