Deep Learning Assesment Project

The goal of this project is to assess your skills in data science and deep learning in general, to see in what way you can contribute to our current team.

- **Requirements:** Usage of Jupyter notebook when using Python for exercises I-III.
- **Recommended tools:** Python, R, javascript
- **Preferred frameworks to be seen:** Tensorflow, Keras, scikit-learn

Important: The goal is to gain insights in how you tackle problems, so don't worry about answering the question right or wrong. Questions will be asked about the thought-process, implementation decisions, expected vs actual results and possibility for improvements. This also means that the following exercises might not have a true solution nor will the resulting models have a high accuracy.

Preliminary work (Scraping and cleaning)

Before starting with the first exercise, data has to be collected. Instead of providing clean and ready data, we're going to scrape this from a website and clean it ourselves.

Quora.com is a **Q&A** website where people can answer questions, much like stackoverflow. The best answer is usually voted to the top, so the best answers can be found easily.

1. Visit Quora.com and look around. Plan your scraping strategy.
2. Choose 10 various topics of interest. Preferably not too 'mathy' and too related to each other.
3. For every topic: Scrape 1000 questions and their best answers. (if there are less than 1000, scrape all of them)
4. Clean and prepare the data for use in data exploration.

Exercise I (Classification)

The first exercise consists of classifying answers to topics. The model you'll be building should be able to classify an unseen answer to a certain topic. Instead of returning 1 topic, return the *top 3 matching topics*, given an answer.

Make sure that you validate your model on unseen data, not used in the training session.

Exercise II (NLP / Cognitive)

The second exercise is tougher than the first one. We'll be applying natural language processing to deep learning. The general goal is to provide an answer for the following question:

Is it possible to generate a question from the answer or vice versa.

Do the following for this exercise:

1. Explain in your notebook how you'd approach both ways and how in theory it could work.
2. Choose either way (answer->question or question->answer) and implement, *tinker* and explore with deep learning models what can be done with the limited data and time you got.

Exercise III (Exploration & Visualizations)

The last exercise is about data exploration and visualization.

Imagine you're working for Quora.com as a consultant and they want to know **more (insights) about the questions and answers for the 10 topics you chose**.

Given your scraped dataset, try to find some (generic?) insights and visualize them. It's a bonus if you're familiar with visualizing data in interactive environments (D3.js, plot.ly, Rshiny, ...).

Timing

The moment you have received the link to this page, we expect an confirmation e-mail from you. From that moment on, you'll have 2 weeks to complete this exercise and submit your project to us.

We will discuss the project internally and provide you with our feedback. If we are intrigued by your submission, you will be invited to our office to defend your case and you'll have a chance to meet up with our project team and therefore your potential colleagues.

If you are in trouble, feel free to contact us for us. In the end, we might become colleagues!

Good luck and happy data hunting !