

# Allan Dong

✉ allan.w.dong@gmail.com ☎ 917 514 1513 📍 Oakland Gardens, NY 11364 in /in/allandong 🌐 allanwdong

## + projects

### Public Health: Socioeconomic Factors in Infant Mortality

Analyzed county-level data for over 240 features across ten years. Used a regularized Linear Regression model to help identify the most significant features related to Infant Mortality.

### Meta-Research: Topics in Cancer Stem Cell Research

Performed topic modeling on 70,000+ Cancer Stem Cell papers from 1980 to 2017 to reveal historical trends of research topics/techniques on a year-to-year basis using word vectorization and clustering with Non-negative Matrix Factorization (NMF).

### Research: Automatic Determination of Cancer Gene Classes

Used word embedding on annotated genetic variant database to classify 9 categories of cancer pathogenicities. Resampled data using SMOTE and oversampling, used Naive Bayes and Logistic Regression to reach 78% accuracy.

### Public Health: Consumers vs Professionals

Acquired Yelp reviews and Health Inspection results on 100,000+ restaurants in New York City, used Linear Regression to determine if/which health code violations were correlated with consumer reviews.

## + experience

### Metis

*Data Scientist*  
2017 to 2017

Metis is an immersive data science bootcamp that combines real-world projects with on-site instruction to develop skills in statistical analysis, machine learning algorithms, feature engineering, advanced database management, and data communication and visualization.

### Rutgers - Office of Advanced Research Computing Newark, NJ

*Data Analyst - Intern*  
2016 to 2017

Pulled, cleaned, and used NLP techniques on Tuberculosis gene mutation data to create a search engine to allow search by related features of gene mutation descriptions.

### Tactile Navigation Tools

*Engineering Technician*  
2014 to 2015

Wrote and won a grant for \$25,000 from the Promobilia Foundation to fund the research, development, and testing of medical devices to aid the visually impaired.

Worked on two projects simultaneously:

- Eyeronman Vest: Oversaw technical team of 3-5 - built Proof of Concept and 2 iterative Prototype models.

- CumbaCane: Acted as product manager - went from developed Prototype model to the final Market model.

### AA Wendon LLC

*Medical Transcriber*  
2011 to 2014

Transcribed medical reports for outpatient cardiology diagnostic/imaging centers and personally located, communicated with, and negotiated deals with centers/radiologists for services.

At peak, was transcribing several hundred reports per months.

New York City, NY

New York City, NY

New York City, NY

## + summary

Data Scientist with roots in healthcare and medicine, and 5 years of experience working at startups. Looking to pivot from analyzing data as a scientist to using machine learning and deep learning in industry.

## + skills

**Microsoft Office:** Word, Excel, Powerpoint, Access

**Programming Languages/Tools:** Python, Pandas, scikit-learn, SQL, PostgreSQL, MongoDB, Amazon Web Services, Spark

**Data Visualization:** matplotlib, Seaborn, D3, Bokeh

**Machine Learning:** See Projects

**Natural Language Processing:**  
See Projects

## + education

Rutgers University, Graduate School of Biomedical Sciences  
Masters Biomedical Sciences  
2017

Washington University in St. Louis  
B.A. Economics 2012