{Propulsion}

# Optimizing job ad titles at Jobcloud using Google BERT and neural networks

David Furrer & Adrian Imfeld

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD ▸jobs.ch | jobup.ch

# David Furrer

- MSc Chemistry ETHZ
- Novartis
- MA Economics HSG
- Data Science

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# Adrian Imfeld

{Propulsion}

- Cognitive Neuroscience / Artificial Intelligence, MSc
- Co-Founder / Lead Web Developer at Bright Answer
- Moving into Data Science / Machine Learning

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD jobs.ch | jobup.ch

# Business problem

Idea: Good job titles yield more views for job ads

Identify characteristics of high-quality job titles

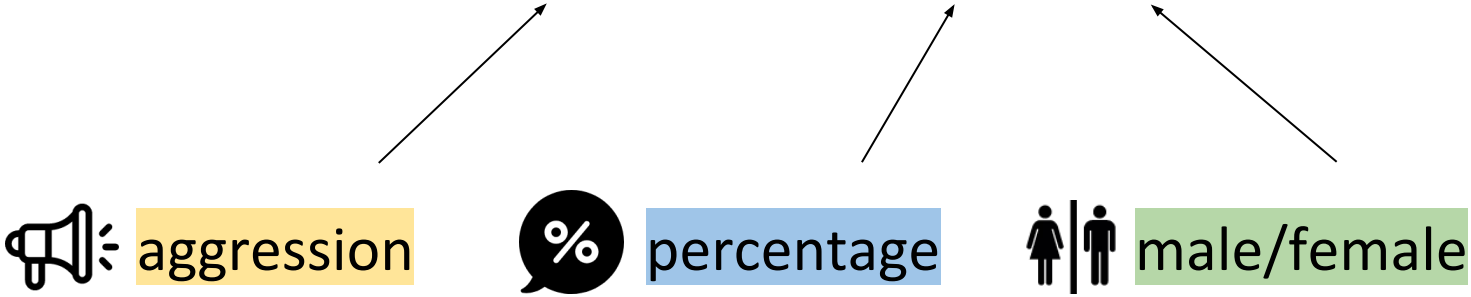
Autocomplete tool: suggest good job title variants

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

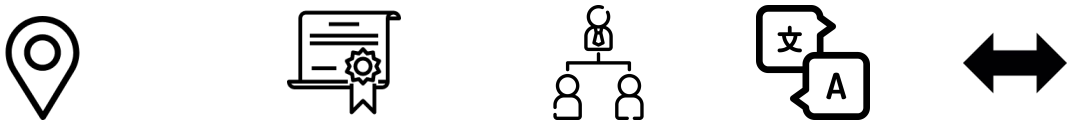
# Feature Engineering

{Propulsion}

Elektro-Fachmann als BETRIEBS-ELEKTRIKER 100% (w/m)




