Elaine Gao

Email : zhg071@ucsd.edu

Mobile : +1-858-952-2534

Address: 9450 Gilman Dr 80075, La Jolla, CA 92092-0100

PROFICIENCY

- **Analytical Skills**: Excellent analytical problem-solving skills (ETL, Descriptive Analyzing, Data Visualization, A/B testing)
- **Analyzing & Modeling Tools**: Proficiency using R, Python (NumPy, Pandas, Scikit-learn, etc), SQL, Tableau
- **Big Data Experience**: Experience with big data technologies (Hadoop, pySpark, Hive, etc)
- **Machine Learning Experience**: General Linear Regression, SVM, Tree Models, Penalized Regression, K-NN, Clustering Methods, Naive Bayes, Bootstrap, Cross Validation, PCA, Ensemble Method, NLP, etc

EDUCATION

- **University of California, San Diego** — San Diego, CA
  Master of Science in Statistics; GPA: 3.2/4.0 — *Sep. 2017 – Dec. 2018*
  - **Focus**: Linear Regression, Time Series, Mathematical Statistics, Hypothesis Testing, Survival Analysis
- **Zhejiang Gongshang University** — Zhejiang, China
  Bachelor of Art in Economic Statistics; GPA: 3.25/4.0 — *Sep. 2013 – July. 2017*
  - **Focus**: Sampling, Time Series, Applied Statistics, Logistic Regression, Probability, Statistical Inference

EXPERIENCE

- **Zhejiang Merit Interactive Network Technology Co., Ltd.** — Zhejiang, China
  Data Scientist Intern — *Feb 2017 - Aug 2017*
  - **ETL & Analysis**: Extracted, transformed and cleaned users profile data from Hive, analyzed the distribution, behaviors patterns and segmentation of users by exploratory and clustering methods with python. Displayed the analysis with Microsoft Excel or Tableau.
    Case example: user's salary tag, obsessed with car or not, obsessed with luxury goods, etc
  - **Rule-based Algorithm**: Developed, validated, optimized and implemented rule-based machine learning algorithms. For example, detected potential illegal trading locations of stolen smart phones, promoted a cooperation with a government department.
  - **Classification Models**: Fine-tuned and implemented user interest prediction models using pySpark and python with data imported from other companies. For example, applied Naive Bayes Model to new dataset, which improved the original gender prediction accuracy by 8%
  - **Recommendation System**: Trained a multi-class model based on tens of millions of users' historic behaviour-records and items' content information to predict the users' interest in a certain new audiobook, the empirical validation set error was 23%, promoted the cooperation with the biggest audiobook company in China.
- **China Yiwu Commodities Market Company** — Zhejiang, China
  Data Analyst Intern — *Jan 2016 - Aug 2016*
  - **Data Processing**: Managed and collected commodities inventory data from surveys with 14 team members, applied sampling methods, cleaned and updated the commodities Unicode database with new commodities with SPSS.
  - **Analyzing Task**: Analyzed the categories of commodities and their sale with R, explored the main trend of the sales of different categories commodities.

PROJECTS

- **Movie Recommendation System**: Built a movie recommendation system using hybrid system using python, achieved a rating MSE of 5.78 with rating scale 1-100.
  - **Technologies**: : Collaborative Filtering Algorithm(Cosine Similarity), Content-based Approach(Linear Regression), Rule-based Method
- **Text Mining**: Mining the text with entity extraction method, built topic segmentation model to classify transaction types of China Telecom's customer service data using bag-of-words and Tf-Idf algorithm, achieved an accuracy of 85%.
  - **Technologies**: : Word Tokenization, Entity Extraction, Topic Segmentation, Sentiment Analyzing, Bag-of-word, Tf-Idf, Random Forest
- **Social Survey & Analysis**: Analyzed the impact of new traffic rule on Hangzhou's residents attitudes and lifestyles using Python and SPSS, provided recommendation for rule improvement.
  - **Technologies**: : Descriptive Analyst, Clustering Method, Logistic Regression, Clustering Sampling and Simple Random Sampling, Association Rule Learning