LUCHAO QI

Data Scientist
Baltimore, MD

+1 (443) 839-9129 | luchaoqi.email@gmail.com
https://www.linkedin.com/in/LuchaoQi/
https://luchaoqi.com/

Summary

Data scientist familiar with gathering, cleaning, and organizing data for use by technical and non-technical personnel. Advanced understanding of statistical, algebraic, and other analytical techniques. Highly organized, motivated and diligent with significant background in predictive analytics.

Employment

Data Scientist Intern 09/2020 – 12/2020
Ekohealth, Oakland, CA

- Helped build the product Eko-core, an FDA-cleared digital stethoscope attachment device, saving monthly cost for patients with arteriovenous fistula (AVF)
- Spearheaded a project to build the prototype of an audio-based dialysis fistula assessment algorithm detecting stenosis, which helps secure $295,881 in SBIR funding from the National Institutes of Health (NIH)
- Productionalized customer-facing python-based analysis pipeline using AWS cloud services
- Implemented Fast Fourier transform (FFT) algorithm on audio signals for feature engineering based on frequency domain
- Constructed machine learning models (acc: 73.68%, AUC: 0.85) detecting stenosis caused by AV fistula

Research Assistant 08/2019 – 05/2020
The Johns Hopkins Data Science Lab, Baltimore, MD

- Spearheaded the project focusing on association analysis between lifestyle patterns, physical activity, and body mass index (BMI)
- Migrated data from SAS to R and performed EDA using dplyr and tidyverse
- Trained convolutional neural networks (CNN) using Keras for BMI prediction with 25.45 mean squared error (MSE)
- Decreased the data dimensionality using principal component analysis (PCA) and improved prediction by 23% training a generalized linear model (GLM)
- Hosted R Shiny website comparing machine learning algorithms (PCA, k-means, UMAP, and t-SNE) & visualized clustering results using ggplot2 and plotly
  Demo: https://github.com/LuchaoQi/Shiny_clustering

Projects

Natural Language Processing: Amazon Reviews 09/2019 – 12/2019
Use of Machine Learning to Detect Fake Amazon Fine Food Reviews
Demo: https://www.kaggle.com/luchaoqi/amazon-review-rating-prediction

- Processed Amazon Food Review data using pandas, NumPy, and dfply in Python
- Tokenized unstructured text of user reviews using NLTK; converted text to vector using bag-of-words models with scikit-learn
- Predicted customer ratings using logistic regression with 0.94 AUC
- Improved negative reviews detection by 3% using random forest

Skills

Programming Languages

Python, R (RShiny), SQL, Shell scripting

Data Visualization

Tableau, Matplotlib, Seaborn, ggplot2, plotly

Packages

Pandas, NumPy, SciPy, NLTK, scikit-learn, Tidyverse

Frameworks & Platforms

PyTorch, TensorFlow, Keras, Hadoop, AWS

Machine Learning & Deep Learning

GLM, Random Forest, SVM, PCA, CNN, LSTM

Education

Johns Hopkins University
Baltimore, MD 05/2020
Master of Science in Engineering
Biomedical Data Science
GPA: 3.6/4.0

Northeastern University
Liaoning, CN 05/2018
Bachelor of Science
Biomedical Engineering
GPA: 3.8/4.0