Location, qualification, CEO, language, length




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD ▸jobs.ch | jobup.ch

# What makes a good job title?

{Propulsion}

## Top 7 features for predicting view count

Contract %

Package B

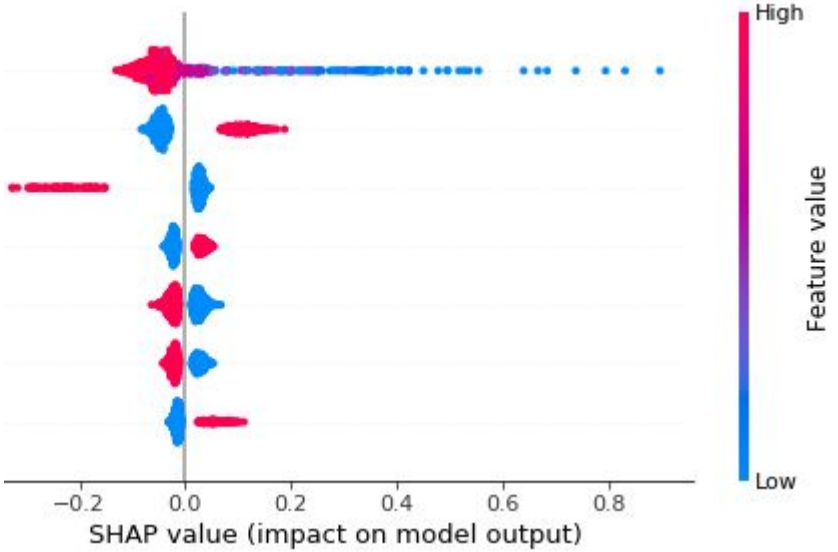
Package D

Title percent

City: Other

Package C

City: Zurich




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD ▸jobs.ch | jobup.ch

# Google BERT

{Propulsion}

- State of the art pretrained language model for NLP
- Trained on wikipedia and BookCorpus (11k books)

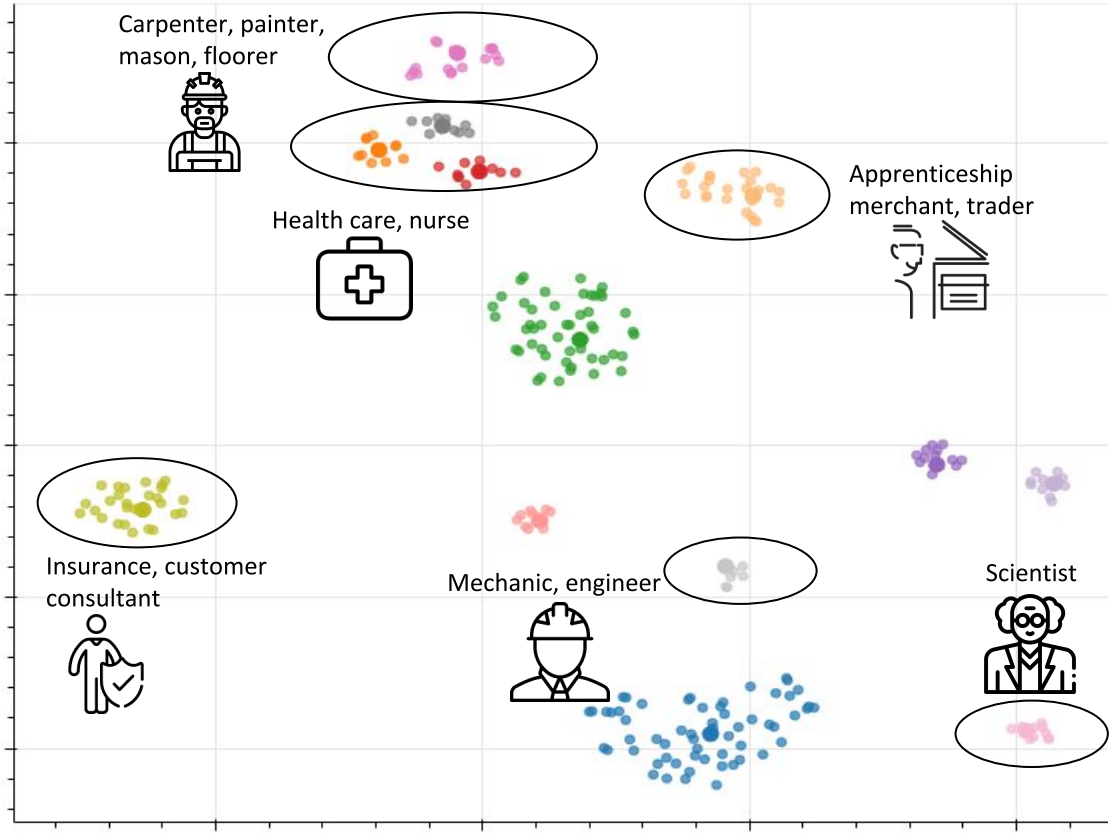
| Model | # Parameters |
|---|---|
| BERT-Base<br>BERT-Large | 110M<br>340M |
| GPT-2 | 1.5B |
| Honey Bee Brain | ~1B synapses |

