Edward **Yu**

US Citizen · Machine Learner · Software Developer

Columbus, OH
614-707-3495
YuEdward4@gmail.com

My years of experience developing, presenting, and moving new products to customers as a Chemist led me to the world of Data Analysis. I hope to use my extensive analytical background along with garnered experience in my current Data Analytics Masters program to pivot into a Data Science role.

Education

Georgia Institute of Technology
M.S. in Data Analytics — *Atlanta, Georgia* — *2021 - Present*

- Expected graduation Aug. 2023

We Can Code IT
Certificate of Software Development — *Columbus, Ohio* — *2022*

Marietta College
B.S. in Biochemistry — *Marietta, Ohio* — *2003 - 2007*

Skills

Languages: Python, R, Hadoop, Spark, JavaScript, D3, Tableau, HTML, CSS, SASS, SQL, Java, GitBash, ARENA
Technical: TDD, OOP, React, REST, Data Cleaning, Hyperparameter Tuning, Visualization, Imputation, Regression (Linear, Multiple, Poloynomial, Random Forest, SVR), Classification (KNN, SVM, Random Forest), Clustering (K-Means, Hierarchical), ANOVA, LASSO

Projects

Exploration and visualization of large-scale online video uploads metadata — *2022*

This collaborative project, inspired by the "Butterfly Effect" podcast, sought to visualize trends of user-generated tags on videos hosted by a major video website. User-generated communities and their evolution over time was visualized.

- 3-million video URLs scraped from online corpus utilizing scrapy python library subsequently cleaned and transformed for analysis.
- Anomaly detection via angle-based-outlier allowed easy comparison of anomalous tags, top 50 anomolies visualized.
- Community detection and visualization using user-generated tags using both Leiden and Louvain clustering algorithms.
- Visualizations presented in both RMD utilizing R, and D3 via AWS server.

Oil tanker simulation and optimization — *2022*

Inspired by the notorious "Ever Given" debacle which caused massive supply chain issues around the world. ARENA simulation software was used to quantify and optimize ship de/berthing, un/loading, and travel times utilizing a single tug traveling to and from berths to harbor.

Predicting NFL draft selection using NFL combine data — *2021*

Using available NFL Combine data and NFL Draft Selection data drew insights on which measured attributes contributed to draft pick status for NFL prospects.

- Deep cleaning and pruning, MICE imputation of joined data with respect ot preventing bias.
- Variable selection using LASSO allowed for classification of individuals with specific attributes into different position rankings.

Work Experience

ASK Chemicals LP
Dublin, Ohio — *2016 - Mar. 2022*

Research and Development Chemist

- Implemented data dashboard to track plant manufacturing defects in contrast with past performances, optimizing a previously entirely manual process, utilizing R, R-Shiny, Excel, CentOS VM.
- Performed cost analysis on industry specialty products, involved testing of new chemical compositions and optimization of formulations, utilized design of experiments and statistical comparisons in MATLAB and R software.
- Developed advanced Excel templates to aid in laboratory requests, progress, and results. Further implemented into MS Access database to allow for tracking and advanced analysis.
- Advanced 3D-printing manufacturing process using RFID, Proxmark tools, resulting in $5MM USD realized cost savings.
- Consistently employed Six Sigma methodology, experimental design, statistically rigorous analysis.

Product Specialist — *2014 - 2016*

- Constructed and administered MS Access database for inventory control and safety data sheet management. Conformed to OSHA standards for hazardous chemicals control.
- Developed a method of high-throughput, automated time series analysis using R, reducing technician processing times from 10 minutes to 10 seconds.