LEON YIN

hello.leonyin@gmail.com
(781) 439-2875
yinleon.github.io

RECENT WORK EXPERIENCE

THE ORCHARD
January 2016 to Present

Data Science Engineer

I am proud to have demonstrated the value of data science and machine learning in the music industry with the following projects:

- ○ *PySpark* to transform data using user-defined functions, join metadata from *Redshift*, and write to *AWS S3*.
- ○ Robust batch jobs from YouTube and Spotify API using *Airflow*.
- ○ Feature engineering, model selection, validation, and prod deployment of random forest for binary classification using *MySQL, Pandas, Scikit-Learn, Google API, and Airflow*.
- ○ Dimension reduction and K-means clustering of albums, artists, and labels using *Scikit-Learn*.
- ○ *Keras* and *Theano* convolution neural network to access the quality of an album based on artwork (WIP).
- ○ Generated microsite and automated data modeling in Looker.
- ○ Point on all internal workflows for structured and semi-structured data wrangling using *Pandas and MySQL to S3*.

NASA GISS
July 2015 to August 2015

Climate Science and Oceanography Intern

Worked on a Fortran numerical model and performed linear regression, jackknife analysis and 2D-interpolation on global seawater measurements using *Pandas, NumPy, SciPy, Matplotlib.*

October 2015 to August 2016

Web Development Consultant (Part-time)

Built QC pipeline for data and an interactive *d3.js* map[1] Research published at the American Geophysical Union[2]

HOBBY PROJECTS

CONGRESSIONAL ACCOUNTABILITY 2017

I know nothing about US Gov't, so I started with congressional ideology using DW-Nominate[5], and wrote a Python wrapper for ProPublica Congress API. Next steps TBD.

FEDERAL FUND TRACKER 2016

NSF records are not stored in a format conducive to analysis at scale. I parsed XML-formatted NSF awards into *SQLite* databases, and traced the funding history of institutions, topics, and PIs using *Seaborn and Plot.ly*[3]. I also created a faulty network graph of investigators based on award co-authorship using *d3.js*.

PALMER LTER INORGANIC ANALYSIS 2015-2016

The poles are oft referred to as canaries of climate change. I performed spatial-temporal analysis on 15-years of shipside measurements along Antarctic's Western Peninsula to understand how the marine carbon cycle— which generates 50% of atmospheric O_2, will respond to climate change[4]

KAGGLE COMPETITIONS 2017

Thanks to the fast.ai MOOC, I was able to quickly deploy state of the art deep learning using Keras and Theano. I jumped into competitions to identify different species of fish, and tumor malignancy based solely on images!

FUNDED PROJECTS

DEAN'S UNDERGRADUATE RESEACH FUND 2015

Designed an experiment on the physiological effects of Ocean acidification on oysters.

NYU GREEN GRANTS 2015

Wrote and produced a short film inspired by Portlandia about composting[6]

EDUCATION

NEW YORK UNIVERSITY

BS in Chemistry, 2011 - 2015
Minor in Computer Science and Math

COMPUTER SKILLS ^{with years of experience >1}

Languages	Python[4], MySQL[3], Bash, HTML, Fortran
Programs	Matlab[2], AWS, Spark, Airflow, Jupyter[2]
OS	Mac OS X, Linux/Unix

LINKS

[1]	bit.ly/d18o_d3js	Bl.ocks page
[2]	bit.ly/agu_poster_v2	AGU poster
[3]	bit.ly/nsf_notebook	Jupyter Notebook
[4]	bit.ly/palmer_notebook	Jupyter Notebook
[5]	bit.ly/nominate_notebook	Jupyter Notebook
[6]	bit.ly/1_mans_trash	Vimeo Video