Reference: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (*Oct 2018)*

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD ▸jobs.ch | jobup.ch

# Clustering job titles with BERT embeddings




- Job titles are encoded as **BERT embeddings** (vectors)
- BERT embeddings **represent meaning** of the job titles
- BERT embeddings can be used to **cluster similar job titles**

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# Interactive App Demo

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

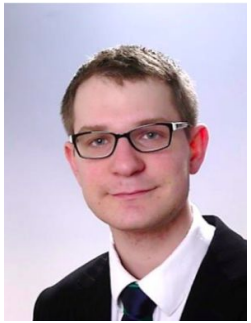
# Outlook

{Propulsion}

- Improve model (e.g. unemployment rate, company name)
- Implementation at Jobcloud

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# Contact



linkedin.com/in/david-furrer

furrer.david1@gmail.com

github.com/davidfurrer

medium.com/@fulowa

linkedin.com/in/aimfeld/

aimfeld80@gmail.com

github.com/aimfeld



© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# Approach

Data exploration

Data cleaning

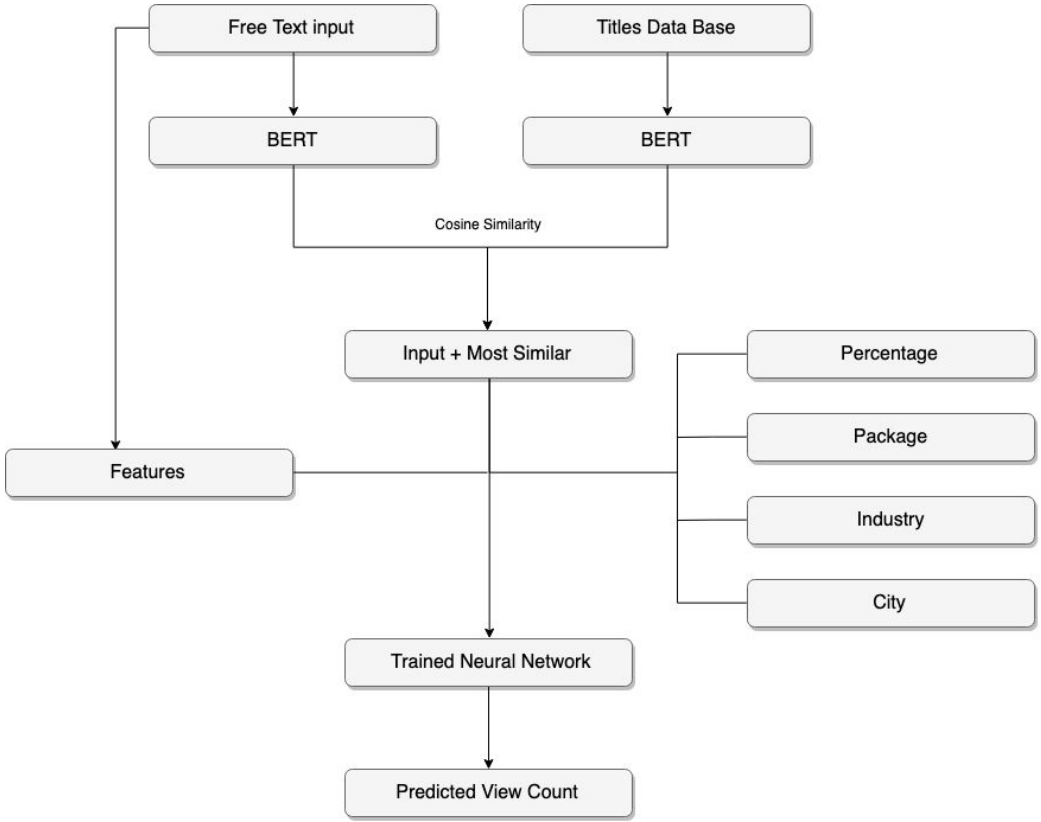
Feature engineering

Model to predict views

Interactive app

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# Interactive App Demo




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# BERT =
# Bidirectional Encoder Representations from Transformers.

