

Bike Sharing Prediction Project Description:

Predict the total number of bicycle users on an hourly basis.

<https://www.kaggle.com/marklvl/bike-sharing-dataset/downloads/Bike-Sharing-Dataset.zip/1>

(Overview: <https://www.kaggle.com/marklvl/bike-sharing-dataset/home>)

Submission format: One Jupyter notebook (3_Bike-Sharing_GroupN.ipynb)

Please do *not* submit:

- A zip file
- A link to Google CoLab
- A file with the wrong extension
- A Python script

Task description:

- **Training data:** whole 2011 and first 3 quarters of 2012.
- **Test data:** 4th quarter of 2012. *Do not fit your models with these data! They should just be used to see how good/bad your model predictions are.*
- **Error metric:** R2 score (scikit-learn's default for regression).
- **Features to use:** at least the ones present in the data (except for cnt). *Do not use both casual and registered columns, as cnt=casual+registered (you may use one, but not both).*
- **Groups:** default groups for this term. You can split work as you consider best, but make sure *each and every member is able to explain details on what was done throughout the project, even not their part.*

Grading criteria:

As explained in the syllabus, the project report (the submission in the form of a Jupyter notebook) weighs 30 % of the overall grade of the subject, while the presentation amounts for 10 %. For this task there are no detailed, explicit steps to be followed, but broad elements expected to be present.

Report (10 points):

- Exploratory Data Analysis (descriptive analytics) (4 points)
 - Ensuring data quality (correctness, consistency, missing values, outliers...).
 - Plotting clear and meaningful figures.
 - Giving insights on what seems relevant for prediction and what does not.
 - Bonus points for:
 - Studying the influence of combinations of features (not just individual features).
 - Checking possibly redundant variables via correlations.
- Machine Learning (predictive analytics) (5 points)
 - Choosing sensible models (linear and non-linear).
 - Tuning model parameters with validation (either with a fixed validation set or with cross-validation).
 - Obtaining accurate predictions in test (measured with R2 score).

- Plotting predictions vs. reality for additional insights.
- Bonus points for:
 - Plotting validation results to justify further choices (parameter ranges, other validations...).
 - Following an incremental approach (baseline models first, then more complex models, then combining models...).
 - Iterating the whole process if incorporating additional data.
- Work description (1 point):
 - Work division and planning
 - Steps finally followed
 - Results obtained after those steps
 - Conclusions and further work

Exposition in class (general overview of your project and results, approx. 20 minutes per group):

- Clarity, conciseness, quality of content & delivery
- Answering final questions (posed by teacher and/or other groups)