

Youssef Samaan

438-404-7439 | ysamaan2001@gmail.com | [linkedin.com/in/youssefsamaan](https://www.linkedin.com/in/youssefsamaan) | github.com/YoussefSamaan

EDUCATION

McGill University

Montreal, CA

B.Sc. in Computer Science, Minor in Statistics

Sep. 2021 – Apr. 2025

- Courses: Software Design, Algorithm Design, Databases, Operating Systems, Data Structures, Applied Machine Learning, Reinforcement Learning, Natural Language Processing, Distributed Systems, Object-Oriented Programming, Probability, Statistics, Linear Algebra, Stochastic Processes, Discrete Mathematics.

EXPERIENCE

Ericsson

Montreal, CA

Machine Learning Intern

January - August 2024

- Architected and implemented a scalable end-to-end data pipeline that extracts data for different customers from a datalake, parses, transforms, compresses, and stores them on Amazon OpenSearch and S3 buckets.
- Designed data pipelines to handle 100+GB of data per day for one customer for 10+ customers using PySpark.
- Optimized data access for 100+ radio experts, radio engineers, and ML specialists, reducing retrieval time by 99%.
- Utilized Agile methodologies to iteratively improve data processing workflows.

DermBiont

Boston, US

Software Developer and Data Scientist

May - August 2022

- Engineered a program that processes corrupt input files with multiple dictionary files, parses data, and produces a comprehensive report while efficiently classifying information into corresponding sheets.
- Developed scripts to identify and rectify over 25% discrepancies between Excel files data and database values, ensuring the integrity and accuracy of the data.
- Migrated 1TB of data from Dropbox to SharePoint and eLab to Benchling via APIs, ensuring 100% data integrity.

PROJECTS

Generalization and Preprocessing for Sarcasm Detection | *nltk, scikit-learn, PyTorch, TensorFlow, Transformers*

- Tested model generalization across a diverse dataset using nltk, scikit-learn, PyTorch, TensorFlow, and Transformers, finding that LLM-based classifiers with CoT (ChatGPT-o1) significantly outperformed alternatives while classical models maintained similar accuracy (17% difference) with faster training.
- Evaluated various preprocessing techniques, showing that minimal preprocessing (lowercasing and punctuation removal) maintained competitive performance while reducing training times by 30-99%.

Effect of Noisy Rewards on RL-Agent Performance | *Python, NumPy, Matplotlib, PyTorch, Gymnasium*

- Tested four reinforcement learning strategies (Q-learning, Expected SARSA, DQN, DDQN) on Gymnasium environments (Cart Pole and Acrobot) using different reward noise distributions, demonstrating that controlled noise accelerates convergence and helps agents find optimal policies faster.

Machine Learning Projects | *Python, Pandas, NumPy, TensorFlow, HuggingFace, SimpleTransformers*

- Implemented Linear Regression, Logistic Regression, and KNN for analyzing UCI datasets.
- Designed a Multi-Layer Perceptron with regularization and various activation functions and combined it with a CNN (TensorFlow) to classify CIFAR-10 images.
- Created a Naive Bayes classifier from scratch and assisted in fine-tuning a BERT model (HuggingFace) for sentiment analysis on the IMDB Movie Review Dataset, including analysis of its attention matrix.

RESEARCH EXPERIENCE

McGill University

Montreal, CA

Research Assistant in Reinforcement Learning & NeuroAI

Sep. 2024 – Present

- Investigating sparse reward RL algorithms (PPO, SAC) in AnimalAI environments to mimic biological learning.

TECHNICAL SKILLS

Languages: Python, Java, C/C++.

Tools & Frameworks: Pandas, NumPy, TensorFlow, PyTorch, scikit-learn, HuggingFace, Transformers, Matplotlib, Seaborn, Plotly, Gymnasium, stable-baselines3, PySpark.

Awards: 30th place ICPC NENA, Dean's Honor's List 4x, Outstanding Achievement, Poker bot Competition 2x 1st.

Hobbies: Solving hard problems, Trying new food, Swimming, Pingpong, Drawing, Movies/Anime/Shows, Traveling.