IE 7374 PROJECT PROPOSAL

Topic: Bike Sharing

Group 8

Atharva Pawankar

Paresh Choudhari

Rohan Kadam

Syed Hussain

PROBLEM SETTING AND DEFINITION:

Bike sharing systems are a new version of conventional bike rentals in which the entire process of joining, renting, and returning the bike is automated. These systems allow users to borrow a bike from a certain location and return it to a different location. There are currently around 500 bike-sharing schemes operating throughout the world, with over 500 thousand bicycles. Because of their importance in traffic, environmental, and health concerns, these systems are generating a lot of interest nowadays.

It's critical to make rental bikes available and accessible to the general public; otherwise, providing the city with a consistent supply of rental bikes will become a huge worry. The most important component is predicting the number of rental bikes that will be needed at each hour. Therefore, the objective of this project is to predict bike rental count hourly or daily based on the environmental and seasonal settings in year 2011-2012 using several Machine learning techniques and algorithms.

DATA SOURCES:

Link to the UCI repository:
https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset
The dataset contains 17389 instances and 16 attributes, as well as the hourly and daily counts of rental bikes in the Capital bikeshare system during 2011 and 2012, as well as weather and seasonal data.

DATA DESCRIPTION:

The following variables contain bike sharing attributes information:
- 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])
- 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

METHODS:

This project aims to train model and evaluate how the model performs on the given data using testing and training of variable proportions based on how we can build machine learning models. For this project we will build regression machine learning models, which will predict bike rental count hourly or daily based on the environmental and seasonal settings using several regression models.

We will implement the following regression techniques:

i) Linear Regression model
ii) Lasso Regression model
iii) Decision Tree model

However, we will compare the predictive performance of this model with the other models to find if the proposed model fits better than the rest or not.