Baitong Liu

213-544-5575   baitongl@usc.edu
Los Angeles

EDUCATION

University Of Southern California   Los Angeles, CA

M.S. Applied Data Science - Viterbi School of Engineering   Jan 2021 - May 2023

GPA: 4.0/4.0; Foundations of Data Management(A), Machine Learning for Data Science(A), Analysis of Algorithms(A)

Donghua University   Shanghai, China

B.S. Statistics - Schoole of Science   Sep 2016 - Jun 2020

GPA: 3.7/4.0; Regression Analysis(A), Bayesian Statistical Inference(A-), Non-parametric Statistics(A), Data Analysis with Python(A)

Scholarship: Special Scholarship - Contest, Academic Scholarship, Community Scholarship

ACADEMIC PROJECTS

Machine Learning Models Implemented on Real-World Dataset based on Scikit-Learn   Los Angeles

Academic Projects   Sep 2021 - Nov 2021

- Preprocessed and visualized data by Numpy, Pandas, Matplotlib, and Seaborn; applied SMOTE to deal with imbalanced data.
- Implemented supervised and unsupervised methods (KNN, Regression, SVM, Tree-based Method, K-means, etc) by scikit-learn packages.
- Constructed semi-supervised learning, passive learning, and active learning methods by L1-penalized SVM.
- Applied CV, Regularization, Recursive Feature Elimination modify models to prevent overfitting.
- Evaluated models by ROC, AUC, Confusion Matrix, MSE, etc.

Detecting COVID-19 with Chest X-Ray using PyTorch   Los Angeles

Kaggle Dataset Project   Jan 2022 - Jan 2022

- Conducted custom dataset, image transformations, and prepared data loader.
- Created and trained ResNet-18 model and apply it to a Chest X-Ray Radiography Dataset based on PyTorch.
- Visualized custom dataset and result of predictions by Matplotlib.

PROJECT EXPERIENCE

Natural Language Processing Program   Shanghai, China

Research Assistant of Mathematics and Statistics   Jan 2019 - Nov 2019

- Collaborated team members to learn the knowledge of probability and statistics required for machine learning, based on Statistical Learning Methods, MLaPP, ESL; produced more than 10 notes and slides
- Studied the mathematical problems in papers related to Statistical Learning Methods, derived the theories
- applied statistical knowledge to the university-level project practice of the natural language processing project team

Mathematical Contest In Modeling and Interdisciplinary Contest In Modeling   Shanghai, China

First Prize Team Leader   Dec 2018 - Jan 2019

- Preprossed about 200,000 pieces of data through Excel, python, and MATLAB, including data cleaning, data visualization, and descriptive statistical analysis
- Performed Principal component analysis (PCA) through SPSS, selected main influencing variables after dimensionality reduction , and optimized model by stepwise regression analysis
- Designed criterion of drug analysis throughK-means clustering method
- Organized team members to write the paper "Analysis of the Opioid Crisis and Strategies"

SKILLS, CERTIFICATIONS & OTHERS

- **Knowledge:** Conceptual Understanding of Classical Statistical Machine Learning Theory
- **Certifications:** DATA ANALYSIS TO UNDERSTAND CARDIOVASCULAR DISEASE
- **Languages:** Python(Machined Learning Framework), SQL, Excel
- **Student Work:** Course Producer(Mentor) of Machine Learning for Data Science