Nimish Magre
617-817-0148 | nimishmagre@gmail.com | Cambridge, MA
LinkedIn | GitHub

EDUCATION AND HONORS

Northeastern University Boston, MA
Master of Science in Electrical and Computer Engineering August 2022
Concentration: Computer Vision, Machine Learning and Algorithms GPA: 3.9/4.00

- Coursework: Deep Learning, Applied Probability and Statistics, Database Management
- 2-year AWS Open Data Sponsorship for Typography Analysis dataset *(TMNIST, TMNIST_Glyphs)*

Australian National University Canberra, Australia
Bachelor Of Engineering (Honors) July 2019
Concentration: Mechatronic Systems GPA: 5.75/7.00

SKILLS

- **Programming Languages:** Python, C++
- **Data Analysis:** SQL (MySQL), Pandas, NumPy, SciPy
- **Data Visualization:** Matplotlib, Seaborn
- **Python ML Libraries:** TensorFlow, Pytorch, Scikit-Learn, SHAP, H2O
- **Tools:** MATLAB, Git, Jupyter, Anaconda, Visual-Studio, Docker, AWS-S3, AWS-EC2, LangChain
- **Machine Learning:** Regression, Classification, Clustering, Neural Networks

WORK EXPERIENCE

Northeastern University (Contract Role)

Course Development Assistant (Program Structures and Algorithms) **Mar 2024-Present**

- Assist the course professor to co-author a comprehensive book on Algorithms and Programming Structures.
- Provide technical assistance to the university's online course development team and Coursera to seamlessly integrate course content onto the Coursera Platform.
- Lead the development of a RAG application tailored for the Algorithms course, facilitating student interaction and engagement with course specific content.

Squark Inc. (no-code AI SaaS and API for prediction/forecast models)
Data Scientist (Backend Python Development) **Dec 2022-Mar 2024**

- Utilize Python and supporting tools/libraries to build feature extraction and data curation pipeline, improve model explainability and modify codebase to efficiently manipulate large data.
- Adapted the Word2Vec algorithm to obtain word vectors for each text column and transformed the vectors with the average aggregation method to obtain vector representation for each row.
- Communicated model interpretability through SHAP (SHapley Additive exPlanations) summary graphs, enhancing the impact of each feature on the trained models.
- Developed a weighted-averaging technique to articulate Variable Importance in H2O-based Stacked-Ensemble models, contributing to improved model understanding and decision-making.
- Optimized data handling by transitioning from Pandas to the faster Polars library, adapting the codebase to efficiently download, store, and clean training data, enhancing overall data processing speed.

Harvard Medical School (Massachusetts General Hospital)
Machine Learning Researcher (Co-Op) **Jan 2022-Jun 2022**

- Implemented a self-supervised denoising network to denoise MRI scans (notably contributed to reducing the time required for patients to complete an MRI scan).
- Demonstrated superior performance by achieving average Peak Signal-to-Noise Ratio (PSNR) results consistently 1.5dB higher than relevant single-image denoising methods when dealing with Additive White Gaussian Noise.
- Evaluated the model's robustness using a real-world noisy dataset (PolyU), attaining an average PSNR value of 37.52 dB. This achievement aligns closely with the performance of the best-performing model in the field, which recorded a PSNR value of 37.55 dB.

PROJECT EXPERIENCE

AI Skunkworks (Northeastern University)
Graduate Machine Learning Researcher (Arxiv**)** **Dec 2021-Dec 2022**

- Designed a python-based pipeline to generate a synthetic dataset of 565,292 MNIST-style grayscale images representing 1,812 unique glyphs in styles of 1,355 Google-fonts for typography analysis.
- Authored an introductory paper for the dataset on the *Arxiv* platform and trained a custom CNN-based classification model to identify non-Latin glyphs with ~97% accuracy.
- Inspected glyph identities and eye-tracking tools to effect real time mapping of type (fonts) to cognitive properties; generated "Panose" number identities used to characterize fonts for the entire dataset.

Northeastern University
Siamese network for object-tracking (GitHub**)** **Jan 2021-Apr 2021**

- Modified the template patch on the ResNet-50 based Siamese Box Adaptive Network (SiamBAN) for single object-tracking using correlation between initial, t-1 and search frames to detect occlusion and improve inference-time when multiple object instances were present.
- Displayed differentiation of target object by model when partial occlusion and multiple instances of the object were present with qualitative testing on sample VOT-2018 video dataset.

KEY PERSONAL PROJECTS WITH GITHUB HYPERLINKS

- ML-based Data Insights for NGO: Performed synthetic data generation, sentiment analysis on volunteer feedback and used data from a volunteer's visit log to generate scikit-learn based classification/regression models.
- LSTM-based single instrument music generation: Represented guitar and violin samples as midi files to train the LSTM-based RNN model to predict after every 100 notes. For music generation, supplied a sequence of random 100 notes with a random offset value to generate 500 output notes.
- ML Research Paper Reviews: Comprehensive review of recent papers that focus on Advances in Computer Vision and techniques to induce sparsity with Big Data.

HONORS AND AWARDS

- ANU College of Engineering and Computer Science International Partnership Scholarship
- Silver medal on Kaggle for typography dataset *(TMNIST, TMNIST_Glyphs)* and 2-year data hosting sponsorship by AWS