

Paarth Tandon

Seattle | 📞 401-497-4218 | ✉️ ptandoncs@gmail.com | 🌐 [GitHub](#) | [paarthtandon.com](#) | [in](#) [LinkedIn](#)

Skills

- Python | Pandas | NumPy | Matplotlib | PyTorch | Tensorflow | Jupyter | SQL | NoSQL | JavaScript | Node | Hadoop | Julia | R | Spark | Git
- Machine Learning | Natural Language Processing | Transformers | Statistics | Data Visualization | Clustering | Streaming Algorithms
- AWS | GCP | Elasticsearch | EC2 | S3 | BigQuery | Docker | API Design | Unit Testing | OOP | Backend | Linux | Terminal | LaTeX | Agile

Experience

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.
- Applied **transformer** based models to anti-cheat using **multi-terabyte**, highly imbalanced datasets.
- Crafted performant **SQL** queries to uncover anomalous events occurring in player data spanning **trillions of rows**.
- Carried detections from **ideation to production**. Worked with data engineers to productionize models using **Docker** and **Airflow**.
- Adapted and implemented **modern machine learning research** to tackle **unsolved problems** in the anti-cheat space.
- Technologies used include Python, NumPy, Pandas, Matplotlib, PyTorch, Docker, Airflow, GCP, AWS, and Git.

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.

Web Developer

UMass Art History

Remote

06/2020 - 05/2021

- Modernized online coursework which was originally written in Adobe Flash. Used **HTML**, **CSS**, and **JavaScript** to recreate them.
- Redesigned over **30** pages to include interactive activities using **HTML 5 Canvas**, **WebGL**, and **KonvaJS**.

Education

Master of Science

University of Massachusetts

Amherst, MA, USA

08/2022 - 07/2023

- Major in Computer Science, 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

Bachelor of Science

University of Massachusetts

Amherst, MA, USA

08/2019 - 05/2022

- Major in Computer Science, 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**.
- **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**.
- **DS Algos**: Implemented various DS algorithms using **Julia**. Currently in-progress, implemented **two-level hashing** and **bloom filters**.

Leadership

- **President**: ACM Machine Learning Club | Ran weekly meetings | **Recruited 150 members** | Technical workshops | Discussions on AI ethics
- **Course Instructor**: Thinking with Machine Learning | Discussions on industry, research, ethics | Introduced freshmen to machine learning
- **Course Assistant**: Introduction to Algorithms | Held office hours | Graded homework and exams | Answered questions on course forum