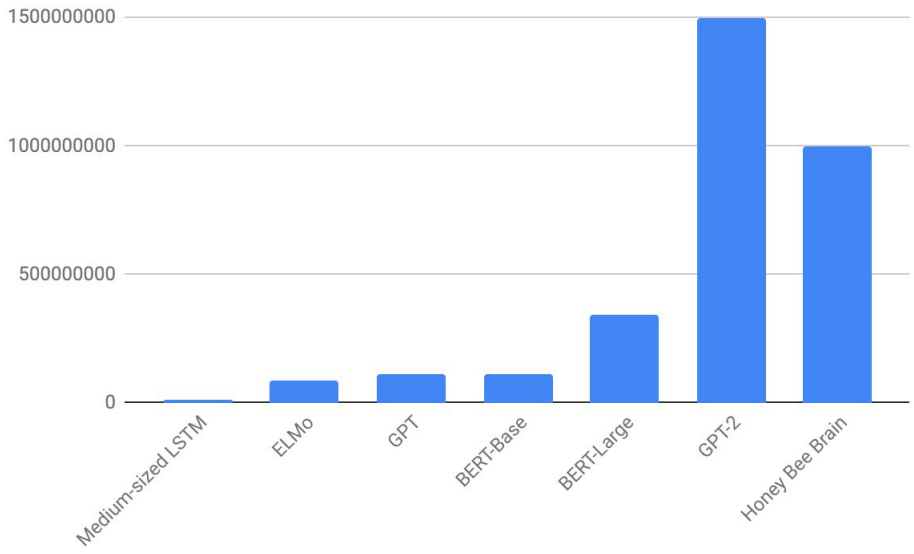
{Propulsion}

| Model | # Parameters |
|---|---|
| Medium-sized LSTM | 10M |
| ELMo | 90M |
| GPT | 110M |
| BERT-Base<br>BERT-Large | 110M<br>340M |
| GPT-2 | 1.5B |
| Honey Bee Brain | ~1B synapses |

