Daniel **Shin**

275 South St. Apt 19B, New York, NY 10002
443.860.1380 | danieldshin@gmail.com | dshincd.github.io

SUMMARY

A skilled and curious data scientist with 4 years of research experience in the biomedical sciences. Experienced in powering through complex problems in short timeframes using industry level data science tools. Strong foundation in statistical analysis, machine learning and creative problem solving.

PROFESSIONAL EXPERIENCE

Metis Data Science Bootcamp | New York, NY | April 2015 - Present
Data Scientist

- Participated in 12-week intensive program focused on building skills used in data science industry
- Completed 5 full-length data science projects using a breadth of statistics and computing tools

Projects

1. Created a web-app that functions as a similarity search engine for startups
2. Built a pipeline for scraping job descriptions on Indeed and performing text analysis & clustering
3. Implemented a classifier to predict heart disease using historical patient data
4. Predicted international box office results using multiple linear regression
5. Identified NYC subway stations with highest commuter traffic using open source MTA data

Weill Cornell Medical College | New York, NY | May 2013 - April 2015
Research Technician

- Optimized a pipeline for efficient collection and processing of biological samples
- Created a system for organizing sample inventories and presented weekly summary reports

NYU School of Medicine | New York, NY | July 2011 - February 2013
Research Technician

- Lead experimentation on a NIH-funded grant under supervision of the Director of Cardiology
- Designed experiments, collected data, and analyzed results using statistical methods
- Presented results and progress to team members and collaborators in weekly lab meetings

SKILLS

- ☑ Statistical analysis and machine learning algorithms
- ☑ Software development and data analysis in Python, R
- ☑ Data collection (API, web scraping) and storage (Postgres, MongoDB, Hadoop, Spark)
- ☑ Data transformation (reshaping, cleaning, type casting, filtering)
- ☑ Scientific Python packages: NumPy, SciPy, scikit-learn, stats models
- ☑ Supervised learning: regression, classification, ensemble methods, regularization, GLMs
- ☑ Unsupervised learning: clustering, topic modeling, dimensionality reduction
- ☑ Natural Language Processing: nltk, text classification using Naive Bayes, vectorization
- ☑ Data visualization: matplotlib, d3.js, ggplot2

EDUCATION

New York University
Bachelor of Arts in Biochemistry (2007-2011)

Johns Hopkins University
Coursera Data Science Specialization (2014-2015)