Ericka Wu

(650)665-0660
ericka.wu@columbia.edu
http://www.erickawu.github.io

EDUCATION

Columbia University **Aug 2017 - May 2021**

Dual B.A. in Statistics and Computer Science: GPA 3.78, Major GPA 3.90

Relevant Coursework: Causal Inference, Distributed Systems, Artificial Intelligence, Machine Learning, Statistical Inference, Data Structures, Algorithms, Data Visualization, Probability, Databases

Programs and Activities: Peer Health Exchange, MIT Grand Hack, Columbia Music Performance Program, Alpha Kappa Psi, Data Science Society

SKILLS

Programming: Python, Java, Javascript, Scala, SQL, Golang, R

Technologies & Frameworks: Luigi, Spark, PostgreSQL, HTML/CSS, Kubernetes, Kafka Connect

Visualization & Design: d3, Tableau, Sketch, Figma

EXPERIENCES

Columbia Data Science Institute | Data for Good Scholar **Oct 2020 - Present**

- Building a probabalistic record matching algorithm to match and track migration of African North Americans

Datadog | Data Engineering Intern **Jun 2020 - Aug 2020**

- Built and scheduled ETL pipeline using Luigi and Spark that consolidates new metrics metadata weekly from Postgres tables with over 50 million rows, refactoring legacy code and removing a two-day long query
- Created and deployed Kubernetes cronjobs to monitor lag for data creation in AWS and GCP
- Implemented a Luigi script that streamlined the manual process of deleting and replenishing of timeseries data of multiple resolutions, improving on-call experience for team members

ServiceNow | Software Engineering Intern **May 2019 - Aug 2019**

- Wrote server-side Javascript to integrate maintenance as an attribute for software entitlement records
- Designed and implemented an iOS feature that allows users to scan and receive hardware assets without a purchase order
- Created a software contract parser with named entity recognition during internal team hackathon

Lionbase | Data Science Team **Jan 2019 - May 2019**

- Built a baseline resume parsing and scoring pipeline trained on in-house recruitment and interview data
- Pioneered a recommendation system to suggest candidates to companies based on technical skill and cultural fit

Columbia University Biomedical Engineering | Research Assistant **Dec 2017 - May 2019**

- Trained neural network to track mouse paw and pupil movements in 2D videos to observe and visualize behavioral patterns
- Constructed a data preprocessing pipeline that compresses, color corrects, and exports raw brain imaging videos
- Visualized and auralized types of mouse brain activity by attributing sets of piano chords to normalized matrices representing location and intensity of pulses

PROJECTS

AINetwork (RidgewayPartners & LionBase)

- Invitation-only data science recruitment platform with a dual portal web application for both companies and applicants
- Participated in client calls, discussed product features, and managed deadlines for successful deliverables

The Face of the Average European Portrait at the Metropolitan Museum

- Used OpenCV to detect facial landmarks, align features, and perform Delaunay triangulation to generate an average face
- Created visualizations communicating the aesthetic and artistic changes of the human face over time using d3

Prediction of Clinical Response to Anti-PD-1 and Anti-CLA Immunotherapies in Cancer

- Trained a stack of denoising autoencoders to create patient representations from genetic mutation and pathway data
- Evaluated representations by classifying patient response to immunotherapy. Representations reduced patient data by 98% while maintaining accuracy

Inter-Metropolitan Migration in the United States

- Designed and built an interactive, scroll-driven narrative that investigated the relationship between metropolitan-to-metropolitan migration and real personal income