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

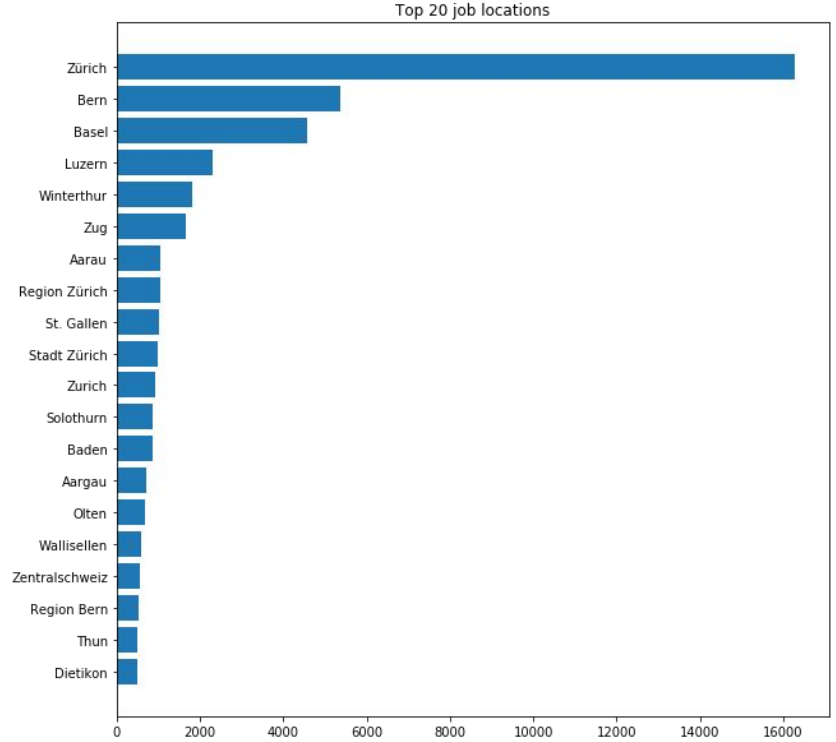
JobCLOUD ▸jobs.ch | jobup.ch

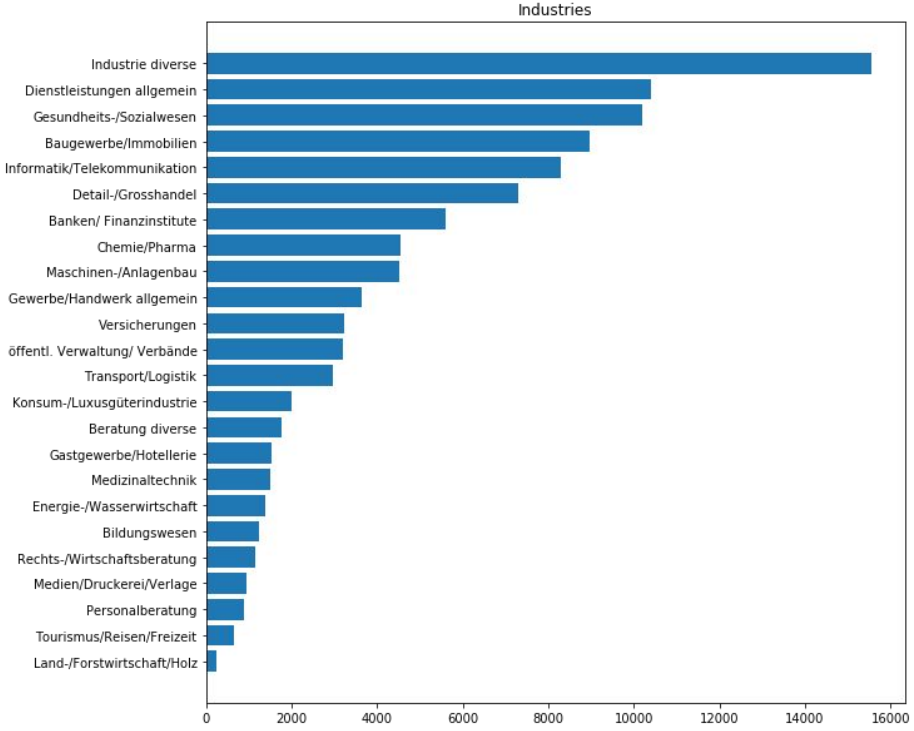
# Backup




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

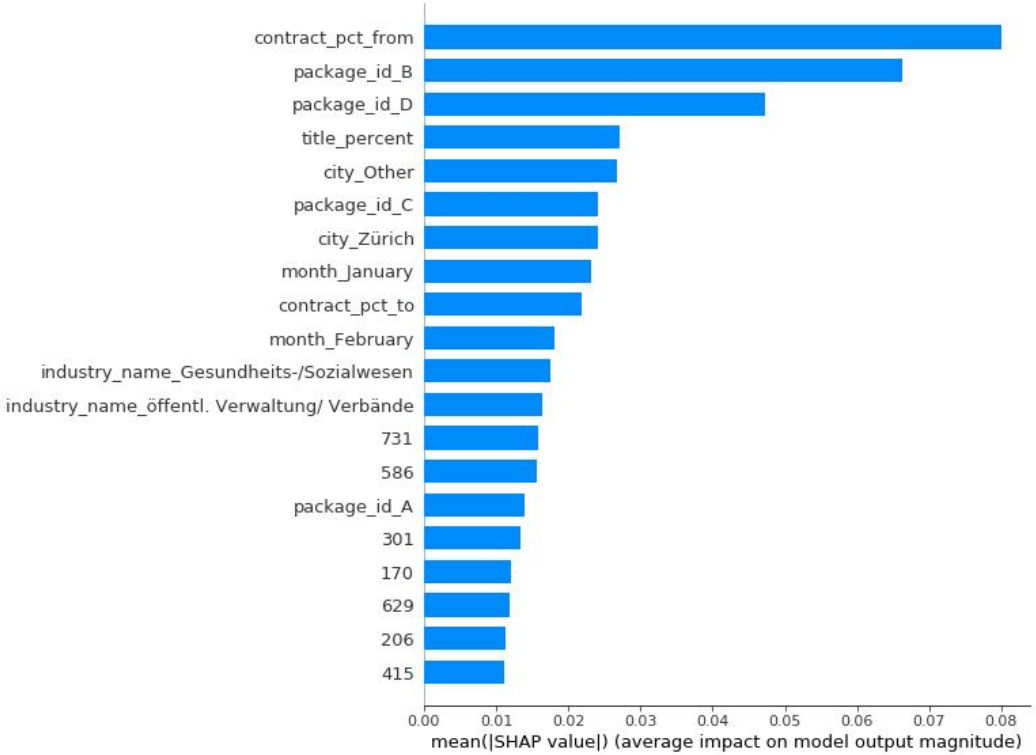
# Data Description




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# SHAP - Feature importance




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# Embeddings

- Can be used for words, sentences or paragraphs
- BERT embeddings have 768 dimensions

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

# App gif

choose a city
Zürich

choose an industry
Industrie diverse

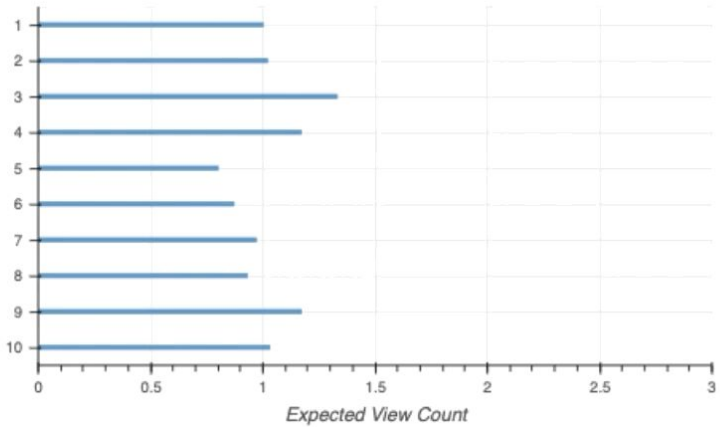
Enter Your Job Ad Title here
Data Scientist

choose a package
A

choose from %
100

choose to %
100

| # | title | pred | title_no |
|---|---|---|---|
| 0 | Data Scientist | 1 | 1 |
| 1 | Data Scientist (80-100%) | 1.020 | 2 |
| 2 | Data Scientists (m/w) 80-100% | 1.330 | 3 |
| 3 | Data Scientist/-in (80% bis 100%) | 1.170 | 4 |
| 4 | Data Scientist Auditor/in | 0.800 | 5 |
| 5 | Data Scientist Expert (m/w) | 0.870 | 6 |
| 6 | Senior Data Scientist | 0.970 | 7 |
