

William E. Lewis

Email: willeslewis@gmail.com

Phone: 501-680-1310

Key Skills and Highlights

Programming	<i>Python</i> (pandas, matplotlib, bokeh, beautifulsoup, scikit-learn, tensorflow, keras), Spark, SQL, and Git. Prior experience with <i>MATLAB</i> , <i>R</i> , <i>Mathematica</i> and <i>C++</i>
Data Science	Web scraping, cleaning data, exploratory analysis, data transformation (standard-normal scaling, t-SNE, SVD, PCA), image/time-series analysis, data visualization, dashboard maintenance, statistical inference (A/B, ANOVA), regression and classification models (k-means, GLMs, boosted trees), deep learning (ANN, CNN, RNN architectures)
Mathematics	Bayesian statistics, matrix factorization, eigenvector/eigenvalue problems, convolution, fast Fourier transforms, short-time Fourier transforms, Monte-Carlo
Achievements	National Barry M. Goldwater Scholarship (2011) 5 peer-reviewed publications in 3 distinct subfields 5 oral presentations: 3 national conferences, 1 international conference, 1 colloquium
Other	Experience working within the Agile framework Scrum Code review and unit tests for python and SQL code Individual and collaborative research with up to 5 team members Mentorship of 2 undergraduate research students

Education

- Ph.D. Physics** University of Colorado May 2018
- *Strongly Interacting Fermi Gases: Hydrodynamics and Beyond*
- Data Science Fellow** The Data Incubator February 2018
- *Churn Prediction in Music Streaming Subscriptions*
- M.S. Physics** University of Colorado August 2015
- B.S. Mathematics, Physics *Summa Cum Laude*** University of Arkansas May 2012
- *Polarization and propagation characteristics of cylindrical vector beams*

Work Experience

- Data Analyst at Jana Mobile** March 2018 - Present
- Identified assumptions of A/B tests were too restrictive thereby discovering $\mathcal{O}(\$1000)$ per day savings
 - Utilized terabytes of high-dimensional data in analyses and feature engineering for predictive models
 - Prototyped a gradient boosted tree model to predict if online purchases would be returned

Research Vignettes

- Fluid dynamics in quantum gases** May 2015 - December 2017
- Utilized MLE in *Python* to fit models to fluid simulation data using AIC for model selection
 - Reduced data output size from $\mathcal{O}(1\text{GB})$ to $\mathcal{O}(10\text{kB})$ in *C++* Lattice Boltzmann simulations
 - Worked with analytical and numerical methods of fluid dynamics
 - Made experimentally verifiable predictions for novel dynamics
- Nano-imaging and spectroscopy** July 2012 - May 2015
- Applied WMA to characterize statistical properties of laser fluctuations in *MATLAB*
 - Conducted short-time Fourier analysis of time-series data using *MATLAB*
 - Fit regression models to experimental data in *MATLAB*
- Polarization Structure of Laser Light** January 2010 - May 2012
- Engineered features to intuitively describe properties of a thirteen-dimensional data set
 - Related theoretical results to experimentally measurable quantities (e.g. laser power)

Additional Activities

Deep learning specialization October 2018

- Deeplearning.ai Coursera 5 course specialization with weekly python programming assignments
- Deep neural networks for image and time-series data (*e.g.* YOLO, VNN, LSTM, Attention Networks)

Board gaming and data science

- Own a library of about 50 board games and enjoy organizing weekly game nights
- Scraped comments and ratings of more than 500 games on *boardgamegeek.com*
- Coded a collaborative filtering based recommendation algorithm in python using scraped ratings

Data Rescue Boston February 2017

- Volunteer hackathon hosted at MIT for preserving data from U.S. government web pages
- Developed web scraping scripts in python with selenium and requests to scrape pdfs and parse html/xml