**Machine Learning Essentials: Lab 4**

**Introduction to regression models**

**1. Rental Bike Share Prediction**

1.1 Problem Statement

Bike sharing systems are a new generation of traditional bike rentals where the whole process from membership, rental and return back has become automatic. Through these systems, users are able to easily rent a bike from a particular position and return back at another position. Currently, there are about over 500 bike-sharing programs around the world which is composed of over 500 thousand bicycles. Today, there exists great interest in these systems due to their important role in traffic, environmental and health issues. Apart from interesting real-world applications of bike sharing systems, the characteristics of data being generated by these systems make them attractive for the research.

The goal here is to build an end-to-end regression task. Here the user will provide the data and the result will be given by the best performing hyper tuned Machine Learning model. The user will also get privileges to choose the deployment options.

1.2 Dataset description

Both hour.csv and day.csv have the following fields, except hr which is not available in day.csv

- instant: record index

- dteday : date

- season : season (1:winter, 2:spring, 3:summer, 4:fall)

- yr : year (0: 2011, 1:2012)

- mnth : month ( 1 to 12)

- hr : hour (0 to 23)

- holiday : weather day is holiday or not (extracted from [[Web Link]](http://dchr.dc.gov/page/holiday-schedule))

- weekday : day of the week

- workingday : if day is neither weekend nor holiday is 1, otherwise is 0.

+ weathersit :

- 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 : Normalized temperature in Celsius. The values are derived via (t-t\_min)/(t\_max-t\_min), t\_min=-8, t\_max=+39 (only in hourly scale)

- atemp: Normalized feeling temperature in Celsius. The values are derived via (t-t\_min)/(t\_max-t\_min), t\_min=-16, t\_max=+50 (only in hourly scale)

- hum: Normalized humidity. The values are divided to 100 (max)

- windspeed: Normalized wind speed. The values are divided to 67 (max)

- casual: count of casual users

- registered: count of registered users

- cnt: count of total rental bikes including both casual and registered

1.3 Dataset link

<https://drive.google.com/file/d/1wo3EKBSQksYw6VDNZlLYR9NackKo5YAt/view?usp=share_link>

**2. Exercise**

2.1 Implement following models:

1. Simple Linear regression

2. Multi linear regression

3. Simple polynomial regression

4. Multi polynomial regression

5. Lasso regression

6. Ridge regression

7. Decision tree regressor

8. KNN regressor

2.2 Calculate MSE and R2 score and find the best model.