Youssef Samaan

438-404-7439 | ysamaan2001@gmail.com | linkedin.com/in/youssefsamaan | github.com/YoussefSamaan | Website

Education

McGill University Montreal, CA

B.Sc. in Computer Science, Minor in Statistics — GPA 3.8/4.0 *Sep. 2021 – Apr. 2025*

- Relevant courses: Functional programming, Software Design, Algorithm Design, Databases, Data Science, OS, Applied Machine Learning, Reinforcement Learning, Natural Language Processing, Distributed Systems, Algorithms and data structures, Software engineering project, Theory of Computation, Object oriented programming, Probability, Calculus 3, Linear Algebra, Discrete Structures, Statistics, and Stochastic processes.

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 scrum methodologies through daily stand-ups and iterative development, leveraging regular customer feedback to drive continuous improvement.

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.
- Automated file and folder renaming in 500+ HTML files, improving hyperlink functionality by 35%.
- Developed scripts to identify and rectify over 25% discrepancies between Excel files data and database values, ensuring data integrity.
- Migrated 1TB of data from Dropbox to SharePoint and eLab to Benchling via APIs, ensuring 100% data integrity.

Projects

News Coverage of Taylor Swift | *Python*

- Used APIs and open coding to analyze media coverage trends of Taylor Swift, focusing on sentiment and topics.

Splendor | *Python, Java, SQL, Docker, Maven, Spring Boot*

- Collaborated with 5 team members to design, document, implement, and test a board game called Splendor.
- The game was Implemented using Python for the GUI and Java, Maven, and Spring Boot for the backend.
- Integrated LobbyService for player registration and login, utilizing SQL for storage and Docker for for startup.

Watched Movies | *React, Javascript, HTML, CSS*

- Deployed a website on Netlify using React, HTML, CSS, and JavaScript, that displayed the movies, Netflix shows, and anime that I enjoyed watching.

Risk-Free Trading Website | *Python, Django, HTML, CSS, Javascript, Bootstrap*

- Developed a free stock trading simulator web application using Django and deployed it on Heroku.
- Integrated Finnhub's real-time stock price API

Event Management System | *Java*

- Designed a Festival Management System that enabling seamless creation of Festivals, Concerts, Workshops, Galas, Screenings, and Coming Soon Events. Added extra functionality such as calculating profit and filtering.
- Implemented multiple software design patterns and principles, including encapsulation, flyweight, visitor, and polymorphism, while rigorously testing the code with 100% coverage

Risk-Free Trading Website | *Python, Django, HTML, CSS, JavaScript, Bootstrap*

- Developed a Django-based stock trading simulator web application for risk free practice.
- Integrated Finnhub's real-time stock price API to provide up-to-date market data, enhancing realism.

Solar System Simulation | *MATLAB*

- Programmed a realistic solar system model in MATLAB by using the velocities and gravitational force to calculate their new position. Used a non-elastic simulation for when two entities collide with each other.

Machine Learning Research Projects

Generalization and Preprocessing for Sarcasm Detection

- LLM Benchmarking: Benchmarked NLP model generalization (LLMs, Classical ML) on multiple sarcastic datasets; LLMs (ChatGPT-o1) showed 17% accuracy gain while classical models trained faster.
- NLP Preprocessing Optimization: Demonstrated minimal text preprocessing (lowercase, punctuation removal) maintains performance, reduces training time (30-99%), highlighting word sequence importance in sarcasm detection.

Effect of Noisy Rewards on RL-Agent Performance

- Novel RL Reward Mechanism Study: Evaluated the impact of controlled reward noise (normal/uniform distributions) on the convergence and performance of Q-learning, Expected SARSA, DQN, and DDQN algorithms across Gymnasium environments (Cart Pole, Acrobot).
- Convergence Acceleration Discovery: Discovered that strategic introduction of reward noise accelerates RL agent convergence and facilitates faster discovery of optimal policies across varying noise levels.

Machine Learning Algorithm Implementations

- Algorithm from-scratch Development: Developed and implemented core ML algorithms from scratch, including Linear Regression, Logistic Regression, KNN, and Multi-Layer Perceptron with regularization and different activation functions in Python, demonstrating a deep understanding of fundamental ML principles.
- Advanced Model Building & Analysis: Built a CNN in TensorFlow for CIFAR-10 image classification and fine-tuned BERT (Hugging Face) for sentiment analysis, including attention matrix analysis, showcasing expertise in modern DL architectures and NLP techniques.
- Project 1: Used linear and logistic regression, and KNN, to analyze the Energy Efficiency and Qualitative Bankruptcy data sets from UC Irvine.
- Project 2: Used MLP class it along with CNN class from TensorFlow to classify image data from the CIFAR-10.
- Project 3: Used NBC and fine-tuned BERT using Huggingface for IMDB sentiment analysis.

Research Experience

McGill University Montreal, CA

Researcher in Reinforcement Learning & NeuroAI *Sep. 2024 – Present*

- RL Algorithm Innovation: Investigating sparse reward RL algorithms (PPO, SAC, TD3) in AnimalAI environments to mimic biological learning processes, aiming to enhance algorithm efficiency and applicability in novel domains.

Technical Skills

Languages: Python, Java, C/C++, JavaScript, HTML/CSS, SQL/NoSQL.
Tools & Frameworks: Django, React, Next.js, Express.js, Gymnasium, Pandas, NumPy, Spring Boot, TensorFlow, PyTorch, scikit-learn, HuggingFace, Transformers, Matplotlib, Seaborn, Plotly, Git, AWS, Docker, Postman, REST, TypeScript, Linux, JUnit, Bootstrap, tailwindcss, stable-baselines3, PySpark,

Awards & Hobbies

Awards: 30th place ICPC NENA, Dean's Honor's List 4x, Outstanding Achievement, Poker bot Competition 2x 1st place, 2x 3rd place.
Hobbies: Solving hard problems, Trying new food, Swimming, Pingpong, Drawing, Movies/Anime/TV shows, Traveling.