|  |
| --- |
| **Jonathan D. Groth, PhD** |

**www.linkedin.com/in/jonathan-groth-phd jon.groth1@gmail.com | (337) 581-2789**

**github.com/jgroth1 Princeton, NJ**

**Summary**

* Data Scientist with 12 years’ experience creating solutions with data analysis. With expertise in experimental design, predictive analytics, time series forecasting, and natural language processing; creating clear stories with data in Python, MatLab, JavaScript/HTML.
* Excellent communication skills with the ability to present complex ideas to experts and novices. Experience of 2 years of Teaching, 10 conference presentations, 4 peer reviewed articles, protocol design, and employee training.
* Experienced at developing data analysis methods, statistical inference, and machine learning techniques, including feature engineering, algorithm design. Resulting in 10 conference presentations and 2 peer reviewed conference papers and 2 journal articles.

**Projects**

Commodity Forecasting:

* Performed time series forecasting on coffee Commodity Prices that resulted in an average error of $7.60.
* Compared Lasso and Ridge regression with weather and market predictors and applied ARIMA on the errors to get the confidence interval.
* Found that L1 regularization out performed L2 as L2 overfit the training data.

Natural Language Processing Project:

* The aim of this project was to create a system that a user could upload a resume or CV and the system would automatically select the best job postings to apply for and display the skills needed for for selected jobs.
* Used gensim’s phrases model to extract the common bigrams contained in the corpus.
* Used gensim’s word2vec and doc2vec vector representations of the corpus to determine the most similar job postings to input resume.
* Applied Kmeans clustering with cosine distance to segment the word2vec vectors and extract candidate skills.
* Used candidate skills and a random selection of non-skill phrases to train and test a SVM classifier with a resulting 99% accuracy on the test data.
* Applied the SVM model to the top selected job postings to display skills required for each document.

**Experience**

Data Science Bootcamp – Rutgers University **05/18 – 10/25**

* Designed a time-series forecasting model to predict the commodity price of coffee using Lasso regression and ARIMA.
* Data munging and analysis using python.
* Data visualization and storytelling using python and java script.
* Data collection and parsing using SQL, APIs, and web scraping.
* Machine Learning.

Data Analyst / Research Scientist – Melior Discovery  **08/17 – 10/17**

* Developed new experimental models to be used in the preclinical biomarkers of pharmaceuticals. Resulting in new marketable pharmaceutical testing models.
* Developed algorithms and software for the analysis of data. Resulting in new analysis procedures that enable new experimental paradigms to be brought to market.
* Performed feature selection and feature engineering extracting more robust predictors for analysis.
* Developed exclusion criteria for animal models using analysis measures.

Instructor - New Jersey Institute of Technology  **09/15 - 05/16**

* Communicated complex concepts to a novice audience in a 25 to 32 person lecture and lab course on introduction to engineering concepts and analysis for biomedical engineering and electrophysiology.
* Mentored students on best courses to take, identifying a concentration track, and how to find and choose research and internship opportunities.
* Problem solved student’s lab activities in programing and electronic design then guided them in developing solutions.
* Managed 25 to 32 students in an instrumentation lab course directing students in how to design, prototype, and validate the instrumentation and troubleshoot problems.

Graduate Researcher – New Jersey Institute of Technology **09/08 - 05/15**

* Developed and programmed Machine learning systems for brain computer interfaces using both primary motor cortex and cerebellar signals to forecast arm movements.
* Worked with Team members to improve experimental design and increase integration between different experiments being done on the project.
* Helped other lab members in their problem solving of other projects and helped develop communication schemes to better present information and methods.
* Developed data analysis techniques for the analysis of electrophysiological time series signals from the cerebellum and primary motor cortex resulting in two publications.
* Designed experiments and created protocols for training and experimenting on rats and stimulation and analysis techniques for cutaneous, muscular, cortical, and motor evoked potentials.

**Education**

PhD biomedical engineering - New Jersey Institute of Technology 05/15

MS Biomedical engineering - New Jersey Institute of Technology 05/08

BS Biomedical engineering - Louisiana Tech University 02/05

**Techniques and Software**

|  |  |  |  |
| --- | --- | --- | --- |
| Python SciPy Stack | SQL | Web scraping | Supervised ML |
| JavaScript | NoSQL | Restful APIs | Scikit-learn |
| HTML/CSS | Excel/VBA | Unsupervised ML | TensorFlow |