Allen Lau

Contact: (484) 855-9707 | Email: allenlau3@outlook.com | Address: Queens, New York, 11356
Developer Portfolio: https://lau-allen.github.io | LinkedIn: http://www.linkedin.com/in/allenlau2019

Education

The City College of New York — August 2022 – May 2024
Master of Science, Data Science & Engineering | GPA: 3.97/4.00

Purdue University — August 2015 – May 2019
Bachelor of Science, Mechanical Engineering | GPA: 3.80/4.00

Projects

❖ **Modeling and Extraction of Forest Features** — March 2023 – Present
- Researched and applied Neural Radiance Field (NeRF) deep networks using *GCP Compute Engine* to model forests and extract important features to gauge the health of forest ecosystems, showcasing proficiency in computer vision, data processing, and statistics.
- Achieved notable results, including spatial pose errors of 0.11 meters, trunk diameter errors of 10.9%, and a depth map RMSE of 0.23 meters, benchmarked against ground-truth data acquired from LiDAR SLAM systems.

❖ **Sentiment Analysis ETL Pipeline** — February 2024 - March 2024
- Designed and implemented an ETL pipeline, utilizing *Prefect Workflow Orchestrator*, extracting financial data from 100+ websites per run, transforming raw HTML data into cleaned formats, and loading it into *AWS S3* buckets for sentiment analysis.
- Attained high performance and scalability by deploying *Docker* containers on *AWS ECS* through *Terraform*, leveraging parallelized computation (Dask, Asyncio), and storing transformed data in *AWS Redshift*, achieving an average pipeline latency of 23 seconds.

❖ **Big Data Analytics of Food Insecurity** — May 2023 – May 2023
- Conducted an in-depth analysis of the food insecurity problem in NYC, utilizing *GCP Dataproc* to run Apache Spark MapReduce-based logic to achieve the quick and efficient parsing of 10 GB data (9.2 million rows) in less than 2 minutes.
- Leveraged *Apache Spark* Resilient Distributed Datasets (RDD) and DataFrame implementations to construct scalable logic for execution on Google Cloud Dataproc's Apache Hadoop YARN framework.

❖ **Search Engine ETL Pipeline** — April 2023 – May 2023
- Engineered a pipeline, in Python (Requests, Asyncio, BeautifulSoup), to query search engines results, scrape HTML text data, and store the transformed, processed text and metadata in a *MySQL* database at an average rate of 5 webpages per second.
- Designed and implemented a user-friendly GUI with the *Flask* Web Framework, CSS/HTML, and a keyword ranking algorithm to streamline the querying of new search engine results and accessing stored data from the RDBMS.

❖ **Sign Language Interpreter** — March 2023 – May 2023
- Utilized computer vision techniques, in *Python* (OpenCV, MediaPipe), for hand detection and tracking via a camera, coupled with machine learning algorithms (Scikit Learn, TensorFlow) to achieve real-time classification of signed letters with over 90% accuracy.

❖ **Hospital Mortality Classification Mode** — November 2022 – December 2022
- Trained a K-Nearest Neighbors model, in *Python* (Pandas, NumPy, Scikit-Learn), for predicting patient mortality events with an accuracy of 88%, after applying techniques for feature selection like random forest methodology.

Professional Experience

❖ **Graduate Research Assistant**, The City College of New York (New York, NY) — March 2023 – Present
- Collaborated with research professors to build a pipeline with the goal of modeling forest environments via images, obtaining photorealistic renderings, and extracting structural characteristics like tree dimensions via point clouds.
- Orchestrated detailed literature reviews, communicated relevant research findings, and managed weekly progress meetings to ensure the timely completion of project milestones.

❖ **Technical Solutions Engineer**, Epic Systems Corporation (Verona, WI) — August 2019 – June 2022
- Partnered with three healthcare systems to identify and manage projects, ensuring successful implementation and operationalization of functionality with Epic's Electronic Medical Record (EMR) and clinical application.
- Acted as a subject-matter expert for the Wound Care app to provide expertise and drive the success of the product by leading product review initiatives and discussions with more than 15 healthcare systems across the United States.
- Resolved and troubleshoot 500+ system build issues to maintain the system health of the EMR and to address the workflow concerns of stakeholders, while collaborating and communicating with a wide variety of functional teams.

Technical Skills

❖ **Programing Languages:** Python, SQL, R, MATLAB
❖ **Data Engineering:** Database Management, Workflow Orchestration, ETL (Extract, Transform, Load), Big Data Analytics, Distributed Computing, Infrastructure-as-Code (IaC), Dashboarding
❖ **Data Science:** Data Preprocessing, Exploratory Data Analysis, Feature Selection, Dimensionality Reduction, Machine Learning, Natural Language Processing, Computer Vision
❖ **Software Tools:** Google Cloud Platform (Compute Engine, BigQuery, Dataproc, Bucket), Amazon Web Services (Elastic Container Service, Redshift, S3), Prefect, Terraform, Docker, Apache Spark, Git, dbt, MSSQL, MySQL