| 7 | Data Scientist Imaging | 0.930 | 8 |
| 8 | Data Analyst | 1.170 | 9 |
| 9 | Data Science Expert | 1.030 | 10 |




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD jobs.ch | jobup.ch

# Insights about using BERT

{Propulsion}

- Improved results by removing noise
- Tricky to use
- Needs a lot of computing power, RAM

| | title | title_clean |
|---|---|---|
| 0 | Softwarearchitekt / Projektmanager (m/w) - All-in-One Datenmanagement in Design, Produktion und Qual | softwarearchitekt projektmanager all one datenmanagement design produktion und qual |
| 1 | Prozessingenieur Lasertechnologie - Industrialisierung innovativer Fertigungstechnologien | prozessingenieur lasertechnologie industrialisierung innovativer fertigungstechnologien |
| 2 | Softwarearchitekt / Projektmanager (m/w) - All-in-One Datenmanagement in Design, Produktion und Qual | softwarearchitekt projektmanager all one datenmanagement design produktion und qual |
| 3 | Projektleiter (m/w) - Werkzeug- oder Maschinenbau | projektleiter werkzeug oder maschinenbau |
| 4 | Fachverantwortlichen Metrologie - Produkteentwicklung und -validierung | fachverantwortlichen metrologie produkteentwicklung und validierung |

