How Manhattan and Toronto neighborhoods are different?

Presented by Artem Ramus

Background

Neighborhoods in cities can be characterized in different ways. In capstone project of IBM Coursera Data Science Certification, New York Manhattan and Toronto neighborhoods were classified by venues using clustering model k-means.

Introduction

There are distinguishable similarities and differences between neighborhoods in Manhattan and Toronto based on various venues that are located in the neighborhoods. But, how Manhattan and Toronto neighborhoods are different?

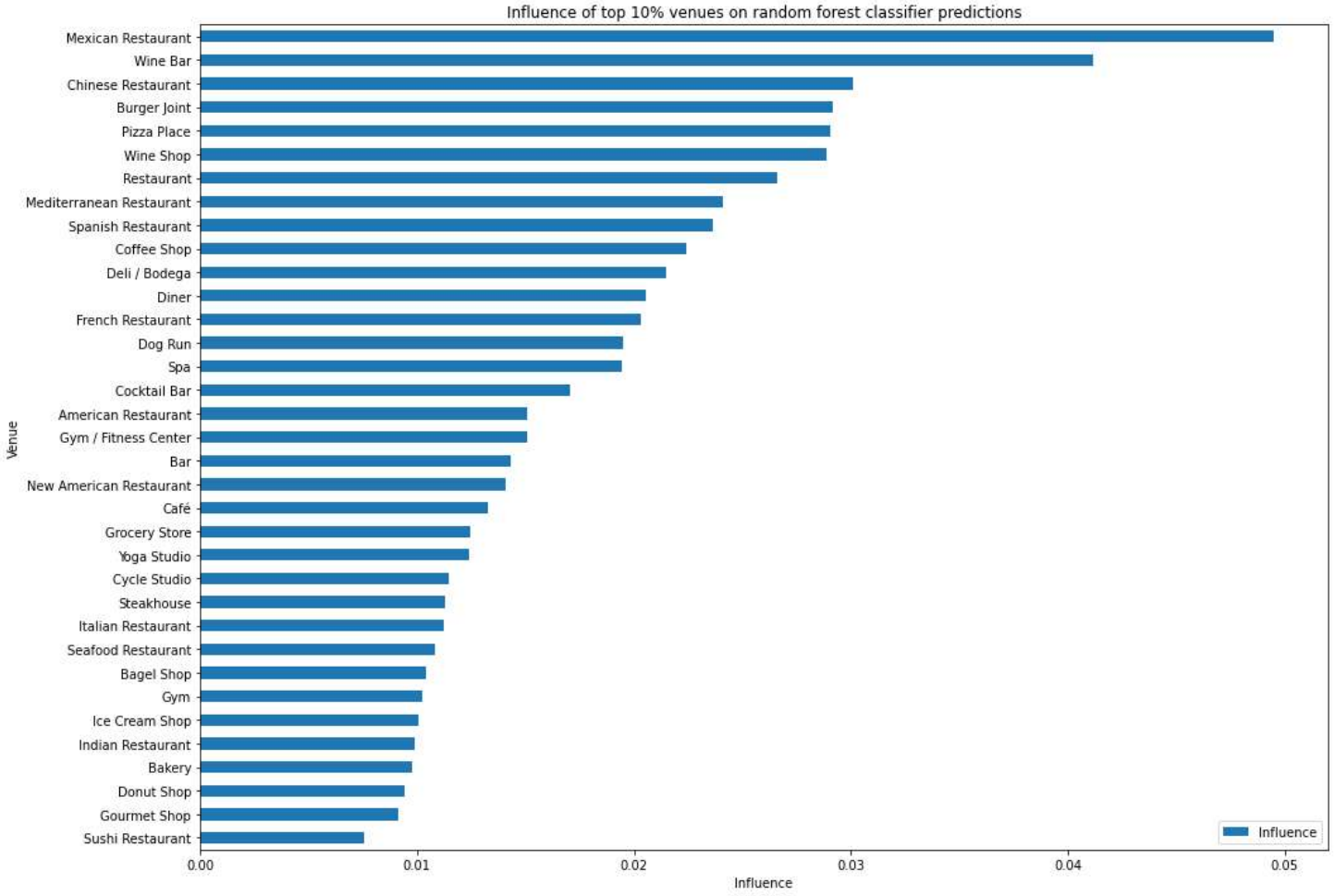
Methodology

- The data set built from two parts, "Manhattan venues frequency" and "Toronto venues frequency".
- The data set is checked for duplicates, null values and homogeneous features.
- Train-test split, normalizing the model with standard scaler and training of KNN, logistic regression, decision tree and random fores models.
- 5-fold cross validation score is calculated for all the models.

Feature selection

The most influencing feature between the top 10% influencers is that 31 of them, this is 90%, are related to food.

There is restaurants, bars, cafes, coffee shops, and other food-related stores. The rest 10% of the top 10% influences are related to sports, and they are mostly studios.



Performance of the Model

- Performance of the four models was estimated based on the mean value of Jaccard score of 5 cross validation.
- The random forest model made the highest score prediction of 87%.

Summary and Conclusions

- It is interesting to mention that most prominent features that differentiate between Manhattan and Toronto neighborhoods are all about the food and sports-related activities.

- This might indicate that each city has it’s own food habits and preferences and favorite sports activities.

The end

Thank you for your attention!