Alan (Hongyao) Shao

hongyaoshao@gmail.com | (617) 840-4101 | www.linkedin.com/in/hongyaoshao

EDUCATION

Boston University **Boston, MA**
Bachelor of Science in Data Science & Business Administration Expected May 2027

- **GPA:** 3.84 / 4.0
- **Honors:** Dean's List, all semesters

DOMAIN KNOWLEDGE & TECHNICAL SKILLS

Domain knowledge: Data Processing & Visualization, Experimental Design, Statistical Modeling, Website Development, Supervised Machine Learning, Web Scraping, Sentimental Analysis, Feature Engineering
Technical Expertise: Python (Pandas, NumPy, Scikit -Learn, Matplotlib, Seaborn, SciPy), Rust, Java, LaTeX, SQL, TensorFlow, Tableau

WORK EXPERIENCE

BU Faculty of Computing & Data Science **Boston, MA**
Course Assistant Jan 2025 – Present

- Graded and provided detailed, constructive feedback on coding efficiency, algorithm design, and mathematical proofs to enhance student understanding for CDS DS 320, Algorithms for Data Science.
- Led group tutoring sessions, explaining complex topics such as dynamic programming, graph algorithms, and probabilistic methods, adapting explanations to different learning styles.

Hongfu Investment **Shanghai, China**
Investment Analyst May 2025 – Aug 2025

- Utilized web scraping with Python (BeautifulSoup) to extract and structure financial data of 20+ tech firms, including advanced materials and data labeling companies, for fundamental analysis, processing and clearing the data to assess market position, growth drivers, and risks to facilitate investment decisions.
- Leveraged SQL to clean and prepare large datasets of raw financial data (over 1,0000 rows), and utilized Python to calculate revenue CAGR, assess margin trends, and determine valuation multiples, which enabled rigorous peer benchmarking and informed strategic analysis.
- Conducted competitive analysis by benchmarking the short-video strategies of over 100 peer investment firms and developed interactive dashboards to visualize key findings, which led to a successful initial expansion plan.

DeFiner Labs **San Mateo, CA**
Data Analyst Dec 2024 – Feb 2025

- Conducted in-depth research on various cryptocurrencies and tokens, identifying market trends and growth opportunities through comprehensive on-chain analysis.
- Utilized SQL to transform over 75,000 rows of raw blockchain data into actionable insights, creating interactive visualizations in Dune Analytics to track key metrics like account type and transaction volume.

ACADEMIC PROJECTS

Progressive Mass: Home Rule Petition Sep 2025 – Dec 2025

- Built an end-to-end data pipeline using the MA Legislature Public API to collect, clean, and structure 10 years of MA Home Rule Petition data, producing a standardized dataset with petition details, timeline, and outcomes.
- Analyzed petition lifecycles by mapping legislative workflows, identifying approval and failure patterns, and developing a refiling fatigue analysis dashboard to quantify repeated filings and their impact on success rates.
- Delivered findings through story-driven presentations to the client and stakeholders, translating complex legislative analytics into clear insights and visual narratives to inform policymaking and organizational strategy.

Emotion Recognition from Audio Using CNN-Based Spectrograms Feb 2025 – May 2025

- Developed a Python/TensorFlow pipeline that transforms raw audio into mel-spectrograms, utilizing Librosa for feature extraction; Implemented robust data augmentation, including pace, pitch, and noise manipulation, to train resilient convolutional neural networks for audio emotion classification.
- Conducted a structured, methodical hyper-parameter sweep to find the optimal configuration for separate models trained on speech and song, systematically adjusting key parameters like the learning rate, the number of layers, and batch size, which significantly improved F1 score by 50% than baseline.
- Developed and delivered end-to-end notebooks with real-world audio inference and clear documentation, which enabled rapid deployment in human – computer interaction and call-center analytics workflows.

Delivery Performance Dynamics Analysis Oct 2023 – Dec 2023

- Conducted statistical testing using OLS regression, ANOVA to ensure model accuracy and assess key correlations between personnel performance, delivery times, and customer satisfaction.
- Utilized Python to develop machine learning models, including Random Forests, to predict delivery time and analyze factors impacting overall efficiency and delivery performance, achieving over 0.85 F1 score.