WILD CODE SCHOOL

Bikes in demand

Project Week 04

Demand prediction is a very important task for Data Scientists. To dive into that subject, you will look at a dataset from Washington based “capital bikeshare”. The company provided two years of hourly rental data from 2011 and 2012 to Kaggle to design a challenge.

Objective

You are provided with hourly bike rental data for the first 19 days of each month. Your task is to design and tune a machine learning regression model to predict the demand for bike in the rest of the days of the months.

The Challenge

This project was part of a Kaggle Challenge.
You can find the test and train data set as well as descriptions of the features there.

1. Data Wrangling & Cleaning
3. Exploratory Data Analysis
4. Feature Engineering (including polynomial feature expansion)
5. Train-Validation-Split
6. Train the model
7. Validate the model (calculate the RMSLE for the training als well as the validation set)
8. Run your model on the test set and upload your results to Kaggle

Bonus: Regularize the model and test how you can avoid overfitting with it

Deliverables

A clean jupyter notebook with comments on the code and a clear structure.
It should present and describe each step of the process and contain a variety of plots for the data exploration and your findings as well as comparisons of the algorithms.

You upload your jupyter notebook to the student folder (till Friday 9am)
In the presentation on Thursday you have 5 min to talk about one of these:

- one challenge in the project and how you solved it
- one ressource you found really helpful
- a python command or library that helped you a lot

Good Luck
with your work!