

linkedin.com/in/drradan
github.com/DrRadan

Skills:

Research

Experimental design, hypothesis testing, model validation and benchmarking, predictive models, statistical inference

Machine learning

Supervised & unsupervised algorithms, feature engineering, dimensionality reduction, programming (Python, R, bash), database queries (SQL)

Healthcare

Clinical decision support, personalized medicine, population health, biochemistry, epigenetics, genomic sequencing

Awards:

- Best paper award 2019 (Studies in graduate & postdoctoral education)
- Distinguished citizen award 2019 (City of New Propontis, Greece)
- OpenData data science award 2018 (NYC office of the Mayor)
- Best Paper award 2015 (Epigenetic & chromatin)
- Young Biochemist Award 2008 (UK biochemical society)

Thought leadership (select invitations):

- Towards Data Science (2020)
- IBM Thought Leaders podcast(2019)
- TMLS conference, (2019)
- AI Law PLI institute (2019)
- The Data Skeptic podcast (2014)

Published research:

<https://bit.ly/2HvUaWs>

EXPERIENCE

Data scientist

H2O.ai (CA, USA)

9/2019 - Present

- Developing innovative solutions for value-based clinical decision support. Coordinating a team of three data scientists towards deadline-driven project completion.
- Technical account manager for three healthcare Fortune500 companies.
- Consulting customers in healthcare to help solve a broad spectrum of industry problems: insurance, pharma manufacturing/commercial, clinical trials, molecular diagnostics, patient-facing data products.

Principal data scientist & owner

Common Sense Analytics, LLC (NY, USA)

3/2019 - 6/2019

- Consulted and built a data analysis pipeline for the personalized medicine initiative of a large children's cancer hospital in the US South (subcontractor of Cloudera Inc.).

Research scientist (applied data science)

Langone Health & Courant Institute of Mathematical Sciences, New York University (NY, USA)

9/2016 - 8/2019

- Designed an interdisciplinary research plan (departments of Medicine, Population Health, Computer Science, Engineering), for the integration of health data from various sources in the prediction of childhood obesity in the US using machine learning.
- Designed and implemented a self-contained research project using Twitter as a source of data for the prediction of childhood obesity. Led a team of five data science graduate students towards its successful completion (github.com/DrRadan/TwitterHealth).

Postdoctoral research scientist

*Center for Genomics and Systems Biology, New York University |
Center for Epigenomics, Albert Einstein College of Medicine (NY, USA)*

9/2008 - 5/2016

- Developed the experimental design and quantitative basis for a novel calibration method of genomic sequencing data. Created the foundation for the experimental acquisition of absolute quantities of RNA molecules per cell. Led implementation of the host laboratory research plan. (NYU)
- Developed a quantitative approach for the detection of novel, rare structures in the human genome using genomic sequencing. Led implementation of the host laboratory research plan. (AECOM)

DATA SCIENCE TRAINING

Visiting bioinformatician fellow

NCBI, National Institutes of Health (MD, USA)

10/2017 - 11/2017

- Designed a prototype analytical pipeline for the integration of genetic, clinical, and behavioral data in predictive models of complex disease (github.com/NCBI-Hackathons/Complex_Phenogeno).

Data science fellow

Fast Forward Labs (NY, USA)

6/2017 - 9/2017

- Benchmarked pertained embeddings models for Natural Language Processing.

Data science fellow

Insight Data Science (CA, USA)

6/2016 - 8/2016

- Built an OS web app using NYC's OpenData. Used experimental design to overcome limitations of the datasets (github.com/DrRadan/Plan_t_wise).

EDUCATION

Ph.D. in Cellular and Molecular Biology, University of Edinburgh, U.K.

M.Res. in Biomolecular Sciences, University of York, U.K.

B.Sc. in Biology, Aristotle University of Thessaloniki, Greece.

Professional statement

Data scientist and quantitative healthcare professional with expertise that ranges from molecular research, to experimental design, to machine machine learning

Having a strong molecular biology experimental background, I have embraced since early in my training quantitative analytics to better understanding cellular and human biology. During my PhD in epigenetics, working with professor Adrian Bird at the university of Edinburgh, I was the first person in the team to adopt probe-based microarray assays and built my own computational pipeline for their analysis. As a postdoctoral researcher in New York City, I have continued working in the interface of predictive analytics and molecular biology, studying cell-cycle related epigenetic phenomena and the factors that affect cell growth rate, with applications ranging from aging to cancer. During this time I created and contributed to the development of three new methodologies that combine appropriate experimental design and computational analytics: RNA:DNA IP-seq, RATE-seq, and a maximum likelihood-based calibration method for RNA-seq experiments. The approach in the former methodology received the award "Best Paper of 2015" in the Journal Epigenetics and Chromatin.

Motivated by the early successes of machine learning and its promise in personalized medicine solutions, I honed in on my machine learning skills by completing a series of fellowships. The first project I completed in this journey during my Insight fellowship, successfully combined hypothesis testing together with machine learning to create an end-to-end solution pipeline, and was later awarded the OpenData Data Science award by the NYC office of the Mayor. At Fast Forward Labs, a machine learning research company (now acquired by Cloudera), I expanded my analytics toolbox, by working with natural language processing and text-based quantitative analysis methods. Finally, as a visiting bioinformatician fellow at the NIH, I combined behavioral, genetic and demographic information to prototype a disease prediction pipeline, in a project that I had designed and proposed to NCBI. It is significant that the approach allows us to combine data that capture both "nature" and "nurture" factors, whose non-linear interaction, is well understood, gives rise to the phenotypic (appearance) variability in humans and all living organisms.

At New York University, as a researcher with a focus in applied data science, I worked on preventative personalized medicine approaches using electronic medical record data, and explored additional data sources as data augmentation tools. More specifically, focusing on childhood obesity and in collaboration with pediatricians at the School of Medicine, we demonstrated that machine learning can be successfully used to predict childhood obesity years before the first diagnosis is possible, enabling effective and targeted preventative interventions. Moreover, in a parallel line of work I initiated, we showed that Twitter posts and their metadata contain information (proxies) about obesity rates at the regional level, opening the way for such data to be used as augmentation sources in personalized medicine algorithms.

As a principal data scientist in my company "Common Sense Analytics" and in my current role at H2O.ai I pursue exposure to a variety of healthcare use cases, and help create meaningful solutions for the industry. Importantly, with my focus continuing to be on the application of machine learning to personalized medicine, I lead the development of AI-driven clinical decision support solutions for H2O.ai in the form of self-contained applications. In these approaches, the emphasis is on instance-level model interpretability, to empower physicians in taking the best course of action for the patient.

I look forward to discussing the Data Scientist position and my qualifications with you in more detail. I'm available to talk at your soonest convenience.

Best regards,
Niki Athanasiadou, MRes, PhD