KAIZHI LU

2063846950 | Seattle, WA | kaizhilu@uw.edu | www.linkedin.com/in/kaizhilu | Lucas1500.github.io

SUMMARY OF QUALIFICATIONS

Experience data scientist with strong foundation in statistics and solid programming skills in Python and SQL. Actively looking for data scientist position. Proven track record of projects in machine learning and data analytics.

EDUCATION

University of Washington (GPA: 3.80/4.00)   **Seattle, WA**
-Master of Science in Biostatistics Capstone   Expected Mar 24
-Bachelor of Science in Economics & Statistics: Data Science   06/2021
- Magna cum laude

SKILLS

Programming: Python, R, SQL, Excel, Tableau, HTML, CSS, JS
Machine Learning Techniques
- Exploratory Data Analysis (EDA), Experimental Design, Hypothesis Testing, A/B Testing
- Principal Component Analysis (PCA), Regularization, Feature Engineering, Model Evaluation
- Logistic Regression, Random Forest, K-Means, Gradient Boosting

PROFESSIONAL EXPERIENCE

Beijing Tellhow Intelligent Engineering Co., Ltd.   **Beijing, China**
Data Technology Department Intern   08/2021-04/2022
- Collaborated with 8 analysts in compiling the bidding document of "Management Cockpit for ZGC Group" by researching 20+ software system functions, resulting in a total 100+ pages documents
- Improved the management cockpit's evaluation of the smart city's operating and monitoring system by 5% through formulating 7 first level indicators

Orient Securities Investment Banking Co., Ltd.   **Beijing, China**
TMT Analyst Intern   07/2020-01/2021
- Updated the target company's 2020 Semi-Annual Financial Report through analyzing financial statements and business summary reports to attain data, advancing the overall progress by 8%
- Generated 4 tables and mind maps in Excel and Xmind to visualize the target companies' net growth trend, forming template summary tables to increase productivity to do tasks by 10%

PROJECT WORK

Glove Efficacy Evaluation Research   **Seattle, WA**
Advisor: Prof. Diana M Ceballos (UW)   01/2023-Present
- Apply R software to create validated database by generating 1500+ logarithmically distributed imputed values
- Perform exploratory data analyses to examine the relationship between sensor loadings and glove materials

Statistical Data Science for Public Health and Biomedicine
Advisor: Prof. Steven Ma (Yale University)   08/2021-10/2021
- Applied techniques such as up-sampling and AUC of the ROC curve to deal with and evaluate the imbalanced dataset
- Trained a penalized logistic regression model using 10 indicative biomarkers to predict HCV infected stage
- Authored "Logistic Regression in Biomedical Study" independently, included in the proceedings of BLSME 2022

Banking Customer Churn Prediction and Analysis
- Developed algorithms to predict customer churn probability based on labeled data via Python programming
- Trained supervised machine learning models including Logistic Regression, Random Forest and K-Nearest Neighbors, and applied regularization with optimal parameters to overcome overfitting
- Evaluated model performance of classification via k-fold cross-validation technique and analyzed feature importance to identify top factors that influence the results