YUEYUE HELENNA YIN

New York, NY | 757-947-7995 | yy2294@cornell.edu

linkedin.com/in/yueyueyin

EDUCATION

Cornell Tech (Cornell University) New York, NY   May 2026
MS in Information Science
Honors: Merit Scholarship, Related Company Fellowship
Relevant Courses: Machine Learning Engineering, Networks & Markets, Applied Machine Learning, Algorithms

College of William and Mary, Williamsburg, Virginia   May 2023
Bachelor of Science in Data Science | Major GPA: 3.95
Relevant Courses: Databases, Data Visualization, Deep Learning, Natural Language Processing, Adv. Applied ML

TECHNICAL SKILLS

Programming Language:	Python(Pandas, Numpy, Tensorflow, Pytorch, Sklearn, Scipy), SQL, R, C#, YAML
Tools:	Git, AWS, ArcGIS, Tableau, Apache Airflow, HTML, CSS, Hadoop, Spark

PROFESSIONAL EXPERIENCE

Uniqlo, Data Strategy, Data Science Intern, Shanghai, China   May 2024 - Aug 2024
- Developed forecasting models to predict future in-store traffic across multiple dimensions using ensemble techniques combining time series model(Prophet) and LightGBM. Achieved an error rate (assessed using MAPE) of around 4%
- Enhanced forecasting efficiency by 95% and accuracy by 8% compared to the previous manual forcasting process, providing robust foundations for setting retail sales goals, resulting in an average 5% increase in sales achievement rate
- Streamlined the forecasting process with an automated data pipeline using Airflow to manage and schedule tasks. Configured the pipeline to run either on regular schedule for fixed-length or on-demand for custom forecasting periods
- Conducted ad-hoc analyses to meet business needs using SQL & Python. Notable projects include analysis on customer retention following store closure, leading to 12% increase in customer retention rate through locating core customers
- Created interactive dashboards with Tableau to analyze sales events, provide sales and marketing insights for stakeholders

Institute of Integrative Conservation, Research Analyst, Remote/Arlington, Virginia   Feb 2022 - Nov 2022
- Spearheaded comprehensive statistical analysis on correlation between spatial datasets using Python & ArcGIS, computed various correlation coefficients to evaluate substitutability, provided alternative solutions for datasets difficult to update
- Leveraged spatial autoregressive models, logistic regression, and random forest to predict correlations between datasets, selected potential features impacting results through literature review, engineered features, and achieved 92.3% accuracy
- Regularly maintained, updated team databases, performed data extractions, produced visualizations for over 10 reports

ADDITIONAL EXPERIENCES

Break Through Tech, AI Studio Course Support, New York,NY   Aug 2024 - Present
- Guide student teams through machine learning projects sponsored by major companies (Google, JPMC, Pfizer, etc)
- Develop project scopes with student teams, track progress, manage timelines to ensure successful and on-time delivery
- Review and evaluate technical project and code notebooks, providing actionable feedback and strategic guidance

PROJECTS

Identify Gender Stereotypes in Movie Script with Natural Language Processing (Python)   Mar 2023 - May 2023
- Developed tools with NLP algorithms to assess gender stereotypes in movie scripts with named entity recognition to predict character gender based on context and LDA topic modeling to assess the complexity of topics discussed
- Developed web scrapers to retrieve movie scripts from IMSDb, utilized tools like Selenium and Beautiful Soup for information extraction and data cleaning, cleaned and prepared over 200,000 pieces of movie conversations for traning

Job Search Assistance Tool Development (Python, SQL)   Feb 2022 - Mar 2022
- Developed job search assist tool with NLP to match user resume, preferences, job descriptions from job search websites
- Ranked job openings based on candidates' qualification and provided feedback to users, optimizing job search process
- Set up and configured SQL databases using AWS RDS, normalized data tables, automated the process of writing service data into the database and fetch info from databases to present to users