Yupeng Li

Los Angeles, CA | 323-382-7306 | yupengli531@gmail.com | https://www.linkedin.com/in/yupengli531/

EDUCATION

University of Southern California Los Angeles, CA
Master of Science in Analytics August 2023-May 2025
Stony Brook University Stony Brook, NY
Bachelor of Science August 2019-May 2023
Double Major in Applied Mathematics & Statistics and Business Management with specialization in accounting

PROFESSIONAL EXPERIENCE

State Grid Qingdao Economic Research Institute Qingdao, China
Data Analyst Internship May 2023-July 2023

- Built an end-to-end ETL and forecasting pipeline using MySQL and Python (Pandas, NumPy) to integrate, clean, and validate multi-year electricity-demand data across 10 districts, improving processing efficiency by 35%
- Designed and implemented ML models (Random Forest, XGBoost) to predict regional electricity load growth, increasing forecast precision by 12% and enabling data-driven infrastructure planning
- Created interactive Tableau dashboards to visualize KPIs (energy demand, EV adoption), forecasting a 57.7% surge in demand for electric vehicles and charging facilities in the next five years

ACADEMIC PROJECTS

University of Southern California Los Angeles, CA
Machine Learning-Based Mortality Prediction for Sepsis-Associated Liver Injury(SALI) February 2025-May 2025

- Built an automated data-to-model pipeline using MySQL and Python to extract and preprocess a total of 1,157 ICU records with 250K+ time-series medical measurements from 3 large ICU datasets (MIMIC-IV, MIMIC-III, eICU) to predict 28-day mortality
- Developed 8 machine learning models, with XGBoost achieving the best AUROC of 85.56%, improving performance by approximately 8.3%
- Enhanced model transparency through SHAP and LIME analyses, identifying five key predictors for early stratification and clinical decision support
- Conducted statistical analyses in Python (Mann–Whitney U, chi-square tests) to assess cohort comparability and feature associations, ensuring dataset representativeness and modeling robustness

Sleep Improvement Satisfaction Prediction October 2024-December 2024

- Partnered with a 3-member team to scrape and merge 1,695 melatonin and zolpidem reviews from public healthcare websites using Python (BeautifulSoup, Pandas, Requests) to parse HTML content for sentiment analysis
- Preprocessed and balanced text data through spaCy NLP, SMOTE, and TF-IDF vectorization, ensuring consistent rating scales and class representation
- Fine-tuned BERT, RoBERTa, and DistilBERT models via Hugging Face transformers for sentiment classification, achieving 91.2% accuracy, providing data-driven insights into sleep aid satisfaction trends and side effect patterns

Machine Learning–Based Mortality and Readmission Prediction in Type II Diabetes September 2024-December 2024

- Led a 5-member team to extract and consolidate 14K+ ICU records from MIMIC-III using SQL on Google Cloud Platform, integrating 45 initial features for 3-, 30-, and 365-day mortality and 30-day readmission prediction
- Engineered time-windowed features (min, max, mean, median) and implemented robust preprocessing pipelines for both mortality and readmission prediction tasks
- Co-developed and tuned ML models (AdaBoost, XGBoost) for mortality and readmission prediction, achieving an AUROC of 89.07% for 3-day mortality (11.4% improvement), supporting early risk detection and data-driven readmission prevention

Credit Card Approval Dashboard Analysis | [GitHub Link] February 2024-March 2024

- Performed exploratory data analysis (EDA) on 438K+ applicants using bivariate and univariate analysis to detect skewness, correlations, and data-quality issues
- Developed two interactive Tableau dashboards with interactive filters to visualize trends and KPIs (approval rate, credit limit distribution), revealing an extremely low approval rate of 0.04%
- Presented data-driven recommendations that improved model precision and reduced bias across applicant segments

RELEVANT SKILLS

- **Technical Tools**: Python, SQL, MySQL, Tableau, Power BI, R, Looker
- **Technologies**: Azure Databricks, Google Cloud Platform (Vertex AI, BigQuery), Jupyter Notebook, Google Colab, MS Excel (PivotTables, Macros, VLOOKUP), Windows, macOS, Linux, GitHub, Google Applications, Photoshop
- **Frameworks:** scikit-learn, AUROC, XGBoost, AdaBoost, spaCy, SHAP, LIME, Pandas, NumPy, SMOTE, TF-IDF, TensorFlow, Keras, Matplotlib, Seaborn, BeautifulSoup, Requests, PCA, Hugging Face Transformers