HAOMIN LIN

humaslin97@gatech.edu | (404)-433-0697 | https://www.linkedin.com/in/humaslin | Atlanta, GA

EDUCATION

Georgia Institute of Technology, Atlanta, GA — Expected May 2021
Master of Science in Computational Science and Engineering — GPA: 4.0
Tianjin University, Tianjin, China — Sep 2015 – July 2019
Bachelor of Engineering in Optoelectronic Information Science and Engineering (TJU & NKU Organized) — GPA: 3.6

TECHNICAL SKILLS

Areas of Expertise: Data Analysis, Data Visualization, Machine Learning, Natural Language Processing
Languages and Technology: Python, C, SQL, PHP, R, D3.js, Tableau, SciKit-Learn, PyTorch, Tensorflow, Spark, Hadoop

EXPERIENCE

Teaching Assistant, Georgia Institute of Technology — Jan 2021- Present

- Advise **NLP-related** course projects and instruct students in data analysis with **Spark** and **deep learning** frameworks.

Research Assistant, Georgia Institute of Technology — Sept 2020 - Present
Supervised by Prof. Munmun De Choudhury

- Studied how media portrayed immigrants and caused corresponding effects on society throughout the years.
- Collected text data from newspapers and more recent data via Webhose.io API in **Python** and data of real-world events concerning immigrants from GDELT via **Google BigQuery**.
- Implemented a **word2vec** model with NLP toolkits like **spaCy** to extract images of immigrants constructed by media.
- Ran **regression analysis** on the transformations of immigrant images and real-world events concerning immigrants.

Research Assistant, Northwestern University — July 2020 - Sept 2020
Supervised by Prof. Ágnes Horvát

- Investigated how online media on COVID-19 differed across three platforms including news, blogs, and discussions. The work is summarized in a paper to be published on *Journal of Quantitative Description: Digital Media*.
- Pre-processed over 3 million entries of data to reduce their features to a 120-day series with **Pandas** in Python.
- Applied windowed time lagged cross correlation analysis to present temporal variation in the amount of coverage and sentiment scores with designed visualization in **Tableau** and **Python**.
- Conducted cross-platform **ANOVA** and **post hoc Tukey** tests to validate linguistic disparities in different platforms.

Research Assistant, Indiana University Bloomington — May 2020 - Aug 2020
Supervised by Prof. Xiaojing Liao

- Extracted information of 4,989 products, 2,943 vulnerabilities, plus other information from structured/unstructured text data in 1,316 network security reports and transformed features of vulnerabilities into more generic words.
- Used **regular expression matching** and **POS tagging** to refine the results and then structured the data for storage.

RELATED PROJECTS

Automatic Options Trading Generation Via Distributed Deep Reinforcement Learning

- Designed a system to advise on trading actions based on **Q-learning** in reinforcement learning.
- Built the designed model with **PyTorch**, trained by historical option prices, achieving 10% revenue by month.

Revealing Gendered Language in Job Descriptions

- Scraped over 300,000 text data from Indeed.com to investigate usage of gender languages in online job postings.
- Trained classifiers with **Ensemble Learning** in **Scikit-Learn** to predict salary levels of jobs with 87% accuracy.
- Categorized job descriptions into 16 industries by counting the frequencies of keywords from each industry.
- Visualized the usage of gender language in **Tableau** to present the difference between industries and salary levels.

Community & Digital Archives Project

- Produced a plugin in **PHP** with VADER sentiment analysis lexicon to flag negative comments in the database at backend so that administrators can look into them. Over 80% warnings in a test phase are deemed necessary.
- Trained a model based on **Transfer Learning** in **Tensorflow** to identify maps images with accuracy over 90%.

PUBLICATIONS

- Dambanemuya, H.K., **Lin, H.,** and Horvát, E-Á. Characterising Online Media on COVID-19 during the Early Months of the Pandemic. *Journal of Quantitative Description: Digital Media*. Forthcoming.