Chi Yun Chen

☎ +1 202 868 0865 | ✉ chiyun.chen001@gmail.com | ⬚ cychen116 | in chi-yun-chen | 🌐 website

EDUCATION

MS, Data Analytics, CS track ◆ The George Washington University — *Washington, D.C., US*
- Courses: Machine Learning, Statistics, Big Data, Database Management, Algorithms, Marketing Analysis, etc. — *Expected Dec 2022*
- Honors: Fellowship funding & Best ML/AI approach prize in 2021 GW Hackathon.

BS, Food Science and Biotechnology ◆ National Chung Hsing University — *Taichung, Taiwan*
- Courses: Statistic Methods and Data Analysis, Game Theory, etc. — *Jan 2016*

EXPERIENCE

Data Scientist Intern — *California, US*
Beagle Technology Inc. — *May 2022 – present*
- Collaborate with robotic team to build a real-time YOLOv4 object detection model and optimize recall by 16% with PyTorch.
- Develop data infrastructure and schema of RDB with Python and Google Cloud SQL, resulting in a 20% ETL processing time reduction.
- Manage Google Cloud Storage and create a dashboard with PostgreSQL to deliver quantitative analysis for cross- team project management.
- Program automatized data pipeline, such as pre-labeling function, for image segmentation model in computer vision model with Python.

Data Analyst, I.T. Administration — *Washington, D.C., US*
The George Washington University — *Feb 2022 – Aug 2022*
- Collected, organized, and wrangled the I.T. data among 6,000 staff, then validated data integrity with Excel and Python.
- Analyzed I.T. data with Excel Pivot Table, meanwhile generating a performance dashboard with insight to support project management.

Data Intern — *Virginia, US*
iChemLabs LLC. — *Jul 2020 – Sep 2021*
- Evaluated thousands of chemical compound data and wrote testing cases to validate with Java.

Research Assistant, Pharmaceutical Department — *Taipei, Taiwan*
National Yang Ming University — *Jul 2018 – Aug 2019*
- Designed biotechnology experiments and statistics model, then analyzed data with Excel and R on nanoparticles in cancer therapy.
- Leveraged experimental data to develop research methods in custom-made quantitative frameworks.

Project Manager and Data Analyst — *Taipei, Taiwan*
Innovation Open House — *Jan 2016 – Jul 2017*
- Established a relational database from scratch and systematized data storage by setting up a NAS system for a startup company.
- Scoped product data with SQL and Excel, then reduced 20% turnaround times by creating an automatic performance evaluation system.
- Designed and maintained KPI dashboards with Excel and Tableau among 20 interns to support team management and led the orientation.
- Utilized statistical methods and Excel pivot table to execute marketing campaigns and increased engagements of the FB Fan Page by 30%.
- Partnered with marketing team to inspect website performance data with GA and executed SEO, resulting in 74% sessions increased.
- Implemented A/B testing for video product improvement, resulting in a 200% improvement in average view duration.
- Coordinated and led a cross-functional team 3 times for a live-stream event that owns tens of thousands of audiences.

SKILLS

Programming Languages:
- Proficient: Python *(NumPy, pandas, SciPy, TensorFlow, PyTorch, scikit-learn, etc.)*, SQL.
- Advanced: R *(Dplyr, Knitr, Shiny, RMarkdown, etc.)*.
- Prior Experience: Java, JavaScript, HTML.

Languages:
- Professional/ Native: English, Chinese.
- Intermediate: German.
- Beginner: Spanish.

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

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

PROJECTS

American Sign Language Translation Approach by Machine Learning with Python — *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 with Python — *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 with ML/DL models.