


Andrew Follmann

Data Scientist

 [follperson.github.io](https://github.com/follperson)

 [follperson](https://github.com/follperson)

 [LinkedIn](#)

 follmann.andrew@gmail.com



EXPERIENCE

THOMSON REUTERS SPECIAL SERVICES | DATA SCIENTIST

July 2020 – Current | McLean, Virginia

- Consult with internal clients about proof of concept research and analysis and customer needs.
- Lead team in Kaggle-like competition to identify and classify fishing vessels in the open ocean on massive images, using SAR satellite data, coming in 7th.
- Built bespoke anomaly models using statistical mixture models, in a large web-app, served through docker and flask.
- Implement extreme classification models for hashtag prediction on Twitter.
- Lead Ethical AI working group, presenting on current topics, mitigation mechanisms, and facilitating company wide conversations.

THE OAKLEAF GROUP | SENIOR ANALYST

December 2015 – July 2019 | Bethesda, Maryland

- Developed ETL routines in Python and SQL for data cleaning, aggregation, and quality control, for use in litigation support and auditing projects.
- Lead small technology team, coordinating development schedules from management and clients.
- Built image data pipeline for mortgage loan documents enabling 180 person team and doubling of YoY revenue.

PROJECTS

SHIP IDENTIFICATION ON OPEN OCEAN | PYTHON (keras, numpy)

2021

- Model for predicting ships in the open ocean in gigapixel Synthetic Aperture Radar (SAR) images. Achieved 7th globally, and 4th in the USA, awarded \$4,500.
- Model was a multi-resolution Constant False Alarm Rate (CFAR) algorithm for object detection, and a Convolutional Neural Network for classification.
- Open source release forthcoming

MIXTURE MODELS | PYTHON, (scipy, numpy)

2021

- Mixture Modeling software package, estimating parameters using Expectation Maximization and BIC.
- Several uni-variate distributions and a few multi-variate distributions
- Open source release forthcoming

FOLLZAM | PYTHON (bs4, tweepy)

2019

- Music fingerprinter, like Shazam. Built as a final project for statistical computing.
- Project Page: <https://github.com/follperson/follzam>

SMART ALARM CLOCK | PYTHON (flask, circuit-python, pyaudio)

2018

- Raspberry Pi based alarm clock, built in python with a Flask front end.
- Slow alarm intensity raising alarm with led lights and USB speaker audio.
- Project Page: https://github.com/follperson/smart_alarm

SKILLS

PROGRAMMING

Proficient:

Python • R • SQL

Experienced:

HTML • L^AT_EX • bash

LIBRARIES/Frameworks

Proficient:

pandas • numpy • scipy • flask
scikit-learn

Experienced:

networkx • tensorflow •
statsmodels

TOOLS/PLATFORMS

Git • Docker • Linux • WSL •
Snowflake • AWS Sagemaker,
S3

EDUCATION

CARNEGIE MELLON UNIVERSITY

MASTER'S IN STATISTICS

August 2019 - May 2020 | Pittsburgh, PA

Cum. GPA: 4.24 / 4.33

Capstone: Estimating Electricity

Congestion Pricing with Regularized

Vector Autoregression

OBERLIN COLLEGE

BACHELORS OF ARTS IN

MATHEMATICAL ECONOMICS (HON.)

Aug 2011 - May 2015 | Oberlin, OH

Cum. GPA: 3.5 / 4.0

Thesis: Geographic Distribution of Urban

Public Services: A Case Study of

Sanitation and Law Enforcement

Expenditure in New York City