|  |  |  |
| --- | --- | --- |
| **LUCHAO QI** | **Research Data Scientist** | <https://www.linkedincom/in/LuchaoQi/> |
| 3111 N. Charles St. #4C | lqi9@jhu.edu | <https://luchaoqi.github.io/> |
| Baltimore, MD 21218 | 443-839-9129 | <https://github.com/LuchaoQi> |

# SKILLS

|  |  |
| --- | --- |
| **Programming:** Python, R, SQL, Batch Scripting | **Packages:** NumPy, Pandas, Tidyverse, NLTK, Keras, PyTorch |
| **Data Science:** A/B testing, Hadoop, Kaggle | **Data Visualization:** Tableau, Matplotlib, Seaborn, ggplot2, plotly |
|  | **Machine Learning:** GLM, Random Forest, SVM, PCA, CNN |

# EDUCATION

THE JOHNS HOPKINS UNIVERSITY, Baltimore, MD

**M.Sc.Degree Biomedical Engineering & Biomedical Data Science:** Expected May, 2020 **GPA:** 3.7

NORTHEASTERN UNIVERSITY, Boston, MA

**B.Eng. Degree in Biomedical Engineering:** August, 2018 **GPA:** 3.9

# SELECTED PROJECTS

**Reinforcement Learning on Super Mario Bros (NES)** Mar 2020 – Apr 2020

**AI that learns to play Super Mario Bros using Deep Q-Network (DQN) in TensorFlow**

**Demo:** <https://github.com/LuchaoQi/Reinforcement_Learning>

* Built reinforcement learning environment using OPENAI GYM and emulated NES using Nes-Py.
* Designed a convolutional neural network (CNN) model with 5 hidden layers as an agent in TensorFlow.
* Trained the agent using deep Q-learning and reduced training time by 20% using Adam optimizer.
* Completed different levels of Super Mario Bros successfully without death which was twice as fast as averaged human players.

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

**Detection of suspicious or fake Amazon product reviews using machine learning**

**Demo:** <https://www.kaggle.com/luchaoqi/making-predictions-over-amazon-recommendation-data>

* Extracted Amazon Food Reviews data from Kaggle and cleaned data using PANDAS, NUMPY and DFPLY.
* Tokenized unstructured text of user reviews using NLTK for feature construction.
* Converted text to vector using bag-of-words model (unigram/bigram) with SCIKIT-LEARN.
* Predicted customer ratings using logistic regression with 0.94 AUC.
* Reduced prediction error by 3% using random forest to improve detection of abusive reviews.

**Investigation of Yelp user funnels, Key Performance Indicators (KPIs)** Jan 2019 – Mar 2019

**Performance analysis of Yelp users & restaurant using SQL**

**Demo:** <https://github.com/LuchaoQi/Yelp_Data_Set_SQL>

* Programmed web crawler to scrape and parse unstructured data from Yelp using Xpaths, BeautifulSoup.
* Created a database using MySQL workbench and imported ~10 GB data file into the database.
* Visualized geographic distribution of restaurants with average ratings using Tableau.
* Created metrics (bracket retention, DAU/MAU) to measure customer engagement and made suggestions for ways to improve upon KPIs via A/B testing.

**LUCHAO QI**

# WORK EXPERIENCE

THE JOHNS HOPKINS UNIVERSITY, Baltimore, MD Nov 2019 – Jan 2020

**Data Science Research Assistant, Data Science Lab**

**Survival Analysis of Time-Series Data Using Python and R**

* Used DPLYR and TIDYVERSE to clean data in National Health & Nutrition Examination Survey (NHANES).
* Reduced dimensionality of data using PCAto capture essence of the data.
* Selected features using tree-based model, AIC/BICto achieve better predictive performance of model.
* Constructed a spectral-based convolutional neural network (CNN) on 3000 patients using Kerasto predict mortality with 71% accuracy.
* Improved mortality prediction accuracy to 86.45% using regularized logistic regression.
* Hosted R shiny website comparing PCA, k-means, UMAP, t-SNEand visualizing clustering results using ggplot2 and plotly. (demo: <https://luchaoqi.github.io/Shiny_clustering/#1>)

**Bloomberg School of Public Health** Summer, 2019

**Data Analyst Intern**

**Association Analysis Between Lifestyle Patterns & Body Mass Index (BMI) via Linear Model**

* Wrangled time-series data of 32971 subjects and built pipeline to front-end dashboard using MySQL.
* Explored user distribution on Hadoopusing MapReduceto maximize dataset value.
* Trained a generalized linear model (GLM) to predict user BMI with 46.07 mean squared error (MSE).
* Reduced prediction error by 13% using ANOVA and feature engineering method (normalization, Random Forest) through 10-fold cross validation.
* Identified statistically significant (p-value < 0.5) impact of lifestyle patterns on BMI to encourage the performance of multiple good health behaviors.

PAUL C. LAUTERBUR LAB **at SIAT,** Shenzhen, CHINA Nov 2016 – Jan 2017

**Senior Researcher**

**EMG Signal Pattern Recognition for Hand Gestures Using Spectral Analysis**

* Designed, constructed and assembled EMG data acquisition system for recognition of arm activities.
* Converted time-domain data of 200 gestures into frequency domain using fast fourier transformto denoise signal.
* Classified different hand movements using support vector machines (SVMs) with 82% accuracy.
* Improved accuracy by 3% in training a neural network, providing insight for medical rehabilitation systems.

# PUBLICATIONS

**Qi L**, Zhang Q, Tan Y, et al. Non-contact High-frequency Ultrasound Microbeam Stimulation: A Novel Finding and Potential Causes of Cell Responses. *IEEE Trans Biomed Eng* 2019.

**Qi L**, Zhang Q, Lam KH, et al. Calcium fluorescence response of human breast cancer cells by 50-MHz ultrasound microbeam stimulation. Presented at 2017 IEEE International Ultrasonics Symposium (IUS), 6-9 Sept. 2017