

[MY SPACE](#)[HOME](#)[CHALLENGES](#)[FAQ](#)[CONTACT](#)[LOGOUT](#)

Challenge Data

Defect Prediction on production lines by VALEO

[Introduction](#)[Public leaderboard](#)[Private leaderboard](#)

Description



Dates

Started on Jan. 6, 2020

Challenge context

Valeo is a French global automotive supplier headquartered in France, listed on the Paris Stock Exchange (CAC-40 Index). It supplies a wide range of products to automakers and the aftermarket. The Group employs 113,600 people in 33 countries worldwide. It has 186 production plants, 59 R&D centers and 15 distribution platforms. Its strategy is focused on innovation and development in high-growth potential regions and emerging countries. Valeo ranked as France's leading patent filer from 2016 to 2018.

Challenge goals

The goal of the challenge is to predict defect on starter motor production lines. During production samples assembly, different values (torques, angles ...) are measured on different mounting stations. At the end of the line, additional measures are performed on two test benches in order to isolate defects. As a result, samples are tagged 'OK', 'KO'. We would like to design a model that could identify such defects before the test bench step.

Data description

ID = PROC_TRACEINFO = it's a unique code given to the product. Example : I-B-XA1207672-190701-00494.

- XA1207672 is the reference.
- 190701 is the date: here 01st of July of year 2019.
- 00494 is the unique code given to the product, whatever it happens, the product will have this id number frozen forever.

This number is increased by 1 each time we process a new product, every 12s. So for example : I-B-XA1207672-190701-00495 is the next product.

Inputs : Input features are measures collected on different assembly stations with the sensors or devices connected to Programmable Logic Controllers which are storing all of them to keep the full quality traceability. (Examples : OP070_V_1_angle_value, OP120_Rodage_L_value, etc...).

Output : This is the result value of OP130 (test bench). Value 0 is assigned to OK samples (passed) and value 1 is assigned to KO samples (failed). This is the combined result of multiple electrical, acoustic and vibro-acoustic tests.

The target is to find the best prediction : $\text{Output} = f(\text{inputs})$. The dataset contains 34515 training samples and 8001 test samples.

Benchmark description

We expect a AUROC more than 0.675 which can easily be obtained with a basic Naive Bayes classifier. This AUROC value had been obtained using some techniques for unbalanced classes handling and not only with the challenge metric. That's the main reason why our AUROC value of 0.675 is different from the benchmark value (0.5904).

Public metric

roc_auc_score from scikit-learn.
[scikit-learn metrics](#)

Files

[x_train](#)
input data of the training set
[y_train](#)
output data of the training set
[x_test](#)
input data of the testing set
[random_submission_example](#)
a random submission csv file example.

The challenge provider



Automotive

[PROVIDER WEBSITE](#)

Contact

For any questions regarding the challenge, you can contact the challenge provider at cda.challenge-valeo.mailbox@valeo.com



[team](#) | [terms of use](#) | [legal notice and privacy policy](#)