

John Miller

JOHNWMILLR@GMAIL.COM | JOHNWMILLR.COM | PORTLAND, OR

EXPERIENCE

EXTRAHOP – CYBERSECURITY THREAT DETECTION AND ANALYSIS

Senior Data Scientist

MAY 2024 - PRESENT

DATA SCIENTIST III

SEP 2021 - MAY 2024

- Led development of a completely novel entropy-based ransomware detector, coordinating across multiple teams
- Constructed an AWS Glue pipeline capable of processing billions of records for domain and IP reputation scoring
- Wrote and reviewed customer-facing production code for ML-based real-time threat detection in network data
- Owned end-to-end development and deployment of an internal data analysis package for the data science team

BILLUPS – GEOSPATIAL ANALYSIS FOR MARKETING AND ADVERTISING

DATA SCIENTIST

FEB 2020 - JUL 2021

- Overhauled company's campaign measurement pipeline, reducing necessary manual work from days to hours
- Implemented an automated PySpark workflow, decreasing computation time by 20% and minimizing human error
- Completed advertising campaign attribution and lift studies for dozens of clients across the United States

ZOOMINFO – DATA ENRICHMENT AND CLASSIFICATION FOR BUSINESS SALES INTELLIGENCE

DATA SCIENTIST

OCT 2018 - FEB 2020

- Collaborated with data engineers, product managers, and executive leadership to develop and implement machine learning and deep learning models for automated data cleaning and enrichment
- Trained a regression model to predict company employee counts, improving private company estimates by 15%

GARVIN IMAGE LAB – AUTOMATED DIAGNOSTIC TOOLS FOR DISEASES OF THE RETINA

MASTERS FELLOW

AUG 2016 - JUN 2018

- Developed novel combinations of shape analysis and image processing techniques for disease classification
- Trained random forest classifiers to determine the cause of optic disc swelling with 86% accuracy

HUMAN SPINAL CORD RESEARCH LAB – NEUROSCIENCE RESEARCH AND EXPERIMENT DESIGN

RESEARCH ASSISTANT

SEP 2014 - AUG 2016

- Designed, conducted, and analyzed experiments investigating the therapeutic effects of spinal cord stimulation
- Wrote and maintained Matlab codebase for the research team's analysis and visualization workflows

PROJECTS

[LYRICSGENIUS](#) – PYTHON WRAPPER FOR DOWNLOADING SONG LYRICS AND METADATA FROM GENIUS.COM

FOUNDER & PRINCIPAL DEVELOPER

- Actively maintaining package on GitHub for over 30k users, coordinating open-source collaboration
- *Trucks and Beer* music lyrics analysis project awarded "best data-driven stories of 2018" by *The Pudding*

[MUSCLE MIDI MUSIC](#) – ARDUINO / IPHONE DEVICE FOR TRANSLATING MUSCLE ACTIVITY INTO ELECTRONIC MUSIC

SOLE DEVELOPER

- Designed and built an electrical circuit and wrote software that generated and controlled music via muscle signals
- Winner out of 200+ entrants in the "Sensors Contest 2017" on Instructables.com

[MUSICIAN MAKER](#) – COMPUTER-CONTROLLED SYSTEM ALLOWING NON-MUSICIANS TO IMPROVISE EXPRESSIVE MUSIC

RESEARCH SCHOLAR

- Wrote multi-threaded Python code to generate computational music, interfacing with novel hardware instruments
- Selected as a finalist in the 2012 Guthman New Musical Instrument Competition at Georgia Tech

SKILLS

CODE: Python, pandas, scikit-learn, SQL, PySpark, Matlab, NumPy, Jupyter, Flask, Matplotlib, seaborn, Altair

TOOLS: AWS (SageMaker, Glue, DynamoDB, Athena, RDS, Lambda, S3), Databricks, Git, GitHub/GitLab, Docker, Linux

TOPICS: Machine learning, image processing, signal processing, natural language processing, data visualization

EDUCATION

UNIVERSITY OF IOWA – MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

2018

GOSHEN COLLEGE – BACHELOR OF ARTS IN PHYSICS, MINOR IN ENGLISH

2014