**LUCHAO QI**

Baltimore, MD [luchaoqi@luchaoqi.com](mailto:lqi9@jhu.edu) | (443)-839-9129

<https://www.linkedin.com/in/LuchaoQi/> <https://luchaoqi.com/>

**PROFESSIONAL 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.

**TECHNICAL PROFICIENCIES**

* Programming: Python, R (Shiny), SQL, Shell scripting
* 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

**PROFESSIONAL EXPERIENCE**

**Data Scientist Intern**, Ekohealth, Berkeley, CA 08/2020 - Present

* Help build the product Eko-core, an FDA-cleared digital stethoscope attachment device, saving monthly cost for patients with arteriovenous fistula (AVF)
* Spearhead a project to build the prototype of an audio-based dialysis fistula assessment algorithm detecting stenosis, which help secure a $295,881 in SBIR funding from the National Institutes of Health (NIH)
* Maintain the database loaded to Amazon S3 bucket using AWS
* Implement Fast Fourier transform (FFT) on audio signals for feature engineering based on frequency domain
* Constructed a random forest model (acc: 73.68%, AUC: 0.85) detecting stenosis caused by AV fistula

**Research Data Scientist**, The Johns Hopkins Data Science Lab, Baltimore, MD 08/2019 - 04/2020

* Spearheaded the project focusing on association analysis between lifestyle patterns, physical activity, and body mass index (BMI)
* Migrated data in SAS transport file format from external databases (National Health and Nutrition Examination Survey) using R and processed data using dplyr and tidyverse
* Decreased the data dimensionality using principal component analysis (PCA) and predicted user BMI with 46.07 mean squared error by training a generalized linear model (GLM)
* Achieved a 13% error reduction rate utilizing random forest and nested ANOVA (F-test) on principal component groupings

**Data Analyst Intern**, Johns Hopkins University, Baltimore, MD 04/2019 - 07/2019

* Executed and managed research project on survival analysis of accelerometer time-series data
* Created a convolutional neural network (CNN) using Keras to predict the 5-year mortality with 71% accuracy
* Improved the accuracy to 86.45% by implementing a regularized logistic regression model using principal component scores
* Hosted R Shiny application comparing machine learning algorithms (PCA, k-means, UMAP, and t-SNE) & visualized clustering results using ggplot2 and plotly

**EDUCATION**

**Johns Hopkins University**, Baltimore, MD 05/2020

Master of Science in Engineering Degree - Biomedical Data Science (GPA: 3.6/4.0)

**Northeastern University**, Liaoning, China 05/2018

Bachelor of Science Degree - Biomedical Engineering (GPA: 3.8/4.0)