Case study data mining applied in data analysis as a marketing strategy for vegan and vegetarian restaurant

Valeria de Jesús Castañeda García [a], *Karina Raya Díaz* [b], *Graciela Félix López* [c]

[a] Universidad de Tijuana, castaneda.vcg@gmail.com, México.
[b*] Universidad de Tijuana, k.raya@udetijuana.edu.mx, México.
[c] Universidad de Tijuana, ingenierias@udetijuana.edu.mx, México.

Keywords: *Clustering, data mining, K-Means, marketing.*

Abstract:

Marketing identifies unrealized needs and desires; define, measure and quantify the size of the identified market and the potential profit; For this, it is essential to capture, store and analyze consumer information. Big data can be analyzed to obtain information that in the long term leads to better business decisions and strategies, but it could improve customer relationships and optimizing the operation of the company?

The objective of this research is to apply a data mining technique that allows a large volume of data to be analyzed in detail, with the purpose of extracting the most relevant information to implement a digital marketing strategy to retain customers by generating clusters and the classification of them, to identify the attributes that allow developing a marketing strategy appropriate to customer retention. The technique to be applied is the K-Means in which it is an unsupervised algorithm, when applied to a set of data that show the attributes that define the consumer profile, identifying the centers obtained will generate a modeling of the customer profile with the purpose of implementing a more precise marketing strategy for customer retention for vegan restaurants. The application of data analysis tools is important for PyMES because it helps determine customer profiles to improve marketing strategy. The importance of developing the project lies in supporting the learning of new trends in marketing tools through the analysis of big data using grouping techniques based on data mining, thus encouraging its introduction in PyMES. Likewise, the project will contribute to support the themes developed, which will allow a more detailed analysis of customer profiles that cannot be identified through traditional statistical analysis.

The methodology used was knowledge discovery databases known by its acronym in English as KDD, which considers the six procedures that are listed below:
1. Analysis of a data set of vegetarian and vegan restaurants.
2. Data selection: In this stage, relevant data are selected for the analysis of the restaurants.
3. Preprocessing: It consists of the preparation and cleaning of the extracted data. data cleansing for the database.
4. Transformation: It consists of the preliminary treatment of the data, transformation and generation of new variables from the existing ones. Aggregation or normalization operations are performed.
5. Data mining: It is the modeling phase where intelligent methods are applied in order to extract previously unknown, potentially useful and understandable patterns hidden in the data.
6. Interpretation and evaluation: The patterns obtained are identified and an evaluation of the results obtained is carried out.

Basic clustering algorithm of k-means is defined as follows:
Step 1: choose the number of clusters k.
Step 2: Make an initial selection of k centroids.
Step 3: Assign each data item in s to its closest centroid (in this way, the k groups form one for each centroid, where each group consists of all the data items assigned to that centroid).
Step 4: For each group, make a new selection from its centroid.
Step 5: Go back to step 3, repeating the process until the centroids do not change (or some other convergence criterion is met).

In this case study, two scenarios will be analyzed which consist of the analysis of a data set of vegan and vegetarian restaurants, which will be classified considering the number of followers and likes of your Facebook page. The algorithm applied was K-means, which classifies each of the data records. As results, six groups of twelve restaurants evaluated were obtained for the first vegan restaurants scenario, and for the second scenario were classified eight clusters to vegetarian restaurants.