Christine Miller

Seattle, WA christine.mf.miller@gmail.com 517 944 0896
github.com/cmfmiller linkedin.com/in/cmfmiller

SKILLS

Programming Languages: Python, SQL, R
Packages : Python: pandas, NumPy, sklearn, plotnine, geopandas, tensorflow, matplotlib
R: tidyverse, ggplot2, Rstan, rspatial
Data Science Methods: Stochastic Modeling, Time Series Analysis/Forecasting, Predictive Modeling, Machine Learning, Cloud Computing, Data Mining, Unstructured Data
Tools: AWS, GitHub, Markdown

EXPERIENCE

Data Science Fellow, *Insight, Seattle, WA* May 2020 - Present

- Designed a web app to provide hikers with live updates and forecasts of parking availability at trailheads using yolo3 to detect cars in webcam images (Python).
- Created robust data pipeline linking AWS and S3 to continuously scape 5000+ images/week from webcams at trailhead parking lots and update current conditions and forecasts in real-time (Python).

Postdoctoral Researcher *University of California, Davis, CA* Jan 2019 – May 2020

- Redesigned regulatory data collection protocols to reduce groundwater nitrate contamination from manure fertilizers.
- Constructed dataset summarizing the management practices of California dairies from unstructured public records using regular expressions (R).
- Utilized Bayesian time-series analysis to show that farmers can underestimate their nitrogen application by 40% by using 2-week old measurements (R).

Graduate Student Researcher *University of California, Davis, CA* Sep 2013 – Dec 2019

- Established extent of uncertainty in farmer measurements of nitrogen fertilization to guide redesign of groundwater quality regulations in California.
- Built stochastic simulations to show that 70% of dairy farmers underestimate the amount of nitrogen they apply to forage crops, calling for data quality improvements (R).
- Organized 5-person research team to improve accuracy of farmer collected data by up to 50% over industry standards using resampling algorithms and exploratory analyses (R).
- Combined Monte Carlo and numerical groundwater models to generate 50 GB datasets to stochastically quantify nitrate leaching as lead statistician and programmer on an interdisciplinary team (Python, R).

Data Quality Intern *Climate Corporation, San Francisco, CA* Jun 2017 – Sep 2017

- Predicted yields of corn and soybean to fill gaps in yield maps created when data collection machinery malfunctioned to provide farmers with complete harvest information.
- Used satellite imagery to develop multiple regression models to lower RMSE of corn and soybean yield predictions by up to 30% compared to interpolation models (Python, R).

EDUCATION

Ph.D. ANIMAL BIOLOGY *University of California, Davis, CA* Dec 2019

M.S. STATISTICS *University of California, Davis, CA* Jun 2019

B.A. BIOLOGICAL SCIENCES *Wellesley College, Wellesley, MA* May 2011