Minh Hieu Do – 288414

THE DATA MINING PROJECT

Bike Sharing Demand

Data Description:

The hourly rental data spanning two years is provided.

The training set is comprised of the first 19 days of each month, while the test set is the 20th to the end of the month.

Data Fields:

datetime - hourly date + timestamp

season - 1 = spring, 2 = summer, 3 = fall, 4 = winter

holiday - whether the day is considered a holiday

workingday - whether the day is neither a weekend nor holiday

weather –

1: Clear, Few clouds, Partly cloudy, Partly cloudy

2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

temp - temperature in Celsius

atemp - "feels like" temperature in Celsius

humidity - relative humidity

windspeed - wind speed

casual - number of non-registered user rentals initiated

registered - number of registered user rentals initiated

count - number of total rentals

Link to data: https://www.kaggle.com/c/bike-sharing-demand/data

Features are created separately for casual and registered users.

Used Methods:

Linear Regression Model

Regularization Model – Ridge

Regularization Model – Lasso

Ensemble Models - Random Forest

Ensemble Model - Gradient Boost

Evaluation Method: RMSLE Scorer

As can be seen, the RMSLE of **Gradient Boosting** is the minimal value among five models which means it performs better than the others.