

Chunlei Liu
100 S Fell Avenue, Normal, IL

Email : chunleiliu1995@gmail.com
Mobile : 2177217672

EDUCATION

- **University of Illinois at Urbana-Champaign** Champaign, IL
Master of Science in Statistics; GPA: 3.86 Aug 2017 – May 2019
- **Zhejiang Gongshang University** Hangzhou, China
Bachelor of Economics in Economic Statistics; GPA: 3.68 Aug 2013 – Jun 2017

WORK EXPERIENCE

- **GenInvo Inc.** Bloomington, IL
Solution Analyst Intern July 2019 - Present
 - Applied **Name Entity Recognition** to improve the efficiency of **Clinical Document De-Identification** component of Shadow Software, leveraging **Spacy** library in **Python**
 - Collected 2000 De-Identification training records to prove the concept of machine learning models' capability, achieved cross validation accuracy up to **98 percent**
 - Collaborated with the product development team using **Azure Server** and **Git**, incorporated the **AI** feature into the Shadow platform, designed the **Dashboard** with **R Shiny** and added user interaction with **JavaScript**
- **University of Illinois at Urbana-Champaign** Champaign, IL
Research Assistant Jul 2018 - Dec 2018
 - Utilized **Python** to scrape speech text files from Securities and Exchange Commission(SEC) website
 - Collaborated with 3 teammates to clean 2059 speech text files leveraging **NLTK** library and labelled sentiments of text files
 - Used **Bag of Words** model to transform tokens into vectors and applied **sentiment analysis** on text files

PROJECT EXPERIENCE

- **Airline Data Analysis Project** Champaign, IL
Data Science Foundation Apr 2019 - May 2019
 - Implemented **Map Reduce** with **Unix Command Line**, calculated summary statistics on **1 million** airline records in **Hadoop Distributed File System**
 - Controlled data flow with **Pig** and analyzed the data with **Hive** within **HUE** interface, calculated summary statistics of departure delay in different segmentation
 - Visualized the summary statistics with **R** using **box plot**, **scatter plot**, **treemap plot** and **network plot**, explored state difference, weekday difference, month difference of the departure delay
- **Lending Club Loan Project** Champaign, IL
Statistical Learning Aug 2018
 - Classified whether a customer will be able to repay the loan based on the historical data with **1 million** records and 30 variables
 - Cleaned and visualized the data within **R**, utilized **over-sampling** technique to solve the imbalanced data set
 - Operated **logistic regression**, **KNN** and **Random Forest** on the data, tuned the model based on cross-entropy loss, achieved loss down to **0.447** with Random Forest
- **Real Estate Market Project** Champaign, IL
Applied Regression and Design Dec 2017
 - Predicted house price on the historical data with 2930 rows and 84 variables, finished **data cleaning and feature engineering**
 - Established **Lasso** regression and **random forest** on the data, decided the best hyper-parameters with cross validation, achieved RMSE of logarithm sales price down to **0.13**
 - Designed a **Dashboard** with **R Shiny**, provided data exploration, data visualization, personalized prediction and real-life google mapping functionality

SKILLS

- **Software:** Python, R, SAS, MySQL, Hive, JavaScript, Google Analytics **Language:** English, Mandarin
- **Certification:** SAS Base Programmer, SAS Advanced Programmer, Udacity Web Developer Nano Degree