

## Mini-Project 1: Seoul Bike Sharing Demand Forecasting

Download Data:

[https://www.dropbox.com/s/msdt8r9c7f1bi7a/SeoulBikeData\\_MP1.csv?dl=0](https://www.dropbox.com/s/msdt8r9c7f1bi7a/SeoulBikeData_MP1.csv?dl=0)

### Data Description

Abstract: The dataset contains count of public bikes rented at each hour in Seoul Bike Sharing System with the corresponding Weather data and Holidays information for 365 days.

Number of Instances: 8760

Number of Attributes: 14

Data Source :<http://data.seoul.go.kr/>

SOUTH KOREA PUBLIC HOLIDAYS. URL: [publicholidays.go.kr](http://publicholidays.go.kr)

**Note:** The dataset for the Mini-Project has been modified from the above source. Use only the data provided for the mini-project.

### Data Set Information:

Currently Rental bikes are introduced in many urban cities for the enhancement of mobility comfort. It is important to make the rental bike available and accessible to the public at the right time as it lessens the waiting time. Eventually, providing the city with a stable supply of rental bikes becomes a major concern. The crucial part is the prediction of bike count required at each hour for the stable supply of rental bikes.

The dataset contains weather information (Temperature, Humidity, Windspeed, Visibility, Dewpoint, Solar radiation, Snowfall, Rainfall), the number of bikes rented per hour and date information.

### Attribute Information:

Date: year-month-day

Rented Bike count - Count of bikes rented at each hour

Hour - Hour of the day

Temperature-Temperature in Celsius

Humidity - %

Windspeed - m/s

Visibility - 10m

Dew point temperature - Celsius

Solar radiation - MJ/m<sup>2</sup>

Rainfall - mm

Snowfall - cm

Seasons - Winter, Spring, Summer, Autumn

Holiday - Holiday/No holiday

Functional Day – NoFunc (Non Functional Hours), Fun (Functional hours) of the bike sharing

#### Relevant Papers:

[1] Sathishkumar V E, Jangwoo Park, and Yongyun Cho. 'Using data mining techniques for bike sharing demand prediction in metropolitan city.' Computer Communications, Vol.153, pp.353-366, March, 2020

[2] Sathishkumar V E and Yongyun Cho. 'A rule-based model for Seoul Bike sharing demand prediction using weather data' European Journal of Remote Sensing, pp. 1-18, Feb, 2020

#### Mini-Project Tasks

Overall: Develop a good model for rented bike demand forecasting.

#### Sub-tasks:

1. Handle missing data in Temperature by different techniques and compare accuracy of model
  - a. Simple Imputation
  - b. Regression with other weather variables
  - c. Regression with time
  - d. Nearest Neighbor
  - e. Any other technique
2. Handle hour of day categorical variable with cyclical feature handling techniques (e.g., do feature transformation with trigonometric functions)
3. Build one model for all seasons and build separate models for different seasons. Is the overall accuracy improving by this procedure?
4. Explore whether precipitation (rainfall or snowfall  $> 0$ ) is a strong predictor for rental. Build a separate model for precipitation and no precipitation. Compare and report advantages/disadvantages.
5. Do more exploration and report any strong predictors and if building a separate model is useful.

You must do the above sub-tasks or give justification on why you did not perform some. You are free to do any other sub-tasks to make the mini-project more interesting.

#### Mini-Project Evaluation

A 10 min presentation by your team (will be scheduled around the due date), including Q/A.