YICHEN ZHAO

Union City, CA, 94587 | C: (972)-375-7525 | yzhao131@dons.usfca.edu | linkedin.com/in/yichenzhao1

EDUCATION

University of San Francisco — M. S. in Data Science — *Exp. Jun 2023*
Relevant Coursework: *Linear Regression Analysis, Data Acquisition, Relational Database (PostgreSQL), Exploratory Data Analysis and Visualization*

University of Pennsylvania — M. A. in Applied Mathematics and Computation Science — *May 2022*
Relevant Coursework: *Machine Learning, Introduction to Optimization Theory, Applied Probability Theory*

Georgia Institute of Technology — B. S. in Mathematics, Applied Mathematics — *May 2020*
Graduated summa cum laude

WORK EXPERIENCE

Data Science Intern, Dagshub Inc. – Developer Relations — *Nov 2022 – Present*

- Construct a deep neural network to predict next word given a sequence of text using RNN and LSTM, track the experiments using MLflow, and deploy the model on Dagshub.
- Create a Colab notebook introducing how RNN and LSTM perform on next word prediction and how MLOps features provided by Dagshub interact with the ML pipeline for the purpose of attracting potential users.

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

- Set up office hours for courses: Real Analysis, Complex Analysis, Probability Theory.

PROJECTS

Game Result Prediction

- Trained four supervised learning models to predict the result of a League of Legends game as the game proceeds to different stages (pre-, early-, mid- & late-game), using Python XGBoost and Scikit-learn package. The models achieved 52.57%, 72.71%, 96.65%, and 97.10% accuracy, respectively.
- Deployed the model on an interactive program in Python to ingest ongoing game data and produces live predictions with probability. (Github)

Recognition of Facial Expression

- Trained two classification models to recognize over 30000 images of human faces and classify the facial expressions into seven major classes: built a convolutional neural network and achieved 55% accuracy using PyTorch; constructed a modeling pipeline that combines PCA and SVM classifier, and achieved 46% accuracy.
- Discussed reasons for underfitting when classifying sad, disgust and fear faces: human faces with such negative emotions have similar configurations. (Report)

Text Analysis of Documents in a Corpus

- Trained a TF-IDF vectorizer using Python Scikit-learn package to determine summarizing keywords with highest TF-IDF scores for each text document of a corpus.
- Implemented text pre-processing by parsing the corpus in XML format using Python ElementTree library and tokenizing text using Python NLTK package.

Vehicle Routing Problem

- Examined, in a theoretical context, the performance of three commonly used algorithms on this problem: classic heuristic algorithm, insertion heuristic algorithm, and savings algorithm.
- Implemented and analyzed three algorithms respectively on the pickup and delivery problem and the time window problem using Google OR-Tool in Python. Identified the optimal algorithms for each of the two problems using parallel grid search. (Github)

A Simple Call of Duty Game

- Built a simple classical game called Call of Duty using Java. On a randomly generated map, the player is equipped with two types of weapons and aims to destroy all buildings before ammunition is depleted.
- Designed an in-game user console for the player to easily read and track game progress. (Github)

SKILLS

Programming: Python (Pandas, Matplotlib, Scikit-learn, PyTorch, Word2Vec, SpaCy, BeautifulSoup, Flask, PySpark), Java, PostgreSQL, MySQL, JavaScript, HTML, CSS, Git, DVC, MLflow

Other: Machine Learning, Deep Learning, NLP, Image Processing, MLOps, Data Acquisition, Statistics