Danlu Liu

GitHub: https://github.com/liudanlu001 • LinkedIn: www.linkedin.com/in/danluliu1993
Tel: 573-823-6626 • Email: dltb9@mail.missouri.edu

OBJECTIVE

Enthusiastic data science research seeking for an internship data scientist position in 2020 to support business growth and success by leveraging my 5 years of experience in data analytics and machine learning.

EDUCATION

Ph.D. student in Computer Science, University of Missouri, Columbia, MO (GPA 3.98/4.0) 08/2017-12/2021

M.S. in Computer Science, University of Missouri, Columbia, MO (GPA 4.0/4.0) 08/2015-05/2017

B.S. in Applied Mathematics, China University of Petroleum, Tsingtao, China 09/2011-06/2015

EXPERIENCE

- **Graduate Research Assistant**, University of Missouri, Columbia, MO
  1) Exploratory data mining for subgroup cohort discoveries and decision support on EHR/EMR dataset.
     - Developed a novel subgroup discovery method slice and dice thousands of potential subpopulations and prioritize potential cohorts.
     - Recommended hundreds co-occurring risk factors patterns of each subgroup and provide interventionable insights.
     - Applications:
       SFARI Autism data: 6 novel subgroups discovered among 2,591 probands' data with 15 phenotypes and 10000 genotypes.
       Cerner HealthFacts glaucoma data: a list of explainable patterns discovered for glaucoma prediction among 131,257 patients' data with at least 12-year longitudinal history.
       T1D Exchange Clinic Registry data (2010-12; 2015-17): Multiple risk factors discovered between paternal T1D and maternal T1D.
       The Cancer Genome Atlas (TCGA) data: Significant risk factors discovered among 625 breast cancer RNA-Seq data.
  2) Build a novel indexing structure for risk pattern mining, which is 100× faster compared to the state of the art.
     - Performed gene expression analysis between 555 cancer patients and 100 normal breast tissues.
     - Identified 10 significant single/combined risk factors between Triple-negative and Luminal A cancer types.
  3) Conducted BigData ecosystem processing and analysis with large-scale GeoSpatial data
     - Building a large database for 300+ million data points all over United States by using HBase and Hadoop.
     - Performed large-scaled geospatial calculation by using R and Spark.
- **Graduate Teaching Assistant, University of Missouri, Columbia, MO**
  1) Assisted with CS4050-Design and Analysis of Algorithms course. (50+ students/semester) 05/2017-12/2019
  2) Assisted with CS4410-Theory of Computation course. (40+ students/semester) 01/2016-12/2019

SKILLS

Programming:	Scala, Python, C/C++, Java, R
Tools:	Spark, Hadoop, HBase, Hive, RStudio, Jupyter, MATLAB, Tableau, Git, Linux
Statistical Skills:	Multivariate Analysis, Generalized Linear Model (GLM), Time Series, Bayesian Analysis, Machine Learning, Association Rule Mining, Natural Language Processing (NLP), Recommender System

PUBLICATIONS (2 MORE ON LINKEDIN)

D. Liu, W. Baskett, D. Beversdorf and C. Shyu, "Exploratory Data Mining for Subgroup Cohort Discoveries and Prioritization," in IEEE Journal of Biomedical and Health Informatics. (IF 2018: 4.217).

M. Raju, **D. Liu**, F. W. Fraunfelder and C. Shyu, "Discovering multifactorial associations with the development of age-related cataract using contrast mining," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, 2017, pp. 2297-2299.