Tyre Inc. R&D

The Research & Development team of Tyre Inc. has been testing a number of new products for potential introduction to the market. For each tyre, a number of attributes have been collected about the tyre characteristics and the testing conditions, as well as the outcome of the test, i.e., success or failure. As member of the team in charge of data science you are in charge to create a machine learning task to predict the failure of a tyre and describe which attributes are the most relevant for outcome.

Dataset description The dataset contains a list of tested tyres. The target and attributes information are described below.

- Number of instances: 3000
- number of attributes: 15
- Target variable: "failure":
  - 0: test was successful
  - 1: test ended with tyre failure

Attributes information

Attribute	Type	Notes
vulc	Numerical	Vulcanization level
perc_nat_rubber	Numerical	% natural rubber
weather	Numerical	Weather condition index: 1: best, 0: worst
perc_imp	Numerical	% impurities
temperature	Numerical	Temperature in °C
elevation	Numerical	Elevation in meters above sea level
perc_exp_comp	Numerical	% experimental compound
diameter	Numerical	Diameter in cm
tread_type	Categorical	
tyre_season	Categorical	0: winter tyre, 1: summer tyre
month	Categorical	Month of testing
tread_depth	Categorical	
wiring_strength	Categorical	
tyre_quality	Categorical	0: standard, 1: premium
add_layers	Categorical	Number of additional inner layers

1 Important dates

- 5 Dec. → Training set (ind. vars + target) uploaded to WeBeep
- 19 Dec. → Test set (ind. vars, **NO** target) uploaded to WeBeep
- 20 Dec. 23:59 CET → **Submission deadline**

2 Submission: requirements and procedure

Requirements

- Vector of predictions for the test set
  - .txt file
  - One row per obs, as many rows as the test set
- Report detailing all the steps of your analysis
  - .pdf file
  - This file must include the step-by-step methodology (i.e., pre-processing, visualization, training, testing, etc.) that you have followed to develop your model, and illustrate the motivation behind your choices (e.g., for selected approach).
  - Can be a commented notebook (with the outputs).
- Source code to reproduce the results you obtained. This is the code that you used to obtain your model. Comments must be included to ensure that the code is easy to follow.

Procedure

The submission procedure consists of an email for each student (regardless of group membership) to `andrea.mor@polimi.it`. The email must contain an attachment in the following format:

```
#.zip
├── #.txt ..... File of the forecasts
├── #.pdf ..... Report
└── #.[ipynb/py/etc.] ..... Source code/file
```

with `# = CODICEPERSONA(8digits)_surname_name`, e.g., `19958400_danzig_george`.

Again, even if you can work in groups, the submission is individual. Students can only submit their assignment (i.e., you cannot submit the assignment for other members of your group).

3 Notes

- The assignment score is valid for **1 academic year**, i.e., until the exam of September 2023.
- Groups are formed independently by students and can be of **3 students at most**.
- Any submission that does not respect the guidelines (submission after deadline, empty file, wrong student code) **will not be graded**.