

# Challenges overview

Challenge

Challenge Desjardins

## DESCRIPTION

# Default payment prediction (Challenge Desjardins)

## Description of the challenge

Desjardins believes it is essential to serve its members and customers fairly and to build trust in its AI by demonstrating responsibility and inclusivity in order to be unbiased towards specific sub-populations or individuals defined by sensitive attributes such as gender or ethnicity. It is therefore committed to developing and implementing ethical and fair AI algorithms, standards and practices. We therefore systematically explore existing approaches or tools to exploit. Many approaches, notably adverse learning or representation, are based on neural networks, which are sometimes more difficult to implement.

We would like to develop an in-processing approach to bias mitigation, i.e. acting directly on the training of a supervised model, such as a boosting ensemble model. The approach would be based on adversarial learning. In this type of learning, a so-called adversarial model, often based on neural networks, attempts to learn the sensitive attribute from the predictions of the target variable at each iteration of the boosting model (see for example:

<https://arxiv.org/pdf/1911.05369.pdf>). The latter incorporates the loss function of the adversary model into the total loss function in such a way as to minimize the ability of the adversary model to predict the sensitive attribute while attempting to predict the target variable itself. We would like to develop an approach where the adversarial model is based on tree methods, which are more practical to implement in-house than neural networks.

## What you will accomplish

The aim of the challenge is to predict whether the customer will pay his credit card next month.

## Criteria for judgement

The score on the dashboard is the harmonic mean between the f1 score and the fairness score.

To learn more about fairness, please visit <https://codeml.ca>. The calculation we use to measure the fairness of your model is the normalized disparate impact.

The presentation of the method used by the winners will be requested. The code must also be provided.

Accuracy

F1\_score

Fairness

Score

**Good luck, have fun!**