Jalan Plumbum 7/103,
Shah Alam 40000, Selangor

WEI ZHE LIU

GitHub | LinkedIn | Malaysian

+1(765) 701-9494
lwz1030@gmail.com
Personal site

Education

West Lafayette, IN, United States **Purdue University** **Aug 2020- May 2024**

- Bachelor of Science in Data Science, May 2024, In-major
- Coursework: Data Mining and Machine Learning, Large Scale Data Analytics, Information Systems, Statistics For Data Science

Technical Experience

Autonomous Vessel Design Team Member for AIMM ICC Competition **January 2024- May 2024**

- Design and programming of a fully autonomous low-profile vessel for AIMM ICC, with a focus on computer vision tasks.
- Orchestrated research and creation of a robust dataset for classification, addressing variations in lighting, backgrounds, and angles. Implemented Python script with YOLOv8 for effective task execution.
- Demonstrated technical prowess and contributed to the success of the project, enhancing skills through feedback and experience.

nanoHUB Data Analyst **August 2023- December 2023**

- Enhanced user experience on nanoHUB, a leading online platform for nanotechnology, through refining the classroom cluster detection system with advanced algorithms.
- Implemented an incremental approach, focusing on early data usage to predict emerging trends in classroom dynamics and proactively identifying behavior patterns for optimized user experience and resource allocation.
- Played a key role in code development, statistical analysis, and data visualization for the clustering algorithm, presenting findings to the team in weekly meeting with mentors.

Lightning Wildfire Research Lab (Purdue EAPS X NASA) **January 2023- December 2023**

- Employed statistical models to examine satellite data and identify the correlation between dry lightning and wildfires counts.
- Improve wildfire danger prediction by training machine learning models on meteorological, vegetation, and human-related data.
- Analyze each environmental parameter's predictive power, specifically for lightning and humidity, and compare data with traditional and Machine Learning-based fire danger metrics.

Projects

Kaggle Bank Lending Prediction

- Participated in a community Kaggle competition focused on making lending (binary classification) for a bank, ranked 12th place among a total of 130 classmates
- Conducted data exploration and cleansing for modeling, including handling missing values, removing outliers, and encoding categorical variables
- Trained and evaluated multiple classification models, including Random Forest, Gradient Boosting, and AdaBoost, using a variety of performance metrics

Forecasting Global Active Power Using LSTM Neural Network and Time Series Analysis

- Performed time series analysis on the household electric power consumption dataset from the UCI Machine Learning Repository, uncovering a yearly seasonality and weekly periodicity in the global active power.
- Developed and trained a LSTM network to forecast the global active power time series, using mean squared error as the loss function and Adam optimizer
- Visualized the predicted values and the actual values in plots, validating the model's performance and showcasing its potential for practical applications

Senior Design Project- Indiana Corn Yield Predictor

- Created Led the Implementation of machine learning methodologies to forecast state-level corn production, orchestrating the data pipeline encompassing extraction, cleaning, modeling, and refinement phases
- Designed the integration of corn yield, weather, and soil data from diverse sources, ensuring data compatibility and coherence to construct a robust dataset tailored for machine learning training
- Directed a team of five in code development, research, technical report writing, and presentation delivery to mentor, ensuring collaboration throughout the project life cycle

Languages and Technologies

- Python, R, GitHub, SQL, Power BI, Java, Roboflow, MySQL, Microsoft Office, ChatGPT, Google Cloud, Docker
- Big Data & Machine Learning: scikit-learn, PyTorch, Keras, OpenCV, Matplotlib, Spark, Hive, Hadoop, YOLOv8
- Data Science pipeline (cleansing, wrangling, visualization, modeling and interpretation)