Cities and Restaurants

Ahmet Oruc
oruc.ahm@gmail.com

Sergen Topcu
sergentopcu08@gmail.com

Denizkaan Araci
denizkaanaraci@gmail.com

Machine Learning Project, Department of Computer Engineering
Hacettepe University, Ankara, TURKEY

Abstract

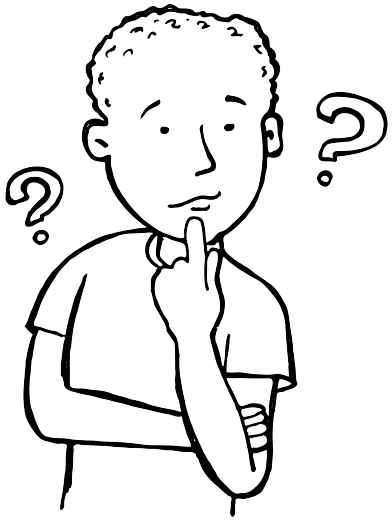
In this paper, we offer a method for restaurants that customers like, and in doing so, they use the points given by the customers to the restaurants they have visited before. With this information, restaurant recommendations will be made by establishing similarities with the features of the restaurants. To take this guess, we took into account the customer's point of view on restaurants and the properties of restaurants. For this, we try to use machine learning algorithms and use the best algorithm.

1. Introduction

People often prefer to eat outside for reasons such as having a good time, relaxing, socializing. Choosing the right restaurant has an important place in terms of the beautiful part of the day. With time being limited and valuable, this method will make it possible to make faster and more accurate choices.

There are many studies and projects related to this problem. Some of them are close to the restaurant, some of them have a taste of food and some of them are calculated according to other criteria and they are in restaurant suggestions. The work on this topic is ongoing. At the same time, new machine learning projects and new methods are being identified and helping to develop on this topic.

There are many features that will make restaurant preferred by customers. For example food tastes, proximity, the speed of service, alcohol or smoking, car parking. These features are some of the features that customers consider when choosing a restaurant. In our project called "Cities and Restaurants," we want to develop a method that estimates the most suitable restaurants that customers would like to take into consideration.



Our first goal when developing this system is to suggest locations where customers will appreciate. In addition, restaurants are expected to improve their services and add new features in order to get higher scores.

2. Related Work

A lot of research has been done on the recommendation project we are working on. In these researchers, it is emphasized firstly which algorithms should be worked in an effective and harmonious manner. Studies on the use of algorithms in [1] and[9] sites have been done. In these studies, the two most important approaches are content-based filtering and collaborative filtering. We have been working on [7] and [9] for k-Nearest Neighbor, which we decided to use. Weighted Naive Bayes Algorithm [8] and [6] have been studied.

When we review the "Yelp" dataset we have decided to use, we see the design work of the students working on this dataset and Yelp[4]. Approaches such as Clustering,

Neighborhood model, Combined Weighted Mean have been used in Business and User ratings are known, Business or User ratings are known and business and user ratings are unknown studies.

We could not find dataset because it was very difficult to collect data on this topic. Apart from the yelp dataset "Yelp" dataset we used, we could not find such a big dataset on the Internet. However, you can find a small dataset like [5] that is used in a similar Kaggle-based competition (Restaurant Data with Consumer Ratings).

3. Methodology

The machine learning algorithms we have decided to use in this project are weighted k-Nearest Neighbor and Naive Bayes algorithms. In this algorithm, we will first look at the customer's restaurant reviews and the properties of the restaurant that they went to. Later, we will try to make an estimate by comparing the information we gathered from the customers we want to present the restaurant with the reviews made by other customers. We will look at attributes and features of restaurants when collecting this information.

If we look at Weighted k-Nearest Neighbor Algorithm;[2]

A refinement of the k-NN classification algorithm is to weigh the contribution of each of the k neighbors according to their distance to the query point Xq, giving greater weight Wi to closer neighbors. This can be accomplished by replacing the final line in the algorithm by

$$F(x_q) = \arg\max_{v \in V} \sum_{i=1}^{k} w_i \ \delta(v, f(x_i))$$

where the weight is

$$w_i = \frac{1}{d(x_q, x_i)^2}$$

(in case Xq exactly matches one of xi, so that the denominator becomes zero, we assign F(Xq) to be f(Xi) in this case.

For the version of k-NN for real-valued output the final line of the algorithm will be:

$$F(x_q) = \frac{\sum_{i=1}^{k} w_i f(x_i))}{\sum_{i=1}^{k} w_i}$$

If we look at Naive Bayes Algorithm;[3]

In machine learning, naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

Given:

- Class prior P(Y)
- d conditionally independent features X1,Xd given the class label Y
- For each Xi feature, we have the conditional likelihood P(Xi—Y)

Naïve Bayes Decision rule:

$$f_{NB}(\mathbf{x}) = \arg\max_y \ P(x_1, \ldots, x_d \mid y) P(y)$$
$$= \arg\max_y \prod_{i=1}^{d} P(x_i|y) P(y)$$

4. Experimental Setup

In this work, we use a limited version of Yelp[5]'s restaurant-based service. In the original case of the data set, the characteristics of the companies and the comments and scorecards of users who have received service from these places are kept. We will use the information of 51,613 restaurants that are in the category of this data set. We will remove the data with missing information to prevent noise from this data. We will also convert the ratings for these restaurants into the appropriate data format for their users.

Binary classification algorithms such as weighted k-NN, logistic regression and Naive Bayes are planned to be used. These algorithms will use the features that restaurants have as features. These features have been used indefinitely in the lobe and we have identified 10 features that are often found within them. We will transform our data set into one form in the direction of these properties. We will use the information of the restaurant which has at least 50% of the properties we have determined. Missing features of restaurants that we have specified will be considered false. We will also split the scores of users and reviews into features.

5. Future Work

Until now, we have decided in our project that we will use the "Cities and Restaurants" restaurant forecasting algorithm. We did research for this and decide to use logistic regression, k-NN, and naive-Bayes algorithms. In the next step, we will code the algorithms we have decided to use and calculate accuracy. According to this accuracy, we will observe and choose which algorithm will give higher accuracy. After this section, we will study how we can increase this accuracy. In summary, we have completed the planning phase of the project and are moving on to the implementation phase.

References

[1] Big Data Behind Recommender Systems. https://indatalabs.com/blog/data-science/big-data-behind-recommender-systems#XTewQDKDdLvqIqlZ.99. Author: Valeryia Shchutskaya.

[2] Distance weighted k-NN algorithm. http://www.data-machine.com/nmtutorial/distanceweightedknnalgorithm.htm.

[3] Hacettepe University BBM406:Fundamentals of Machine Learning - Naive Bayes Classifier. https://web.cs.hacettepe.edu.tr/~aykut/classes/fall2017/bbm406/slides/l8-bayes-classifier.pdf. page:20.

[4] Recommender Systems Designed for Yelp.com. https://www.math.uci.edu/icamp/summer/research/student_research/recommender_systems_slides.pdf. Authors: Naomi Carrillo,Idan Elmaleh,Rheanna Gallego,Zack Kloock,Irene Ng,Jocelyne Perez,Michael Schwinger and Ryan Shiroma.

[5] Yelp Dataset. https://www.yelp.com/dataset.

[6] M. Ghazanfar and A. Prugel-Bennett. An improved switching hybrid recommender system using naive bayes classi? er and collaborative filtering. 2010.

[7] M. Jahrer, A. Töscher, and R. Legenstein. Combining predictions for accurate recommender systems. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 693–702. ACM, 2010.

[8] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

[9] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.