Chia-Hui (Anita) Shen

Tucson, AZ 85718

chshen@ucdavis.edu || (213)590-1148 || linkedin.com/in/chia-hui-shen || https://aenni0409.github.io

EDUCATION

M.S. Statistics	University of California, Davis	GPA3.71/4.0	12/2017
Exchange Student	Engineering Science, Osaka University (OU), Japan	GPA3.60/4.0	09/2015 - 08/2016
B.S. Mathematics	National Cheng Kung University (NCKU), Taiwan	GPA3.50/4.0	06/2016
B.B.A. Statistics	National Cheng Kung University (NCKU), Taiwan	GPA3.75/4.0	06/2016

WORK EXPERIENCE

Data Analyst (R&D), Metropia Inc., Tucson, AZ 04/2018-Present

- Leveraged personalized incentives to change commuters' departure time by applying socio demographic features, trip histories, and users' app engagement on the machine learning approaches with 80% accuracy
- Improved user experience and product features by designing experimental design, including nested design and the Latin square design
- Analyzed user behavior to make business decision by collaborating with product and marketing
- Applied missing value imputation approaches—missForest to handle missing values of users' app usage data.

Bioinformatics Intern, Genomic Health, Redwood City, CA 06/2017-09/2017

- Extracted and manipulated over one terabyte NGS genetic dataset from National Institutes of Health (NIH)
- Developed and implemented optimization algorithms to tumor burden tracking with up to 90% patient coverage and less than 50,000 base pair panel size
- Improved time complexity by over 50% of the developed algorithms with parallel computing

RESEARCH EXPERIENCE

Joke Recommendation System, Department of Computer Science, UC Davis 09/2017 - 12/2017

- Assisted in SQL database management and transformed informative features for feature selections
- Applied NLP and sensitive analysis on text from joke contents with Python and Google Cloud Platform
- Built the SVD and random forest model in Python to predict and recommend top 10 jokes for users

Spam Email Detection, Department of Statistics, UC Davis 04/2017 - 06/2017

- Extracted information by collecting and cleaning over 100k raw text emails using Python
- Applied Natural Language Processing, including tf-idf and Latent Semantic Analysis, on feature engineering
- Compared and applied 6 machine learning and deep learning models including logistic regression, kNN, Neural Networks, and others on classifications with pipeline module on anomaly detection, reaching 97% accuracy

Causal Inference in Boston Housing Data, SHIMODAIRA Laboratory, OU, Japan 09/2015 - 08/2016

- Simulated the estimation of correlation for the causal effect in grouped and high-dimensional variables
- Estimated causal structures based on Independent Component Analysis and other factor analysis approaches
- Predicted strengths of relations between variables and applied causal inference approaches

SKILLS

- **R** : Data manipulation, model expression, statistical test (dpylr, sqldf, car, glmnet, ggplot, ggmap, shiny)
- **Python** : Pandas, Numpy, Scipy, Scikit-learn, Matplotlib, Nltk, JSON, and Tensorflow
- **Database** : MySQL, MSSQL, Porstage SQL
- **Others** : Periscope, Tableau, Plotly, Linux, Unix, Git, Excel, SAS (Based Programmer 9.0), Matlab