

Chi-Yun Chen

☎ +1 (202) 868 0865 | ✉ chiyun.chen001@gmail.com | 📱 cychen116 | 🌐 chi-yun-chen

EDUCATION

MS, Data Analytics ♦ The George Washington University

Washington DC, US

- Courses: Machine Learning, Statistics, Big Data, Database Management, Algorithms, Marketing Analysis, etc.
- Honors: GPA 3.8, Fellowship & Best ML/AI approach prize in 2021 GW Hackathon.

Expected Dec 2022

BS, Food Science and Biotechnology ♦ National Chung Hsing University

Taichung, Taiwan

- Courses: Statistic Methods and Data Analysis, Game Theory, etc.

Jan 2016

EXPERIENCE

AI Research Fellow

Remote, US

AI For Good Foundation

June 2022 – present

- Improve word embedding NLP model of Climate Trend Scanner Project by noise reduction classifier.
- Establish an object detection model of wildlife and an individual identification model for field cameras to study animal behavior.

AI/ML Intern

California, US

Beagle Technology Inc.

May 2022 – present

- Design a relational database and draw the schema, which organizes the information of thousands of images for company's dataset.
- Construct YOLOv4 object detection model of plant and fruit real-time recognition for agriculture machines.

IT Administration Data Project Assistant

Washington DC, US

The George Washington University

Feb 2022 – present

- Collect and wrangle the computer data among 6,000 staff, generate a performance dashboard with insight, and validate data integrity.

Data Intern

Remote, US

iChemLabs LLC.

Jul 2020 – Sep 2021

- Evaluated thousands of chemical compound data and wrote testing cases in the Java environment.

Pharmaceutical Research Assistant

Taipei, Taiwan

National Yang Ming University

Jul 2018 – Aug 2019

- Designed biotechnology experiments and analyzed data with Excel and R on nanoparticles in cancer therapy.

Product Manager and Data Analyst

Taipei, Taiwan

Innovation Open House

Jan 2016 – Jul 2017

- Developed SQL database from scratch and systematized data storage by building NAS for a startup company.
- Implemented A/B testing for video product improvement resulting in a 200% improvement in average view duration.
- Inspected website performance data with Google Analytics and executed SEO, resulting in an increase of 74% of sessions.
- Analyzed product data with SQL, then created an automatic performance evaluation system resulting in a 20% turnaround time reduction.
- Utilized statistical methods and Excel pivot table to execute marketing campaign and increased engagements of the FB Fan Page by 30%.
- Designed and maintained KPI dashboards among 20 interns and led new employee orientation.
- Collaborated and led a cross-functional team 3 times for a live-stream event owns tens of thousands of audiences.

SKILLS

Programming Languages: Python (*NumPy, pandas, SciPy, TensorFlow, PyTorch, scikit-learn, etc.*), R (*Dplyr, Knitr, RMarkdown, Shiny, etc.*), SQL.

Data Science & Miscellaneous Technologies:

- Data Science Pipeline: ETL Pipeline, Cleaning, Wrangling, Exploratory Data Analysis, Modeling, Interpretation.
- Data Visualization: Python (*matplotlib, seaborn*), R (*ggplot2*), Tableau, QGIS.
- Statistics Analysis: Regression Analysis, ANOVA, PCA, Time Series, A/B Testing, MATLAB, GA, Excel.
- Others: Machine Learning, Deep Learning, Natural Language Processing, Text Processing, Topic Modeling.

Big Data & Database Management: MySQL, MongoDB, Neo4j, MLaaS (*AWS, Google Cloud*).

Language: English, Chinese, German.

PROJECTS

American Sign Language Translation Approach by Machine Learning

bit.ly/SL_medium

- Regulated motion capture and computer vision recognition module with OpenCV and Mediapipe libraries.
- Built Deep an Encoder-Decoder Seq2Seq Model and Convolution 3D Classification deep learning with PyTorch.

Analysis of COVID-19 Topic on Facebook/Twitter in LDA Evaluation Metrics by PCA

bit.ly/LDA_dan_21

- Extracted data from open API and deployed topic modeling and Latent Dirichlet Allocation (LDA) on NLP studies with Python.

Exploratory Data Analysis Repository Using Statistical and Machine Learning Methods

bit.ly/EDA_github

- Developed data pipeline and ETL process for various datasets, which includes time series, statistics analysis, and prediction modeling.

Analytic SQL Database Management and Recommendation Function of Yelp Dataset

bit.ly/yelp_github

- Restructured dataset schema into AWS RDS and Python SQL libraries, then constructed recommendation function by ML/DL models.