YUAN DU

Melbourne, FL 32940

https://yuan-du.com/ • https://www.linkedin.com/in/yuaneldaif/

SENIOR DATA SCIENTIST

• 8+ years of experience in Prospective and Retrospective studies, Healthcare outcomes, Medical Research, Clinical Trials
• Techniques: SQL, R (Tidyverse, glmnet, caret, Rmarkdown, ggplot), Python (Numpy, Pandas, Scikit-learn, Matplotlib, Pytorch), SAS, Julia, Spark, Github version control
• Proficient in using statistical tools to conduct data analysis for research publications and business decisions with effective communication to solve business problems and execute solutions
Deep Learning: Generative Models, EBMs, GANs
Machine Learning: Penalized regression (Lasso, Ridge), Support Vector Machine, Decision Tree, Random Forest, GBM, Clustering (K-Means, GMM), Dimension Reduction (PCA, MDS, t-SNE), Recommendation System
Statistical Modeling: Generalized Liner Model (GLM), Mixed model (GLMM), Hypothesis Testing, Propensity Score Matching, Randomization, Power Analysis, Survival analysis, Meta-analysis, Bayesian methods, Time series

EXPERIENCE

Mindoula Health   Melbourne, FL
Senior Data Scientist   *Dec, 2021-present*
• Collaborate with data engineers, data analysts, project manager for projects such as propensity score matching, and build machine learning models on population health projects using Microsoft Azure Databricks, Python, R, SQL.
• Report to VP and Executives directly for technical consulting and discussion.

Research Institute, AdventHealth   Orlando, FL
Senior Statistician   *Jun, 2015-Dec, 2021*
• Collaborate with researchers via statistical consulting service on EHR & other Database (Cerner, Athena, EPIC, Premier, Cancer registry), Power Analysis, Study Design, Machine Learning, Statistical Method on study areas such as Lung cancer, Leukemia cancer, Surgery, Pediatrics, Radiology, Sleep disorder, COPD, CHF, Covid 19, Population Health, etc.
• Lead statistical discussions with a wide variety of internal and external investigators to produce statistical analysis plans and reports for protocols. Lead as a statistical expert during the departmental scientific review process.
• Co-author publications with researchers.
• Mentor researchers, manage junior biostatisticians and analysts by teaching statistical courses, attending journal clubs, and advising on projects.

Rural Health Research Group, University of Central Florida   Orlando, FL
Data Analyst for a 4 year 1.6 million project funded by NIH   *Sep, 2013-May, 2015*
• Organized data for Accountable Care Organizations Impact on Rural Health Clinics (RHCs)
-Reconstructed large healthcare claims data: 7 years datasets for 800 RHCs with 290,000 patients' claims yearly and created calculation algorithms for risk-adjusted outcomes from multiple CMS data sources
• Co-authored two research papers for quality of care including clinical outcomes and being published
• Presented a scientific poster at NIH annual conference in December 2014 at Washington DC

EDUCATION

University of Central Florida   Orlando, FL
Ph.D (part-time) of Big Data Analytics, Statistics and Data Science   *2019-Present*
1st Place Winner - "AFCU (Addition Financial Credit Union) Customers' Churn Analytics Competition 2019" using R
1st Place Winner - "OUC Meter Data Science Competition 2020" using Python & Tableau
Master of Science – Statistics and Data Science, Statistical Computing Data Mining Track   *May, 2015*
 "SAS Certified Base Programmer for SAS 9"; "SAS Certified Advanced Programmer for SAS 9 Credential"

Chongqing Technology and Business University   Chongqing, China
Bachelor of Science in Statistics   *July, 2010*