

ALEXANDRE DALY

Data Scientist

@ alexandredaly@gmail.com

☎ (402) 429-8361

📍 New York City, NY

🐙 github.com/azdaly

🌐 linkedin.com/in/azdaly

EXPERIENCE

Insight Data Science Fellow

New York City, NY

📅 January 2020 – Present

- Developed WeTube, a tool for YouTube creators to quickly visualize their comments section and understand the conversation, reducing the time required to read comments (wetube-insight.com).
- Established a pipeline for summarizing YouTube comments in real-time using BERT for sentence embedding t-SNE for dimensionality reduction, clustering using HDBSCAN, and topic modeling using LDA. This pipeline proved useful in discovering insights missed by YouTube's analytics platform.
- Built the web application with Dash and hosted on an AWS EC2 instance, deployed using AWS Elastic Beanstalk.

Graduate Researcher

Ann Arbor, MI

📅 January 2014 – Present

- Discovered regulatory pieces of DNA that are critical for the proper function of both the pituitary and thyroid gland by analyzing terabytes of High-Throughput Sequencing (HTS) data on the University of Michigan's High-Performance Computing cluster. Validated results both in vivo and in vitro.
- Performed quality control, batch normalization, alignment, signal detection, and data visualization on both DNA and RNA using both published and independently developed algorithms on a previously unexplored cell type.
- Published eight manuscripts and gave six conference presentations on discovering a genetic modifier for thyroid-related deafness, single-cell sequencing of the pituitary, developing a new mouse strain, and discovering an enhancer required for proper pituitary function.

PROJECTS

🐙 SangTK

Developed a model combining LSTM RNNs and logistic regression to improve upon the sequence quality of Sanger sequencing data. The model outperforms field-standard sequence prediction algorithms by 10%.

🐙 UCSC-Track-Prep

Software package that prepares genomic and transcriptomic data for visualization.

🐙 SangUM

Software package for batch processing and download of Sanger sequencing data from the University of Michigan Advanced Genomics core.

SKILLS

Programming Languages

Python Bash SQL

Tools

AWS Elastic Beanstalk AWS EC2
AWS Lambda Docker Git
Snakemake Scikit-learn

Machine Learning

NLP models Neural Networks
Supervised and Unsupervised models

Quantitative Methods

Genome-Wide Association Studies
Likelihood Estimation PCA
Regressions

EDUCATION

University of Michigan

PhD in Genetics and Genomics

📅 Exp May 2020

University of Michigan

Master of Science in Genetics and Genomics

📅 May 2016

University of Nebraska-Lincoln

Bachelor of Science in Biological Systems Engineering

📅 December 2013

AWARDS

- 🏆 1st place at the Structural Birth Defects Meeting Elevator Speech Competition, Bethesda, MD
- 🏆 2nd place at UM Department of Human Genetics Elevator Speech Competition, Ann Arbor, MI