Alexander KLAPHEKE

Boston-based data scientist with linguistic expertise

Cambridge, Mass.

(860) 605–6150

alexklapheke@gmail.com
alexklapheke.github.io

M.A. Linguistics, Harvard University   2019
B.A. Linguistics, Columbia University   2012

EXPERIENCE

Data science fellow, General Assembly   2020

Created several data science reports and applications during 480-hour data science training, covering data cleaning, exploratory data analysis, feature engineering, regression and classification modeling, and machine learning.

Instructor, Harvard University   2014–2017

Taught 5 seminars in logic, semantics, discourse analysis, and philosophy of language.
Received Distinction in Teaching Award in fall 2017 for universally positive feedback.

Research assistant, Columbia University   2011–2012

Wrote BASH & AWK scripts to search 100-million-word linguistic corpus for NLP application.

PROJECTS

Essay scoring   github.com/alexklapheke/...

Built an automatic essay scorer trained on the ASAP corpus.

Reddit   github.com/alexklapheke/...

Built a 70%-accurate prediction model on two similar subreddits, and revealed lexical differences between them, using Python.

Ames, Iowa   github.com/alexklapheke/...

Predicted housing prices using regression models in Python.

SAT vs. ACT   github.com/alexklapheke/...

Revealed insights in standardized test data in Python to find ways to increase nationwide participation.

NPI intervention   DOI: 10.15496/publikation-32629

Gathered linguistic data in 3 languages using Mechanical Turk, and analyzed using linear mixed models in R.

SKILLS

Python   *Matplotlib, NLTK, NumPy, Pandas, scikit-learn*

R   *dplyr, ggplot2, lme4, lsmeans*

Haskell

Scala

BASH

SQL   *MySQL*

Git

LaTeX   *Beamer, TikZ*

HTML5 & CSS

Apache Spark

Amazon Web Services   *EC2*

Amazon Mechanical Turk