# LUCHAO QI

Baltimore, MD  luchaoqi@luchaoqi.com | (443)-839-9129
https://www.linkedin.com/in/LuchaoQi/  https://luchaoqi.com/

## PROFESSIONAL SUMMARY

Technically sophisticated data scientist with comprehensive experience in performing quantitative analysis and data management for analytics studies utilizing state-of-the-art technologies to collect, clean, analyze, predict, and effectively communicate information. Expert in analyzing data, drawing insights, and presenting results in a cohesive, intuitive, and simplistic manner to site and executive management in monthly and quarterly meetings. Possess an in- depth understanding of machine learning algorithms and advanced statistics such as regression, time-series forecasting, clustering, decision trees, exploratory data analysis methodology, simulation, scenario analysis, modeling, optimization, unstructured data analysis, and neural networks. Proactively contributes to multiple projects to support the solution design process and delivers the analytical models and algorithms to achieve business value.

## 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**, Eko, Berkeley, CA  08/2020 - Present
- Head the project building the first prototype of an audio-based dialysis fistula assessment algorithm for data collected using electronic stethoscope
- Maintain the database loaded to Amazon S3 bucket using AWS
- Implement Fast Fourier transform (FFT) on audio signals extracting features based on frequency domain
- Construct a random forest model (acc: 73.68%, AUC: 0.78) detecting stenosis in early stage to reduce the cost by half for patients receiving arterio-venous fistula (AVF)

**Research Data Scientist**, The Johns Hopkins Data Science Lab, Baltimore, MD  08/2019 - 04/2020
- Spearheaded the project focused 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 by using dplyr and tidyverse
- Decreased data dimensionality facilitating principal component analysis (PCA) and also 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 5-year mortality with 71% accuracy
- Improved the accuracy to 86.45% 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)

# LUCHAO QI

## SOFTWARE PORTFOLIO

R Packages
- MRIPCA: Principal component analysis (PCA) on MRI data  https://github.com/LuchaoQi/MRIPCA
- MRIcloudT1volumetrics: Volumetric analysis of MRI data  https://github.com/bcaffo/MRIcloudT1volumetrics

R Shiny Web Applications
- Clustering analysis using K-means, t-SNE, and UMAP  https://github.com/LuchaoQi/Shiny_clustering
- BMI Calculator  https://luchao-qi.shinyapps.io/BMI_Calculator/