

# krusty krab and big data

MHD Khaled Maen, Asem Hamood Al-Abdali, Mohamed Alrefaei,  
Abdirahman Abdullahi  
mk.maen93@gmail.com, asimalabdaly@gmail.com, mhrifay@gmail.com,  
Abdomursal@gmail.com

**Abstract—** Restaurants are on the top of business ideas it falls into the hospitality business category; a restaurant business plan varies from person to person and different types of restaurant are always competing despite what kind of food the restaurant offers but not all the restaurant succeeds especially in their first year. A research group in Dallas city (which consider one of the hottest restaurant marketplaces) determine the rate of startup restaurant first-year failure which is a 23%" failure rate at their first year. Some of the reasons that the restaurant fails is hard competition especially if the location of the new restaurant is loaded with a good restaurant reputation and the lack of a new idea. A good pre-planning strategy and some data and financial analysis help to improve the chances of success. In this project we try to understand and analyze the data and the best Circumstances for a restaurant to success and what are the most favourite food cuisines to the students inside the campus along with the main factors for a restaurant to work from the location and dishes point view.

**Abstract—** big data analytics, restaurants, food choice.

## I. Introduction

Opening a restaurant is a challenging process especially when the target has limited income such as student in this case. Nowadays restaurant businesses are increasing in all around the world. It became a traveling destination where people travel the world sometimes just to try a specific type of food. A study in 2012 showed the convenient access to Halal food, from fine dining to street food, helps Malaysia to tap into the huge Muslims tourist market estimated at US\$126.1 billion a year. Some studies stated that have found one in two millennials consider themselves foodies where also major countries and cities branding itself as a food destination. However, not all the restaurants are crowded with customers or have a good reputation. In fact, a lot of restaurants fails within a few months of opening. restaurants must to have a good pre-planning strategy and some financial analysis where it helps to improve the chances of success. This project aims to understand the requirements of establishing a new restaurant and the best way to start a new food business. knowing what the most ordered dishes to the students inside campus. Moreover, to analyses the collected data to get the main factors for a restaurant to success from the location and dishes point of view [1].

## II. Literature Review

Recently, the main target of marketing shifted from focusing on the brand to customer. In this most of

enterprises put their effort to find the factors that satisfy their customers [2]. In business point of view customer satisfaction is measured by how does a product meet customer expectations [3] . So many researches had been conducted to apply big data in real life restaurants.

Starting with Tama; his study was to identify the factors of customer satisfaction in fast-food restaurant. The data being analyzed by using data mining methods with two machine learning algorithms decision tree and neural network in order to generate rules for managerial and practical implementations in fast-food industry. The proposed method yields more than 80% of predictive accuracy [4].

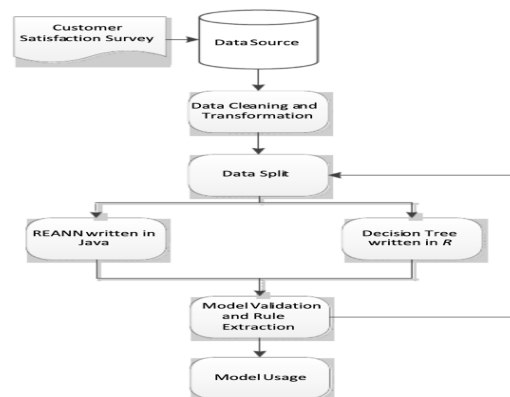


Fig. 1 Diagram of Research Process

Another group were focusing on the student eating habits and the factors that affect it. The study was conducted on 291 university students (141 females, 150 males). The research concentrates on the emotional state of students by identifying life satisfaction scale, multidimensional scale of perceived social support Scale and social emotional loneliness Scale. The algorithms used in this research are Correlations and logistic regression for analyzing the data. Results obtained from logistic regression according to the life satisfaction scale ( $p < .01$ ) being in high or low group based on the eating attitude test scores, there is a significant relationship between the lower dimension of “social support from a friend ( $p < .05$ ) or a special one ( $p < .05$ )” of multidimensional scale of Perceived social support Scale and the lower dimension of “emotional loneliness in romantic relationships” ( $p < .05$ ) of social emotional loneliness scale [5].

### III. Methodology

The main structure of our project is knowing students' food preferences and how they will in the situation of opening a new restaurant. That's why we search for a suitable dataset for our projects.

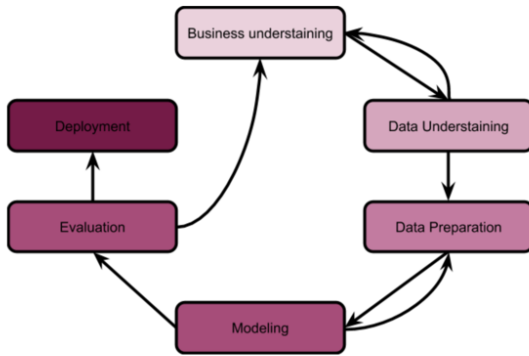


Fig. 2 Data Analytics Lifecycle

#### A. Business understating

The main target of Krusty Krab is university students, so in order to make a success restaurant is to have a menu that satisfy the taste of majority students.

