Huan Shuo Hsu

10027, NY, USA | (347)-727-8391 | huan246shuo@gmail.com | https://www.linkedin.com/in/huan-shuo-hsu | https://github.com/HiMyNameIsBoxxy | https://himynameisboxxy.github.io/

EDUCATION

Columbia University - *M.S. in Applied Analytics*   *Sep. 2024 – Dec. 2025*
- Courses: Machine Learning, Data Modeling, Data Mining, Statistics, Cloud Computing (AWS), SQL, Python

Pratt Institute - *B.Arch. in Architecture | Magna Cum Laude, Dean's List*   *Aug. 2018 – May 2023*

- Courses: Calculus, Physics, Structural Engineering, UX Design

WORK EXPERIENCE

Data Scientist Intern - UX Research   ***Pasadena, CA***
Pulse Protocol   *May 2022 – Sep. 2022*

- **Conducted user behavior analysis** using **SQL**, improving app navigation and **boosting engagement by 15%**.
- Improved **conversion rate** for medical procedures by **10%** using **Tableau** for **exploratory data analysis**.
- Designed and analyzed **A/B tests** for UI features using **Python**, increasing **user retention by 20%**.

Cloud Server Administrator   ***Remote***
Stardust Lab   *Feb. 2020 – May 2022*

- Managed **cloud-based servers** using **OVH** and **AWS** to **host** and **scale game servers**.
- Implemented **security best practices** for cloud infrastructure, including **firewall configuration**, **DDoS protection**, and **data encryption**.
- Monitored **server performance** with **Prometheus** and **Grafana**, ensuring **99.99% uptime**.

Data Analyst Intern   ***Taipei, Taiwan***
Atelier-3   *May 2019 – Sep. 2019*

- Performed **data cleaning**, **data extraction**, **transformation** using **Excel**, **SQL**, and **Python**, ensuring accuracy and consistency of **property data** for analysis.

PROJECT EXPERIENCE

Click Through Rate Prediction with XGBoost

Data Science & Machine Learning   *Oct. 2024 – Nov. 2024*
- Leveraged **XGBoost with Bayesian Optimization** for **hyperparameter tuning**, achieving a significant reduction in RMSE for Click-Through Rate prediction.
- Leveraged statistical tests and importance scores from **linear regression**, **decision tree**, **random forest** to systematically refine **feature engineering**, improving model accuracy and reducing computational complexity.

Live Query of Stock Market Price Change with Kafka

Data Engineering   *July 2024 – Sep. 2024*
- Developed a **real-time data pipeline** to stream stock market data using **Python** and **Apache Kafka**.
- Deployed **Kafka** on **EC2 instances**, processed and **stored stock market data** in **AWS S3**, and implemented process optimization utilizing **AWS Glue** to **catalog the data** for **efficient querying** with **AWS Athena**.
- **Automated the real-time ingestion and processing pipeline with Zookeeper,** ensuring continuous updates and high system reliability.

SKILLS

- Programming Languages: Python, R, C, HTML/CSS
- Tools / Frameworks: SQL, Postgres, NoSQL, Tableau, Power BI, Docker, Tensorflow, Pytorch, Pandas, Matplotlib, Apache Kafka, AWS EC2, S3, Glue, Athena
- Languages: English (Native), Mandarin (Native)