Wei-Lin Hsiao

Wei-LinHsiao.GitHub.io
Houston, Texas

Skills

Languages

Python – *sk-learn, nltk, np, matplotlib*
R – *h2o, shiny, leaflet*
SQL – *MS SQL Server*
Java

Technologies

PySpark
Apache Hadoop
Amazon Web Services
Apache Subversion
Git

Coursework

Statistical Machine Learning
  Regression, classification, kernels, clustering, random forests, ensemble learning

Tools – Data Science
  Databases, text mining, distributed computing, TensorFlow, AWS

Intro. to Comp. Systems
  Concurrency, memory management, linking, exceptions, networking

Extracurriculars

Duncan Innovation Space
Studio Head

Leading creation of recording studio, securing $20,000 of audio equipment transactions

ACLU of Rice University
Events Head

Organized events of over 200 attendees, such as rights trainings, voter registration

Education

Rice University | B.A. Computer Science and B.A. Statistics
May 2021 | GPA: 3.97, Presidents Honor Roll

Experience

Rice University Data Sci. Research | Research Assistant
Summer 2018 | Houston, Texas

- Text-mined natural language features (sentiment, PoS tagging, embeddings), evaluated as predictors of unreliable news
- Automated pipeline to transform article data into network structure of similarity scores
- Implemented clustering (k-means, spectral, mixture) for classification of articles and detection of community structure
- Achieved 85.8% classification accuracy with an F1 score of 0.845 on articles

National Defense Medical Center | Assistant Programmer
Summer 2016 | Taipei, Taiwan

- Designed test cases for a browser-based distributed sequence alignment system
- Implemented Smith-Waterman and Needleman-Wunsch algorithms for 41% speed improvement over initial version

Personal Projects

Houston Demolition Map - Gentrification
R – *shiny, leaflet* | Python | Google Maps API

- Built interactive map of Houston, plotting region, income, construction, and demolition features
- Created script to geocode street locations with Google Maps API, parsing 14,000 street names
- Currently gathering data for machine learning model to predict gentrification in each zip code

Political Leaning Detector
Python – *sk-learn* | R – *shiny*

- Built crawler to source 250,000 comments from online forums
- Optimized logistic regression model for binary classification on party affiliation, with TF-IDF vectors as factors
- Analyzed features to indicate words most associated with political leaning
- Achieved F1 score of 0.850, with classification accuracy of 79.1%