

Daniel M. DiPietro

dan.dipietro.pub@gmail.com | www.dandipietro.com

EDUCATION

Dartmouth College

Sept 2018 - June 2022

- GPA: 3.89/4.00
- B.A. in Computer Science and B.A. in Mathematics
- Relevant Coursework: Object-Oriented Programming, Software Design and Implementation, Algorithms, GPU Programming, Computer Architecture, Machine Learning, Artificial Intelligence, Reinforcement Learning, Convolutional Neural Networks, Graph Theory, Discrete Mathematics, Linear Algebra, Probability, Calculus of Multi-variable and Vector-valued Functions, Differential Equations, Applied Mathematics, Complex Analysis

WORK EXPERIENCE

Analyst | New York, NY

August 2022 - present

Jane Street Capital

- Implementing novel statistical/machine learning approaches for a variety of quantitative problems.

Business Development Intern | New York, NY

June 2021 - Aug 2021

Jane Street Capital

- Developed novel machine learning algorithms for outlier detection.
- Implemented performant pipelines to pull, process, and analyze data encompassing tens of millions of rows.

Summer Analyst | New York, NY

June 2020 - Aug 2020

The Blackstone Group

- Summer analyst in the Hedge Fund Solutions group (BAAM) machine learning and data science team.
- Implemented statistical inference pipelines to inform trading decisions.

Quantitative Research Intern | New York, NY

June 2019 - Aug 2019

Rebellion Research

- Constructed long-term equity investing and portfolio rebalancing algorithms in Python using quantitative techniques, outperforming benchmark indices numerous times.

RESEARCH PUBLICATIONS

[1] DiPietro, D. M., & Zhu, B. (2022). Symplectically Integrated Symbolic Regression of Hamiltonian Dynamical Systems. *arXiv:2209.01521*.

- Achieves SOTA performance in learning free-form symbolic governing equations from Hamiltonian dynamical systems by using a risk-seeking policy gradients approach coupled with a number of physics priors, such as preservation of symplectic structure.

[2] DiPietro, D. M., & Hazari, V. D. (2022). DiPietro-Hazari Kappa: A Novel Metric for Assessing Labeling Quality via Annotation. *arXiv:2209.08243*.

- Proposes a novel statistic for assessing dataset label quality via annotator agreement and probability theory.

[3] DiPietro, D. M. (2022). Quantitative Stopword Generation for Sentiment Analysis via Recursive and Iterative Deletion. *arXiv:2209.01519*.

- Presents a novel algorithm for constructing highly efficient stopwords sets; empirically, these sets can shrink dataset size by 64% with only a 2.8% degradation of model accuracy.

[4] DiPietro, D. M., Hazari, V. D., & Vosoughi, S. (2022). Robin: A Novel Online Suicidal Text Corpus of Substantial Breadth and Scale. *arXiv:2209.05707*.

- Using 1.1m scraped social media posts, developed a state-of-the-art corpus of suicidal text.

[5] DiPietro, D. M., Xiong, S., & Zhu, B. (2020). Sparse Symplectically Integrated Neural Networks. *Advances in Neural Information Processing Systems*.

- Developed a novel neural network architecture with embedded physics priors that facilitate the learning of governing equations for Hamiltonian dynamical systems; outperformed current state-of-the-art black-box prediction techniques by an order of magnitude.

[6] DiPietro, D. M. (2019). Alpha Cloning: Using Quantitative Techniques and SEC 13F Data for Equity Portfolio Optimization and Generation. *Journal of Financial Data Science*, 1(4), 159-171.

- Developed several models that outperform the S&P 500 in historical backtesting by identifying equities in each examined portfolio that are overweighted relative to a benchmark weighting approach.

[7] Deng, Y., Zhang, Y., He, X., Yang, S., Tong, Y., Zhang, M., DiPietro, D. M., & Zhu, B. (2020). Soft Multicopter Control using Neural Dynamics Identification. *Conference on Robot Learning*.

- Employs a lightweight neural surrogate model to identify and predict the temporal evolution of a set of geometric variables characterizing an elastic soft body. This model is then incorporated into an LQR controller with an online relinearization scheme.

[8] Fleiss, A., Cui, H., Stoikov, S., & DiPietro, D. M. (2020). Constructing Equity Portfolios from SEC 13F Data Using Feature Extraction and Machine Learning. *Journal of Financial Data Science*, 2(1), 45-60.

- Uses extracted features from SEC filing data to train portfolio generating models with traditional machine learning techniques such as logistic regression and XGBoost.

[9] Liu, A., & DiPietro, D. M. (2020). *Drosera of the New Jersey Pinelands, U.S.A.* Botanical Research Institute of Texas.

- Ecological field research book on the plant genus *Drosera*. Sold out of first press run.

SOFTWARE PROJECTS

Fairer Features

- Implemented a machine learning pipeline of ensembled CNNs and LLMs to ascertain both the representation and depiction of various demographic groups in image datasets.
- Developed an optimization process to reweight datasets and make them fairer; interviewed by YCombinator for 2023W batch.

GPU Conjugate Gradient Solver

- Implemented a GPU-parallelized (CUDA) conjugate gradient solver for positive-definite systems of linear equations.
- Incorporated a mixed precision approach where numerical precision is gradually increased until convergence occurs; decreases run-time by an order of magnitude.

D2 BioSoftware

- Developed image-processing algorithms that count the number of colonies on a bacterial plate, using techniques such as adaptive thresholding, edge detection, and blob detection. Implemented as an Electron Application.
- Used by the University of Pittsburgh's McGowan Institute for Regenerative Medicine.

HONORS AND AWARDS

- Computing Research Association: Outstanding Undergraduate Researcher Award (Finalist)
- Neukom Institute for Computer Science: Outstanding Undergraduate Research in Computer Science Award (1st Place)
- Four academic citations for meritorious performance (machine learning, reinforcement learning, graph theory, GPU computing)
- Thayer Scholar (top 5% of incoming engineering students at Dartmouth College)
- Kemeny Award for Innovative Computing Projects

SKILLS

Programming Languages: Python, C, C++, CUDA, Javascript, Java, HTML, CSS

Technologies: Artificial Neural Networks/Machine Learning (Pytorch, Sklearn), Data Analysis, Data Processing, Data Scraping, Microcontrollers, API Interaction, Bash, UNIX, Apache, SQL, MongoDB, Git, LaTeX, Electron

Professional Skills: Research, Collaboration, Project Management, Public Speaking, Technical Writing