Task Overview

In this competition you will help to create an automation pipeline for consultants.

The task is to classify pairs – (web pages, criteria) to 3 classes: rejected_by_product /rejected_by_function / accepted.

We consider company's services / actions / activities as Function. For example, "manufacturing of cars" -> manufacturing – function, cars – product.

Firstly consultant checks whether the company can be rejected by product, if company is acceptable by product, consultant checks whether the company can be rejected by function.

Only after that, if company is acceptable by product and function, consultant marks it as accepted.

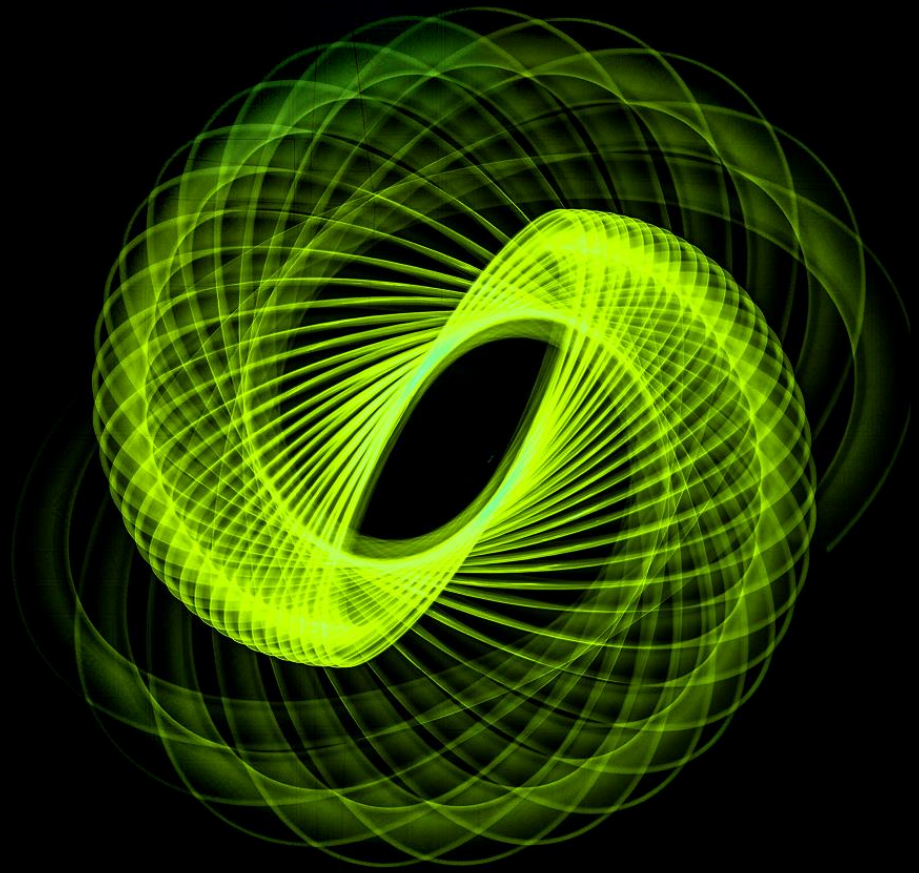
Data

id – unique id of sample.
html – list of html web pages related to this particular company.
text – list of extracted texts form the corresponded html.
keywords - list of relevant words from text (model based).
accepted_function – criteria for acceptance by function.
rejected_function – criteria for rejection by function.
accepted_product – criteria for acceptance by product.
rejected_product – criteria for rejection by product.
target – 0:rejected_by_product, 1:rejected_by_function, 2:accepted

Evaluation

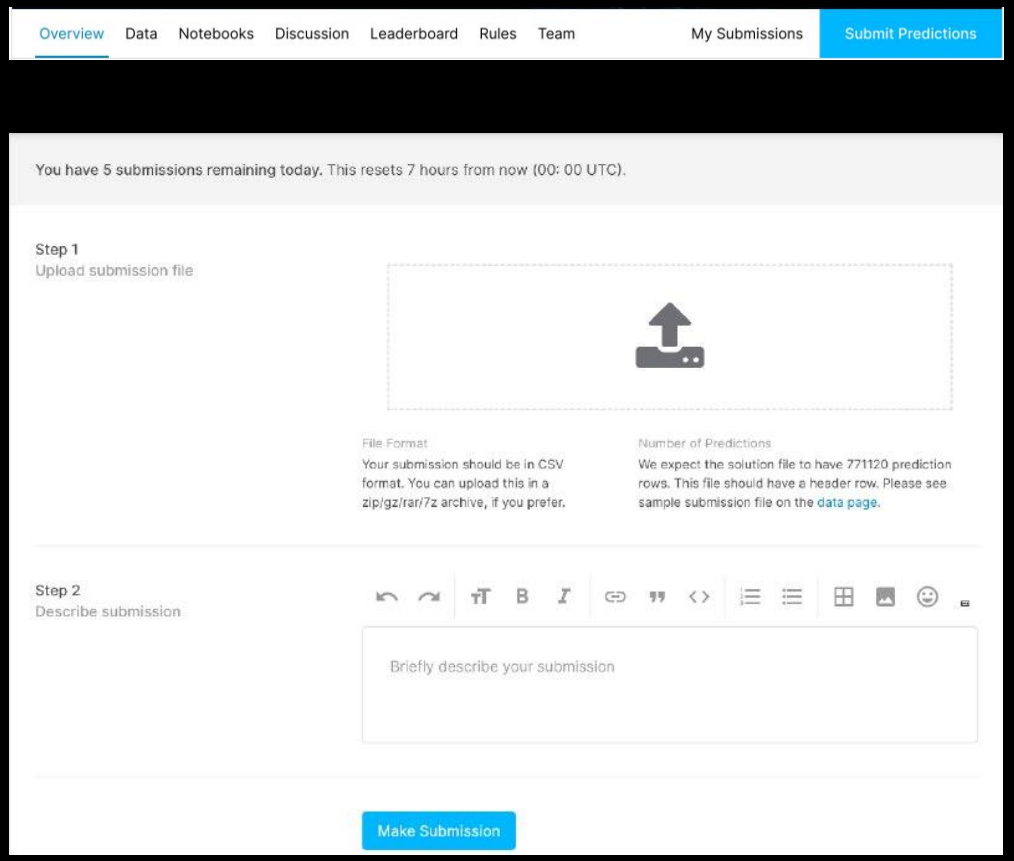
Metric – weighted accuracy.

Weight of each sample depends on it's target:
{rejected_by_product:1,
rejected_by_function:1, accepted:2}



How to submit

Instruction



#1 Step
Go to the "Submit Predictions" tab

#2 Step
Select and download a csv structure file similar to sample_submission.csv from the organizers

#3 Step
Optionally write a comment of the file (model, validation, etc.)

#4 Step
Click on «Make Submissions»