Sherie (Ping-Hsuan) Lin

pinghsuan04@gmail.com | +1 (937) 6722-634 | https://www.linkedin.com/in/pinghsuanlin/
Portfolio: https://github.com/Pinghsuanlin/Portfolio | https://public.tableau.com/app/profile/sherie5055

Summary

Internationally experienced, metric-driven data enthusiast with 3+ years of experience developing and implementing data analytics programs to **optimize operational processes, accelerate project execution, and facilitate automation strategies**. Graduate of Columbia University with a MS degree in Computer and Information Science, **a graded average of 4 /4.** Experienced **handling large data, developing, and deploying machine learning approaches, and collaborating with multi-disciplinary teams to discover business problems and deliver data-informed recommendations**.

CORE COMPETENCY & TECHNICAL SKILLS

Programming Languages: Python, R, C# (Blazor)
Data Visualization: Tableau, Power BI, R Shiny, Streamlit
Data Science Toolkit: Scikit-Learn, NumPy, Pandas 2.0, H2O
Database Platform: Postgres, MongoDB, NoSQL, ERStudio
Big Data Environment: GCP, AWS, E*xcel Solver*, *Pivot Table*
Certificate: AWS Cloud Practitioner

WORK & PROJECT EXPERIENCE

AES Corporation- Data Scientist (AI/ML) – Innovation & Optimization; Dayton, OH Mar 2023 – Present

- *Capacity Planning*: Collected and cleaned data from 161 MW solar sites and performed capacity simulation and forecasting with pvlib, time series and Long-Short Term Memory Network.
- *Production Planning*: Created network diagram and Gantt chart early, late schedule to improve inventory turnover by 12%.
- *Fault Alarm Tracking Dashboard*: Built a data pipeline with Google Cloud Platform to connect 215 sites' real-time data and visualize the results with Power BI, optimizing alarm reporting SOP and saving operational engineers 4hr of work per week.
- *Vegetation Management Database Builder and ML Application*: Researched multiple document studies with LangChain embeddings to fasten data collection process. Conducted PDF scraping with opencv-python, pytesseract, and built customized dashboard. Performed factor, regression analysis on vegetation management methods.

The Ether Group- NYSERDA Program – Energy & *Data Analytics Intern*; New York, NY Oct 2022 – Dec 2022

- *Streamline Data Extraction Process*: Researched and conducted web scraping with Python Selenium package and Excel VBA across data banks related to Climate Mobilization Act and turned data into structured datasets.
- *Building HVAC Data Factor Analysis for Carbon Emission Prediction*: Implemented KNN and multiple Neural Net to program out contributing factors to carbon emission and benchmark ratings. Used Tableau to build audience-informative dashboards.

Protect-Eco Consultancy Co. Ltd - *Environmental Engineering Consultant*; Taipei, Taiwan Aug 2020 – Jul 2021

- *Project Management:* Coordinated with RFMB, Taiwan gov. to track and build financial reports of public recycling funds. Led and monitored 22 local governments' environmental project status updates with 100% timely completion.
- *2015- 2020 Fiscal Year Recyclables Data Performance Analysis*: Assessed past 5-year subsidy rates of 33 disposal article types to stipulate potential levies and subsidies for the next fiscal year. Collected, structured, and cleaned regulated recyclables data to create ad-hoc reporting.
- Bi-annual Recycling Truck Licensing Procurement Research: Researched and formulated bi-annual recycling truck licensing procurement with 10 importers and manufacturers for RFMB.

EDUCATION

Columbia University in the City of New York, New York, NY
MS Applied Analytics 09/09/2021- 08/31/2022
Relevant Coursework: Anomaly Detection, Data Modeling, Cloud Computing), Research Design, Storytelling with Data.

Capstone Project: Global Overview

- Created Tableau dashboard and Excel Pivot Table to identify monthly, and quarterly ROI, CPC, and CPM metrics performance across 26 business brands. Pinned out key-value pair of the optimal advertising strategies to the best ROI with the timestamp.
- Programmed ANOVA, residual matrix to identify significant variables associated with high ROI performance.
- Developed marketing strategy permutations by 33% of ROI optimization with R clustering and regression techniques.

Machine Learning Project: Airbnb

- Applied EDA, visualization, Random Forest, and XGBoost model on 90 variables to predict NYC rental prices and displayed the result with a .Net Blazor website.
- Improved regression model performance by 25% with feature engineering. Filtered the dataset to 26 key contributing factors significantly affecting rental price. Developed ultimate models with an average of 51.4 RMSE.

National Taiwan University, Taipei, Taiwan
BA History (minor in Agricultural Economics) Sep 2015- Jun 2020

Victoria University of Wellington, Wellington, New Zealand
Certificate: International Exchange Student Jul 2018- Jun 2019