

Overview

Dataset link:

<https://drive.google.com/file/d/1m53GMGvefv99ZeTOuhjb5wLkTSwfMTCf/view?usp=sharing>

Dataset description:

<http://archive.ics.uci.edu/ml/datasets/thyroid+disease>

Target variable:

"Class"

Evaluation Criteria

Scoring (15 max):

- EDA: exploration of variables and properties of data with conclusions: 1
- Data preprocessing if needed: 2
 - Handling missing values
 - Dropping unimportant features
 - Data normalization
 - Categorical encoding
- Metrics and cross-validation have been chosen based on target distribution as well as reasoning behind: 3
- Feature importance and hyperparameters tuning
(<http://scikit-learn.org/stable/modules/classes.html#hyper-parameter-optimizers>): 1
- Models
 - Logistic Regression w\wo regularization: 2
 - K-nearest neighbours: 1
- Quality of delivered work:
 - Analytical comments provided: 1
 - The experiment is structured (file is readable, pictures have titles): 1
 - Code is clear (reusable code in functions, comments, code is easy readable): 1
- Extra points for improvements not considered in the criteria: 2

(*)<http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.FunctionTransformer.html>