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD ▸jobs.ch | jobup.ch

# Data set

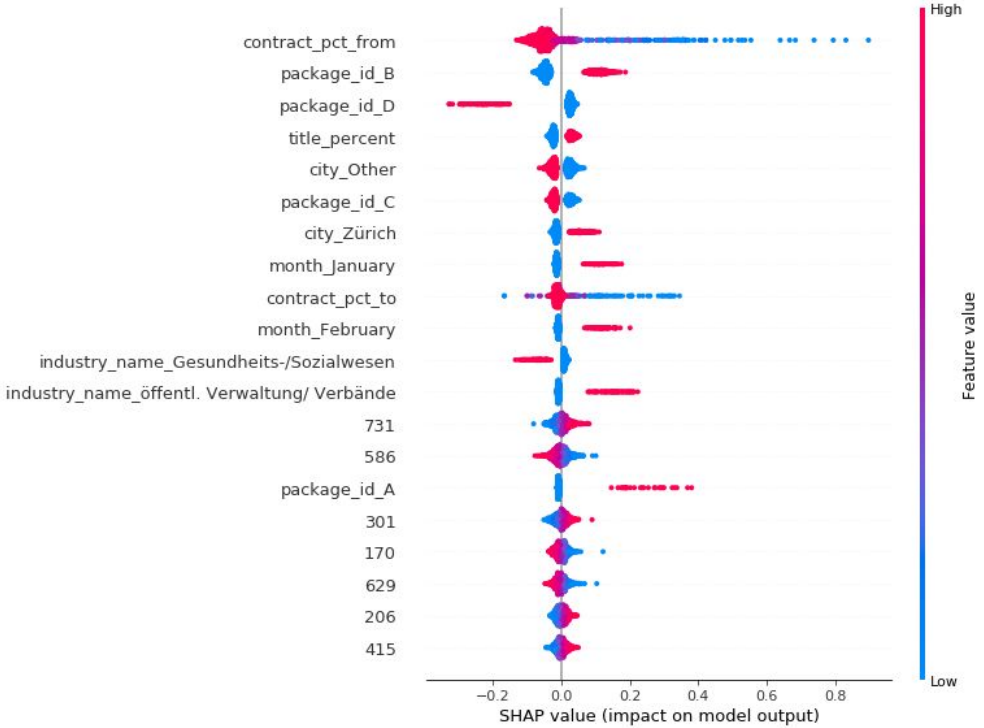
- Dataset
  - Date
  - Contract percent
  - Industry
  - Package
  - Location
  - **Job title**
  - **View count**

© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

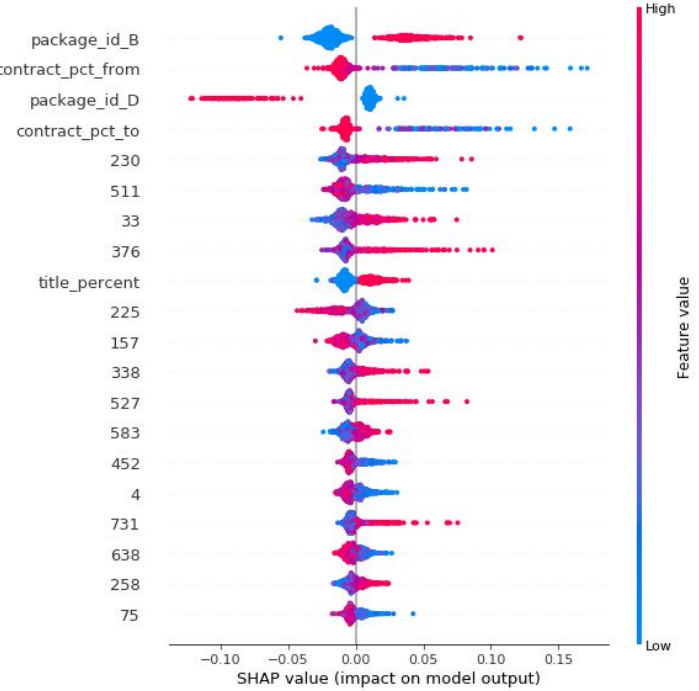
# Feature importance

{Propulsion}

Neural network (r2 = .37)



