Olivia Liau

832-433-2552 | oliau@usc.edu | linkedin.com/in/olivialiau | github.com/oleeveeuh

Education

University of Southern California (Expected Graduation: May 2027) Los Angeles, CA

Master's in Applied Data Science, Bachelor's in Computer Science *(GPA: 3.56/4.0)* *Fall 2024 – Present*

- Courses: Machine Learning, Data Structures & Algorithms, Data Mining, Artificial Intelligence, Probability Theory

Experience

Data Science Research Intern May 2024 - Present

Worcester Polytechnic Institute | *https://github.com/oleeveeuh/WPI-UMASS-TOD-Prediction* *Worcester, MA*

- Co-authored and presented work at *BIOINFORMATICS 2026*, validating approach for drug timing prediction
- Introduced a novel AutoEncoder–CNN architecture to process 20,000+ genes from 146 entities, delivering predictions within 1-hour error with cross-region generalization
- Validated approach with 68% improvement over baseline by systematically benchmarking 700+ model configurations, including 4 dimension reduction methods and 16 regression models

Amazon - Business Analytics Extern Sept 2025 - Present

Extern | *Operational Strategy & People Analytics Externship* *Remote*

- Built an NLP pipeline using GPT-4 and RAG-based thematic analysis to analyze 350+ reviews, achieving 81% classification accuracy and 90%+ retrieval precision through clustering and A/B testing
- Delivered executive-facing insights identifying overtime as the top retention risk, directly influencing updates to warehouse culture and overtime policies

Data Analyst Sept 2025 - Present

Hope Services & BioKind Analytics at USC *Los Angeles, CA*

- Spearheaded Random Forest anomaly detection on 2+ years of pediatric health data covering 248 patients, identifying 42 high-risk treatment cases (7.5% of 896 assessments) with 71% prediction accuracy
- Reduced manual case review time by 60% by automating end-to-end risk stratification pipeline for data preprocessing, feature engineering, model training, and scoring

ML Research Intern May 2025 - Aug 2025

University of Wyoming | *https://github.com/oleeveeuh/MoE-MultiSDAR-PD-Classification* *Laramie, WY*

- Pioneered 90% accuracy on multimodal time-series data and 72% accuracy on medical imaging for Parkinson's Disease detection by implementing ResNet18 CNN with transfer learning on limited sample size
- Overcame constraints through SMOTE data augmentation and feature extraction combining PCA dimensionality reduction and HOG descriptors, improving model generalization by 18% and reducing overfitting by 23%

ML Research Intern, Informatics and Computing in Neuroscience Lab Aug 2024 - Jan 2025

University of Southern California | *https://github.com/oleeveeuh/CURVE-ICON-DBSO* *Los Angeles, CA*

- Proposed and developed predictive model achieving 71% accuracy in forecasting deep-brain stimulation treatment outcomes by engineering 108 multimodal features from EEG recordings and clinical assessments
- Designed SHAP-based explainability framework identifying 10 key neurological biomarkers ranked by predictive importance, enabling individualized predictions to build physician trust for regulatory approval

Projects

RetailPRED: Production ML System | *retail-pred.vercel.app* Oct 2025 - Present

- Engineered end-to-end Airflow pipeline on Google Cloud Platform automating weekly forecast validation and 12-month rolling prediction generation, processing 11 retail categories with 90.9% historical accuracy
- Optimized ensemble model (Random Forest + LightGBM) with 73 engineered time-series features, achieving 10x performance improvement by removing macroeconomic variables through ablation testing
- Deployed dual-interface analytics platform (React dashboard, Tableau BI) delivering explainable forecasts and actionable insights across 2,214 validated predictions

Technical Skills

Languages & Tools: Python, SQL, C/C++, R, Java, Git, React, TypeScript, Tableau, FastAPI, Docker, CI/CD, Airflow, GCP
ML/AI Frameworks: PyTorch, Pandas, NumPy, TensorFlow, scikit-learn, XGBoost, SHAP, HuggingFace Transformers
Specializations: Computer Vision, NLP, Time Series Forecasting, Feature Engineering, Statistical Modeling