#### **OBJECTIVES**

Proven data analytics consultant and fast-learning Data Science Master graduate with experience in data analysis, data visualization and data mining seeking data analyst opportunities.

#### **SKILLS**

Analytics: Python (Pandas, Numpy, Scikit-learn, plotly etc.),

R, MATLAB, SAS, JavaScript, C#

Visualization: D3.js, Qilk Sense, Tableau, Gephi, NodeXL

Database: MySQL, MongoDB, Spark, Database Design

Web: Google Analytics, HTML, CSS, Bootstrap

Business: SAP, bizagi, Agile

### **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

- Built complex model to predict job hopping based on 30 features with more than 20000 observes.
- Leveraged 10-fold cross validation, tried 40+ models and the finalized model is 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.

# **Blue Light Camera Effectiveness Study**

- Studied on blue light camera effectiveness of Rochester, sponsored by Rochester Police Department.
- Researched on overall impact of cameras, list effective and non-effective camera locations using Wilcoxon test etc.
- Implemented Markov Chain Monte Carlo methods to simulate top 5 candidates for new camera locations.

## **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.

# **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' 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 nursing homes 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.
- Created SAS Proc Frequency and Proc Contents output information from RHF datasets for review.