

# DANRONG (ALISON) LI

✉ [dvl5760@psu.edu](mailto:dvl5760@psu.edu) · ✉ [@danrongli](mailto:danrongli) · ✉ [danrongLi](https://www.linkedin.com/in/danrongLi) · 🌐 [Homepage](#) · 🌐 [State College, PA 16801](#)

## 🎓 EDUCATION

---

### Penn State University, State College, PA

08/2023 – Present

Ph.D. in Computer Science and Engineering

Advisors: [Young Kun Ko](#) and [Stefan Canzar](#)

Research focus: Coreset theory, metacell partitioning, and single-cell transcriptomics

- [Coreset-based logistic regression for atlas-scale cell type annotation](#) - In Submission
- Coreset Accelerated Metacell Partitioning (CAMP) for atlas-scale single cell data - In Submission

### New York University, Brooklyn, NY

09/2021 – 05/2023

M.S. Computer Science

Coursework: Algorithmic Machine Learning and Data Science, Statistical and Computational Foundations of Machine Learning, Artificial Intelligence, Human Computer Interaction, Design and Analysis of Algorithms, Principles of Database Systems

### Columbia University, New York, NY

09/2019 – 12/2020

M.S. Industrial Engineering

Coursework: Data Analytics, Optimization Models, Stochastic Models, Machine Learning, Analytics on The Cloud, Business Analytics, Financial Engineering, Simulation

### University of Illinois, Urbana-Champaign, IL

09/2015 – 05/2019

B.S. Industrial Engineering with Honors

Coursework: Analysis of Data, Deterministic Models in Optimization, Industrial Quality Control

## ✍ RESEARCH EXPERIENCES

---

### New York University Brooklyn, NY

08/2022 – 05/2023

Researcher, Advisor: [Christopher Musco](#)

- [Matrix Product Sketching via Coordinated Sampling](#) - Published ICLR2025
- [Sketching Methods for Approximating Matrix Multiplications](#) - M.S. Thesis

### Inky Technology New York, NY

07/2020 – 09/2020

Research Intern

- Constructed a fingerprint-like stylometry feature database with lexical, structural and content properties extracted from email body and header sections to capture senders' unique writing styles
- Evaluated each sender's words' richness standard deviation based on Wikipedia word frequency dataset
- Experimented various feature weights, evaluation metrics and Machine Learning multiclass classification models (Random Forest, Linear SVC, K-Neighbors, Multi-layer Perceptron)
- Communicated weekly progress with company executives through reports and presentations

### Mayco Industries Granite City, IL

09/2018 – 12/2018

Senior Design Research Project

- Performed Python Text Mining with production information to pinpoint major reason for machine downtime
- Created Tableau visualizations to determine the optimal initial feed vendors for Mayco Industries

### University of Illinois Urbana-Champaign, IL

03/2017 – 08/2017

Researcher, Advisor: [Randy Ewoldt](#)

- Built Yield-Stress Fluid design database to allow users to select materials with desired rheological properties
- Developed a matrix of reported Yield-Stress values and tip diameters during extrusions for 3-D printing applications

- Contributed to [manuscript](#): Rauzan, B. M., Nelson, A. Z., Lehman, S. E., Ewoldt, R. H., Nuzzo, R. G., *Adv. Funct. Mater.* 2018, 28, 1707032.

## RELEVANT PROJECTS

---

### **Input Face Image to Output Similar Faces** Brooklyn, NY

01/2022 – 05/2022

Artificial Intelligence Class Project 

- Integrated pre-trained VGG19 model with unsupervised K-Nearest Neighbor model to create baseline
- Improved the baseline searching accuracy with Facenet model and Elastic-KNN approach

### **Web Application to Inquire Complex Database Information** Brooklyn, NY 09/2021 – 12/2021

Principles of Database Systems Class Project 

- Created Entity-Relation (ER) diagram with 8 entities, 7 relationship sets and 6 business rules
- Translated the ER diagram into relational schema and implemented the schema into a relational database on university-hosted server
- Populated the database with web-scraped data
- Designed user-asked complex inquiries and wrote SQL to retrieve them
- Generated a UI with drop-down menu for better user inquiry experience

### **How COVID-19 affected people's mobility around the world** New York, NY

09/2020

Columbia University Data Science Hackathon, Team Bagel, 3rd Place 

- Utilized SQL to obtain mobility reports and hospital confirmed cases statistics from Google BigQuery data warehouse and extracted relevant features to build model inputs
- Web-scraped tweets using Tweepy API from Twitter with keyword COVID-19 and retrieved TextBlob sentiment scores
- Created a prediction algorithm on the next-day mobility with both Linear Regression Model and Machine Learning Models (Random Forest Regressor, Decision Tree Regressor, Gradient Boosting Regressor)

### **Instacart Grocery Orders Analytics** New York, NY

01/2020 – 05/2020

Business Analytics Class Project

- Used 2-segment Weibull Statistical Model to distinguish and identify different user groups for Instacart website in order to customize promotional plans: motivate active consumers, attract new users
- Utilized Linear Regression Statistical Model to capture customer grocery re-order time interval fluctuations in various departments and aisles

### **Amazon Reviews Sentiment Analysis** New York, NY

09/2019 – 12/2019

Data Analytics Class Project

- Processed, filtered and solved the imbalance of Amazon food reviews dataset
- Used Valence Aware Dictionary and Sentiment Reasoner (VADER) to analyze and give scores to reviews
- Created WordCloud visualizations to showcase discrepancies between VADER-reasoned review scores and customer-judged scores
- Created a more reasonable scoring algorithm with machine learning models (Logistic Regression, Ordinal Regression, Long-Short Term Memory, Multinomial Naïve Bayes)

### **Web Application with Open Source Squirrel Data** New York, NY

09/2019 – 12/2019

Python Class Project 

- Created a web application with Django framework and deployed through Google Cloud Platform with functions of importing, exporting, viewing and editing squirrels' characteristics, including habits, gender, age
- Used JavaScript library Leaflet and OpenStreet map data to construct an interactive map with squirrels' locations

## TECHNICAL SKILLS

---

- Programming Languages:** Python, C++, Scala, SQL
- Web/Software Tools:** Google Cloud Platform, Amazon Web Services, Django, Anaconda

- **ML/Analytics:** Scikit-Learn, TensorFlow, Keras, Spark, Hadoop, PySpark, SciPy, Pandas, NumPy, NLTK

## HONORS CERTIFICATES

---

University Graduate Fellowships 	09/2023
New York University Graduate School of Engineering Scholarship	09/2021 – 05/2023
Bernt O. Larson Project Design Award 	04/2019
Richard Baxendale Outstanding Junior Award 	04/2018
Illinois Leadership Certificate	11/2017
Lean Six Sigma Green Belt Certificate	11/2016

## PROFESSIONAL AFFILIATIONS

---

- New York University Tandon Algorithms and Foundations Group 
- Industrial Engineering Honor Society (Alpha Pi Mu)
- Engineering Ambassador of University of Illinois Grainger College of Engineering
- Society for Women Engineers (SWE)
- Columbia University Data Science Society