

# HUI (HENRY) CHEN

[hchen98x@gmail.com](mailto:hchen98x@gmail.com) | [linkedin.com/in/hchen98](https://www.linkedin.com/in/hchen98) | +1 (347) 223-1312

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

Bachelor of Science in Computer Science (CGPA: 3.51/4.00)

Jan 2018 - May 2021

- Concentration in 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

- **Code:** Python, R, SQL, Java, PHP, JavaScript, MapReduce, Linux | **Databases:** MySQL, MongoDB, Firebase
- **Tools:** scikit-learn, Excel, TensorFlow, Keras, pandas, NumPy, Flask, PySpark, Hadoop, NLTK, spaCy, Docker
- **Machine Learning:** EDA, Multiple Linear Regression, Logistics Regression, KNN, SVM, Naive Bayes, Decision Tree, Random Forest, K-Means, Hierarchical Clustering, BoW, TFIDF, PCA, LBP, SIFT, HoG, K-Fold Cross-Validation, CNN, Perceptron, and Natural Language Processing.
- **Visualization:** matplotlib, plotly, seaborn, shiny, ggplot2, streamlit, Apache Superset

## WORK EXPERIENCE

### JobLogic-X Corporation

New York, NY

Data Scientist (Intern)

Feb 2022 - Present

- 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.

### Invisible Hands Inc

New York, NY

Software Engineer (Volunteer)

Aug 2021 - Present

- Maintained and upgraded an existing legacy system of 16.6K+ users in order to prevent potential security breaches.
- Designed and integrated real-time interactive geolocation map for serving 16.6K+ NYC users via the Google API.

### New York Institute of Technology

New York, NY

Graduate Assistant

Sep 2021 - May 2022

- 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](https://hchen98.github.io) (for additional information and projects)

### Image Classification: Feature Selection, Data Augmentation, and Transferred Learning

Feb 2022 - May 2022

- Built discriminative (binary and multiclass) SVM models with LBP, HoG and SIFT handcrafted features and dimensionality reduction PCA, using Nvidia Rapids GPU, achieved 93.3% accuracy and 0.93 AUC.
- Applied image augmentation to enrich the training data and boost the SVM model accuracy by 18.5%.
- **Utilized:** TensorFlow, scikit-learn, Nvidia Rapids GPU, cuml, numpy, opencv, scikit-image, pandas, plotly, matplotlib

### Data Visualization: Job Skillset Seeking

Feb 2022 - May 2022

- Built an interactive data visualization dashboard to better understand the job 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.
- **Utilized:** spaCy, NLTK, Python, plotly, ggplot, matplotlib, R, shiny, pandas

### Scholar Seek: Scholar Recommendation App

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.
- **Utilized:** React Native, Python, Git, JavaScript, MongoDB, Flask, Selenium, AWS EC2, Docker, PyTest, and shell.

### Linear Regression: Airbnb Open Data (NYC)

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.
- **Utilized:** Git, Python, scikit-learn, Folium, seaborn, Scipy, pandas, NumPy, matplotlib, and RapidMiner Studio.