Prachi Rahurkar

3501 Madrid Dr, San Jose, CA- 95132
rahurkap@gmail.com | 669-204-1088 | https://prachirahurkar.github.io/ | linkedin.com/in/prachi-rahurkar/

OBJECTIVE: Looking for full-time roles in Natural Language Processing and Computer Vision (Machine Learning / Deep Learning).

EDUCATION:

Master of Science in **Computer Science**: GPA **3.71 /4** Sept 2018 - Feb 2021
Oregon State University, Corvallis, OR - USA
Relevant Coursework: Deep Learning, Machine Learning, Natural Language Processing, Computer Vision I and II, Intelligent Agents and Decision-Making, Networks in Computational Biology, Software Project Management, Software Engineering Methods

Bachelor of Engineering in **Computer Science**: GPA **8.26 /10** Aug 2014 - May 2018
Thadomal Shahani Engineering College - University of Mumbai, Mumbai, MH - India
Relevant Coursework: Data Structures, Analysis of Algorithms, Parallel and Distributed Systems, Object-Oriented Programming, Data Warehouse and Mining, Cloud Computing, Computer Graphics, Distributed Databases, Artificial Intelligence, Computer Architecture

TECHNICAL SKILLS:

Languages: Python, Java, C/C++, R, Javascript, React, Node.js, jQuery
Libraries: Tensorflow, PyTorch, Transformers, Flask, spaCy, NLTK, Scikit-learn
Databases: SQL, MongoDB
Tools: Docker, Git, Bash script

WORK EXPERIENCE:

Machine Learning Engineer (Part-time) Aug 2020 - present
Memorial Sloan Kettering Cancer Center New York City, NY, USA
- Built a radiology-dedicated language model by unsupervised pre-training of NLP model RoBERTa on radiology text data extracted from medical CT scan reports, and then fine-tuned the language model for multiple downstream tasks.
- Orchestrated the state-of-the-art model for 700K+ reports (which achieved 96% accuracy) and elucidated the remaining team on the latest technologies in NLP, which led to a publication in the cancer research journal.

NLP Data Scientist (Intern) May 2020 - Aug 2020
Memorial Sloan Kettering Cancer Center New York City, NY, USA
- Designed and implemented a complete NLP-ML pipeline that predicts whether a metastatic (mets) disease of cancer is present in a given radiology report; using Python.
- Scaled the pipeline addressing 1 mets task to 13 mets tasks (such as kidney mets, liver mets, lung mets, spleen mets, etc. among others), and from 1K reports to 700K reports, boosting the usability and adoption rate by doctors by 55%.
- This work enabled data engineers in the institution to transition radiology reports into lists of data items, each indicating which condition the patient is experiencing, reducing the look-up time from 40 minutes to 3 minutes for every report.

Algorithm Engineer - Computer Vision (Intern) June 2019 - Sept 2019
KLA - Tencor Corp. Milpitas, CA, USA
- Implemented the Image Quality (IQ) metrics: signal-to-noise ratio, contrast-to-noise ratio, sharpness and rotation in each image (in Python) which provided accurate IQ scores for 94% of the given images, and is 4x faster than the prevalent Matlab pipeline.
- Proposed a no-reference IQ estimation technique which involved: extracting patches of images, extracting features from these patches, and then modeling these features using a multivariate gaussian model.

Research Intern - Machine Learning July 2017 - Apr 2018
Homi Bhabha Centre for Science Education, Tata Institute of Fundamental Research Mumbai, MH, India
- Developed a Handwritten Character Recognition System for the script NaYaNa designed for dyslexic students, using stacked neural networks, along with a training-data generation system.
- Engineered Android application for native mobile users along with a web interface that increased the user satisfaction by 100%.
- Bagged **1st prize** for the system in DJ ASCII, the **state-level** project competition conducted in Mumbai (in May 2018).

PUBLICATIONS:

- Prachi Rahurkar, Matt Olson and Prasad Tadepalli. *"Human Adversarial QA: Did the Model Understand the Paragraph?" NeurIPS 2020 - Workshop on Human and Model in the Loop Evaluation and Training Strategies (HAMLETS). *(**M.S. Thesis**)

PROJECTS:

Brain Networks - Empirical Analysis & Comparison (Fall '20)
- Performed structural analysis on human connectome and chimpanzee connectome data.
- Implemented graphical measures such as centralities, motifs, clustering coefficients, max-flows, analysis of unique edges, etc. among others, and showcased differences between the two brain networks.

Pattern Studio with Deep Generative Adversarial Networks (GANs) (Spring '19)
- Implemented Self-Attention GAN on floral and geometric design pattern datasets to generate new and appealing cloth designs.
- Performed segmentation of the garment selected by the user in the given image to apply the pattern (generated or downloaded) on the clothing segment of the image, and deployed it to an Android application.

Music Generation for Poetry Lyrics (Winter '19)
- A model that takes poetry lyrics as input and generates a suitable melody for it, as the output (in PyTorch).
- Implemented using stacked LSTM layers followed by a simple dense layer with a SoftMax activation function.