David Nicholson

dnicholson329@gmail.com • 412-607-6313 • 11291 Chatterly Loop Apt 104, Manassas VA 20109

Summary

Data Scientist with 13+ years of programming experience and 5+ years of experience in data analytics and visualization. Most of my data analytics experience has utilized language models and document embeddings to gain further insight into biomedical research.

Skills & Proficiency

Github • Python • R • SQL • Data Visualization • Machine Learning • Deep Learning • Transformers • Auto Encoders • Clustering • Natural Language Processing • Topic Modeling • Bayesian Modeling • Data Structures • Data Engineering • GCP Data Engineering • Databases • Vector Databases • Embeddings • Algorithms • Parallel Processing • XML Parsing • Web Development

Professional Experience

Data Scientist
Digital Science & Research Solutions, Ltd.
June 2022 - Present

- Constructed a textual processing pipeline to detect cancer drug treatments using a 64k biomedical document set for pharmaceutical clients.
- Backtested vector databases to determine which database is most optimal in dealing with 10M+ document embedding vectors.
- Utilized a combination of a deep-learning based language model and dimensionality reduction algorithm to uncover research topics and trends for various government funders and clients.
- Maintained dashboards that summarized research results and trends for government funders and clients

Graduate Researcher Scientist
University of Pennsylvania
August 2016 - June 2022

- Designed and implemented parallel processing pipelines that achieved a 3x speed-up when analyzing terabytes of biomedical text.
- Used weak supervision for a 1.5x speed-up when training deep learning models (recurrent neural networks and transformers) to extract biomedical relationships from biomedical text.
- Applied a k-nearest-neighbor model to provide scientists with a web service that identifies a listing of journals linguistically similar to a preprint of interest.
- Applied a time series analysis to discover over 20,000 different timepoints where words have changed their semantic meaning.

Publications

- **Unmasking The Language Of Science Through Textual Analyses On Biomedical Preprints And Published PapersOn Biomedical Preprints And Published Papers**
  Nicholson, D. N. (2022)

- **Changing Word Meanings in Biomedical Literature Reveal Pandemics and New Technologies**
  Nicholson, D. N. Alquaddoomi, F., Rubinetti, V., Greene, C. S. (2023)

- **Characterization of the Genome and Silk-gland Transcriptomes of Darwin's Bark Spider (Caerostris darwini)**
  Babb, P. L., Gregorič, M., Lahens, N. F., **Nicholson, D. N.**, Hayashi, C. Y., Higgins, L., Kuntner, M., Agnarsson, I., Voight, B. F. (2022)

- **Examining Linguistic Shifts between Preprints and Publications**
  Nicholson, D. N., Rubinetti, V., Hu, D., Thielk, M., Hunter, L. E., Greene, C. S. (2022)

- **Expanding a Database-derived Biomedical Knowledge Graph via Multi- Relation Extraction from Biomedical Abstracts**
  Nicholson, D. N., Himmelstein, D. S., Greene, C. S. (2022)

- **Constructing Knowledge Graphs and Their Biomedical Applications**
  Nicholson, D. N., Greene, C. S. (2020)

- **The Nephila Clavipes Genome Highlights the Diversity of Spider Silk Genes and their Complex Expression**
  Babb, P. L., Lahens, N. F., Correa-Garhwal, S. M., **Nicholson, D. N.**, Kim, E.J., Hogenesch, J.B., Kuntner, M., Higgins, L., Hayashi, C. Y., Agnarsson, I., Voight, B.F. (2017)

Education

Doctor of Philosophy (Ph.D.), Genomics and Computational Biology; University of Pennsylvania (Philadelphia, PA)

Postbaccalaureate Program (Penn Prep); University of Pennsylvania (Philadelphia, PA)

Bachelor of Science, Computer Science; University of Maryland Baltimore County (Baltimore, MD)