HUI (HENRY) CHEN

+1 (347) 223-1312 hchen98x@gmail.com linkedin.com/in/hchen98

EDUCATION

New York Institute of Technology New York, NY

Master of Science in Data Science (CGPA: 3.96/4.00) Sep 2021 - May 2022

- **Courses**: Data Visualization, Computational Statistics, Optimization, Big Data Analytics, Machine Learning, Deep Learning

Bachelor of Science in Computer Science (CGPA: 3.51/4.00) Jan 2018 - May 2021

- **Concentration**: Big Data Management and Analytics with the distinction of Magna Cum Laude
- **Google Developer Student Club:** Tech Lead - collaborated with various clubs to host and lead intensive sessions on DevOps and a hackathon on sustainable transport challenges.
- **Courses**: Data Structure, Probability and Statistics, Data Mining, Information Retrieval, Distributed Database Systems

SKILLS

Programming Languages: Python, SQL, R, Java, JavaScript (React Native), PHP, HTML/CSS, MapReduce, Bash
Libraries: Scikit-Learn, pandas, numpy, cupy, matplotlib, seaborn, scikit-image, NLTK, cuml
Tools: Excel, Git, Flask, PySpark, Hadoop, Apache Superset, spaCy, MongoDB, Firebase, Redis, Docker
Data Science: Data cleaning, Data visualization, Data mining, Regression, Classification, Deep Learning, NLP

EXPERIENCE

Data Scientist Intern Feb 2022 - Present
JobLogic-X Corporation *New York, NY*

- Led a team of 6 Data Scientists to design and implement end-to-end recommendation engines for Meet Fresh Inc.
- Undertook Needfinding and analyzed the results from a Product Management perspective in order to identify new products and clients.

Graduate Assistant Sep. 2021 - May 2022
New York Institute of Technology *New York, NY*

- Developed and deployed a Native Land app for bringing awareness of the US indigenous land through AWS Serverless Lambda, DynamoDB, React Native, and Node.JS.

PROJECTS

Personal Site: hchen98.github.io (for additional information and projects)

Image Classification | *Python, TensorFlow, Nvidia Rapids, Scikit-learn, CuPy, opencv* Feb. 2022 - May 2022

- Analyzed the imbalance data of different handcrafted features (LBP, HoG, SIFT) and feature selection (PCA) with SVM for medical and facial expressions recognition using GPU acceleration to achieve 93.3% accuracy and 0.93 AUC.
- Applied image augmentation techniques to enrich the training data and boost the SVM model accuracy by 18.5%.
- Documented the experiment results of stratified cross-validation and feature engineering through Excel, plotly, and Weights & Biases.

Job Skillset Seeking | *Python, R, pandas, plotly, spaCy, NLTK, shiny, ggplot2* Feb. 2022 - May 2022

- Built an interactive data visualization dashboard to better understand the job skill datasets through R, plotly and, shiny.
- Applied spaCy and NLTK tool on job documents to extract keyword data, tokenization, and lemmatization from utils packages to better understand NLP.

Scholar Seek | *Python, JavaScript (React), Web Scraping, AWS, MongoDB, Flask, Docker* Feb. 2021 - May 2021

- Showcased an end-to-end cross-platform mobile app for students to create their profiles and receive personalized content-based filtering (Cosine Similarity) in scholarships, colleges, and majors to the NYIT engineering department.
- Web-scraped 2.7 million rows of semi-structured scholarship data and 3.5k+ rows of unstructured US-college data by using selenium with anti-captcha and real-time authentication.
- Designed and documented RESTful APIs for backend server to enable a secured and encrypted token integration of the client-side devices, recommendation models, and MongoDB through Flask and JWT.

Airbnb Open Data (NYC) | *Python, scikit-learn, Scipy, pandas, Folium, seaborn* Sep. 2020 - Dec. 2020

- Built linear regression models for Airbnb price prediction by examining feature relations, data exploratory analysis, feature engineering, and hyperparameter tuning through Grid Search and k-fold cross-validation.
- Created an interactive map for 44k+ rows of semi-structured data for visualization and analytics through Folium.