



# David López

Machine Learning Expert

Exercise:

Create a model to predict the completion rate (defined as submissions/views ) of a form given the features in the dataset.

You might use any model/library ( Scala/ python preferred )

Real dataset approx 10.000.000 records and growing

- Dataset (Sample 10.000 )=>
  - Csv file with form\_id , submissions,view,features as “-” separated.
- Delivery => API
- Github repo => ideally, we should be able to clone the repository and being able to train and serve

The model with the code and data in the repo and the README file.

Sketch architecture:

- Pipeline running in aws
- Predictions delivered as API call (Should be apro 15000 call for minute at the moment and growing )
- A robust monitoring and update system for the model.

Please describe briefly each stage - choice of tech at relevant stages of the pipeline.

Please point if any weak points of the choices you have made.

Xtra points :

1. Use Spark to solve the exercise.
2. Use a simple - not too DNN to solve the exercise (Not to worry about overfitting)

We expect you to do it within 2-3 business days. If you think you will delay please let us know in advance.

Do not hesitate to contact us if you have any clarifying questions.

Good luck!”