

Skills

Languages: Python, SQL

Python Libraries: Pandas, NumPy, Scikit-Learn/NLTK, Gensim, Surprise, BeautifulSoup, Selenium

Machine Learning: Regression, Classification, Gradient Boosting, Recommender Systems, Natural Language Processing

Visualizations: Matplotlib/Seaborn, Tableau, MS Powerpoint/Excel

Scientific: Data Analysis, Medical Devices, Drug Formulation, DNA/Protein Purification, Cell Culturing, HPLC

Experience

Metis San Francisco, CA
Data Scientist 06/2020 - 09/2020

- Attended a rigorous ACCET accredited project-based data science bootcamp that focused on Python programming, machine learning, data analysis, natural language processing, and communication
- See projects section for details

BioMarin Pharmaceutical Inc Novato, CA
Research Associate 10/2018 - 06/2020

- Performed experimental lab work to quantitatively evaluate drug product stability
- Supported marketing authorization application by performing experiments and writing reports
- Worked with external departments to improve medical device setup and develop drug use manuals
- Presented result summaries and literature reviews both internally and externally

University of Illinois at Urbana-Champaign Urbana, IL
Graduate Research Assistant 08/2013 - 10/2018

- Performed research using neutron scattering at Oak Ridge National Lab to study phase transitions
- Wrote research proposals to Department of Energy and National Science Foundation
- Purified DNA and protein from *E. coli* and *S. cerevisiae* to create fusion-tagged proteins

Projects

Tunable Movie Recommender

- Created a movie recommender with content and collaborative filtering using Doc2Vec and bag-of-words models
- Recommender allows the front-end user to tailor recommendations using up to four parameters
- Developed an application housing 46,000 movies using Streamlit

Identifying Depression in Adults

- Classified depression with 83% recall and 0.94 AUC score using NHANES (CDC) dataset with XGBoost
- Features determining depression strongly aligned with scientific literature: finding that energy, concentration, and appetite were indicators of depression with demographic playing a smaller role
- Developed and deployed a web application using Flask and Heroku

Recognizing Truth and Lies from News Headlines

- Used latent semantic analysis for topic modeling to identify leading news topics for 2020
- Performed dimensionality reduction using t-SNE and used gradient boosting to classify fake/real news
- Web scraped news headlines from Reddit using Pushshift API

Features Leading to a Well-Received Video Game

- Used linear regression to predict user ratings of games and determine features leading to high ratings
- Discovered that indie and simpler games (fewer labeled genres) were well-received
- Dataset was obtained by scraping Steam database using BeautifulSoup and Selenium

Education

University of Illinois at Urbana-Champaign
MS Chemistry 2018

San Jose State University
BS Biochemistry 2013