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

Applied Scientist Intern ◇ *Zillow, Seattle, WA*   May 2018 - September 2018

Improved Recommendation Diversity in Real Estate Recommender System | **Python, Spark, AWS**
· Categorized all listings across the US on Zillow using Bisecting Kmeans algorithm;
· Implemented and extended a personalized diversification algorithm published on RecSys'16 from scratch using PySpark;
· Improved the diversity of recommendation by at least 30% 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 ◇ *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 Algorithm for Infectious Disease Surveillance | **Python**
· Adapted a statistical anomaly detection method Multivariate Exponentially Weighted Moving Average (MEWMA) to a supervised model;
· Applied evolution algorithms/random search to optimize parameters of the above model;
· Developed novel metrics to evaluate the performance of the model;
· Applied stepwise variable selection algorithms to optimize the selection of time series predictors in anomaly detection;
· Detected flu outbreaks in the US with an accuracy 0.9 using the above model with an optimal combination of time series.

Developed Infectious Diseases Surveillance App | **Python, AWS**
· 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**, Srinivasan R, Meyers LA. Early Detection of Seasonal Influenza in the United States. *Under review.*
· 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
· 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% ∼ 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 (∼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