

Skills

Python • Pandas • NumPy • Scikit-Learn • R • command line
 Machine learning • regression • classification • clustering
 Visualization • Matplotlib • Seaborn • ggplot • Tableau
 Experimental design and statistical analysis

Experience

Metis

Data Scientist Fall 2020

Data collection, modeling, and presentation

- Completed a 12-week intensive data science bootcamp
- Performed detailed analysis of model selection, performance evaluation, and parameter tuning for regression, classification, NLP, and clustering algorithms
- Worked with peers in a fully remote environment to design, execute, and present machine learning research projects **See Projects Section**

University of Massachusetts Amherst

Graduate Researcher 2014-2019

Computational and experimental research on human genome evolution

- Identified neural regulatory DNA sequences with accelerated rates of evolution in humans from experimental and bioinformatic datasets
- Functionally tested 100,000 DNA sequences in neurons and neural stem cells using next-generation sequencing
- Scanned the non-coding genome to identify regions under positive selection in humans and chimpanzees

GlaxoSmithKline

Scientist 2004-2011

Applied microbiology support to antibacterial drug discovery programs

- Supported candidate selection of GSK1322322 by generating and characterizing novel *pdf* and *fmt* mutants that confer spontaneous resistance to peptide deformylase inhibitors in *S. aureus*
- Isolated and characterized etoposide resistance mutations in *S. pneumoniae* to help elucidate the structural basis of etoposide and QPT-1 inhibition of DNA gyrase
- Hired, trained, and supervised student interns and contract employees to meet increased project needs

Contact

pizzollojason@gmail.com

(215) 359-5650

<https://www.linkedin.com/in/jasonpizzollo>

<https://github.com/jpizzollo>

Projects

Identify Stress Using Heart Rate Variability

- Built a random forest classifier to distinguish stressful versus restful activities given a time series of HRV
- Associated changes in HRV with actigraphy data to identify sleep disturbances

Topic Modeling from Podcast Transcripts

- Used natural language processing to identify a core set of topics discussed across a corpus of transcripts
- Identified common structural elements across interviews and episodes according to changes in topic weights

Classifying Risk of Credit Card Default

- Constructed KNN, logistic regression, and random forest models to calculate probability of loan default given demographics and credit payment history
- Evaluated the influence of engineered features that capture patterns of debt accumulation over time

Selected Publications

Pizzollo J, Nielsen WJ, Shibata Y, Safi A, Crawford GE, Wray GA, Babbitt CC. Comparative serum challenges show divergent patterns of gene expression and open chromatin in human and chimpanzee. *Genome biology and evolution*. 2018 Mar 5;10(3):826-39. DOI: [10.1093/gbe/evy041](https://doi.org/10.1093/gbe/evy041)

Chan PF, Srikannathasan V, Huang J, Cui H, Fosberry AP, Gu M, Hann MM, Hibbs M, Homes P, Ingraham K, **Pizzollo J**. Structural basis of DNA gyrase inhibition by antibacterial QPT-1, anticancer drug etoposide and moxifloxacin. *Nature communications*. 2015 Dec 7;6:10048. DOI: [10.1038/ncomms10048](https://doi.org/10.1038/ncomms10048)

Education

University of Massachusetts Amherst

Ph.D. Genomics and Bioinformatics 2019

Lehigh University

B.A. Biology 2002