ZHECHENG "Jeff" **SHENG**
Durham, NC 27705
(919) 519-7176 zhecheng.sheng@duke.edu
https://www.linkedin.com/in/zhecheng-sheng/

PROFILE

A creative and adaptive Master's candidate in Biostatistics, enthusiastic about combining statistics and programming skills to explore medical and clinical questions. Strong biology background from undergraduate and intensive statistical and quantitative training during Master's program. Keen interest in statistical programming and modeling process. A team player currently pursuing a full time position as a ***Biomedical Data Scientist*** or a ***Biostatistician.***

EDUCATION

DUKE UNIVERSITY, School of Medicine, Durham, NC
Master of Biostatistics, May 2019
Relevant coursework includes: Introduction to Statistical Theory & Methods I&II, Applied Biostatistical Method, Applied Biostatistical Methods I&II, Introduction to Practice of Biostatistics I&II, Introduction to Statistical Programming I&II, Biostatistician Career Preparation & Development I&II, Survival Analysis, Generalized Linear Model, Data Structure and Algorithms, Statistical Programing for Big Data, Statistical Method for Learning and Discovery, Causal Inference, Longutudinal Study, Case Studies.

Master's Project:
Functional Data Analysis On Association Between Vital Signs And Inhospital Mortality & ICU Transfer
- Performed data wranggling using functional programming in R on massive and dense Electronic Health Records with time stamps
- Conducted functional principle components analysis for exploratory data analysis and descriptive statistics to calculate mean functions, eigen functions and scores for each vital sign.
- Completed variable-domian functional logistic regression to access most informative time period before event occurrence, accounting for different length of stay for hospitalization.
- Tested the bidirectional variations for each vital sign by incorporating quadratic terms into the model.
- Access goodness of fit by calculating AUC for each vital sign.

ZHEJIANG UNIVERSITY, College of Agriculture and Biotechnology, Hangzhou, China
Bachelor of Agricultural Science, June 2017
Relevant coursework included: Calculus I&II, Linear Algebra, Probability and Statistics, Biostatistics and Experiment Design, Bioinformatics

UNIVERSITY OF CALIFORNIA-DAVIS, Davis, CA
Exchange Student/Internship, College of Biological Science, 2016
- Helped develop informatics tools to perform posterior-predictive simulation among different models
- Developed functions in a R packages for visualizing test statistics to demonstrate model adequacy
- Gained proficiency in R programming by designing functions to interpret sequencing data under phylogenetics setting.

EXPERIENCE

DUKE UNIVERSITY, Duke Clinical Research Institute, Durham, NC 2018
Intern, Early Warning Scores (EWS) project
- Data visualization of pre-processed dataset based on different stratifications using tidyverse
- Exploratory data analysis on determining whether Early Warning Score(EWS) is a good indicator for risk of deterioration.
- Formulated statistical analysis plan and generated reproducible reports for collaborators and clinicians
- Performed logistic regression to test odds ratio for different risk factors

DUKE UNIVERSITY, School of Medicine, Durham NC 2019
Course Project, Predicted COCCluster and histological grade of brain tumor
- Preprocessed MRI images with tumor masks for 110 patients
- Performed data augmentation for stacking all the images into a 4D dimension numpy array
- Trained a 2D convolutional neuron network with convolute, pooling, flatten and dense layers to predict both labels
- Used 5-fold cross-validation to accessed model performance and used the parameters from the best fold.
- Calculated AUC of one vs all for each class under multi-classification setting

DUKE UNIVERSITY, School of Medicine, Durham NC 2019
Course Project, Sequencing mapping using probabilistic error model
- Use of probabilistic error model to allow certain number of mismatch between reads and barcodes
- Developed a pipeline that generating sequencing mapping results table implemented in R and C++
- Wrapping up the pipeline into a installable R package in Gitlab
- Written bash script to call the R package and made it work on Duke Computing Cluster

DUKE UNIVERSITY, School of Medicine, Durham NC 2018

Course Project, Predicted Political Party Registration among Durham County Voters

- Tried different machine learning models for predicting party affiliation for residents in Durham County
- Processed unbalanced dataset with up-sampling
- Trained a CART model, a LASSO model and a Random Forest to perform a multi-classification problem, using Grid-Search to find the best hyper parameters
- Used cross-entropy loss to evaluate the performance
- Tree-based models roughly reaches a AUC of 0.80

DIDI CHUXING, Hangzhou, China 2016

Intern, Security Department

- Collected and summarized daily traffic accidents data in three provinces in China
- Analyzed traffic accident data to demonstrate the trend of incident rate in different places using Microsoft Excel and VBA

11th Prefect, Qi Zhen Leadership Club *(2015-2017)*

- Held weekly tea salons and led members in discussion of social problems
- Administered recruitment of new members as well as the maintenance of relationships between members

TECHNICAL SKILLS

Programming Skills: Python(Proficient), R(Proficient), SAS, SQL, Java, Bash, Distributed Computing (Spark)
Domain Knowledge: Machine Learning, Statistical Modeling, Clinical Care, Genetics
Softwares: Adobe, Microsoft, Tableau

ADDITIONAL INFORMATION

Conference: AMIA 2019 Informatics Summit, March, 2019
Student Poster: Analyzing The Association Of Time-Varying Vital Signs With In-Hospital Mortality & ICU Transfer.

Certification: SAS Certified Base Programmer for SAS 9, Sep, 2018
SAS Certified Advanced Programmer for SAS 9, Feb, 2019

Competition: Data Fest 2018

- Managed data on a large job posting dataset from Indeed
- Performed network analysis to create a visualizable web application using RShinny app based on similarity matrix to recommend relevant jobs for job seekers
- Received Honorable Mention as one of six teams