

# 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, 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

■ **Courses:** Data Structure, Probability and Statistics, Data Mining, Information Retrieval, Distributed Database Systems

## SKILLS

**Programming Languages:** Python, SQL, NoSQL, R, Java, JavaScript (React Native), PHP, HTML/CSS, MapReduce, Bash

**Libraries:** Scikit-Learn, pandas, numpy, cupy, matplotlib, seaborn, XGBoost, 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

### JobLogic-X Corporation

New York, NY

Data Scientist Intern (Scikit-learn, Python, Streamlit, XGBoost, opencv)

Feb 2022 - Jul 2022

- Led a team of 6 Data Scientists to design and deploy an end-to-end hybrid recommendation engine with computer vision.
- Undertook Needfinding and analyzed the results from 90+ people in order to identify new products and clients.
- Conducted EDA on the client data with 46 features to investigate trends, outliers, missing data, anomalies, and bias.
- Enhanced training data using image augmentation and over-sampling on multiclass XGBoost to achieve 94% recall.
- Optimized the recommendation engine by using CV to cluster products on handcraft features (HoG, LBP, and SIFT).

### New York Institute of Technology


New York, NY


Graduate Assistant (Node.JS, AWS, Python, JavaScript, React Native)

Sep. 2021 - May 2022

- Showcased an end-to-end mobile app on a conference implemented and deployed through AWS Lambda, DynamoDB, React Native, and Node.Js in order to bring awareness of global native land.
- Implemented an internal tool to perform and boost plagiarism checks through the Cosine Similarity algorithm and resulting in reducing 70% time on plagiarism checking.

## 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 cross-validation and feature engineering through Excel, plotly, Weights & Biases.

**Job Skillset Seeking**  | Python, R, pandas, plotly, spaCy, NLTK, shiny, ggplot2, NLP

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.
- Deployed the backend for production and test environments, performed functional testing, and containerized the recommendation models, RESTful APIs, and web-scrapings to AWS EC2 through Docker.