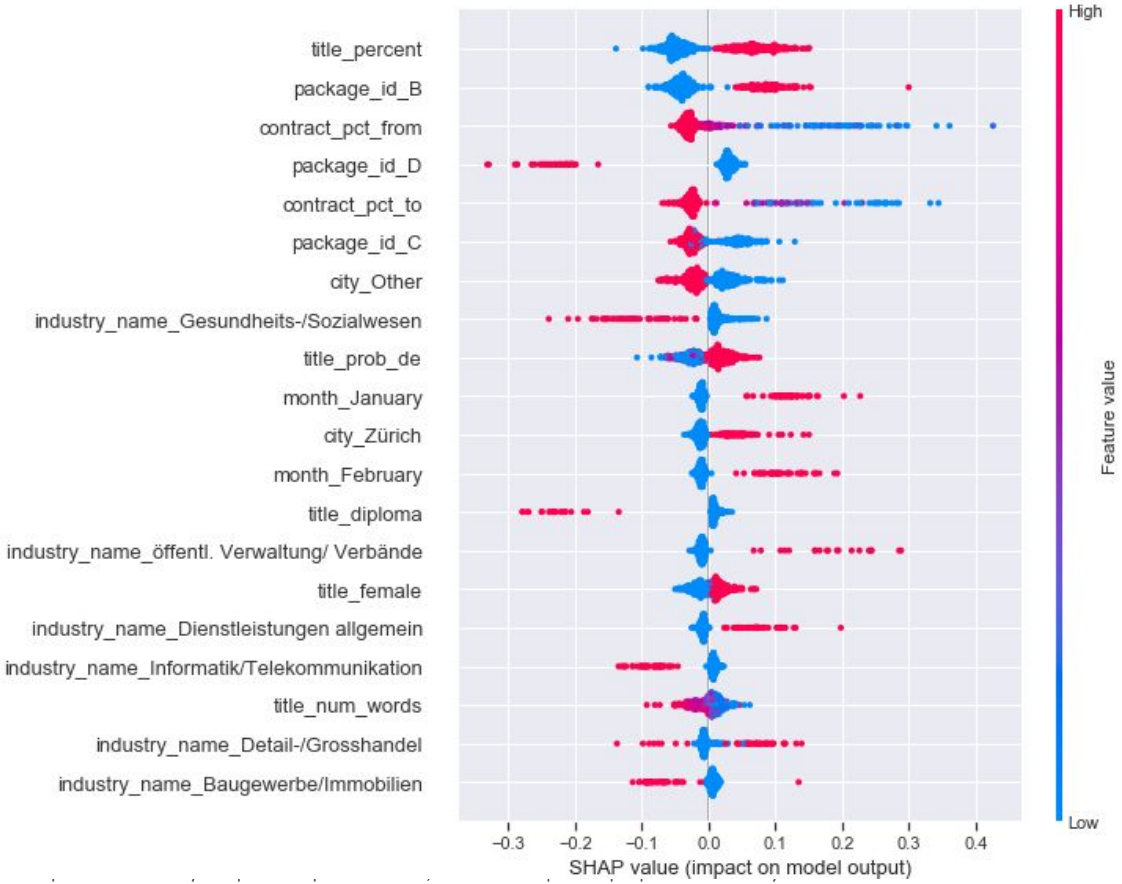

Random forest (r2 = .34)




© 2019 Propulsion Academy AG | Technoparkstrasse 1, 8005 Zürich | www.propulsionacademy.com

JobCLOUD ▸jobs.ch | jobup.ch

# Feature importance




Random forest without BERT embeddings (r2 = .25)