JINGHE ZHANG

Charlottesville, VA • (607)-651-8906 • jz4kg@virginia.edu
https://www.linkedin.com/in/jinghezhang • https://github.com/JingheZ

PROFESSIONAL SKILLS

- Languages: Python, R, SQL
- Libraries: Numpy, Scikit-learn, Pandas, Tensorflow, Matplotlib, Genism, NLTK, Seaborn, dplyr, glmnet, ggplot2
- Technologies & Tools: Spark, HIVE, Jupyter, Git, Atom, Latex, Vim

EXPERIENCE

EDABI (Enterprise Data, Analytics and Business Intelligence), Target Minneapolis, MN
Data Scientist Intern *06/2016 – 08/2016*

- Preprocessed click stream data from target.com using Python and Spark for guest choice modeling and implemented feature learning techniques, NMF and Word2Vec, to construct a meaningful representation of a guest's behavior
- Developed a recurrent neural network, LSTM model, in Tensorflow to predict customers' final purchases using the learned features; evaluated its predictive performance using weighted average precision, recall, and hit rates
- Interpreted the model output using interactive visualization based on t-SNE algorithm and density-based clustering

University of Virginia Charlottesville, VA
Graduate Research Assistant

Representation Learning from Longitudinal Electronic Health Record Data *02/2015 – Present*

- Developed a graph-based framework to represent patients' medical history in large-scale electronic health record DBs
- Evaluated proposed representation learning framework with well-established classifiers, such as SVM, KNN, Random Forest, for the early detection of anxiety and depression of new patients and achieved 1%-5% higher accuracy
- Collaborated with professors, MDs, and two other graduate students in problem definition and data processing

Predictive Modeling on Hyperlactatemia Sepsis *05/2015 – Present*

- Built predictive models for the onset of sepsis using statistical approaches and machine learning algorithms; deployed the logistic regression model into real-time prediction in Epic electronic health record software at UVA health system
- Modeled the real-time probability of recovering from sepsis of critically ill patients based on their physiological characteristics, clinical interventions, and demographic information
- Led a team of six graduate students from engineering and medical backgrounds and conducted weekly presentations and discussions with professors, MD, and heath data scientists

Text Mining on Restaurant Reviews from Yelp *03/2015 - 04/2015*

- Processed millions of reviews from Yelp to construct N-gram vector representation and implemented statistical language models with maximum likelihood estimation and smoothing to retrieve most similar reviews
- Developed and evaluated a text categorization system, including feature selection, Naïve Bayes and KNN classifier with brute force and random vector projection, to distinguish positive and negative restaurant reviews

Graduate Teaching Assistant *01/2014 – 05/2014*

- Hosted office hours, graded assignments, and gave lectures on neural nets for a graduate-level machine learning class

RESEARCH AREAS

- Machine learning and its applications: representation learning, deep learning, predictive modeling, health informatics

EDUCATION

University of Virginia, Charlottesville, VA *08/2013 – 05/2017 (Expected)*

- PhD, Systems & Information Engineering
  - Dissertation Topic: Graph-based Representation Learning from Longitudinal Electronic Health Record Data for Patient Characterization and Prediction of Health Outcomes
  - Coursework: linear statistical models, optimization in machine learning, stochastic process, foundations of neural networks, text mining, information retrieval, big data in software engineering, human machine interface
  - Publications: wrote and published three articles based on thesis research in peer-reviewed conferences and journals

Binghamton University, State University of New York, Binghamton, NY *08/2011 – 05/2013*

- Master of Science, Industrial & Systems Engineering

Hebei University of Technology, Tianjin, China *09/2007 – 06/2011*

- Bachelor of Science, Industrial Engineering