

Machine Learning Specialist

Case study

Hi! Thank you for taking the time to complete this case study. It consists of two parts, one about predicting form's completion rate and the other about classifying form questions according to their intent.

We expect you to return the Case Study within 3-4 business days. If you think you are going to need more time please let us know in advance.

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

The Typeform ML team :-)

Completion rate prediction

One of the main concerns of our customers is creating forms that attain high completion rates, that is the number of form submissions over the total number of form views.

Using the **completion_rate** dataset provided, consisting of unnamed features from published forms, together with their number of views and submissions, we ask you to perform the following tasks:

- Build a supervised model that predicts forms expected completion rates (defined as submissions over views).
- Analyze the results and document the assumptions and modelling decisions.

Requirements

- You can use open source ML libraries (Python or Scala).
- You can use notebooks if you want to, for model prototyping
- Upload your results to a git repository. We should be able to clone the project and run it.

Extra mile

- A simple web API to serve online predictions
- A train+predict pipeline built in Spark

Question intent classification

At Typeform we want to improve our customers' experience when creating their forms. For this reason we have set ourselves to understand what our form creators are asking their users by analyzing the content of the form questions.

Using the **form_questions** dataset provided, containing ~4k form questions, we ask you to perform the following tasks:

- Build an unsupervised model to cluster form questions.
- Analyze the results and document the assumptions and modelling decisions.
- Describe the steps required to build a text intent classification model to tag form questions, starting with a dataset like the one provided above.

Requirements

- You can use open source ML libraries (Python or Scala) and publicly available models in your solution
- You can use notebooks if you want to, for model prototyping
- Upload your results to a git repository. We should be able to clone the project and run it.

Nice to have

- A solution based on a State of the Art language model.