Mingxiang Cai

Email : mcai@eng.ucsd.edu
Mobile : + 1-857-400-4225
Address : La jolla, CA 92037
Links :

Education

- **University of California San Diego** — La Jolla, CA
  Master of Science in Computer Science — *Sep. 2018 – Present*
  - Focus: Statistical modeling, supervised learning
- **Boston University** — Boston, MA
  Bachelor of Arts in Mathematics (Minor in Computer Science) — *Sep. 2014 – Jan. 2018*
  - GPA: 3.94/4.00 (Summa Cum Laude)

Skills

- **Programming Languages:** Python, Java, SQL, C++, R, SAS, Unix Shell, LaTeX
- **Technologies and Tools:** Data stuctures, NumPy, Pandas, SciPy, Statsmodels, Scikit-learn, Tensorflow, NLTK, matplotlib, plotly,Tableau, Git

Experience

- **UC San Diego Extension** — La Jolla, CA
  Researcher — *Oct. 2018 - Present*
  - Assist in prototyping a deep learning algorithm to detect illegal online drug sales on social media.
- **Boston University, Department of Computer Science** — Boston, MA
  Research Assistant — *Jul. 2017 - Dec. 2017*
  - Worked with two other researchers to develop a heuristic algorithm driven by features of participants and tasks to optimize team match in Hackathon events.
  - Liaised with hackathon organizations in Boston area and tested the algorithm in BostonHack 2017, helping over 400 participants get better team match.
- **Converse Product Launches Research: Social Media's Effects**
  Team project supported by Converse, Inc. — *May. 2017*
  - Collaborated with data experts from Converse to extract, clean, and analyze over 150000 tweets related to new product launches, covering over 60000 users.
  - Identified influential twitter accounts by finding social network centers using a centrality metric and detected their content patterns with unsupervised learning of topic modeling and clustering techniques.
- **Massachusetts Real Estate Data Scraping**
  Independent research — *Apr. 2017*
  - Established data pipeline between real estate websites and database through a continuous process of data scraping, structuring and formatting to collect data of over 15000 houses.
  - Trained supervised learning model of random forest to forecast house price and evaluate performance via k-fold cross-validation and achieved average 85.5% accuracy on test sets.
- **Yelp Review Text Mining**
  Independent research — *Mar. 2017*
  - Applied Principal Component Analysis on tokenized text to extract features.
  - Trained a ensemble model combining SVM and regression to predict restaurants star ratings and achieved 90.4% accuracy on test set.