Anna Louise Chen

www.linkedin.com/in/chnanna · https://github.com/acanna2013
xxxx, NY · (xxx) xxx-xxxx · ac.anna2013@gmail.com

Education

Purdue University, West Lafayette, IN — Aug 2021 - May 2025
Bachelor's of Science in Computer Science, Data Science — GPA x.x

- Dean's Honors List, Presidential Scholarship, Machine Learning @ Purdue, Vertically Integrated Projects (Research)

Professional Experience

Software Engineering - Machine Learning Intern | Tesla — Jan 2024 - Jun 2024

- Developed a convolutional neural network from scratch, achieving a final accuracy of 95.73%, improving the initial model by 8%, through hyperparameter tuning and data augmentation techniques in Python.
- Deployed binary and multi classification models onto Amazon AWS S3 buckets for real-time data evaluation.
- Automated the image labeling process using Optical Character Recognition (OCR), utilizing the EasyOCR and PaddleOCR Python libraries, significantly reducing manual work.
- Implemented an OCR model into production with a 99.98% accuracy, enhancing efficiency, scalability, integration by 93%.
- Automated notification systems using email and Mattermost to track model performance and alert employees about inconsistencies/complications.
- Implemented multiple feature enhancements in the team's main application, optimizing backend and frontend changes, resulting in a 7% reduction in loading time.

Data Science Intern | Caterpillar (Cat Digital) — May 2023 - Aug 2023

- Designed and implemented a robust relational database system, creating a well-structured schema for soil data, resulting in a streamlined data access experience for other teams.
- Developed a rules-based classification model to accurately classify different soil types as hard vs. soft, resulting in 99% accuracy.
- Contributed to the company's knowledge base by deploying a database on Chinese soil onto Snowflake through Snowpipe.
- Wrote a Python script that successfully processed and reverse geocoded over 1 million geospatial records, allowing seamless merging with internal databases.
- Utilized Azure DevOps to track and prioritize user stories, features, and bug fixes.

Machine Learning Researcher | NASA (Lightning Wildfire Lab x Prof. Wang) — Jan 2023 - May 2023

- Facilitated comprehensive analysis by developing a Python script using Xarray and NETCDF4 to organize and visualize geospatial data from NASA/NOAA satellites.
- Trained Neural Networks on satellite data to predict the probability of ignition for wildfires in the US.

Data Science Intern | Indiana Family & Social Services Administration (FSSA) — May 2022 - Aug 2022

- Wrote and optimized SQL queries, using Azure Synapse Analytics and Data Factory, to perform data mining on large data-sets, reducing data processing time by 20%.
- Created and published a Tableau dashboard to visualize key metrics about Medicaid providers and their effects on service recipients with intellectual and development disabilities.
- Presented insights and actionable recommendations, derived from data, to stakeholders at the executive level.

Projects

Cirrhosis Analysis | Python, Matplotlib, Scikit-Learn, Google Colab

- Created a model using the decision tree algorithm to predict the patient's status, resulting in 71.43% accuracy.

College Portal Application | Java, JavaScript

- Designed the complete database of the college with functionalities such as courses managements, assignment tracking, and forum discussions.

Skills

- **Technical:** Python, Pandas, Jupyter Notebooks, scikit-learn, Keras, Java, SQL, MySQL, DBeaver, Azure Synapse Analytics, Azure Databricks, Snowflake, Azure DevOps, Tableau, MongoDB, Microsoft Word, Microsoft Excel