In Son Zeng

Statistical data analyst and graduate student instructor with 3+ years of experience in data analytics, natural language processing and statistical consulting using Python, R, SQL and SAS. Seeking to leverage my technical and instructional expertise at the University of Michigan to develop myself into a top scientific researcher.

Skills

❶ Advanced machine learning skills (Regression, LDA, SVM, KNN, RF, AdaBoost, Neural network, K-means, GMM)

❷ Proficient in Python data analytics (Pyspark, Pandas, Scikit-learn, Seaborn, Bokeh, scikit-optimize)

❸ Adept at SQL querying (data aggregation, AB Testing, metrics visualization)

❹ Proficient in natural language processing (anago, spaCy, prodigy, nltk, pytorch, keras)

❺ Effective in statistical consulting (time-series analysis, spatio-temporal data analysis)

❻ Advanced Linear Models techniques (GAM, GEE, GLMM, Random Effect Model)

Software

❶ Experienced in Python programming (Machine learning, Natural language processing, Data mining, Big data querying, Statistical modeling, Visualization)

❷ Proficient in R programming (Statistical computing, Linear, survival and Bayesian models, Machine learning, Visualization)

❸ Adept in Overleaf, Latex and R Markdown (Building courseware, Writing slides)

❹ Strong GitHub management (Creating personal website, Organizing instructional materials and Collaborating projects)

❺ Experienced in Final Cut Pro, Adobe Premiere Pro (Video editing and producing)

Interests

❶ Music Composition

❷ Big-data Analytics

❸ Volunteer Activities

Personal Info

Address
1783 Weatherstone Drive,
Ann Arbor, MI 48108, U.S.A.

Phone
734-882-9536

E-mail
insonz@umich.edu

GitHub
https://github.com/son520804

LinkedIn
https://www.linkedin.com/in/in-son-zeng-566b9b148/

Work Experience

2019-03 - present

Data Analyst

Groundspeed Analytics, Inc.

- Developed automated text-labeling system to remove PII label insurance claims using BI-LSTM-CRF model, Regex and spaCy package, resulting in a 97% F-1 score
- Provided visual instruction using Lucidchart and organized 150+ page operational documentation at Jira and Confluence website to streamline Data Analyst's workflow
- Conducted quality assurance and audit for 10000+ policy, loss run and exposure files

2018-06 - 2019-03

Research Assistant in Michigan Model of Diabetes (MMD)

University of Michigan, Biostatistics Department, Supervisor: Wen Ye

- Participated in building, simulating and debugging the MMD with six sub-models with 350+ variables and functions, through agent-based modeling in Java-based Anylogic
- Synthesized knowledge from 10+ recent clinical reports to calibrate the risk factors of type 2 diabetes in the MMD
- Specialized in compiling a 50-page MMD user manual detailing the complete tutorial of MMD to support the clinical evaluation in University of Michigan Health System and disease prevention in Public Health departments

2018-09 - 2018-12

Graduate Student Instructor

University of Michigan, Statistics Department, Professor: Jack Miller

- Conducted student-centric and interactive tutorials during weekly and remote office hours
- Wrote 23 class notes and 3 exam reviews using Latex and Overleaf to organize the course materials and summarize the key concepts taught during the lectures, and videotaped the solution of extra review questions (uploaded via GitHub and YouTube)
- Graded approximately 200 students' assignments bi-weekly and wrote 8 homework feedbacks to address major misconceptions
- GitHub Courseware: https://github.com/son520804/STATS-412-Revision-Materials

Education

2017-09 - 2018-12

University of Michigan, Ann Arbor, M.S. Statistics

- Specialization: Agent-based Modeling, Bayesian Modeling, Statistical Consulting
- Projects: Opioid Prescription Analysis, Measuring Latent Political Ideal Points

2013-09 - 2017-06

University of Macau, B.S. Mathematics

- Specialization: Mathematical Modeling, Agent-based Modeling, Theoretical Statistics
- Scholarships: Macau Foundation Scholarship, Scholarship for Outstanding Academic Achievement, Admission Scholarship

Courses

❶ Python data analytics, Statistical computing and consulting in Health Analytics (Mastered data mining and visualization skills, and conducted 10 big-data studies with written reports)

❷ Machine learning and Bayesian modeling and inference (Familiarized with supervised and unsupervised machine learning algorithms and Bayesian neural network in Python and R)

❸ Stochastic Processes, Time-series, Regression Analysis (Studied the application of Linear modeling and Spatio-temporal analysis in Python and R)

Certificates

2017-04 Meritorious Winner at 2017 COMAP Interdisciplinary Contest in Modeling

2016-11 The Second Prize at China Undergraduate Mathematical Contest in Modeling

Conferences

2018-07 2018 Symposium on Big Data, Human Health and Statistics

2018-04 2018 Michigan Student Symposium for Interdisciplinary Statistical Sciences (MSSISS)