FROM LONDON TO TOKYO:

Olympics Games Impact on Retail Businesses Using Classification Methods

Daniel Bi, Karen Zheng, Leni Wen, Bill Zhuo

Highlights

- Adopted large amount of realistic retail data with insights on business success
- Utilized CatBoost classifier to reach high auc of 0.897 while having fast training time
- Provided prediction for future Olympics Games

Problem

We aim to answer the question about what are the determinants of success of a retail business in the host city of the Olympics Games.

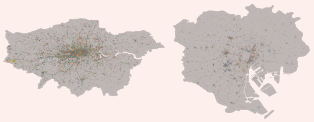
Solution

We use London Olympics data to understand the dominant features in retail business success and apply our model to predict Tokyo Olympics' corresponding outcomes.

Data

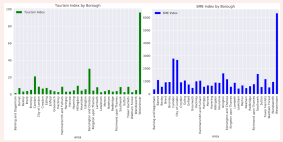
These data are processed to provide spatialfeatures for the classification problem. The following are venue distributions in London and Tokyo.

- Foursquare Global Check-in Dataset
- 2,000,000+ Check-in Data
- 100,000 + Retail Venue Data



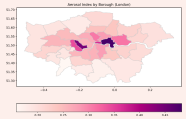
These data can provide additionaldimensions to the success of a retail business.

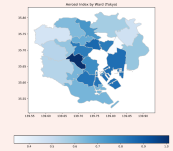
- Borough/ Ward Data
- Business count (SME Index)
- Tourism popularity data (Tourism Index)



Environmental friendliness of businesses are considered in the modeling process.

- NASA Remote Sensing Data
- Aerosol data





Model

After testing 7 classification models, we eventually choose to use CatBoost to train our selected features in the London Olympics setting to understand important features and proceed to prediction under Tokyo Olympics setting.

Training Features

- Stadium Distance
- Accessibility Measure
- SME Index
- Tourism Index
- Aerosol

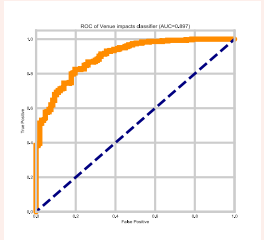
CatBoost Model

CatBoost is a gradient boosting based classification algorithm which is more robust and solves the bias problem in gradient boosting

Feature Importance & Probability Score

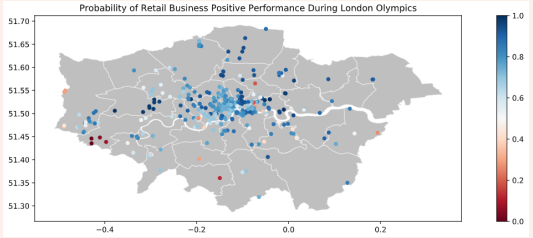
Classification Results LONDON 2012

Model Accuracy



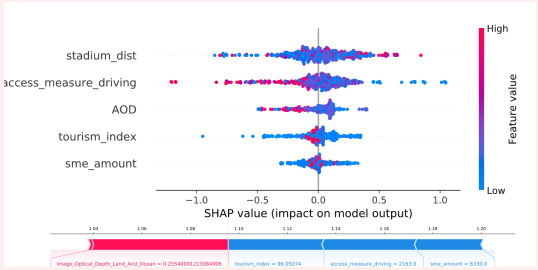
AUC = 89.7 %

Validation Summary



We note that the spatial features provide clear indications towards the success of a retail business.

Feature importance Summary



Red means more negative influence on model output, and blue means more positive influence on model output. Thus, AOD plays more negative influence on model output,while others play more positive role.

Prediction Results TOKYO 2021