DAVID LAING

LinkedIn: https://www.linkedin.com/in/laingdk
Website: https://laingdk.github.io/
davidkendalllaing@gmail.com
1-778-834-1745

Summary

I am an aspiring **data scientist** with an **interdisciplinary education**. I have experience as a **statistics tutor**, a **physics research assistant**, a college-level **technical writing instructor**, and a **business ethics intern**. Through **machine learning**, **data visualization**, and **clean code**, I am committed to enabling sound business decisions.

Education

Master of Data Science, 2016 – 2017
University of British Columbia
GPA: 4.0

Honours Bachelor of Arts & Science, 2010 – 2014
McMaster University
GPA: 3.7

Master of data science capstone project

My three teammates and I are building an instructor-facing **dashboard** for UBC's massive open online courses. The dashboard will help instructors to discover patterns across the structures of their courses, with a special focus on course elements that are beneficial or detrimental to students' learning. Personally, I am writing scripts to **wrangle**, **mine**, and **visualize** the data from course discussion forums, using **natural language processing**.

Relevant work experience

Business Ethics Intern, Fall 2016
Ethical Systems

- I conducted a literature review on **risk assessment** and **decision-making** in the context of **business ethics**. I also researched and wrote a report on file storage options for the Ethical Systems' internal documents.

Technical Writing Instructor, 2015 – 2016
Niagara College Canada

- I planned and delivered over 350 hours worth of lessons and workshops on **technical writing** and **argumentation**. I taught 150-170 students per semester.

Statistics Tutor, Fall 2013
McMaster University & University of Guelph

- I met bi-weekly with two students to help them understand concepts and prepare for tests in ARTSSCI 2R03 **Applied Statistical Inference** and STAT*2080 **Introductory Applied Statistics**, respectively.

Physics Research Assistant, Summer 2012
McMaster Physics Department

- I coded 2,000 physics homework problems for LON-CAPA, an online learning platform, in **Perl**. I also created a tutorial to help instructors develop **animations** in **Adobe Flash**.

Tools

- Python	- numpy, pandas, sklearn, ntlk, matplotlib, many more
- R	- tidyverse, ggplot2, shiny, tidytext, ts, many more
- SQL, LaTeX, RStudio, Jupyter, Amazon Web Services, Travis CI, Docker, bash, Makefile	

Interests

- Reading	- Philosophy, psychology, economics, literature
- Music	- Hip hop, funk, jazz, alternative
- Exercise	- Squash, windsurfing, swimming, longboarding

Relevant assignments from coursework

Machine learning & statistics

- Classified handwritten digits using **convolutional neural networks**, with **Keras** (**Tensorflow backend**) and **EC2** on **Amazon Web Services**.
- Built and tested **recommender systems** using **collaborative filtering** and hybrid models, using sklearn.
- Used **forward selection**, **backward selection**, and **recursive feature elimination** to identify relevant variables from among hundreds, using sklearn and bestglm (R).
- Used **regularization**, **model averaging**, and **Bayesian methods** to prevent overfitting.
- Implemented k-means, k-means++, and k-medoids for **clustering** unlabelled datasets, using R.
- Implemented the **expectation-maximization** algorithm for created mixed models, using R.
- Used **cross-validation** to select among multiple models, using sklearn.
- Compressed and reconstructed images of faces via **principal-component analysis**, using R and sklearn.
- Used **factor analysis** to discover latent variables driving variation in observed data.
- Used **Latent Dirichlet Allocation** (LDA) for **topic modelling**, using Python.
- Created **Markov chain models** of words and natural language.
- Fit and interpreted linear models, mixed effects models, generalized linear models, generalized additive models, LOESS models, splines, and robust models.
- Used **multiple imputation** to deal with missing data.
- Used **Markov-Chain Monte Carlo** (MCMC) for Bayesian models of baseball data, using rjags (R).
- Performed classical hypothesis tests, exact tests, and permutation tests on many datasets.
- Applied **Bonferroni** and **Benjamini-Hochberg** corrections to account for multiple comparisons.
- Modelled **temporal and spatial data**.

Programming

- Built and automated the **testing** of R and Python packages, with **continuous integration**, using TravisCI, pytest (Python), and testthat (R).
- Used **dynamic programming** to write a seam-carving algorithm for image dimension reduction.
- Implemented standard searches of lists and graphs (insertion, binary, merge, depth-first, breadth-first).
- Analyzed **time and space complexity** of algorithms.
- **Scraped** data from static web-pages.
- Conducted analysis of data queried from the **Twitter API**, using tweepy (Python) and twitteR (R).
- Analysed historical word frequencies in the Google n-grams dataset, using **Simple Storage Service** (S3) and **Elastic Map Reduce** (EMR) on Amazon Web Services.

Data visualization

- Built interactive **Shiny apps** to explore encodings and subsets of US crime data, as well as demographics of US colleges.
- Created static **visualizations of network data** (arcplots, network graphs, adjacency matrices), such as Twitter hashtag networks and character networks in Jane Austen novels.
- Created static **visualizations of spatial data** (choropleth maps, dot maps), such as crime rates in Vancouver and earthquakes across the west coast of North America.

Databases & data wrangling

- Queried from **SQLite databases** and a **Google BigQuery** database.
- Designed database schemas in SQLite.
- Queried from XML files and JSON files.
- Applied **k-anonymity** and **l-diversity** to protect against privacy threats.
- Used dplyr to manipulate dataframes, using window functions, grouping functions, and mutations.
- Used standard joins (inner join, left join, full join, etc) to combine datasets with common variables.
- Used **regular expressions** to clean messy text data.