Rappi Machine Learning Engineer

Challenge

The Challenge

Based on the well known Kaggle problem "Titanic: Machine Learning from Disaster" and the jupyter notebook provided as a zip, we would ask you to develop two process:

● Create a pipeline for training a classifier.

● Create an API.

Pipeline for training a classifier

Create a pipeline for training a binary classifier. The input for this training can be found in the mentioned jupyter notebook. The output should be a binary classifier model and exported somewhere so it can be used from the API.

Your solution should cover the following points:

● Binary classifier as described above (you can use the code provided in the notebook)

● Automatized pipeline execution

● Pipeline profiling (CPU and RAM usage)

It would be great if your solution covers the following points (optional):

● Improved implementation of the classifier provided in the notebook

● Logs configuration

● Virtual environment configuration

● Dockerized solution

● Data version control

Create an API

After exporting the classifier as a binary, it is required to use it for real time predictions. With the mentioned objective in mind, you would need to create an API that receives a list of passengers and predicts for each of them if they're likely to survive or not.

Your solution should cover the following points:

● API as described above

● Proper error handling (http status code, message, unhandled exceptions, etc)

● API monitoring and alerts

- API profiling (CPU, RAM usage, maximum requests handled) and how it would scale

- Unit testing of model's predict function

It would be great if your solution covers the following points (optional):

- Download model and load it on API deployment

- Logs configuration

- Binary model issues tracking

- A/B Testing so two models can be deployed and compete between them

- Dockerized solution

Results delivery

1. Implementation code folder/repository and instructions to run it.

2. You may use any language, tools, and cloud services you want.

3. You will have a 30 minutes presentation to show your work.

4. We expect to receive your work within <> days since we sent you the challenge.

Documentation

We encourage you to write down every step down the road. Design decisions, references you used, difficulties and insights, and partial results. As mentioned, not all the points are mandatory but as mentioned in the respective sections it would be great if your solution covers them all.

If you have any doubts, feel free to contact us at osvaldo.renteria@rappi.com