Shawn H. Xu

sx131@georgetown.edu • 925-858-6938 • https://www.linkedin.com/in/shawn-xu22

San Ramon, CA, 94582

Skills

Programming Languages: Python, JavaScript, SQL, R programming, MATLAB, HTML
Tools: Numpy, Pandas, Scikit-Learn, Quarto, Shiny, Visual Studio, Jupyter Notebook, Microsoft Office Suite, Git/Github

Work Experience

Data Analyst – *Data Annotation Tech* *May 2024 – Present*

- Engineered code-related prompts to Large Language Models (LLMs).
- Ensured high standard of conversational, coding, and mathematical AI performance by conducting detailed quality assurance on chatbot responses based on accuracy and reasoning.
- **Techniques:** Python, Prompt engineering, Reinforcement Learning, LLMs, Error analysis

Student Researcher/Intern – *Dana Farber Cancer Inst./Harvard Medical School* *May - July 2022*

- Studied structural functions of degraders through functional proteomics and genomics to profile synthetic degraders.
- Aided in the discovery of molecular glues that paved a new way in pharmacology to kill cancer cells.
- Co-authored a publication on Nature in relation to my work: https://doi.org/10.1101/2023.02.14.528208

Projects

Skin Cancer Detection with Image and Tabular Data - https://github.com/shawnhxu/SCD_with_3D-TBP

- Applied 3D total body image and tabular data onto modern Convolution Neural Network (CNN's) and Decision Trees to distinguish malignant skin lesions from benign ones.
- Employed under and over-sampling to counter the imbalancing of classes in the image dataset.
- Incorporated an Ensembling strategy to consider all machine learning predictions.
- Recorded above 90% accuracies in models.
- **Techniques:** PyTorch, CatBoost, LightGBM, Residual Network (ResNet), MobileNetV2, Mobile Neural Architecture Search (MNAS) Network, EfficientNetB4, DenseNet, Ensembling, Under/Over Sampling, Feature Engineering, Cross-Validation, Partial Area Under ROC Curve (pAUC)

Predicting Lumbar Spine Degenerative Conditions - https://github.com/mfgeary/dsan6500-project

- Designed data processing pipeline for MRI dataset to be PyTorch-compatible for CNN training.
- Evaluated multiple modern Convolutional Neural Network architectures performance in detecting and predicting Lumbar Spine Degenerative Conditions with 87-95% accuracies.
- **Techniques:** PyTorch, MRI DICOM Image Dataset, CNNs, ResNets, Visual Geometry Group (VGG) Networks

Art Image Similarity Finder App - https://github.com/5cminsuhlim/DSAN6600Proj

- Utilized multiple computer vision techniques in tandem with Deep Learning models to discover the most similar images based on user-chosen image preprocessing steps and user-given input images.
- Engineered a Streamlit App that takes in a user input image and outputs most similar art pieces.
- **Techniques:** Python, Pytorch, Residual Networks (ResNets), CNNs, Deeplake, Streamlit, You Only Look Once (YOLO), Subject Segmentation, Image Embedding, Edge Detection, OpenCV

StarCraft2 Player Performance Analysis - https://github.com/5cminsuhlim/DSAN5300Proj

- Conducted an in-depth analysis on in-game telemetry data and skill rating of StarCraft 2 players.
- Compared machine learning models to predict a player's skill rating (rank) and Actions per Minute (APM).
- Illustrated the complexity of eSports through statistical analysis and predictive modeling.
- **Techniques:** Python, R Programming, ANOVA Testing, Pair-wise T-testing, Logistic Regression, Support Vector Machines (SVMs), Random Forest, XGBoost, SMOTE Balancing, Feature Clustering

Education

Georgetown University, *Washington, DC* *August 2023 – Present*

- M.S. degree in Data Science and Analytics (DSAN). **GPA:** 3.95/4.0

Boston University, *Boston, MA* *Sept 2019 - May 2023*

- B.S. degree in Biomedical Engineering. **GPA:** 3.5/4.0 – Dean's List of Academic Honor.