

Kunlin Hsieh

khsieh10@gmail.com

706-572-0126

Duluth, GA

DATA SCIENTIST

Data scientist candidate skilled in implementing machine learning to solve business problems. Experienced with Python, SQL, Scikit learn, seaborn, matplotlib, and experimental design.

Skills

Data Analysis: Machine learning, predictive modeling, clustering algorithms, A/B Testing

Programming: Python, SQL

ML Packages: Keras, Scikit-learn, Tensorflow, Pandas, NumPy

Data Collection: survey design, web scraping, API calls, data wrangling, imputation

Data Visualization: Matplotlib, Seaborn

NLP Packages: Gensim, SpaCy, NLTK

Productivity: Git/GitHub, Word, Excel

Recent Projects

Final Capstone:

The purpose of this project is to forecast energy consumption from a data set obtained from Kaggle. Traditional statistical and supervised learning approaches and Facebook Prophet have been used and compared. The accurate predicted results can be beneficial for power companies by either expand or downsize their power plants and eventually reduce their operating cost.

Supervised Capstone:

The goal of this project is to predict the house price in the data set obtained from Kaggle competition. Three features elimination methods in combination with six machine learning algorithms have been used to perform the supervised learning analysis. The predicted results can be used by real estate companies to provide more accurate house price estimate to their customers.

Unsupervised Capstone:

Ten articles with similar topics from 10 different authors are obtained from Gutenberg Project. Bags of Word technique was used to extract features from the data set. Unsupervised learning and supervised learning approaches are used to classify the book author. The accurate classification can be utilized by online book reader provider for their customers to search for their favorite books.

Education

THINKFUL

Jan 2019

Data Science Program

- Completed intensive data science program with a focus on Python, mathematical tool-sets, statistical analysis, and big data techniques including machine learning.
- Learned industry best practices and standards by collaborating several hours every week with a senior data scientist.

Virginia Tech

MS, Engineering Mechanics

May 2007