Pulkit Madaan

Email: pmadaan2@jhu.edu | madaanpulkit.github.io
in/pulkit-madaan | madaanpulkit | Pulkit Madaan

Education

Johns Hopkins University
MSE Computer Science
CGPA: 3.97/4.0
Aug 2022 – May 2024

Indraprastha Institute of Information Technology Delhi
B.Tech in Computer Science and Applied Math
Aug 2016 – Aug 2020
CGPA: 9.23/10.0
Best Academic Performance in Major
Thesis: Better Representations - Deep Mean Shift Clustering

Skills

Languages: Python, Java, C++, Bash, R
Other: PyTorch, HuggingFace, LangChain, Pinecone, Jupyter, Git, Torchtext, Streamlit, Numpy, scikit-learn, PyTorch Lightning, Detectron2, Docker, Hydra, NNI.

Experience - Industry

Associate ML Scientist - I Jul'20 - Jul'22
Wadhwani Institute for Artificial Intelligence
Agriculture Team, Core ML Team [Code]
- Successfully executed multiple deployments reaching 10k+ users across 10+ states.
- Brought down the Mean Absolute Error from 10+ to <2.
- Developed and open-sourced a flexible & generic Object Detection ML pipeline. [Code]
- Solution won the Global Change Award 2022 [article].

Experience - Academia

Graduate Research Assistant-NLP Aug'23 - Ongoing
CLSP @ JHU x Amazon AGI
Advisor: Anjalie Field
- Anonymizing long-form unstructured text medical notes while maintaining research fidelity.
- Working with Child Protective Services to help them redact and replace sensitive information automatically with an ML pipeline to anonymize and share data effortlessly.

Graduate Research Assistant-CV Nov'22 - Aug'23
MINDS @ JHU
Advisor: Benjamin Haeffele, Matthew M. Ippolito
- Built Domain Adaptative ML Library for Malarial Parasite detection and life-stage prediction.
- Improved F1-score from 56% to 89% with the domain adaptive approach making the process viable for use on unseen unlabled datasets.

Graduate Research Assistant-NLP Jan'23 - Dec'23
CLSP @ JHU
Advisor: Philipp Koehn
- Building Language Agnostic Sentence Representations without the use of parallel data.

Mitacs Globalink Research Intern May'19 - Aug'19
UQAM
Advisor: Dr. Fatiha Sadat
- Improved multilingual translation of low-resource languages by BLEU score points. Accepted at WILDRE-5 (part of LREC 2020)

Publications

- White, J., **Madaan, P.**, Shenoy, N., Agnihotri, A., Sharma, M., & Doshi, J. (2022). A Case for Rejection in Low Resource ML Deployment. ArXiv preprint `arXiv:2208.06359`. [Accepted at Challenges in Deploying and Monitoring ML Systems Workshop - NeurIPS 2022] [LINK]
- **Madaan, P.**, & Sadat, F. (2020, May). Multilingual Neural Machine Translation Involving Indian Languages. In Proceedings of the WILDRE5–5th Workshop on Indian Language Data: Resources and Evaluation (pp. 29-32).[LINK]
- **Madaan, P.**, Maiti, A., Anand, S., & Mittal, S. (2019). Deep Mean Shift Clustering. [preprint] [LINK]

Academic Service

- Program Committee - RANLP 2023.
- Reviewer - ACM TALLIP 2022.
- Programme Committee - DravidianLangTech-ACL 2022.
- Course Assistant - Statistical Analysis at JHU.
- Course Assistant - Gateway Computing: Python at JHU.

Select Projects

Anime GPT Chatbot
- Built a chatbot with LangChain prompt templates, HuggingFace google-flan-t5-xl model on Anime data vectors with HuggingFace Transformers indexed in Pinecone.

[HuggingFace, Transformers, LangChain, PineCone]

Flow Based Generative Models: GLOW
Course: Probabilistic Graphical Models [Code] [Slides]
- Conditioned GLOW in different generation and conversion tasks (eg: replacement to vocoders, GANs)

[PyTorch, Colab]

SaNJI: Satellite eNhanced Judicial Irrigation
- Predicting soil factors from satellite spectral images to recommend near real time irrigation recommendation.

[PyTorch, Landsat. Torchvision]

Benoit: Better English Noisy Audio Transcripts
Course: Machine Learning [Code] [Slides]
- Trained a denoising seq2seq autoencoder on top of Wav2Vec 2.0 for grammatically correct ASR.

[PyTorch, TorchAudio, TorchText, Colab]

Awards

- Part of Wadhwani AI team that won the HUL, Google and MyGov India's AI for Agriculture Hackathon with a 1 Million INR grant.
- Best Academic Performance in B.Tech. CSAM [Branch Topper | Gold Medalist] **[cert]**
- Dean's Academic Excellence Award for 2 consecutive years: 2017-18, 2018-19 **[cert]**

Relevant Courses

CS: ML System Design, AI Ethics, Computer Vision, Causal Inference, Machine Translation, Speech Recognition, Reinforcement Learning, Digital Image Processing
Math: Differential Geometry, Calculus on $\mathbb{R}^N$, Stochastic Processes, Statistical Inference, Linear Optimisation, Real Analysis, Abstract Algebra