Scrape 2000 reviews for Jean Georges

DF = (restaurant\_name, michelin\_stars, yelp\_stars, review\_text)

Tokenize sentences and words

Compare words to a food corpus?

Evaluate Sentiment

**Problem Definition**

The objective of the problem definition is to understand and clearly describe the problem that is being solved.

1. Problem Description
   * What is an informal description of the problem?
     1. How does one know what the best and worst items are at Michelin star rating?
   * Why does the problem need to be solved?
     1. So people can have an educated idea on what to eat at dope restaurants.
   * What assumptions do you have about the problem?
     1. Many people eat the same 10 items and those that venture from the standard items
     2. Many Michelin star restaurants have price-fix, rotating seasonal menus and few menu items.
   * How would you solve the problem?
     1. Ball out so hard
2. Data Description
   * How do you obtain the data required? Public dataset, scraping or other methods?
     1. Scraping
   * What constraints were imposed to select the data?
   * Define each attribute in the provided dataset.

**Data Analysis**

The objective of data analysis is to understand the information available that will be used to develop a model.

*Data Selection* - Consider what data is available, what data is missing and what data can be removed.

*Data Preprocessing* - Organize your selected data by formatting, cleaning and sampling from it.

1. What data types are the attributes?
2. Are there missing or corrupted values?
3. Review the distributions of the attributes, what do you notice?
4. Review the distributions of the class values, what do you notice?
5. Review the attribute distributions with class values in the histograms, what do you notice?
6. Review pairwise scatter plots of attributes, what do you notice?

**Data Preparation**

The objective of data preparation is to discover and expose the structure in the dataset.

*Data Transformation* - Transform preprocessed data by engineering features using scaling, attribute decomposition and attribute aggregation

1. Normalize the dataset
2. Standardize the dataset
3. Square the dataset
4. Discretize attributes (if integer)
5. Remove and/or replace missing values (if present)
6. Create transforms of the dataset to test assumptions raised in the Problem Definition

**Evaluate Algorithms**

The objective of evaluating algorithms is to develop a test harness and baseline accuracy from which to improve.

1. Explore different classification algorithms
2. Design and run a spot­check experiment
3. Review and interpret the algorithm rankings
4. Review and interpret the algorithm accuracy
5. Repeat process as needed