#### B. Data set

The data set is a collection of a 61 attributes include information on food choices, nutrition, preferences, childhood favorites, and other information from college students (GPA, Gender, breakfast, calories chicken, calories day, calories scone, coffee, comfort food, comfort food reasons, comfort food reasons coded, cook, comfort food reasons coded, cuisine, diet

current, diet current coded, drink, eating changes, eating changes coded, eating changes coded1, eating out, employment, ethnic food, exercise, father education, father profession, fav cuisine, fav cuisine coded, fav food, food childhood, fries, fruit day, grade level, Greek food, healthy feeling, healthy meal, ideal diet, ideal diet coded, income, Indian food, Italian food, life rewarding, marital status, meals dinner friend, mother education, mother profession, nutritional check, on off campus, parents cook, pay meal out, Persian food, self-perception weight, soup, sports, Thai food, tortilla calories, turkey calories, type sports, veggies day, vitamins, waffle calories, weight)[6].

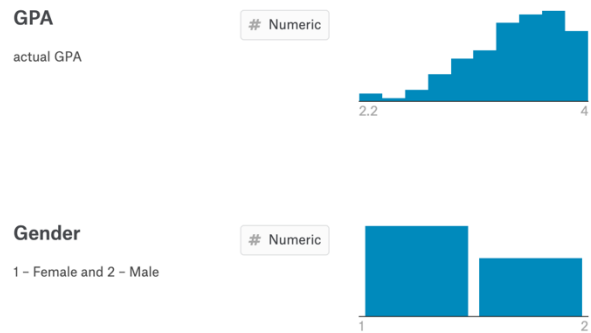


Fig. 3 Visualization of some data set attributes

### C. Data preprocessing

In order to use the data set many operation has to be done, starting with data encoding by transfer nominal variables into numerical variables then filling the null column values with proper value in our case we filled the null values with the median of the column that belongs to, after that non related column should be dropped which was 13 columns in this case.

### D. Algorithm

Since our problem is unsupervised machine learning clustering K-Means tend to be suitable for the task. K-Means is one of the most popular clustering algorithms, it stores K centroids used to define the clusters [7]. A point is considered to be in a particular cluster if it is closer to that cluster's centroid than any other centroid. K-Means finds the best centroids by alternating between (1) assigning data points to clusters based on the current centroids (2) choosing centroids (points which are the centre of a cluster) based on the current assignment of data points to clusters and that can be done after many iterations as shown in Fig.2.

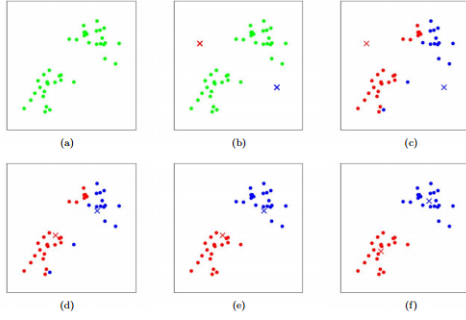


Fig. 4 K-Means Visualization

## E. Evaluation

Unlike supervised learning where there is a ground truth for model performance evaluation, clustering algorithms including K-Means don't have a solid evaluation metric that can be used to evaluate the model outcome. However, K-Means require  $k$  as input and there is no right answer in terms of clusters' number, some methods are useful to give an intuition about  $k$  [8]:

### 1. Elbow Method

Elbow method helps to find out what a good  $k$  number of clusters based on the sum of squared distance (SSE) as shown in Fig.4.

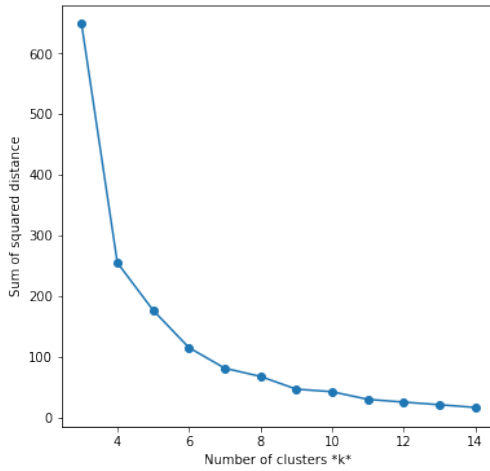


Fig. 5 Elbow Method

In this case the target is to have a big cluster of students so the selected value of  $k$  is 3, it might not be the optimal value, but it should be acceptable for this kind of tasks.

