Paarth Tandon

401-497-4218 | ptandoncs@gmail.com | GitHub | paarthtandon.com | LinkedIn

Skills

- Python | Pandas | NumPy | MatPlotLib | PyTorch | Tensorflow | Jupyter | SQL | NoSQL | JavaScript | Node | Julia | R | Git | Statistics
- Machine Learning | Natural Language Processing | Transformers | Anomaly Detection | Unsupervised Learning | Generative Models
- AWS | EC2 | S3 | GCP | BigQuery | ElasticSearch | Docker | Airflow | Linux | Reinforcement Learning | LLM Fine-tuning | Spark

Experience

Data Scientist — **Freelancer** — *Remote* — **04/2023 - Present**

- **Arex Life Sciences (Biotech):** Developed a proprietary signal processing algorithm that classified biological samples with **near 100% accuracy**. Designed a user-friendly interface for laboratory technicians to import and manage sample data using a **NoSQL** database.
- **Intorqa (Digital Forensics):** Exploring NLP and network analysis methods to identify key influencers in extensive, unreliable networks of textual interactions, enhancing the detection of pivotal bad actors in digital forensic investigations.

Data Scientist — **Bungie** — *Bellevue WA* — **06/2023 - 10/2023**

- Built data-driven **anti-cheat** for Destiny 2. Automated detection pipeline that **increased bans by 20%** through the detections I created.
- Trained custom image and vector-based **transformer models** using highly imbalanced, **3+ terabyte** datasets.
- Crafted and optimized (**2x speedup**) SQL queries to uncover anomalous events occurring in player data spanning **trillions of rows**.
- Optimized Security Analyst workflows using a query which corroborated evidence against a cheater, **speeding up investigations by 5x**.
- Worked with data engineers to productionize a high throughput vision transformer using **Docker**, **AWS Batch Compute**, and **Airflow**.
- Implemented **Autoencoders**, **GANs**, and **Diffusion Models** to detect anomalies using techniques such as **density estimation**.

Data Science Intern — **Moody's Analytics** — *Remote* — **06/2022 - 10/2022**

- Worked on NewsEdge, an **NLP** news analytics service used by companies in finance, publishing, and for corporate awareness.
- Developed a novel algorithm for real time event detection, replacing a previously unusable feature. The algorithm was built using **Python**, **Pandas**, **NumPy**, and **Pytorch**.
- Achieved event labeling speeds of under **3 ms per story**, while also improving label specificity and accuracy over previous attempts.
- Leveraged **AWS**, **S3**, **ElasticSearch**, and **EC2** cloud computing technologies to process stories for the real time event detection feature.
- Improved **language detection**, related stories, and the **automatic summarization** features by applying state of the art NLP models.
- Collaborated with Software Engineers and Data Engineers to prepare features for production; my contributions will enter production by the end of 2022.
- Applied software development best practices using **Git**, **Agile**, **Jira**, **Confluence**, **unit testing**, and extensive **documentation**.

Data Science Intern — **Ribbon Communications** — *Remote* — **06/2021 - 06/2022**

- Applied dimensionality reduction (**PCA**) and **clustering** using **Scikit-Learn** and **MatPlotLib** to answer business questions in the domain of telecommunications. Areas of interest include anomaly detection and error correlation.
- Built an automatic schema matcher that was able to correctly match columns on over **200** table schemas with over **99% accuracy**, using a combination of **NLP** and traditional methods.

Education

MS in Computer Science — **University of Massachusetts** — *Amherst, MA, USA* — **08/2022 - 07/2023**

- Data Science Focus, 3.9 GPA
- Highlighted Courses: Reinforcement Learning, Systems for Data Science, Visual Computing, Advanced NLP, Algorithms for Data Science, Data Science in R Mathematical Statistics, Ethics in Computation

BS in Computer Science — **University of Massachusetts** — *Amherst, MA, USA* — **08/2019 - 05/2022**

- 3.61 GPA
- Highlighted Courses: Machine Learning, Natural Language Processing, Data Visualization, Artificial Intelligence, Database Management, Search Engines, Data Structures, Algorithms, Statistics, Discrete Math, Multivariable Calculus, Linear Algebra

Projects

- **TrashGPT:** Fine-tuned **LLaMa** on the Trash Taste podcast. Generated realistic interactions and rendered them using **speech generation**.
- **Personal MusicGen:** Fine-tuned **MusicGen**, a transformer based **music generation model**, on a personal dataset of music I liked.
- **Pokémon Battle AI:** Applied **Deep Q Learning** using **PyTorch** to train a Pokémon AI, winning against a greedy AI in over **80%** of battles.
- **DreamPop:** Used the **Spotify API** to scrape a large dream pop playlist. Created a dream pop classifier with **82% acc** using **Scikit-Learn**.
- **Search Engine:** Implemented a search engine in **Python** using tokenization, **PageRank**, inverted index, query likelihood, and **MapReduce**.
- **Discord Bot:** Created using **Python** for a server of **70 members**. Included activity tracking and minigames. Data logged on **PostgreSQL**.

Leadership

- **President:** ACM Machine Learning Club | Ran weekly meetings | **Recruited 150 members** | Technical workshops | Discussions on AI ethics
- **Course Instructor:** FYS 191: Thinking with Machine Learning | Discussions on industry, research, ethics | Introduced freshmen to ML