Yuhan Yao

yuhan_yao@g.harvard.edu | 917-495-4827 | Portfolio Website | GitHub | LinkedIn | Medium Blog

EDUCATION

Harvard University, School of Engineering and Applied Sciences, Massachusetts, US — Expected graduation date: 12/2023
Master of Science in Data Science

New York University, NYU Shanghai, Shanghai, China — 09/2018 - 05/2022
Bachelor of Science in Data Science (AI track); Minor: Mathematics — Cumulative GPA: 3.94/4.0
Study Abroad: NYU Abu Dhabi from 09/2021 to 12/2021; NYU New York from 01/2022 to 05/2022

PROFESSIONAL EXPERIENCE

Data Science Intern, PayPro Global, Remote at Ukraine Office — 02/2021 - 05/2021
▪ Utilized Python to implement K-means and Hierarchical Clustering methods to transform customer recency, frequency, and monetary values, which segmented customers into 3 target groups and engineered 8 features for further prediction
▪ Constructed and finetuned a customer lifetime value prediction XGBoost model with 85%+ accuracy, assisting the marketing team to refine customer target strategy
▪ Created and designed a Power BI data visualization report featuring 16 plots on webpage template performance, enabling the frontend team to debug hidden template errors and optimize template functionality
▪ Presented business insights to CEO and team leader and wrote requested executive summary detailing model mechanism and suggested marketing strategy for senior leadership

Artificial Intelligence Intern, Shanghai Hyron Software Co., LTD, Shanghai, China — 06/2020 - 08/2020
▪ Preprocessed and extracted structural information from driver's license photos using Python OpenCV and trained a 95%+ accurate CRNN model to recognize numbers, dates, and 7000+ Japanese characters, saving time and manual labor for DMV
▪ Implemented an automated pose detector of abnormal behavior for AirPods factory safety check using YOLOv5 and Resnet18 models, preventing theft and larceny
▪ Collaborated with 7 other team members and successfully delivered 2 fully-deployed AI products to clients in 3 months

Data Scientist & Software Developer, Coopsight, LLC, Shanghai, China and Abu Dhabi, UAE — 04/2020 – 01/2022
▪ Managed and maintained data extraction and modification of Firebase NoSQL database, ensuring stable and smooth data connection between frontend and backend
▪ Built and tested 6 web pages using React.js and Tailwind.css, with features including embedded Google map, company searching, user matching, enhancing user experience
▪ Tested business hypotheses and product-market fit by cold-calling 50+ people and interviewing 20+ potential customers and industry experts to propose data-driven solutions for product iteration

RESEARCH & PUBLICATION

Undergraduate Student Researcher, *NYU Shanghai, Shanghai, China* — 05/2021 - present
Received Dean's Undergraduate Research Fund; Paper to be published in September; Advisor: Prof. Zhibin Chen Demo | GitHub
▪ Proposed a tailored Genetic Algorithm based on Python to solve a black-box optimization problem and devised an improved shuttle bus schedule, which reduced cost by 6.82% while satisfying students' demand
▪ Formulated a real-life vehicle scheduling problem into 2 variations of Spatio-temporal networks and constructed a non-closed form objective function with 3 real-life constraints
▪ Led a team of 2 students and 1 shuttle service supervisor and organized bi-weekly meetings with advisor

History Beyond: Approaches to Messy Digitized Archival Documents, Shanghai, China — 08/2020 - 09/2020
Research Assistant; Advisor: Prof. Heather Ruth Lee — Project Website | GitHub
▪ Implemented Optical Character Recognition using Python OpenCV and Google Tesseract to recognize English words in ancient fonts, digitalizing and preserving historical documents
▪ Utilized Python Pandas package to wrangle and organize tabular data with 18,000+ entries from Chinese Restaurant Database
▪ Presented research outcome to NYU Shanghai Chancellor, Provost, Dean and Professors

