KAI LIU

kai.liu@utexas.edu • 512.917.6781
GitHub: github.com/KaiUT • LinkedIn: www.linkedin.com/in/kai-liu-utaustin

EDUCATION

Ph.D. in Computational Biology ◇ *The University of Texas at Austin, Austin, TX* — Expected May 2019
M.S. in Microbiology ◇ *Huazhong Agricultural University, Wuhan, China* — Grad. June 2013
B.S. in Biotechnology ◇ *Huazhong Agricultural University, Wuhan, China* — Grad. June 2011

WORK EXPERIENCE

Machine Learning Intern ◇ *QuintilesIMS, Plymouth Meeting, PA* — June 2017 - August 2017

Predicted Quality of Investigators in Future Clinical Trials | Python, Spark

- Predicted outliers of investigators per Key Risk Indicator using distribution based approach;
- Built multiple machine learning models (Lasso Regression, Neural Network, Random Forests) to predict the quality of investigators in a future study, which is one of the core projects in the investigator recommender system.

Graduate Research Assistant ◇ *The University of Texas at Austin, Austin, TX* — December 2014 - Present

Developed a Supervised Anomaly Detection Model for Infectious Disease Surveillance | Python

- Adapted a Multivariate Exponentially Weighted Moving Average (MEWMA) method in biosurveillance to a supervised model;
- Combined the model with stepwise variable selection algorithms to optimize data source selection in infectious diseases surveillance;
- Retrospectively detected seasonal influenza outbreaks in US with an accuracy 0.9 using an optimal combination of data sources.

Developed Infectious Diseases Surveillance App | Python

- Implemented a regression model and the supervised MEWMA model in the application;
- Built up data pipeline to automate the process of retrieving and cleaning data related to 526 infectious diseases from CDC, RSS feed, Google Trends, Wikipedia, Twitter etc;
- Integrated the App into Cloud Ecosystem in collaboration with a software engineer.

PUBLICATIONS

- **Liu K**, Miller JC, Meyers LA. Effects of Directed and Clustered Contact Patterns on Infectious Disease Dynamics. *In preparation*.
- **Liu K**, Srinivasan R, Ertem Z, Meyers LA. Optimizing Early Detection of Emerging Outbreaks. *Submitted*.
- Castro LC*, Fox SJ*, Chen X, **Liu K**, Bellan SE, Dimitrov NB, Galvani AP, Meyers LA. Assessing Real-time Zika Risk in the United States. *BMC Infectious Diseases*. DOI: 10.1186/s12879-017-2394-9.

PERSONAL PROJECTS

Being Involved in Building an Open Source Software to Detect Lung Cancer | Python, TensorFlow August 2017 - Present

- Contributing to improving the 3D Convolutional Neural Network that identifies locations of nodules in scans;
- Contributing to improving the 3D Convolutional Neural Network to find the boundaries of nodules in scans.

Developing a R Package for Big Data Analysis | R & Rcpp — December 2016 - Present

- Implementing following algorithms in the package: Stochastic gradient descent using line search and quasi-Newton methods to determine step size · The lasso · The proximal gradient method · Laplacian smoothing solved by sparse Cholesky/LU, the Gauss-Seidel method, the Jacobi iterative method, and conjugate gradient method · Graph fused lasso solved by Alternating Direction Method of Multipliers (ADMM) · Sparse matrix factorization.

Predicted the Direction of Exchange-Traded Fund (ETF) movement | Python — April 2017 - May 2017

- Retrieved nine historical ETF sector datasets from Yahoo Finance;
- Implemented Logistic regression, Ridge & Lasso regression, and Artificial Neural Network to predict the direction of ETF movement;
- Achieved an accuracy of 55% $\sim$ 60% for predicting nine ETF sectors movement; and the trading strategy based on my prediction outperforms baseline strategies.

Predicted Yelp Rating Based on User Review Enhanced Collaborative Filtering | R — September 2015 - December 2015

- Extracted user opinions from restaurants dataset from Yelp ($\sim$10GB) using Stanford coreNLP tool;
- Developed a new Collaborative Filtering-based method to improve the accuracy of user's rating prediction and solve the sparseness of dataset by combining item's features and user opinions from all reviews;
- Improved the prediction accuracy by 4.23% compared to the traditional KNN method, and the coverage is 100%.

SKILLS

Programming — Fluency in Python(NumPy, SciPy, pandas, scikit-learn), R, Git · Familiar with MATLAB, Linux, LaTex · Experience in SQL, Hadoop, Spark, TensorFlow, C++

Machine Learning — Deep Neural Network · Regression with regularization · Support Vector Machine · Random Forests · Hidden Markov Model · Clustering · Time series and dynamic models · Frequent Pattern Mining · Natural Language Processing · Image Processing