



# Technical exercise - Data scientist

As part of our recruitment process, we'd like to ask you to develop an ML classifier.

## Goal

Your goal is to build an ML algorithm that automatically classifies emails by themes.

To do that, we provide you a dataset containing:

- A text file containing all the data about the email

- The 13 following classification labels:
  1. regulations and regulators (includes price caps)
  2. internal projects -- progress and strategy
  3. company image -- current
  4. company image -- changing / influencing
  5. political influence / contributions / contacts
  6. california energy crisis / california politics
  7. internal company policy
  8. internal company operations
  9. alliances / partnerships
  10. legal advice
  11. talking points
  12. meeting minutes
  13. trip reports

You can download the dataset here:

 giskard\_dataset.csv 10410.2KB

## Requirements

We want you to predict the classification label using at least the following features:

- The content of the mail
- The subject of the mail
- Date time

Feel free to add all additional features you might think of.

## Bonus: Web Interface

Create a web interface that returns the classification label predicted by your algorithm for given email features as input.