Open Source Swarm Intelligence Robotics Research, Shanghai, China — 03/2019 - 12/2019
Research Assistant; Advisor: Prof. Rodolfo Cossovich — Hackaday
▪ Designed, 3D-printed, and tested infrared positioning device of robot "Swarmesh" and built scalable and decentralized swarm intelligent robots from scratch
▪ Compared Swarmesh with Kilobot, Jasmine and R-one to identify the limitations of existing Swarm robot systems
▪ Published and presented paper Framework for Present Swarm Robotic Systems and New Implementations to Increase Scalability at SWARM 2019: The 3rd International Symposium on Swarm Behavior and Bio-Inspired Robotics in Japan

SCHOLARSHIPS & HONORS

Summa Cum Laude	05/2022
Good Wood Global Scholar, NYU Shanghai	2021 - 2022
Dean's Undergraduate Research Fund, NYU Shanghai	05/2021
Dean's List 2019-2021, NYU Shanghai	2019 - 2021
Presentation at Academic Conference (PAC) Grant, NYU Shanghai	11/2019
Global Quintessence Scholarship ($39,000 awarded for top 1%), NYU Shanghai	2018 - 2022

ACTIVITIES

Technical Writer, Towards Data Science and Towards AI (Medium Publications)	07/2021 - Present
Open-source Dataset Contributor, Kaggle	08/2020 - Present
Senior Student Worker, Career Development Center, NYU Shanghai, Shanghai, China	09/2019 - 05/2020
Treasurer, MUBOtics Club, NYU Shanghai, Shanghai, China	03/2019 - 12/2019
Marketing Leader of Yanghuo Project, Match Edu (NGO), Shanghai, China	06/2019 - 08/2019

DATA SCIENCE COMPETITION & PROJECTS

Bechdel Test: Comparing Female Representation Metrics in Movies, Data Analysis & Visualization Project GitHub
Showcased quantitatively that more females on set can translate into better female representation on screen; Visualized trend of female representation evolution using Python Seaborn; Obtained Bechdel scores of 9,300+ movies over 150 years using API. improvement advice on creating a superior metric.

NL2SQL: BERT-based Model for SQL Generation, Natural Language Processing Final Project GitHub
Designed and built a BERT-based slot-filling classification model that converted questions in Chinese into SQL statements, which enabled non-programmers to interact with SQL databases effortlessly in Q&A scenarios.

SAC Bias Reduction: Clipped Double Q vs Multi-Step Method, Reinforcement Learning Final Project GitHub
Tested and compared clipped double Q, multi-step method, and the combination of them in terms of SAC bias reduction efficacy for algorithm optimization.

Chinese Traffic Sign Recognition Based on Annotated Street View Images, Machine Learning Final Project GitHub
Created and trained a self-designed artificial neural network on 6000+ images and tested it against VGG16 and Resnet50 to accurately classify 58 categories of Chinese traffic signs.

Video Streaming Platform: Which Has Better Shows, Regression & Multivariant Data Analysis Project GitHub
Utilized Minitab to conduct a two-way ANOVA analysis on the effects of video streaming platform and genre on TV show ratings; Concluded that platform and genre are two main effects with no interaction effect.

Online Air Ticket Reservation System, Databases Final Project GitHub
Designed and implemented a relational database using MySQL; Created a frontend user interface using HTML, CSS, Bootstrap, and JavaScript to connect with the backend Python Flask app.

1st Place in Eleme Delivery Analytics Under COVID-19, Kaggle Data Science Competition Kaggle | GitHub
Built, compared, and optimized various Machine Learning models such as logistic regression and XGBoost to pinpoint relevant factors conducive to predicting couriers' decisions and behavior; Narrowed down the error of delivery time prediction to around 1.5 minutes.

SKILLS

Data Science: Python (NumPy, Pandas, Matplotlib, Plotnine, Scikit-learn), R (ggplot2, tidyverse), SQL, Firebase NoSQL
AI: Python (Pytorch, OpenCV), Machine Learning, Computer Vision, Reinforcement Learning, Natural Language Processing
Software Development: HTML, CSS (SASS/SCSS, Tailwind), JavaScript (React.js), Python Flask
Language: Chinese (Native), English (Professional Work Proficiency)