Chia-Hui (Anita) Shen

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, Woebot Labs., San Francisco, CA 05/2019-Present

- Refining 20% training dataset from false positive results by calculating semantic similarity using BERT
- Evaluated NLP model performances by applying inter-rater reliability tests and power analysis
- Created product metrics dashboard and ad-hoc analysis to improving user engagement by 28%

Data Analyst (R&D), Metropia Inc., Tucson, AZ 04/2018-04/2019

- Applied pattern recognition and peak-detection algorithm on time series data to discover commuters' daily travel pattern and time flexibility to change users' habitual travel schedule and increase system benefit by 20%
- Designed real-time anomaly detection on BI dashboard to monitor the abuse of incentive system
- Built personalized incentives engine to target core users using hierarchical models with 80% accuracy
- Developed carpooling matching system with 86% matched using network analysis and matching algorithm

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

- Extracted and manipulated over one terabyte NGS genetic data from National Institutes of Health (NIH)
- Developed and implemented optimization algorithms with machine learning concepts to tumor burden tracking with up to 90% patient coverage and less than 50,000 base pair panel size

Research Experience

Google Cloud & NCAA ML Competition 2018-Men's, Kaggle Competition 02/2018 - 03/2018

- Applied collaborative filtering on ten-year NCAA Tourney data and player information for feature engineering
- Implemented TrueSkill Ranking System to estimate the dynamic ranks for each possible match-ups over time
- Forecast the outcomes of all possible match-ups by applying XGBoost with 87% accuracy

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

- Integrated jokes through APIs into SQL database by applying ETL process and cron jobs
- Applied NLP and sensitive analysis on text from joke contents with Python and Google Cloud Platform
- Built SVD and random forest model to predict and recommend top 10 jokes for users with 80% accuracy

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

Skills

- **Scripting Language** : Python, Node.js
- **Analytics Tool** : R, SAS
- **Data Visualization and BI Dashboard:** Tableau, Plotly (D3.js), Periscope, Data Studio
- **Database** : RDBMS (MySQL, MSSQL, Porstage SQL), NoSQL (DynamoDB, MongoDB), BigQuery
- **Others** : AWS, Serverless, Linux, Unix, Git, Excel, Matlab, Firebase