

This repository contains the code needed to replicate the results found in Levin-Konigsberg and Schubert (2020), “Deep Learning Deep Skills: Firm Differences in Job Tasks and Stock Market Performance”.

Authors:

Gabriel Levin-Konigsberg
Gregor Schubert

Dataset: BGC proprietary data—not publicly available.

Project description: Even though the BLS statistics may classify an Administrative Assistant at Goldman-Sachs in NYC and one at a construction company in Wichita, KS under the same category, the tasks each perform may differ substantially. There’s no clear mapping between a job classification to the skills required in a specific job. We intend to predict the complexity that a job requires in different dimension (e.g. leadership, social skills) based on the job description in internet postings.

Algorithms:

apply_model.py: Used to predict the complexity of an occupation using a posting, given a previously estimated file.

bert_sample_to_file.py: Use to sample from a file containing posting and OCC data, applying BERT and getting the features that are the input for our main architecture.

implementation_bert.py: Used to train our model from a list of files containing OCC data and BERT features.

implementation.py: deprecated

nn_implementation.py: deprecated

processing_xml/parallel_processing_xml_no_dropbox.py: Takes the raw data in xml format and produces a zip file containing a cab file with our variables of interest.

processing_xml/parallel_processing_xml.py: deprecated

processing_xml/processing_xml.py: deprecated

word_clouds/word_clouds_leader.py: Creates the word clouds for leader characteristics.

word_clouds/word_clouds_socskills.py: Creates the word clouds for social skills characteristics.