

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 ◇ <i>The University of Texas at Austin, Austin, TX</i>	Grad. August 2019
M.S. in Microbiology ◇ <i>Huazhong Agricultural University, Wuhan, China</i>	Grad. June 2013
B.S. in Biotechnology ◇ <i>Huazhong Agricultural University, Wuhan, China</i>	Grad. June 2011

WORK EXPERIENCE

Applied Scientist ◇ <i>Zillow, Seattle, WA</i>	August 2019 - Present
Conducting machine learning research in recommender systems, search ranking and personalization.	
Applied Scientist Intern ◇ <i>Zillow, Seattle, WA</i>	May 2018 - September 2018
Improved Recommendation Diversity in Real Estate Recommender System Python, Spark, AWS	
<ul style="list-style-type: none">· Categorized all listings across the US on Zillow using <u>Bisecting Kmeans</u> algorithm;· Implemented and extended a <u>personalized diversification algorithm</u> published on RecSys'16 from scratch using PySpark;· Improved the diversity of recommendation by <u>at least 30%</u> in offline evaluation, while maintaining comparable accuracy (NDCG, recall, precision) of recommendation, compared with the current recommendation model in production;· Integrated the algorithm into an in-house package used by applied scientists; the work was posted on Zillow AI blog [Link].	
Machine Learning Intern ◇ <i>QuintilesIMS, Plymouth Meeting, PA</i>	June 2017 - August 2017
Predicted Quality of Investigators in Future Clinical Trials Python, Spark	
<ul style="list-style-type: none">· Predicted outliers of investigators per Key Risk Indicator using <u>distribution based approach</u>;· 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 ◇ <i>The University of Texas at Austin, Austin, TX</i>	September 2013 - June 2019
Developed a Supervised Anomaly Detection Algorithm for Infectious Disease Surveillance Python	
<ul style="list-style-type: none">· Adapted a statistical anomaly detection method <u>Multivariate Exponentially Weighted Moving Average (MEWMA)</u> to a supervised model;· Applied <u>evolution algorithms/random search</u> to optimize parameters of the above model;· Developed novel metrics to evaluate the performance of the model;· Applied <u>stepwise variable selection algorithms</u> to optimize the selection of time series predictors in anomaly detection;· Detected flu outbreaks in the US with an <u>accuracy 0.9</u> using the above model with an optimal combination of time series.	
Developed Infectious Diseases Surveillance App Python, AWS	
<ul style="list-style-type: none">· Implemented a <u>regression model</u> 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.	
Assessed Real-time Zika Risk in the State of Texas R	
<ul style="list-style-type: none">· Collaborated with other researchers in developing a <u>branching process model framework</u> that captures variation and uncertainty in Zika case reporting, importations, and transmission;· Applied the framework to assess county-level epidemic risk throughout Texas.	

PUBLICATIONS

- Liu K, Srinivasan R, Meyers LA. Early Detection of Seasonal Influenza in the United States. *Under review*.
- Liu K, Srinivasan R, Ertem Z, Meyers LA. Early Detection of Emerging Outbreaks. Poster presented at: *Epidemics6*. Nov 29-Dec 1, 2017. Sitges, Spain.
- 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

Predicted the Direction of Exchange-Traded Fund (ETF) movement Python	April 2017 - May 2017
<ul style="list-style-type: none">· Retrieved nine historical ETF sector datasets from Yahoo Finance;· Implemented <u>Logistic regression</u>, <u>Ridge & Lasso regression</u>, and <u>Artificial Neural Network</u> to predict the direction of ETF movement;· Achieved an <u>accuracy of 55% ~ 60%</u> 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 (~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), Spark, SQL, R, Git · Experience in Hadoop, TensorFlow, C++

Machine Learning

Rule-based models(GBM, RF etc) · Regression · Neural networks(GCN, CNN, RNN etc) · Other supervised models(SVM etc) · Clustering · Time series models · Natural Language Processing · Image Processing