

Data Science Application – Project 2



Introduction

This project is about the sales of scooter rental. The data include weather and usage pattern features and each observations are recorded every 1 hour. Your task is to forecast the renting demand of scooter by historical data.

File descriptions

train.csv - the training dataset (*Label = count*)

test.csv - the testing dataset (*without label*)

example.csv - a sample submission file in the correct format

feature.pdf – the descriptions for each feature

Performance metric

Rooted mean squared logarithm error (RMSLE)

$$RMLSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [\ln(\hat{y} + 1) - \ln(y + 1)]^2}$$

(n is the total data size, \hat{y} is your prediction and y is the actual label for i)

Grading policy

Professor review: 50%

Peer review: 25%

Accuracy: 25%

Schedule

05/17 Deadline of submission (Before 5:00 p.m.)

05/18 Presentation