HARRY XIONG

Chicago, IL 60615 • Phone: (404) 398-3001 • Email: xiong1@uchicago.edu • LinkedIn • GitHub • Personal Website

PROFILE

- Interests: Data Science, Data Analytics, Behavioral Science
- Relevant Courses: Deep Learning, NLP, Speech Recognition, Python Programming, Data Visualization
- Technical Skills: Python, R, TensorFlow, OpenCV, SQL, AWS, Azure, Tableau, STATA, Excel, PowerPoint
- Certificates: Deeplearning.ai: TensorFlow in Practice, Udacity: Data Analyst Nanodegree

EDUCATION

University of Chicago
Master's in computational social science (applied data science & computer programming) | *June 2021*

Emory University
B.A. in Economics, B.A. in Psychology | *May 2019*

London School of Economics
General Course, Concentration in Economics | *June 2018*

PROFESSIONAL EXPERIENCE

Data Scientist | Whiznook, San Francisco (Remote), US | Oct. 2019 – Present

- Designed, implemented and visualized 9 behavioral data science models for an online team analytics game
- Applied sentiment analysis, topic modeling and key word analysis in NLP with RNN, NLTK and SpaCy to provide instant feedback for all customers
- Directly reported to CEO and CTO and defined the business model, total addressable market and product features as one of the founding members

Data Research Assistant | University of Chicago, Chicago, US | Oct. 2019 – May. 2020

- Gathered, assessed and cleaned 3 million rows of data from Nielsen Analytics consumer datasets to investigate the effect of "Chicago Soda Tax" on consumption behaviors
- Utilized Pandas library and behavioral science theories to detect a 30% drop in soda consumption after the tax
- Visualized the difference-in-difference output using R and Tableau

Data Analyst Intern | Business Big Data (BBD), Beijing, China | May. 2019 – Aug. 2019

- Built a retail risk management XGBoost model as part of a team for a major US bank
- Performed feature selection and hyperparameter tuning to lower the retail default recall by more than 1%
- Extracted and analyzed China's daily reported data across 4 sectors for BBD China New Economy Index (NEI) using SQL

PROJECT EXPERIENCE

Face Mask Detector | Jul. 2020 | Link

- Independently built a face mask detector that supports image and live video face mask detection using TensorFlow and OpenCV
- Applied Convolutional Neural Networks (CNN) and transfer learning architecture to achieve above 0.95 validation accuracy on a data set of 1400 images

Evolution of Neuroscience: An NLP Perspective | May. 2020 | Link

- Analyzed neuroscience and psychology abstracts for the past 50 years using PySpark, AWS EC2 and NLP techniques
- Fine-tuned the abstract data with the state-of-the-art GPT2 model and produced zero-shot text generation output to predict the research direction of Alzheimer Disease

Speech Classification | May. 2020 | Link

- Performed MP3-to-MFCCs transformation 2140 same English speech samples of speakers from 177 countries; trained 7 models to classify genders from speech data
- Found significantly less accurate classification results for African speech samples when models trained on samples of White speakers

NLP Topic Modeling on Netflix Data | Apr. 2020 | Link

- Applied unsupervised models such as k-means, silhouette method, Ward's method, latent Dirichlet allocation topic modeling and network analysis to extract patterns on 6234 TV show and movie descriptions