## 2. Silhouette Analysis

Silhouette analysis can be used to determine the degree of separation between clusters. For each sample it computes (1) the average distance from all data point in the same cluster. (2) the average distance from all data points in the closest cluster. (3) the coefficient.

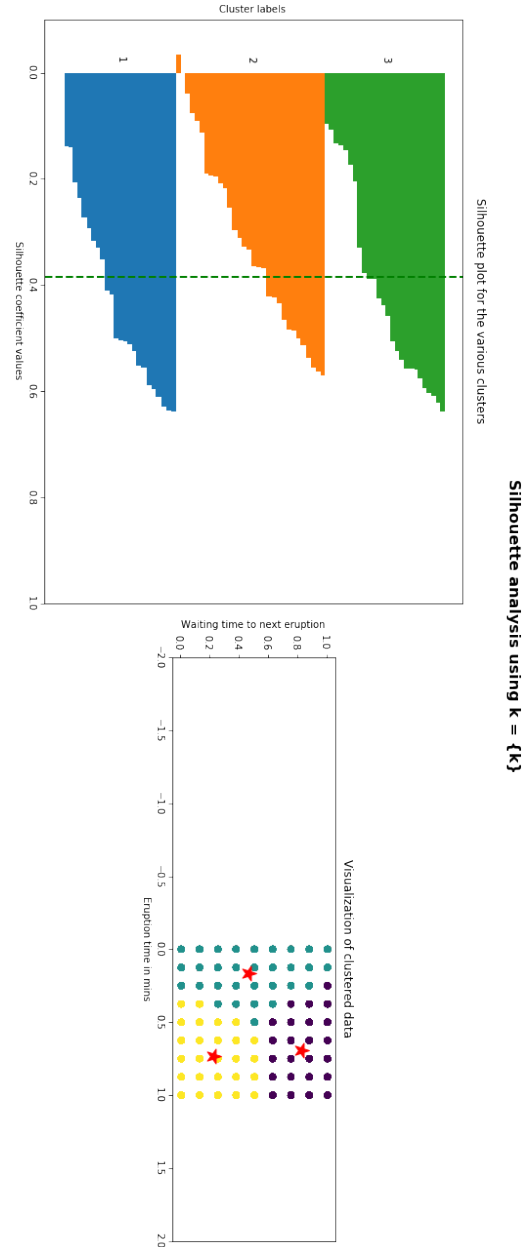


Fig. 6 Silhouette Analysis

## IV. Conclusion

After 10000 iterations with  $k=3$ , K-Means algorithm revealed that the students who are clustered in c3 are 50% of the whole data, in that Krusty krab will target this group and try to add additional meals to approach the one of the remaining groups or both of them.

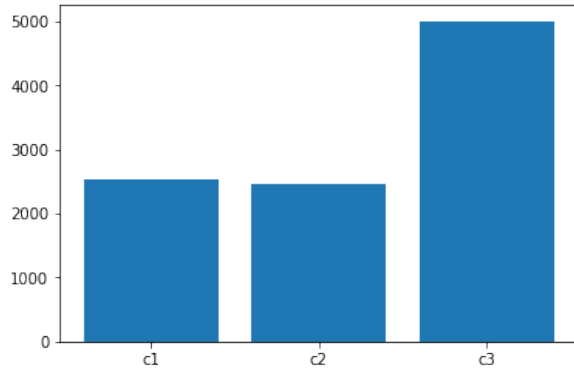


Fig. 7 K-Means outcome

## Acknowledgment

This research was carried out under the course Data Science (CSC 3305), a course under the Department of Computer Science, Kulliyyah of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur. The authors would like to thank Dr. Raini Binti Hassan and the team involved in guiding and facilitating the process in making this paper a success...

## References

- [1] "Global Report on Food Tourism."
- [2] S. Saad Andaleeb and C. Conway, "Customer satisfaction in the restaurant industry: An examination of the transaction-specific model," *J. Serv. Mark.*, 2006.
- [3] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Systems with Applications*. 2009.
- [4] B. A. Tama, "Data mining for predicting customer satisfaction in fast-food restaurant," *J. Theor. Appl. Inf. Technol.*, 2015.
- [5] A. Çekiç, E. Özkamali, and A. Buğa, "The

Analysis of Variables Predicting Eating Habits of University Students," *Procedia - Soc. Behav. Sci.*, 2014.

- [6] "Food choices | Kaggle." [Online]. Available: <https://www.kaggle.com/borapajo/food-choices>.
- [7] "K-Means." [Online]. Available: <http://stanford.edu/~cpiech/cs221/handouts/kmeans.html>.
- [8] "K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks." [Online]. Available: <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a>.