Huachen Shan

New York |hs3478@cumc.columbia.edu|LinkedIn

EDUCATION

Columbia University, New York, NY Sep 2023 - May 2025
MPH in Biostatistics. *Certificate: Advanced Epidemiology. GPA: 3.9/4.0*
Relevant Courses: Regression Analysis, Categorical Data Analysis, Public Health, Data Visualization, Epidemiology
Award: Championship at the Columbia University Biomedical Engineering Hackathon: Sepsis Detection

The Chinese University of Hong Kong, Shenzhen, Shenzhen, China Sep 2019 - May 2023
BS in Applied Mathematics. *Stream: Financial Mathematics. GPA: Top 20%, Dean List Award*
Relevant Courses: Stochastic Process, Machine Learning, Time Series, Optimization, Statistical Inference, Real Analysis

SKILLS

Programming: Python (Numpy, Pandas, Sklearn, Selenium), R, SQL, SAS, MATLAB, Visual Basic
Packages : Power BI, Tableau, Excel (Power Pivot, Power Query), Pytorch, Git, Tensorflow, Matplotlib
Analytics Skills: Machine Learning, ETL Pipeline, Exploratory Data Analysis, Google Analytics, A/B Testing

EXPERIENCE

Shenzhen Research Institute of Big Data, Shenzhen, China Jan 2023 - Aug 2023
Data Analyst Intern

- **Low Vacancy:** Led the development of a prediction system for Vienna Hotels. Deployed a combination of **time series prediction algorithms: Arima and Garch.** Boosted monthly revenue by **12%** with **99.5%** confidence.
- **System Reliability:** Designed and implemented unit testing, integrated testing, and stress testing to ensure the high availability of over **99.99%** for core services. Deployed **Prometheus metrics** to continuously monitor infrastructure health.
- **Monitoring Infrastructure**: Started the monitoring effort and integrated unified logging pipeline and Prometheus metrics to monitor through slack alerts. Slashed system-wide reaction time to production incidents by **83%**.
- **Spark ETL Pipeline:** Developed **ETL pipelines** for five hundred records to be applied to analytics dashboards. Visualize behavioral patterns and user preferences. Effectively help distinguish reward patterns between infrequent and active customers. During promotional events, **KPI** for infrequent customers rose by almost **30%**.
- **Machine Learning Systems:** Developed 6 complex algorithms for the prediction modeling system, each using a combination of models such as **Random Forest**, **Boosting techniques**, and **Artificial Neural Networks**. Effectively fine-tuned over 100k model parameters through the utilization of random search methodologies.

Harvard Medical School, Boston, MA Sep 2022 - Jan 2023
Medical Data Analyst Intern

- **Literature Review:** Reviewed 5+ articles and wrote reports on basic theories of **data computing**.
- **Mortality Prediction:** Performed **data visualization** on non-sensitive electronic medical records to better predict the mortality rate of patients using **Python** and packages. Built **forecasting models** with hundreds of search indices to have **15%** narrower therefore more accurate intervals. Improves prescription refills, and ICU capacity.

PROJECTS

A/B Experimentation Platform | SQL, Python Jul 2023 - Aug 2023
Instacart Shopper Hiring Funnel Optimization

- **Experiment Design:** Designed and implemented an experiment to test different placements of background checks in a flow to maximize qualified candidates. Decided **key metrics** to be collected through front-end events.
- **Decision Science:** Based on **conversion rate (CVR)** and cost **per acquisition (CPA)**, concluded statistically significant evidence to reject the **null hypothesis**. The experiment estimated an improved full-time employee conversion rate of **19.81%** on the site. Provided strategic marketing recommendations for the onboarding process.