# **Kunlin Hsieh**

Data Scientist

(706) 572-0126 khsieh10@gmail.com LinkedIn Profile

#### **PROJECTS**

### Predict House Price (Jupyter Notebook)

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.

# **Classify Book Authors (Jupyter Notebook)**

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.

# Forecast Energy Consumption (Jupyter Notebook)

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.

#### **EDUCATION**

## **Thinkful** — Data Science Program

Aug 2018 - Jan 2019

- 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 in Engineering Mechanics*

#### Aug 2004 - May 2007

#### WORK EXPERIENCE

# **Civil Engineer**

Sept 2007 - Present

- Project Management.
- Collaborate with different professionals to complete projects.

#### **SKILLS**

- Machine Learning
- A/B Testing
- Supervised Learning
- Unsupervised learning
- Time Series Analysis
- Scikit-Learn
- Matplotlib
- Seaborn
- Spacy
- NLTK
- Git/Github
- Python
- Numpy
- Pandas
- SQL