

# LUCHAO QI

Baltimore, MD

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

[lqi9@jhu.edu](mailto:lqi9@jhu.edu) | (443)-839-9129

<https://luchaoqi.com/>

## PROFILE SUMMARY

Technically-sophisticated top performer 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, Bash (Linux)

**Visualization:** Tableau, Matplotlib, Seaborn, ggplot2, plotly

**Packages & Frameworks:** NumPy, Pandas, Tidyverse, NLTK, scikit-learn, PyTorch, TensorFlow (Keras), Hadoop

**Machine Learning & Deep Learning:** GLM, Random Forest, SVM, PCA, CNN, RNN, Reinforcement Learning

## PROFESSIONAL EXPERIENCE

### The Johns Hopkins Data Science Lab, Baltimore, MD

08/2019–04/2020

Research Data Scientist

- Demonstrated mastery in spearheading project focused on association analysis between lifestyle patterns and body mass index (BMI)
- Processed data from the National Health & Nutrition Examination Survey by using dplyr and tidyverse
- Drastically decreased data dimensionality through 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 nested ANOVA (F-test) on principal component groupings
- Recognized for expertise in identifying vital ( $p\text{-value} < 0.5$ ) associations between BMI, age, race, and physical activity level to encourage multiple healthy behaviors

### Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD

04/2019 – 07/2019

Data Analyst Intern

- Executed and managed research project on survival analysis of accelerometer time-series data and also optimized data value through researching on user distribution on Hadoop with MapReduce
- Employed the use of Keras to predict mortality with 71% accuracy and structured a spectral-based convolutional neural network (CNN) on subjects
- Accomplished in implementing regularized logistic regression for achieving an 86.45% increase rate in mortality prediction accuracy
- Piloted the process of using ggplot2 and plotly for hosting R Shiny website comparing machine learning algorithms (PCA, k-means, UMAP, and t-SNE) & visualized clustering results

### Shenzhen Institutes of Advanced Technology, Shenzhen, China

11/2016 – 05/2017

Data Analyst Intern

- Utilized spectral analysis for the project on EMG signal pattern recognition for hand gestures
- Reputed for expertise in designing, structuring and assembling the EMG data acquisition system
- Pioneered the process of using Fast Fourier Transform for denoising time-domain signals of 200 gestures
- Generated an 82% accuracy rate in using support vector machines (SVM) for classifying hand movements
- Delivered extensive and innovative insights for medical rehabilitation systems while increasing the neural network training accuracy by 3%

# LUCHAO QI

## PROJECT EXPERIENCE

### Reinforcement Learning: Super Mario Bros (NES)

02/2020 – 05/2020

AI that Learns to Play Super Mario Bros Using Deep Q-Network (DQN) in TensorFlow

Demo: [https://github.com/LuchaoQi/Reinforcement\\_Learning](https://github.com/LuchaoQi/Reinforcement_Learning)

- Recognized for expertise in utilizing OpenAI Gym; emulated Nintendo Entertainment System for designing and building a reinforcement learning environment through Nes-Py in Python
- Pioneered the process of implementing 5 hidden layers as an agent in TensorFlow for structuring a convolutional neural network (CNN) model
- Drastically reduced training downtime by 20% using Adam optimizer while training the agent on deep Q-learning
- Achieved 2X faster than average of human players with trained agent to complete tasks successfully

### Natural Language Processing: Amazon Rating Prediction

09/2019 – 12/2019

Use of Machine Learning to Detect Fake or Abusive Amazon Product Reviews

Demo: <https://www.kaggle.com/luchaoqi/amazon-review-rating-prediction>

- Exceeded expectations in using pandas, NumPy and SciPy for processing Amazon Food Review data
- Initiated the use of NLTK for feature construction while tokenizing the unstructured text of user reviews
- Proven success in using the bag-of-words model (unigram/bigram) with scikit-learn for converting text to vector
- Forecasted clients' ratings using logistic regression with 0.94 AUC and also steered a 3% increase in negative review detection to identify abusive entities (sellers & reviewers) via random forest

### Investigating Yelp User Funnels, Key Performance Indicators (KPIs)

01/2018 - 03/2018

Yelp User & Restaurant Performance Analysis Through SQL.

Demo: [https://github.com/LuchaoQi/Yelp\\_Data\\_Set\\_SQL](https://github.com/LuchaoQi/Yelp_Data_Set_SQL)

- Excelled in utilizing Xpaths & BeautifulSoup for programming web crawler to scrape/parse unstructured data from Yelp
- Migrated 10GB of data into the database developed using MySQL Workbench
- Determined and visualized the geographical distribution of restaurants with average ratings using Tableau
- Measured customer engagement using well-designed metrics (bracket retention, DAU/MAU) while also conducting A/B testing for exploring other methods for improving upon KPIs

## EDUCATION

### Johns Hopkins University, Baltimore, MD

05/2020

Master of Science in Engineering Degree -- Biomedical Data Science (GPA: 3.65/4.0)

### Northeastern University, Liaoning, China

05/2018

Bachelor of Science Degree -- Biomedical Engineering (GPA: 3.85/4.0)

## SOFTWARE PORTFOLIO

### R Packages

- MRIPCA: Principal component analysis (PCA) on MRI data
- MRICloudT1volumetrics: Volumetric analysis of MRICloud output

### R Shiny Web Applications

- Clustering analysis using K-means, t-SNE, and UMAP: [https://github.com/LuchaoQi/Shiny\\_clustering](https://github.com/LuchaoQi/Shiny_clustering)
- BMI Calculator: [https://luchao-qi.shinyapps.io/BMI\\_Calculator/](https://luchao-qi.shinyapps.io/BMI_Calculator/)