Applied Data Science Project

Intelligent Data Analysis for HR

Value-driven project

Machine Learning algorithms and Natural Language Processing (NLP) applications could support and optimize the work of recruiters during talent scouting and acquisition. During such processes, the first relation between candidates and companies is the job advertisement, and the way they are written is crucial to attracting the best candidates.

An attractive and well-formed job advertisement increases the chances of a better match between candidates and companies, which is in line with the 8th UNSD Goal (Decent Work and Economic Growth).

The stakeholder for this project is INDA, an AI solution for recruitment and talent acquisition. The project proposed to the students may serve as a basis for further developments and could be integrated into INDA.

Data

We will provide two dataset:

- **Recent Job advertisement data** published on online platforms. The data are obtained from the web and are not preprocessed, it will be up to the students to preprocess and clean the data. Each document contains information such as job ads title, company, location, description, seniority, employment, job function, industries, original link, date, domain, location, and location info.

- **Skills data** sample from our private skills database. This data is composed by a list of strings.

Data will be made available in a dedicated Atlas MongoDB database and personal credentials will be distributed to all participants during the project period. A portion of the database will be available for students, for saving intermediate collections or results.

Task

The aim of the project is to develop a solution able to improve the attractivity of job advertisements and it should be used by recruiters during the job advertisements writing process.

The solution is made by several components and will be developed in partnership with us:

- **Preprocessing of job advertisement data**. The project members should deal with missing values and provide a statistical overview of the features.

- Skill extraction from job advertisement description. The project members should use our skill extractor algorithm (or develop their own version) to identify which word in the text is a skill.

- Review of job advertisements with the **recommendation of most attractive skills** to improve the whole attractivity of job advertisements. This task can be split in two subtasks:
  - given a set of skills, identify a set of similar skills (up to the members the definition of the similarity metric);
  - identify, within the set of similar skills, the most attractive ones (up to the members the definition of the attractivity metric).

Light mentoring

We suggest two initial meeting:

- Onboarding in INDA, HR context, and presentation of the data;

- Onboarding in the skill extractor.

During the development of the project we suggest a 1h weekly meeting. We will centralize the communication to one or two INDA data scientists.

Privacy Policy

The data we will provide in this project are not affected by any NDA agreements. Job advertisements and skill data are samples of INDA private database.

Anyway, we invite all the members of the project to not resell any information or data used during the project.

Questions?

Contacts

INDA

+39 011 586 2407
Via Amedeo Avogadro, 20 – Torino
Via Caviglia 11 – Milano
http://inda.ai

Felipe Operti
felipe.operti@intervieweb.it

Luca Colombo
luca.colombo@intervieweb.it

Pierangelo Lombardo
pierangelo.lombardo@intervieweb.it