YICHEN ZHAO

https://shenshenzyc.github.io/yichen-zhao-portfolio/ | yzhao131@dons.usfca.edu | linkedin.com/in/yichenzhao1

EDUCATION

University of Texas at Dallas	Ph.D. in Computer Science	*Exp. May 2028*
University of San Francisco	M. S. in Data Science	*Jun 2023*
University of Pennsylvania	M. A. in Applied Mathematics and Computational Science	*May 2022*
Georgia Institute of Technology	B. S. in Mathematics, Applied Mathematics	*May 2020*

WORK EXPERIENCE

Data Science Intern, Dagshub Inc. – Developer Relations *Nov 2022 – Jun 2023*

Next Word Prediction

- Build a neural network model implementing bidirectional LSTM for next word prediction. Use MLflow APIs to log experiments with parameters, metrics, and model artifact. (Dagshub repo)
- Write two blog posts, one introducing the theory behind RNN and LSTM, and the other explaining the workflow of building this next-word-prediction model with MLOps integration on Dagshub. (Blog post)

Snowflake x Dagshub

- Integrate Dagshub tracking with a Snowflake database, using Git to log SQL queries and DVC to log result tables. Optimize SQL queries to retrieve Lego set data for fine-tuning a Lasso regression model predicting Lego set prices. (Dagshub repo)
- Write a blog post to explain to Dagshub users how they could take advantage of Dagshub tracking features when working with cloud-based database on Dagshub. (Blog post)

Annotation Transfer for YOLO-formatted Annotations

- Create a Colab notebook facilitating the transfer of YOLO-formatted annotations to enable rendering of annotations in Dagshub repositories for explicit display of images with annotations. (Colab notebook)

Teaching Assistant, Georgia Institute of Technology – School of Mathematics *Jan 2019 – May 2020*

- Grade and host office hours for courses: Applied Combinatorics, Complex Analysis, Probability Theory.

PROJECTS

Hand Gesture to Sign Language Alphabet Translator

- Develop an interactive program using Python OpenCV that captures live video from the webcam and translates users' hand gestures into sign language alphabet in real-time.
- Implement two approaches for translation: (1) Construct a Convolutional Neural Network (CNN) to classify cropped hand images, achieving an accuracy of 97.27% . (2) Leverage Google's Mediapipe package to identify hand landmarks and fine-tune a random forest model to classify their coordinates, achieving an accuracy of 92.52%. (Github repo)

Game Result Prediction

- Train four supervised learning models to predict the result of a League of Legends game at pre-, early-, mid-, and late-game using Python XGBoost and Scikit-learn, achieving prediction accuracy of 52.57%, 72.71%, 96.65%, and 97.10%, respectively.
- Deploy the model on an interactive program in Python to ingest ongoing game data and, based on current stage of the game, produce live predictions with probability. (Github repo)

SKILLS

Python-related: Statsmodels, NumPy, Pandas, Plotly, Scikit-learn, OpenCV, PyTorch, Gensim, SpaCy, Flask, BeautifulSoup, Airflow, PySpark, MLflow, Docker, Kubernetes, Metaflow, Streamlit, Evidently
Non-Python programming skills: MySQL, PostgreSQL, MongoDB, JavaScript, HTML, CSS, Java, Git, DVC