



## Data Scientist Assignment

Thank you for showing your interest in Talentwunder and for wanting to be part of Talentwunder's Data Science Team.

This assignment is a chance for you to showcase your technical skills, your creativity as well as your business sense. We hope it will form a good base for discussion further down the process.

### Tasks

#### Typos Correction

Talentwunder's search engine allows their users to find the best candidates across different networks. As a result of the noisy data often found in the source networks, there is much work spent into cleaning and preprocessing data to make it useful for downstream tasks. In this first task you are asked to come up with an approach to detect typos in job titles and correct them. Describe your approach high-level. **No coding needed.**

#### Example

*"Sales Manager"* is an example for a job title that is often misspelled as *"Sales Manger"*. How would you detect and correct this kind of typos?

#### Serving similarity models

Similarity between entities is often used to make recommendations. In this task you are asked to implement a script to train a model that learns an embedding on textual data and to serve the model in a simple web service that returns the top 10 closest terms for a given word.

**Use the dataset made available in the repository at *"data/sample\_data.csv"* for training your model.**

Requirements:

- Implementation should be done in **python**.
- The service should be running as a Docker container.

**Note: The quality of the model is not of importance.**

## **Entrepreneurial thinking**

In the first interview session you were introduced to a few of the features and data products developed at Talentwunder. Think about some data driven product/feature that could be useful in the recruiting domain. What data would you need to realize the project? How would you measure the success of your data product? What kind of model would you use to implement the solution?

**Note:** Talentwunder Products introduced at the interview should be excluded. However, it is not needed to come up with your own product.

## **FAQ**

### **How much time do I have?**

You get 5 days from the time you get this assignment.

### **How much time should I spend on this?**

The assignment should take no more than 6 hours in total. Time is hard to find though.

### **Need any other clarification?**

If you have any other questions, hit reply to the email this assignment was attached to or send us directly to: **oscar@talentwunder.com**

Best of luck