

Zheyang Liu

(929) 786-0747 | zheyang.liu@columbia.edu | New York, NY | www.linkedin.com/in/zheyangliu999

EDUCATION

Columbia University Mailman School of Public Health, New York, NY

Expected May 2023

Master of Science in Biostatistics, Specialization in Data Science

Relevant Coursework: Data Science (R), Biostatistics Method

Shanghai University of Finance and Economics (SUFU) - School of Statistics and Management, China

June 2021

Bachelor of Science Statistics

Relevant Coursework: Data Mining (Python), Machine Learning, Computational Programming (C++), Mathematical Analysis, Probability, Advanced Mathematics, Regression Analysis (R, SAS), Random Process, Database Theory (SQL, C#), EDA (R)

PROFESSIONAL EXPERIENCE

Ping An Technology (Shenzhen) Co. LTD, Shanghai, China

April 2021 – July 2021

Data Scientist Intern

- Used PySpark and SQL to query data, and conduct data mining on user behavior based on business insights
- Optimized feature engineering codes to improve efficiency and complete modeling preparations for core projects
- Use LightGBM, XGBoost and other methods to establish underwriting models (auto insurance, property insurance); Research the feasibility of implementing the GNN auto insurance anti-fraud model

Shengqu Information Technology (Shanghai) Co. Ltd, Shanghai, China

October 2020 – April 2021

Data Analyst Intern

- Responded to data analysis demands of the operation department related to a popular mobile game, using Hive/Spark to query data and produce automatic data report with Plotly on Python and Tableau
- Conducted research on user payment/loss issues to provide data-driven decision support for the operation teams
- Ensembled XGBoost, Random Forest to predict if the user will churn with AUC 0.92

Hylink Digital Solution Co. Ltd Shanghai Office, Shanghai, China

August 2020 – October 2020

Data Analyst Intern

- Wrote Python code to update daily/weekly report and built a new data dashboard on Excel to demonstrate performance of campaign to our client (PayPal) with convenience
- Undertake exploratory data analysis (using Python and Tableau) towards weekly campaign data on several media such as Baidu and Google to offer business insights to our client

RESEARCH EXPERIENCE

Sentiment Analysis of Covid19 Internet Public Opinions | (Graduation Thesis), Shanghai, China

April 2021 – May 2021

- Performed exploratory data analysis based on the text data related to the epidemic situation in Weibo.
- Used methods such as back translation, Bert synonymous replacement, and random replacement for data strengthen.
- Establish and tune BERT and FastText-based SVM and XGBoost models to predict sentiment tendency with AUC 0.89
- Applied voting model to the crawled epidemic-related Weibo data, and analyzed online public opinion of Covid

Empirical Covid19 and market volatility | Research Assistant, Shanghai, China

April 2020 – June 2021

- Determined value for each parameter in SEIRD model through reading plenty of related papers and found the underlying relationship with each variable
- Programmed using Python to simulate data under seven different circumstances
- Recorded findings and optimized scripts to simulate data efficiently and increase detailed markdown and annotation
- Conducted sensitivity test on parameters to give Covid-control policy suggestions

EXTRACURRICULAR EXPERIENCE

Minister, Volunteer Dept of SUFU School of Statistics and Management, Shanghai, China

September 2017 – June 2019

Leader, Field Research & Volunteer Teaching in Sichuan Liangshan Yi Autonomous Prefecture, China

July 2019

SKILLS

Computer: Proficient Python, R, SQL, Intermediate SAS, C/C++, Tableau, SPSS

Language: Chinese & Shanghaiese, English (TOEFL 105), GRE (325)