

Peiyu Wang

oliviawpy960825@gmail.com | (302)-332-6856

Linkedin: <https://www.linkedin.com/in/peiyu-w-a2a804126/> | Check out my projects: <https://oliviawpy960825.github.io/>

DATA SCIENCE WORK EXPERIENCE

Data Science Intern, Mid-Atlantic Permanente Medical Group, Rockville, MD 01/2020 –Present

- Participate in clinical-related topics and provide insights on relevance and significance through JIRA
- Built a Natural Language Processing model from free text doctor notes to ICD-10 codes with SQL, Python, and Python Machine Learning packages including numpy, scipy, sklearn, pandas, nltk and Tensorflow with one month's work which yielded an average AUC of 0.95 on testing set
- Help set up the distributed computing system architecture with Linux, Docker Swarm, Kubernetes, Hadoop and Spark within the organization to scale up machine learning and deep learning tasks.

Data Analytics Intern, Athenium Analytics, Washington, DC 10/2019 –01/2020

- Gathered, verified and conducted analysis of insurance-related data, built processes, models and constructed project base and structure for entity recognition in various messy and free-text data sets with Python

Data Analyst, MilliporeSigma, Rockville, MD 06/2019 – 08/2019

- Worked with business team and designed and implemented 100% of the reproducible, scalable and efficient Data Cleaning and Integration Pipelines with Excel, SQL, and Python to standardize and automate future Data Cleaning and Integration process based on the inputs, demands and feedback of sales and operations teams
- Designed and implemented 100% of the interactive Data Visualization platform with Tableau for intuitive historical data viewing, auditing and utilizing prediction as baseline for future goal setting and waste reducing

RESEARCH PROJECT

Computational Social Science Project on Misinformation on Social Media 09/2019 - Present

- Crawled and queried over 5GB data from Twitter in JSON format based on keywords related to Chemtrails
- Extracted over 180,000 unique users out of the datasets, separated users into suspended, bots, cyborg and human based on criteria defined in previous research and conducted exploratory analysis
- Extracted common words, top topics from tweets with LDA topic modeling, explored the difference in the proportions of malicious URLs by category to analyze the intent of spreading misinformation by user category

EDUCATION

George Washington University, School of Engineering and Applied Science, Washington, DC, **GPA: 3.9/4.0**

Master of Science in Data Analytics, Tracks: Systems Engineering & Computer Science 08/2018-05/2020

University of Delaware, Alfred Lerner Business College, Newark, DE (Dual Degree), **GPA: 4.0/4.0**

Bachelor of Science in Management Information Systems, Minor: Computer Science 08/2016-05/2018

Southwestern University of Finance and Economics, School of Economic Information Engineering, Chengdu, Sichuan, China (Dual Degree), **GPA: 3.8/5.0**

Bachelor of Business Administration in E-commerce 09/2014-05/2018

SELECTED HONORS & AWARDS

- Distinguished Senior Panel for Management Information Systems in 2018 (From University of Delaware)
- Distinguished Junior Panel for Management Information Systems in 2017 (From University of Delaware)