KA HOU SIO
DATA SCIENTIST & ANALYST

Profile

Data scientist and data analyst with hands on experience in data mining, data querying, data analysis, modeling and data visualization. Trained in business and biochemistry with experience in marketing and project management. Demonstrated experience in different technologies, including SQL, PostgreSQL, Python, AWS, Tableau and various machine learning packages. Skilled in design thinking, research, planning and decision making. Known to be a self-starter, and for ability to learn quick, work independently and in a team setting.

Experience

Metis
Data Scientist
San Francisco, CA
July 2018 to Sept. 2018

- Developed five data science projects as part of a highly selective 12-week immersive program.
- Data projects focused on business-applicable outcomes, covering various types of data and applying a wide array of machine learning and statistical modeling techniques.

Projects include:

Shopping Cart Recommender Project
- Analyzed "The Instacart Online Grocery Shopping Dataset 2017" which consists of more than 3 million orders and 200k+ customers.
- Extracted customer purchase behavior from data and built visualizations for communications.
- Built a shopping cart recommender to recommend list of items for shoppers based on their purchase history using Cosine Similarity.

E-commerce Reviews Analysis Project
- Applied Natural Language Processing (NLP) and Vectorization techniques to draw insights from unstructured data using topic modeling, found out what people are talking about the products.
- Used KMeans clustering algorithm, dimensionality reduction, tSNE visualization to find patterns/clusters from different reviews.
- Built different classification models to separate out unsatisfied customer reviews using NLP techniques and customer review text.

Heart Disease Classification Project
- Analyzed Patient info and test results to predict how likely the patient has heart disease.
- Built classification models using KNN, Logistic regression, Decision tree, Random forest, SVM, Naive Bayes, and XGboost.
- Utilized Grid Search CV to tune and found best hyper-parameter for each model.
- Rendered a prototype Flask application for doctors to enter patient's info for prediction.

Rainbow Health
Associate Marketing Manager
Walnut, CA
Oct. 2013 to Sept. 2014

- Determined product pricing by analyzing market trends and costs from suppliers.
- Managed and collaborated with designers and developers to create company e-commerce website to sell 10+ diabetic dietary supplements.
- Led researches on potential diabetic products and potential suppliers for the company.

International Business Support Center
Associate Project Manager
Las Vegas, NV
Oct. 2013 to Sept. 2014

- Devised and managed marketing plans, budgets and materials for clients.
- Managed relationship of Chinese clients, U.S. sales representatives and U.S. distributors throughout the operation.

Contact

- kahousio@gmail.com
- khsio.github.io
- Dublin, CA
- in kahousio
- khsio

Technical Skills

PROGRAMMING LANGUAGES
Python
SQL
R
CSS
HTML

DATA ANALYSIS
Regression
Classification
Clustering
Natural Language Processing
Dimensionality Reduction
Hyper-Parameter Tuning

VISUALIZATION
Seaborn
Bokeh
Tableau
Mode

WEB-BASED EXPERIENCE
AWS
Github
Flask
Panoply

Education

Santa Clara University - Leavey School of Business
MBA 2018
Concentration in Data Science and Business Analytics

University of California, Davis
B.S. Biochemistry & Molecular Biology 2011

DataCamp
2017
Certification of Data Scientist with Python Track