Hui (Henry) Chen

Data Scientist · M.S. in Data Science · New York Institute of Technology

347-223-1312 | hchen98x@gmail.com | hchen98.github.io | hchen98 | hchen98

About Me

Programming Python · Java · JavaScript · NodeJS · MapReduce · Bash

Analytics Data cleaning/visualizing · SQL/NoSQL · Hypothesis Testing · Tableau · Access · Excel · R · Hadoop

Tools Jira · Git · AWS · Docker · Machine Learning · NLP · Deep Learning · Unix/Linux

Work Experience

JobLogic-X Corp.
New York, NY

Data Scientist Intern - Product Team
Feb 2022 - Jul 2022

- Collaborated with cross-functional partners to design and deploy a recommendation engine with computer vision to leverage business metric.
- Assessed the results of Needfinding from 90+ people through a Product Management perspective to identify new products and clients.
- Conducted EDA on the client data with 46 features to investigate trends, outliers, missing data, anomalies, and bias.
- Enriched training data using image augmentation and over-sampling on multiclass XGBoost to achieve **94%** recall.
- Enhanced the recommendation engine by applying unsupervised nearest neighbors to cluster the products through different handcraft features.

New York Institute of Technology
New York, NY

Lead Project Coordinator & Graduate Assistant
Sep. 2021 - May 2022

- Showcased and deployed 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.
- Scrapped and aggregated 4.2k data from multiple sources and performed data cleaning for analysis and visualization on Tableau.
- Implemented an internal tool to automate the grading and plagiarism checking through Git and JUnit and resulting in **70%** time cut-down.

Projects

Img Classification: Feature Selection, Data Aug, and Transfer Learning
Feb 2022 - May 2022

- Analyzed the imbalance data of different handcrafted features (LBP, LBP+PCA, HoG, SIFT) and feature selection (PCA) with SVM for medical diagnosis and facial expressions recognition via GPU acceleration to achieve 93.3% accuracy and **0.93** AUC.
- Built pipelines for image augmentation and fine-tuning to enrich the training data and boost the SVM accuracy by **18.5%**.
- Automated the experiment results tracking of cross-validation and feature engineering using Python, Excel, plotly, Weights & Biases.

Job Skillset Seeking: Data Visualization
Feb 2022 - May 2022

- Built and deployed an interactive data visualization dashboard to better understand the job datasets through R, plotly, and shiny.
- Aggregated data from multiple sources and performed data cleaning to identify the trends.
- Applied spaCy and NLTK tool on job documents to extract keyword data, tokenization, and lemmatization from utility packages to better understand NLP.

Scholar Seek: Scholarship Recommendation
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 in scholarships, colleges, and majors to the NYIT Engineering Department.
- Implemented an automated tool for scrapping and cleaning **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 the 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, testing, and containerized the recommendation models, RESTful APIs, and web scrapings to AWS EC2 through Docker.

nax: Self-driving Vehicle
Feb 2019 - Apr 2019

- Assembled an IoT RC car using a PCA servo driver, Raspberry Pi, and Donkey Car API to control steering speed.
- Showcased the project at NYIT Ventures' Pitch Contest to college faculties and students and won the President Award.
- Collected and trained image data for a CNN autopilot model using physical camera FOV, RC car direction, and simulation software.
- Evaluated and interpreted CNN autopilot model with 11 layers prediction result with 92% accuracy.

Education

New York Institute of Technology
New York, NY

M.S. in Data Science
Sep 2021 - May 2022

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

New York Institute of Technology
New York, NY

B.S. in Computer Science
Jan 2018 - May 2021

- Related Courses: Data Structure, Probability and Statistics, Data Mining, Information Retrieval, Distributed Database Systems