Kyle Willett

Data scientist

Seattle, WA
☏ 651.278.2643
✉ willettk@gmail.com
⌂ willettk.github.io
Github: @willettk

▬▬▬ Experience

2016-present **Research scientist**, *Amazon*, Seattle, WA
- Developed and extended ML algorithms used in Amazon Marketplace Pricing and Fees Analytics. Work includes code deployment, statistical analysis, A/B testing, and integration at industry-scale.

2016 **Fellow**, *Insight Data Science*, Palo Alto, CA (remote)
- Built *AboutFace*, a Python-based application that uses image processing and machine learning to identify and recognize humans in online videos. Packaged as a standalone web-service backed by AWS and delivered as a consulting project to **Muse AI**.
- Intensive sessions in machine learning, algorithms, data structures, SQL, and analytics.

2015-2016 **Lead data scientist**, *Galaxy Zoo*, Minneapolis, MN/Oxford, UK
- Oversaw day-to-day maintenance, backend processing, and preparation of new datasets (in Python) for Galaxy Zoo, a citizen science project with $> 250,000$ participants worldwide.
- Created online data archive (AWS/HTML) for both internal and public access to project publications and scientific results.
- Designed and administered a Kaggle competition (with 326 teams participating) to replicate human classification of galaxy images using convolutional neural networks.
- Lead author on two published papers that constitute major catalog releases.

2011-2016 **Research associate**, *University of Minnesota*, Minneapolis, MN
- Researched structure and morphology of galaxies, including the relative influences of galactic bars, disks, and mergers on supermassive black holes and star formation.
- Created new algorithms for measuring and correcting the effects of image quality on volunteer classification accuracy.
- Oversaw image and data processing and wrote the Python/Mongo reduction pipeline for the Radio Galaxy Zoo citizen science project.

2006-2011 **Graduate research assistant**, *University of Colorado*, Boulder, CO
- Ph.D. thesis research on OH (hydroxyl) megamasers in distant galaxies, including discovery of two new megamasers and numerical modeling of physical conditions at maser sites.
- Extensive observing experience with radio, infrared, and optical facilities, including the Very Large Array, Green Bank Telescope, and Spitzer Space Telescope.

▬▬▬ Education

2005–2011 **Ph.D.**, *University of Colorado*, Boulder, CO
Astrophysical and Planetary Sciences

M.S., *University of Colorado*, Boulder, CO
Astrophysical, Planetary, and Atmospheric Sciences

2001–2005 **B.A.**, *Carleton College*, Northfield, MN
Physics (*magna cum laude*)

▬▬▬ Skills

Languages **Python, Scala, SQL, MongoDB, IDL, LaTeX, Mathematica***   * some experience

Tools numpy, pandas, scikit-learn, matplotlib, AWS, Jupyter, Git, MySQL, scikit-image**, OpenCV**, Spark*, Docker*, HTML*, Flask*