Bingyan(Tina) Liu

(626)-503-6639 | tina116948@gmail.com | bytina923.github.io | U.S. Permanent Resident

Education

University of California, San Diego Sept. 2022 – (expected) Mar. 2026

Bachelor of Science in Data Science, Minor in Economics *GPA: 3.8/4*

- Coursework: Data Structures & Algorithms, Probabilistic Modeling, Machine Learning, Representation Learning, Deep Learning, Principles and Practice of Data Science, Data Visualization, Data Management, Linear Algebra, Statistics, Econometrics, Optimization

Skills

Programming: Python (Pandas, NumPy, Scikit-learn, XGBoost, LightGBM), SQL (PostgreSQL, MySQL), Java, JavaScript, HTML/CSS, JSON
Machine Learning: Regression, Decision Trees, Random Forest, Gradient Boosting, Naïve Bayes, K-Means, PCA
Data Analytics: Exploratory Data Analysis, Feature Engineering, A/B Testing, Hypothesis Testing, Time Series
Data Visualization: Matplotlib, Seaborn, Plotly, Tableau

Projects

Rating Prediction for Amazon Products | *Machine Learning, Natural Language Processing*

- Processed dataset of 490k+ records, performed data cleaning (handling duplicate records, normalizing text, removing outliers) to ensure high-quality input for modeling
- Engineered NLP features using **TF-IDF, Word2Vec, and Doc2Vec**, improving model expressiveness for sentiment and rating prediction
- Built and optimized a **logistic regression model** with hyperparameter tuning (L1/L2 regularization), identifying the best-performing feature set with TF-IDF
- Evaluated model performance with RMSE and R^2, conducted error analysis, and visualized feature importance to extract insights for e-commerce product recommendations

LoL Esports Match Outcome Prediction | *Explotory Data Analysis, Machine Learning, Feature Engineering*

- Conducted **Exploratory Data Analysis (EDA)** to identify key match-winning factors and utilized statistical analysis & visualization for feature selection
- Applied **PCA and Bayesian target encoding** to address the curse of dimensionality and develop models from high cardinality categorical features
- Collaborated with team members to build predictive models with 73% accuracy in forecasting match results, improving decision-making for competitive gaming strategies

LLM-Based Web Agent | *Large Language Model, Prompting Engineering, Computer Vision*

- **Automated browser interaction** and data extraction using **Selenium**, enabling dynamic and scalable web scraping for real-time content analysis
- Designed adaptive prompting strategies for **large language models (LLMs)**, optimizing Planning, Acting, Memory, and Reflection capabilities, resulting in enhanced task performance and inference efficiency
- Conducted comprehensive model evaluation through data collection, test environment simulations, and manual assessments, delivering detailed reports on prompting framework efficacy and task performance improvements
- Delivered detailed reports on prompting framework effectiveness and task performance improvements

Forecast Future YouTube Video Views | *Machine Learning, Predicitve Analytics, Feature Engineering*

- Built and compared multiple machine learning models, including **Linear Regression and Gradient Boosting Decision Tree (GBDT)**, optimizing for **Normalized Mean Squared Error (NMSE)**
- Designed a scalable prediction pipeline with automated data preprocessing, feature extraction, and model evaluation, ensuring accurate and efficient forecasting of YouTube video popularity trends
- Provided actionable insights to optimize content strategy for YouTube creators and marketers, improving engagement-driven recommendations.