Y. Shirley Li

765-430-8920 | li1050@purdue.edu | LinkedIn | GitHub

Summary

Experienced GIS Specialist and Data Scientist with 9+ years in statistical modeling, operations research and data analytics. Skilled in Python, R, and SQL, with expertise in sustainability, transportation and geospatial technologies to drive data-informed decision-making

Experience

Purdue University Libraries and School of Information Studies — West Lafayette, IN
GIS Specialist/Instructor/Analyst — *2016 - 2025*

- Partnered with senior faculty and research leadership to design and deploy data pipelines and scalable geospatial infrastructure, enabling 20+ cross-disciplinary projects
- Developed and deployed interactive dashboards and custom web apps to support geovisualization across campus
- Led workshops and certificate programs, driving adoption of data science and geospatial AI across disciplines

Education

Purdue University — West Lafayette, IN
Ph.D. in Sustainability Engineering and Environmental Engineering — *2020 – 2026*
Graduate Certificate in Applied Statistics — *2018 – 2020*
M.S. in Civil Engineering — *2012 – 2014*

University of Waterloo — Waterloo, ON
Bachelor's in Geomatics, Minor in Computer Science — *2010 – 2012*

Wuhan University — Wuhan, China
Bachelor's in GIS (dual degree) — *2008 – 2010*

Projects

Demand Prediction for Bike Share System Expansion | *ML pipeline, XGBoost, Docker, CI/CD, AWS, ECR, S3*

- Developed ML pipeline to predict station demand using XGBoost/linear regression with 4+ years of trip data
- Engineered spatial features from census, POI, and infrastructure layers, and optimized performance
- Containerized the trained model with Docker and deployed a serverless FastAPI inference API on AWS (ECR, Lambda, API Gateway), integrating S3 for feature storage and CloudWatch for monitoring
- Built an interactive web map (Leaflet hosted on S3/Amplify) for users to receive real-time demand predictions

Solar-Powered Bike Share Station Modeling | *Python, Simulation, Optimization, Energy consumption, LCA*

- Developed a simulation framework in Python for PV-battery charging, discharging, and replacement cycles
- Analyzed variable solar potential and energy consumption, identifying overlooked battery replacement needs
- Recommended optimization strategies that improved energy independence by 30% and reduced downtime by 50%
- Enhanced life cycle analysis (LCA) to more accurately quantify environmental impacts and carbon emission rate

Alien Forest Pest Explorer | *SQL, JavaScript, Dashboard, Stakeholder collaboration*

- Developed the first nation-wide, county-scale interactive dashboards, enabling users to track spread and impacts
- Built data workflows to manage and visualize invasive species datasets for effective forest management and policy
- Created new application features and implemented custom functionality using JavaScript
- Engaged stakeholders to ensure the platform delivers actionable, accessible insights for diverse user needs

Technical Skills

Languages: Python, SQL, R, JavaScript, HTML/CSS, MATLAB
Machine Learning & Data Science: XGBoost, scikit-learn, PyTorch, TensorFlow, regression models, statistical modeling, hypothesis testing, A/B testing, causal inference, metric design, feature engineering, model evaluation
Data Engineering & MLOps: ETL, Docker, FastAPI, Git/GitHub, CI/CD (GitHub Actions), AWS (Lambda, API Gateway, ECR, S3, CloudWatch), Google Cloud Platform, Spark (PySpark), SQL Server, database management
Visualization & Web Apps: Tableau, Power BI, Matplotlib, interactive dashboard design, data storytelling, Leaflet
Geospatial & Remote Sensing: GeoPandas, ArcPy, ArcGIS Pro/Enterprise/Online, QGIS, Google Earth Engine