Jingtian 'Josh' Wang

Data Scientist

13jtjoshua@gmail.com joshjingtianwang.github.io/

joshjingtianwang/ JoshJingtianWang

EXPERIENCE

Metis, *Data Scientist* Jan. 2022 - May 2022

Completed an immersive 5-month data science bootcamp with a strong emphasis on project-oriented skill-building in problem-solving, data wrangling, statistical modeling, machine learning, and communication of deliverables. Project highlights include:

Predicting the Onset of Stroke (Classification):

- Predicted the onset of stroke of a patient based the patient's physical attributes and lifestyle.
- Built an SVM classification model that generates probabilistic prediction of the stroke onset .
- Class imbalance was dealt with using Imblearn. Improved the Brier Skill Score by 10-fold using GridSearchCV.

Public Reception of Elon Musk's Twitter Buyout Using NLP (NLP, Web Scraping):

- Used NLP to understand how Twitter reacted to the Twitter Deal.
- Twitter users were categorized into liberal/conservative based on their Twitter profile descriptions using Regex. Performed topic modeling using LDA and sentiment analysis using VADER.
- Provided insights on how the Twitter Deal potentially affects its user base.

Helping Costco Adapt to the Work-from-Home Trend (Web Scraping, Tableau):

- Devised a business proposal on how Costco should react to the evolving customer demographics.
- Costco store data and US census data were scraped using BeautifulSoup and Selenium. Findings were animated on an interactive dashboard in Tableau.
- Identified multiple changes in the customer demographics due to the Work-from-Home Trend and proposed business opportunities for Costco in the post-pandemic market.

What Makes a Winning NBA team? (Regression, Web Scraping):

- Built an explainable model to find the best strategy to win an NBA game.
- Over 20 years of NBA game data was scraped using BeautifulSoup and Selenium. Applied linear regression and lasso regularization to predict point differentials.
- Determined defense and rebounding as the most important factors to win an NBA game.

University of California, Irvine, *Graduate Student Researcher*, Irvine, CA Aug. 2018 - Aug. 2022

- Extrapolated new cancer gene therapy targets from over 80GB of cancer patient alternative splicing data
- Developed an alternative splicing-based cancer gene therapy
- Optimized a pipeline for RNA-seq data QC, alignment, and differential expression analysis for running on the high-performance computing cluster
- Publications: https://scholar.google.com/citations?hl=en&user=TrF6tCkAAAAJ

EDUCATION

University of Rochester Aug. 2022 - Dec. 2023

M.S. Data Science

Courses: Data Mining, Deep Learning, Computational Intro to Stats, Database Systems, Statistical Machine Learning, Data Science at Scale

University of California, Irvine Aug. 2017 - Aug. 2022

Ph.D. Molecular Biology and Biochemistry

Courses: Foundamentals of Genomics, Intro to Bioinformatics, Regulation of Gene Expression

SKILLS

SUPERVISED MACHINE LEARNING: Linear/Logistic Regression, Tree-based Methods (CART), Ensemble models
UNSUPERVISED MACHINE LEARNING: Dimensionality Reduction, K-means Clustering, PCA
NATURAL LANGUAGE PROCESSING: Topic Modeling (LSA/NMF/LDA), Sentiment Analysis, Text Preprocessing
PYTHON LIBRARIES: Pandas, Numpy, Sklearn, Selenium, BeautifulSoup, PyTorch
DATA VISUALIZATION: Tableau, Matplotlib, Seaborn
OTHER: Python, SQL, R, Bash, Excel, Cloud Computing, Streamlit