SCOTT PENCO

Machine Learning and Data Science Devotee Interested in Making an Impact With Designing Data Driven Solutions Using Mathematical Models and MLOPS. Open to Collaborating With Others and Learning From Professionals.

Montreal, CANADA | scott.penco@alumni.utoronto.ca | LinkedIn | 647-962-9436

EDUCATION

MIT Institute for Data, Systems, and Society (IDSS) - Data Science and Machine Learning

Comprehensive Training in Applied AI, Data Science, and Mentorship. Key Coursework: Applied AI/Data Science, Statistical Modelling and Computation in Applications, Foundations of Statistics, Probability, and Machine Learning. Focus on Integrating Statistical Principles and Computational Methods To Solve Complex Real-World Problems.

Harvard Extension School- Graduate Certification in Bioinformatics

Specialized Education at the Intersection of Computer Science, Bioinformatics, and Statistical Modelling. Key Topics: Advanced Data Structures and Algorithms, Computational Biology, and Statistical Analysis. Gained Hands-on Experience With Biological Datasets, Algorithm Development, and Predictive Modelling.

University of Toronto - HBSc Double Major in Biology and Neuropsychology (Joint Hons.)

Explored Neural Network Models and Their Applications to AI Through a Cognitive Neuroscience perspective.Developed a Strong Foundation in Analytical Thinking, Research Methodology, and Statistical Reasoning. Key Courses: Fundamentals of Statistics, Calculus, Research Design & Analysis.

PROJECTS

Malaria Capstone Project - *MIT IDSS* GITHub Repo
- Designed and implemented a Convolutional Neural Network (CNN) for malaria classification on a dataset of 30,000 images.
- Preprocessed and cleaned image data, building the model with Keras and compared its benchmark against a pre-trained CNN (VGG16).
- Achieved an F1 score of 0.98, outperforming VGG16 by 0.03 in precision and recall.
- Utilized SHAP values to interpret the model and highlight key features influencing predictions.

ML Prediction GITHub Repo
- Developed a predictive machine learning model for Boston housing prices using a linear regression approach.
- Conducted univariate and bivariate data exploration to identify key features influencing housing prices.
- Optimized the model by dropping statistically insignificant features, resulting in a slight increase in R^2 by 0.001.
- Determined model coefficients using Ordinary Least Squares (OLS) to highlight the most impactful predictors.

Data Analysis and Visualization GITHub Repo
- Applied dimensionality reduction techniques (PCA, t-SNE) and clustering on the auto-mpg dataset to accelerate customer queries.
- Visualized reduced dimensions to extract insights and identified three distinct data clusters.

RAG and Fine-tuning of LLMs GITHub Repo
- Scraped and vectorized academic papers for Retrieval-Augmented Generation (RAG) on a Large Language Model (OpenAI).
- Fine-tuned the model using generated prompts and answers in the domain of "Machine Learning for Drug Development."
- Evaluated both models based on proficiency in answering field-specific questions, improving performance through targeted layer adjustments and LoRA.

SKILLS

Data Science & Programming: Python, Numpy, Pandas, sklearn, MatPlotLib, Seaborn, Keras, PyTorch, SQL

Other Certifications: Foundations in Data Science (University of Waterloo), Data-Camp Data Science Certification (concurrent), Public Speaking and Vocal Training (Speech Science).

Languages: Native English proficiency, B2 proficiency in French