ADS2015
Ding Ma, Yuqiao Cen, Yuxiang Zhang, Zeyu Jiang
Pro Tim Savage
Project Proposal

Predicting Rossmann Store Sales

The Restaurant Revenue Prediction was one of the ongoing competitions on Kaggle recently. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. The datasets include historical sales data for 1,115 Rossmann stores. The dataset has much interesting data fields that needs study before we do analysis and predictions. In our project, we mainly focus on using different approaches to predict as the question requires. Before fitting the model, we would like to find correlations between data. 1) Does store sales have a strong correlation with customer numbers? 2) Whether a competitor within little distances indicating a higher sales or lower sales? And in which level of significance does it implies? 3) What's the influence of a competitor opens? 4) Is there a seasonal behavior in the sales data?

With answering these questions, we may have a better understanding of the datasets, and would mainly use two approaches to predict store sales, one of statistical learning, like linear regression, and the method using machine learning methods like random forest. As the team member, I will work with Ding Ma and try to write the code for predicting store sales, using both statistical learning and machine learning methods. Yuqiao Cen will analyze the seasonal and custom effect on the sales and Zeyu Jiang will find out the influence of competitors on the shop.

Kaggle: Rossmann Store Sales: https://www.kaggle.com/c/rossmann-store-sales/data

Random Forest https://en.wikipedia.org/wiki/Random_forest

Leo Breiman. 2001. Random Forests. *Mach. Learn.* 45, 1 (October 2001), 5-32. DOI=http://dx.doi.org/10.1023/A:1010933404324