

OBJECTIVES

Proven data analytics consultant and fast-learning Data Science Master graduate with experience in data analysis, data visualization and data mining seeking Advisor, Data Science position at EMC.

SKILLS

- **Python:** 3-year experience, 4 Kaggle competitions, using Numpy, Tensorflow, XGBoost, sklearn, etc.
- **R:** 4-year experience with statistical learning and visualization, using rpart, glmnet, shiny etc.
- **SQL:** 2-year hands-on experience with enterprise-level MySQL databases, also MongoDB, Spark, Database Design
- Others: SAS, Linux (Bash), MATLAB, PHP, Tableau, Weka, etc.

EDUCATION

University of Rochester

Rochester, NY

MS in Data Science (GPA 3.75/4.0)

Jun 2016-May 2017

Courses: Database Systems, Data Mining, Statistical Learning, Natural Language Process, Random Process, Network Analytics

Beijing Forestry University

Beijing, China

BA in Finance and Computer Science Double Major (GPA 3.8/4.0)

Aug 2012-Jun 2016

North Carolina State University

Raleigh, NC

Academic Program of Statistics

Jun 2013-Aug 2013

Courses: Statistics Methods (SAS), American Culture

PROJECTS**Prediction of Employees' Turnover using LinkedIn data**

- Leveraged 10-fold cross validation to evaluate three different prediction models, including Support Vector Regression, Neural Networks, and Random Forest.
- Analyzed attribute distributions and correlation between job changes and a variety of attributes to draw valuable insights.

American Domestic Airline Network Analysis

- Created traffic grid visualization for major US airports with respect to geometric location.
- Built and evaluated deep learning decision tree models to interpret the airline network.

Blue Light Camera Effectiveness Study

- Studied on blue light camera effectiveness of Rochester for local Police Department.
- Researched on overall impact of cameras, list effective and non-effective camera locations.

Extending compound nouns relations in TRIPS

- Improved relations between compound nouns in TRIPS semantic parses.
- Used features of TRIPS, leveraging supervised classification method, to improve TRIPS's capacity and compatibility of automatically interpreting noun-noun compounds relation.

AuctionBase: Database and the Web Design

- Designed the relational schema for AuctionBase and transformed/loaded a large volume of auction data from JSON files to MySQL database.
- Maintained the integrity of the AuctionBase data with MySQL.

EXPERIENCE

University of Rochester Medical Center

Rochester, NY

Project Assistant

May 2017-present

- Created and characterized resident-sharing-based social networks structure in more than 20000 nursing homes.
- Developed network measures and modeling their association to survey measures, controlling for nursing home characteristics such as the number of beds, using ANCOVA.
- Developed models that can be used to discriminate between different network topologies derived from networks of NHs with specific survey-based medical staff organization.

Geriatrics and Extended Care Data Analysis Center - Veterans Affairs

Canandaigua, NY

Information Service (Contract)

Oct 2017-present

- Developed relational database architecture for future operations of GECDAC, and converted SAS software flat files in a manner to optimize data organization for GECDAC.
- Introduced a new data tool Residential History File (RHF) that summarizes information from Medicare claims and nursing home Minimum Data Set assessments, and the Veterans Affairs equivalents, to track people through health care locations, including NH stays that were not paid for by Medicare or VA.
- Created SAS Proc Frequency and Proc Contents output information from RHF datasets for review.