

GS-DA 1001 Introduction to Data Science

Group Project

Team Member: Weixi Lei(wl1162), Tian Wang(tw991), Mengfei Li(ml4713), Lei Lu(ll2850)

10/15/2014

<https://www.kaggle.com/c/bike-sharing-demand/data>

<https://inclass.kaggle.com/c/predicting-cab-booking-cancellations/data>

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

<http://archive.ics.uci.edu/ml/machine-learning-databases/kddcup98-mld/kddcup98.html>

There are four rough ideas for the project:

(1) Bike Sharing Demand

We are given the training and testing data separately regarding to hourly rental data spanning two years. The task is to predict the total count of bikes rented during each hour and the accuracy is evaluated by the given testing data.

(2) Cab Booking Cancellations Prediction

The dataset comes from YourCabs.com. We are trying to utilize the dataset to build a predictive model for the cab booking cancellations due to unavailability of a car.

(3) Bank Marketing

The data is related with direct marketing campaigns of a Portuguese banking institution. This task is to detect whether a bank product would be subscribed.

(4) KDD Cup

This is the data set used for The Second International Knowledge Discovery and Data Mining Tools Competition. The competition task is a regression problem where the goal is to estimate the return from a direct mailing in order to maximize donation profits.