Matching CVs based on EDISON Data Science Competencies (CF-DS) and advanced text analysis methods
M.F. Maijer

Abstract

Data has started to play an important role in modern data driven research and industry; allowing for deeper understanding of different phenomena. Demand for new types of data science specialists who can support all stages of the data development life cycle; from data production and input, to data processing, visualisation and reporting, is growing. This has caused the gap between the supply and demand of data scientists to widen considerably.

There is a strong need to define a methodology and develop tools for effective job candidate CV assessment against job/employer requirements. The EDISON Data Scientist Framework (EDSF) provides a comprehensive multi-dimensional, multi-domain, and multi-component definition of the relevant data science competencies, skills and professional profiles as part of its Data Science Competence Framework (CF-DS) and Data Science Professional Profiles (DSPP).

This framework classifies the data science related competencies in five different groups, that each contain six competencies. However, to assess a data scientist correctly using the EDISON Competence Framework is a complex task. Document similarity techniques along with regular expressions were used on the CV of a data scientist to create insights into the competencies of the CV's creator.

The aim of this research was to accurately portray the competencies of data scientists by using the CF-DS and to make this tool available and usable for data scientists and recruiters. The proposed solution has three components, a timeline indicating the career path of the data scientist, a graph showing competency scores based on document similarity, and a graph showing competency scores based on the career path.

The best method to get an accurate overview of a data scientist's competencies at this point in time, is to use all three components of the tool that was developed. The developed CV assessment tool and application provides an effective instrument for recruiters to assess CVs which allows them to decide on the most suitable candidates. It is also a useful tool for job seekers and practitioners to assess their competencies and identify a path for professional development.

Using the proposed solution, it has become possible for data scientists and recruiters to start a basic assessment process, which will in turn enable data scientists and even the data science community itself to develop. The project has delivered a working code that is included in the ongoing EDSF community development project: https://github.com/EDISONcommunity/EDSFapps.