3111 N Charles Street, Allston 4C **Luchao Qi** Linkedin: [linkedin.com/in/LuchaoQi](https://www.linkedin.com/in/LuchaoQi/)

Baltimore, MD 21218 [lqi9@jhu.edu](mailto:lqi9@jhu.edu) (443)839-9129 Github: <https://github.com/LuchaoQi>

**EDUCATION**

MSE, Bioengineering and Biomedical Engineering, *The Johns Hopkins University* May 2020(expected)

Visiting student, Biomedical Engineering, *University of Chinese Academy of Sciences* Sep 2016 - Aug 2018

B.S., Biomedical Engineering, *Northeastern University* Sep 2014 - Aug 2018

**WORKING EXPERIENCE**

**Data Science Researcher, The Johns Hopkins Data Science Lab** Baltimore, MD | Sep 2019 - Present

*Investigation of associations between lifestyle patterns and body mass index (BMI) via generalized linear model*

* Extract time-series data of 32971 subjects from database using **SQL**
* Manipulate features using **normalization** to enhance machine learning pipelines
* Explore user distribution on **Hadoop** using **MapReduce**
* Train generalized linear model (**GLM**) to predict user BMI with 46.07 mean squared error (**MSE**)
* Improve features using **hypothesis testing, Pearson correlation** and **Bonferroni correction**
* Reduce mean squared error by 13% based on results from feature improvement and achieve the conclusion that there are significant associations between unhealthy lifestyle patterns and BMI

**Data Analyst Intern, Johns Hopkins Bloomberg School of Public Health** MD | May 2019 – Aug 2019

*Survival analysis in time-series data using Python, R*

* Cleaned National Health and Nutrition Examination Survey (NHANES) data using **dplyr**, **tidyverse** in R
* Reduced dimensionality of NHANES data using **PCA**
* Selected features using **tree-based model**, **AIC/BIC** to increase the performance(哪方面的performance有所提升？是指后面的accuracy还是其他的比如有关效率，内存优化？) of model
* Trained a neural network using **Keras** to classify patient mortality with 71% accuracy
* Improved the classification accuracy to 86.45% using **regularized logistic regression**
* **实验结论？**

*Classification of patient age based on volumes of brain structures* Baltimore, MD

* Trained a support vector machine (**SVM**) model on 1207 structured data to predict patient age with 0.46 accuracy
* Improved classification accuracy by 5% using **multinomial regression**
* Hosted R shiny website to visualize results of **PCA**, **k-means**, **UMAP** and **t-SNE** using **ggplot2**, **plotly**

**Senior Researcher, Paul C. Lauterbur Lab at SIAT** Nov 2016 - Jan 2017

*EMG signal pattern recognition for hand gestures using spectral analysis*

* Designed, constructed and assembled EMG data acquisition system for arm activities recognition
* Converted time-domain data into frequency domain using **fast fourier transform** and denoised signal
* Classified different hand movements using support vector machines (SVMs) with 82% accuracy
* Improved accuracy by 3% training a **neural network**
* Published a paper in Transactions on Biomedical Engineering and presented a poster in 2017 IEEE IUS

**SELECTED PROJECTS**

**Amazon product review rating prediction** June 2019 – Aug 2019

*Prediction of Amazon product reviews’ ratings using machine learning in Python*

* Extracted Amazon Food Reviews data from Kaggle and cleaned data using **pandas**, **numpy** and **dfply**
* Tokenized unstructured text of user reviews using **scikit-learn** and **nltk** for **NLP algorithms**
* Predicted customer rating categories using **logistic regression** with 0.94 AUC
* Reduced prediction error by 3% using **random forest**
* **项目结论？**

**Investigation of Yelp user funnels, KPIs** Nov 2018 - Jan 2019

*Proposed performance analysis of yelp users & restaurant using SQL*

* Wrote **web crawler** to **scrape data** from yelp using **Xpaths**, **BeautifulSoup** in Python
* Created a database using **MySQL workbench** and imported ~10 GB data file into the database
* Performed **metrics analysis** (bracket retention, DAU/MAU) using SQL to measure customer engagement
* Visualized geographic distribution of restaurants with average ratings using **Tableau**
* **项目结论？**

**SKILLS**

* **Technical Skills:** SQL, R, Python, Shell Scripting, Tableau, SPSS, Scikit-learn, Keras, A/B testing, Natural Language Processing, Hadoop, HDFS, AWS
* **Languages:** Japanese (Basic), Chinese (Native)