YUER ZHOU

Stanford, CA 94305

📞 (1)650-683-2612 ✉ nina2zhou@outlook.com 🔗 linkedin.com/in/yuer-zhou 🔗 https://github.com/yuer2021

Education

Stanford University   **Jan 2024**
Master of Science, Environmental Engineering (Concentration: Environmental Data, Statistics and Modeling)   *Stanford, CA*
- **Coursework:** Machine Learning, Natural Language Processing with Deep Learning, Data Analytics, Probabilistic Models, Data Management and Data Systems, Data Structures, Algorithm Design and Analysis, Object-Oriented Programming, Imaging With Incomplete Information

Sichuan University   **Jun 2021**
Bachelor of Science, Environmental Science   *Sichuan, China*

Experience

Meituan   **June 2023 – Present**
Machine Learning Engineer Intern   *Beijing, China*
Recommender System:
- Engineered and deployed a **Java-based** landing page recommendation engine for diverse business units, **serving 10M+ DAUs**. Conducted iterative refinements through **online A/B testing**, resulting in a 5.2% uplift in UV-CTCVR.
- Implemented end-to-end pipeline setup, including multi-path retrieval, ranking, and post-ranking. Developed and launched retrieval models based on Popularity, Collaborative Filtering(ItemCF), Two-Tower DNN with Self-Supervised Learning(SSL) techniques.
- Optimized the ranking model by developing 50+ additional online features, leveraging Customized Gate Control(CGC) model combined with Co-Action Network(CAN) for feature interaction, achieving a 2.8pp offline AUC improvement.

Real-Time Bidding:
- Implemented 48-hour and 7-day post-click CVR models, transitioning from training with launched to all impressions. Integrated delayed feedback and 150+ key features, achieving +5.8% in customer acquisition and -5.3% in CAC in online A/B tests.
- Utilized Spark-based **ETL** tasks in Scala for data processing and storage on HDFS. Conducted feature engineering and implemented models like XGBoost, Wide&Deep, MMoE, and ESMM with **TensorFlow**, with the highest offline AUC increase of 1.8pp.

Augmented Reality Navigation Application Development   **Jan 2023 – Jun 2023**
Stanford University Partnership with Mercedes-Benz   *Stanford, CA*
- Led the design, prototyping, and iterative testing of an AR Navigation app, informed by user-focused research, feedback and co-design sessions, unveiling novel AR automotive use-cases in collaboration with Mercedes-Benz's AI and Digital Vehicle team.
- Utilized Figma for prototyping, Unity for 3D simulation and real-world AR implementation, and Xcode for iPhone compatibility
- Presented to Mercedes-Benz executives in Germany, garnering positive feedback from corporate professionals and target users.

Projects

RDF-Text-Story: An Integrated Workflow for Controllable Short Story Generation | *Pytorch, HuggingFace* **Jan 2023**
- Implemented an innovative two-stage workflow to transform Resource Description Framework (RDF) triples into short stories, harmonizing structured information and open-ended storytelling, significantly enhancing narrative coherence and creativity.
- Fine-tuned **Transformer models(T5 and GPT-2)** on preprocessed datasets, experimenting with different hyper-parameters, prompt engineering strategies, adapted modules, and decoding algorithms, resulting in a 200% improvement in the BLEU score.
- Integrated the optimized ensemble model with **image-generation diffusion models** to produce an innovative platform for interactive and immersive storytelling experiences.

Long-term and Short-term Prediction and Interpolation of PM 2.5 Measurements | *Python, Pytorch* **Sept 2022**
- Forecasted short and long-term PM 2.5 levels in Foshan and Tianjin, China, using data from static and mobile sensors attached to moving vehicles, addressing pervasive pollution issues and aiding local air quality management.
- Pre-processed data using techniques like FFT, Lomb-Scargle Periodogram, seasonal decomposition and Gaussian kernel smoothing
- Developed a **Gaussian Process Regression** model leveraging Matern 52 and Exponential kernel functions for short-term and long-term PM2.5 predictions, demonstrating robust performance by achieving an average RMSE of 3.07 and 1.34 respectively.
- Built an **LSTM** model that effectively identified significant trends, with a notably lower RMSE of 1.84 and 0.76 respectively.

Predicting Fluorescent Lifetime Imaging(FLIM) via Traditional ML and Deep Learning | *Python, Keras* **May 2022**
- Collaborated with Stanford's Kasevich Group, addressing the challenge of unmixing multiple lifetimes in multi-spectral FLIM, enabling real-time and parameter-free processing of complex fluorescence lifetimes.
- Synthesized a 3D voxel dataset by convolving exponential decays with a determined impulse response function, activated by MNIST images, to simulate FLIM data and thus circumventing the need for large experimental training datasets.
- Built a **3D CNN** model with independent fully convolutional prediction heads to reconstruct 2D spatial maps of fluorescence lifetime parameters, with test MSE being 0.018. Generated image visualizations for visual validation.

Modeling of Population Change in California | *Python, SQL, BigQuery*   **Nov 2021**
- Executed **complex SQL queries** on multiple BigQuery datasets, conducted detailed feature engineering and created visualizations for trend analysis, yielding deep insights of population change in California counties. Identified seven most important features.
- Built three multi-class classification models, namely **Logistic Regression, Gradient Boosting Classifier and DNN Linear Combined Classifier**, to predict future population change intervals of all CA counties, with the highest **F1 score** of 0.44.

Technical Skills

Programming Languages: Python, Java, Scala, SQL, C++, R, Matlab
Machine Learning: TensorFlow, Pytorch, Keras, Scikit-learn, NumPy, SciPy, Pandas, Matplotlib, NLTK
Big Data technologies: Hadoop, Spark, Hive
Databases: MySQL, MongoDB, Oracle, PostgreSQL, Redis
Tools: Git/GitHub, Vim, Docker, Linux, AWS, JUnit, Jupyter Notebooks, IntelliJ, VS Code, PowerBI, Tableau