

## + WORK EXPERIENCE

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

- Assisted in developing the product EKO-CORE, an FDA-cleared digital stethoscope, saving around \$900 monthly cost for hemodialysis patients
- Led project to design machine learning algorithms detecting vascular stenosis in fistula, which resulted in 73.68% accuracy and 0.85 AUC
- Productionalized customer-facing analysis pipeline in Python using AWS cloud services (S3, EC2, SageMaker) to improve team communication
- Prepared regulatory submissions for FDA; successfully secured \$295,881 in SBIR funding from the National Institutes of Health (NIH)

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

- Led project to analyze associations between demographic patterns, physical activity, and body mass index (BMI)
- Extracted 10K+ time-series SAS data from CDC in R and manipulated data using dplyr and tidyverse
- Developed convolutional neural networks (CNN) using Keras for BMI prediction with 25.45 mean squared error (MSE)
- Reduced data dimensionality using principal component analysis (PCA); improved prediction by 23% training a generalized linear model (GLM)
- Hosted R Shiny website performing cluster analysis; visualized clustering results using ggplot2 and plotly

Demo: [https://github.com/LuchaoQi/Shiny\\_clustering](https://github.com/LuchaoQi/Shiny_clustering)

## + PROJECTS

**Reinforcement Learning: OpenAI Gym** 02/2020 – 05/2020  
AI that Learns to Play Super Mario Bros Using Deep Q-Network (DQN)  
Demo: [https://github.com/LuchaoQi/Reinforcement\\_Learning](https://github.com/LuchaoQi/Reinforcement_Learning)

- Developed DQN-based convolutional neural network (CNN) model as an AI agent using TensorFlow
- Accelerated network training by 30% training model parallelly with JAX
- Achieved 2X faster than average of human players with the trained agent completing tasks successfully

**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
- 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 (R Shiny), SQL, Shell scripting

### Data Visualization

Tableau, Matplotlib, Seaborn, ggplot2, plotly

### Packages

Pandas, NumPy, SciPy, NLTK, scikit-learn, dplyr, tidyverse

### Frameworks & Platforms

PyTorch, TensorFlow, Keras, Hadoop, AWS, Kubernetes

### 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

- JHU Fellowships (20% acceptance rate)

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

- Dean's List