Haitian(Enoch) Liu

haitian.job@gmail.com | (347) 925-3966 | linkedin.com/in/enoch-liu | U.S. Citizen

EDUCATION

Harvard University <div align="right">Cambridge, MA</div>
MS, Data Science (Cross-registration at MIT) <div align="right">*Aug. 2022 - May. 2024*</div>
Relevant courses: Computer Vision, Natural Language Processing, System Development, Probability and Statistics

University of California, Berkeley <div align="right">Berkeley, CA</div>
BA, Applied Mathematics; BA, Data Science <div align="right">*Aug. 2019 - Dec. 2021*</div>

EXPERIENCE

Harvard Business School [*Database | Python | MySQL | Stata | JavaScript | HTML*] <div align="right">Cambridge, MA</div>
Research Associate on Impact Investments Project <div align="right">*Feb. 2024 – Present*</div>
- Optimized ETL pipelines for data transformation and consolidation in PCRI database, resolving discrepancies from varied data types across 30+ data vendors over 100,000+ private capital companies, achieving 98% data integrity
- Redesigned and developed Private Capital Research Institute (PCRI) website via HTML5, JavaScript and CSS

Nock Lab, Harvard University [*LLM | PyTorch | HuggingFace*] <div align="right">Cambridge, MA</div>
Research Assistant Supervised by Ph.D. Yaniv Yacoby and Taylor McGuire <div align="right">*Sep. – Dec. 2023*</div>
- Curated the first benchmark dataset for detecting LLM-generated text in mental health surveys through human subjects simulation using LLMs, text augmentation and rigorous quality validation
- Developed tree-based, transformer-based, and few-shot in-context learning models for LLM text detection, enhancing survey data quality and achieving 96.2% test accuracy
- Presented at *Women in Data Science Conference 2024* to hundreds of professionals from industry and academia

YunExpress USA [*Python | GeoPandas | Tableau | HTML | Google Apps Script*] <div align="right">Vernon, CA</div>
Data Scientist <div align="right">*Apr. – Aug. 2022*</div>
- Developed a custom website for the Operations department, boosting data processing efficiency by 20%
- Built a pallet recycling tracking system and automated email solution, reducing costs by $4 million annually
- Utilized data analysis and visualization techniques (e.g. Tableau dashboard, GeoPandas) on 24 million parcel records to improve the package delivery operations, resulting in a 1.2 days average reduction in delivery time

Department of Mathematics, UC Berkeley [*PDE | MATLAB*] <div align="right">Berkeley, CA</div>
Research Assistant for Prof. Di Fang on Bifurcation and Clustering of Biological Networks <div align="right">*May. – Dec. 2021*</div>
- Designed numerical algorithms of biological networks models, especially on leaf venation
- Built an efficient semi-implicit marching algorithm that preserves gradient flow structure
- Calculated Fréchet derivative for the gradient flow of the constraint optimization problem

PROJECTS

Sentiment Analysis for Tinder Review using Natural Language Processing [*NLP*] <div align="right">Feb. – May. 2023</div>
- Utilized various NLP techniques (e.g., simple Recurrent Neural Networks, Bidirectional Long Short-Term Memory, pre-trained word embeddings (GloVe) on Bidirectional Gated Recurrent Units, simple transformer, BERT, and Roberta) to categorize the sentiments expressed in Tinder app reviews, achieving accuracy of 94%

MRI Brain Tumor Image Diagnosis and Classification [*CV | TensorFlow*] <div align="right">Feb. – May. 2023</div>
- Implemented and built our own deep convolutional neural network (e.g., VGG Network, Residual Network, and Google Inception Network) to classify MRI images into meningioma, pituitary, glioma, and no tumor. Evaluated model performance by confusion matrix and F1 score, achieving test accuracy of 97%

Mortality prediction in General Population [*Survival Analysis | Cox Regression | Elastic Net*] Oct. – Dec. 2022
- Constructed predictive models leveraging Regularized Regression (e.g., Ridge, Lasso), Principal Component Regression, Random Forest, Boosting, Neural Networks, and survival prediction techniques in the National Health and Nutrition Examination Survey (NHANES) Follow-up Study, achieving an accuracy of 83%

SKILLS

Programming Languages: Python, Java, SQL, R, JavaScript, HTML, CSS, MATLAB, LaTeX
Platform & Tools: Git, Unix/Linux, Docker, Tableau, Jupyter Lab, R Studio, VS Code, PyCharm, IntelliJ
Libraries: PyTorch, TensorFlow, HuggingFace, pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, D3.js, Bootstrap
Mentorship: Teaching Fellow for *AC 207: System Development* at Harvard University
Entrepreneurship: Founded Triple Six Group, supplying beauty products to 3000+ global SMEs via digital marketing