Li Yuan
403 Village at Vanderbilt, Unit 104, Nashville, Tennessee, 37212
TEL: (510)-606-2089 | Email: li.yuan@vanderbilt.edu

EDUCATION

Sichuan University (SCU: rank #27 in CHN) **Chengdu, CHN**

- ✧ ***B.S. in Statistics*** *Sept.2015 - June 2019*
- ✧ ***Major Statistics GPA:*** *3.73/4.0 (Junior year);* ***Overall GPA:*** *3.23/4.0 (WES)*
- ✧ **Awards:**
  - Second Class Academic Progress Scholarship *2016-2017 & 2017-2018 Academic Year*
  - Successful Participant of The Mathematical Contest in Modeling *Jan. 2019*

Vanderbilt University (Vandy) **Nashville, Tennessee**

- ✧ **M.S in Data Science (in progress)** *Aug.2020 - May 2022*
- ✧ **Fall semester** *(GPA: 3.85/4.0)***:** *Aug.2020 - Dec.2020*
  - **Programming & Simulation: A**
  - **Exploratory Data Analysis: A-**
  - **Survey of DS Applications: A-**
  - **Probability & Stat Inference: A**

University of California, Berkeley (UCB) **Berkeley, CA**

- ✧ ***Berkeley Global Access Program (Start - Discover)****(GPA: 3.594/4.0)* *Aug.2019 - May 2020*
- ✧ ***Berkeley Summer Session (2020)(Numerical Analysis & Linguistic Data)****(GPA: 3.471/4.0)* *June - Aug.2020*

Yale University *(GPA: 4.0/4.0)* **Berkeley(online)**

- ✧ ***Yale Summer Session (2020)(Multivariable Calculus*** *&* ***Intro Computing & Programming)*** *May - June 2020*

Winter Camp at Harvard University and Columbia University **New York & Boston**

Training Course Advised by a Harvard Professor, Zoya Kinstler (without credits)

- ✧ *Two-week Training Program on Innovative Technology & Business Analytics* *Feb. 2018*
- ✧ *Awarded 2018 Science and Technology Innovation Program Certification*

Academic Experience

Build and Merge two existing databases, protein structure databank and kinetics databank **Online at Vandy**

Advised by Prof. Zhongyue(John) Yang from Chemistry and RA. Bailu(Lucy) Yan from Vanderbilt *Sept.2020 - Present*

- ✧ Used PDB file from raw labotory records to extract desired data fields, I used python with varied packages forming a user-defined class to realize extracting and use of Data generating in Lab
- ✧ Used Data API from websites to directly extract data info which is not available in the PDB file
- ✧ Found the relations between PDB API and UniProt API, and build the relationship database which serves enzyme computation platform

Single cell RNA sequence analysis (Genomics Analysis) **Online at Vandy**

Advised by Prof. Todd Giorgio and Dr. Charreau Bell from Biomedical Engineering at Vandy *Nov.2020 - Present*

- ✧ Used R and Package Seurat to Normalize, Filter, Reduce Dimension using PCA, Cluster to find tumor cell in mouse breast cancer sample(big data) and
- ✧ Found out which kind of cells express lotta Olfactory Receptors, focusing on macrophages(M0, M1, M2)

Real-time Prediction and Big Data Analysis in NBA (In Progress) **Online**

Part Time Research Assistant, Advised by Yang, Jiaxi Ph.D. from Columbia University *Nov.2019-Present*

- ✧ Expect to mine data and build machine learning models to predict NBA games in real time
- ✧ Fetched data from online resources of NBA games and summarized statistics of players, teams and games over time to derive important features
- ✧ Trained in Machine Learning including logistic regression models, survival analysis, classifiers and random forests, KNN, Clustering models and Neural networks

Cathaypath Institute of Science **Shanghai, CHN**

Trainee, Advised by Prof. Joseph Chang from Yale University *July 2019-Aug.2019*

Coursework: *Probability and Bayisan Statistics Inference and Model Selection (A)*

CIS Summer Academic Program: Model Selection Research

- ✧ Explored different Bayesian approaches to model selection, including AIC, BIC, and Laplace Approximation, derivated the Laplace Approximation method and simulated its performance in R
- ✧ Used R programming to realize the comparison of different estimators, proposed new estimators for computing the marginal likelihood, and used MCMC/JAGS and importance sampling, with 10% precision improvement
- ✧ Analyzed hot hand phenomenon using JAGS/Gibbs

The Rise of Machine: Support Vector Machine

Graduation Thesis, Advised by Prof. Song Enbin from Sichuan University *Nov.2018-May 2019*

- ✧ Conducted literature review of Support Vector Machine
- ✧ Utilized Convex Optimization and Lagrange Multiplier to compute the linear least hyperpanecs
- ✧ Used Projection to Kernel Space to solve non-linear classification, and wrote codes to realize the algorithm in R
- ✧ Did data analysis on a real dataset using SMV with R, and found a ideal machine to classify different flowers and make predictions about new observations, and got 98% accuracy

Spread of Characteristics of the Synthetic Opioid and Strategy for Countering the Opioid Crisis

Team Leader, Mathematical Contest in Modeling *Jan.2019*

- ✧ Adopted biostatistics methods in analyzing the synthetic opioids data of five states in USA from 2010 to 2017 to study the using status of opioids in America, used Excel to clean data and conducted polynomial regression fitting for data screening
- ✧ Utilized R in time series analysis and special series modeling to predict the trend of drug use, applied genetic algorithm and biostatistics methods to build models for infectious diseases

Statistics Computing Algorithms with Matlab *Mar. 2018-June 2018*

- ✧ Adopted Matlab to realize statistical algorithms, including random number generation, matrix transformation, matrix factorization, conditional least squared regression etc, improved codes and simplified them

Time Series Analysis with R *Mar. 2018 - June 2018*

Team Leader | *Three datasets with time going*

- ✧ Used high-dimensional time series methods to analyze three datasets, including Robot Position, Stock in New York, Los Angle Air pollution
- ✧ Adopted cross-validation methods and lasso, Elastic Net methods and got the good fitness of three datasets

Linear Models to Housing Prices at Boston Suburb with R *Oct. 2017 - Jan. 2018*

Team Leader | *Boston Housing Price with some variables*

- ✧ Analyzed Boston housing price dataset to discover the influence of on the values of houses in Boston suburb
- ✧ Built linear regression models, best subset regression, ridge regression and lasso methdos with R for predicting housing prices
- ✧ Concluded that the lasso method gave best prediction with the least Test MSE

INFORMATION

Computer Skills: C/C++, R, Python, Matlab, SQL, SAS, LaTeX, Markdown, JavaScript, Data Structure and Database