Paarth Tandon

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

Skills

- Python | Pandas | NumPy | Scikit-Learn | PyTorch | Tensorflow | Jupyter | SQL | NoSQL | JavaScript | Node | Hadoop | Julia | Spark | Git
- Machine Learning | Natural Language Processing | Data Visualization | Clustering | Streaming Algorithms | Multiprocessing | Statistics
- AWS | Cloud Computing | ElasticSearch | EC2 | S3 | API Design | Unit Testing | OOP | Backend | Linux | Terminal | VS Code | LaTeX | Agile

Education

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

- Major in Computer Science, Data Science Focus
- Highlighted Courses: Reinforcement Learning, Systems for Data Science, Visual Computing, Advanced NLP, Mathematical Statistics, Ethics in Computation

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

- Major in Computer Science
- Highlighted Courses: Machine Learning, Natural Language Processing, Algorithms for Data Science, Data Visualization, Artificial Intelligence, Database Management, Search Engines, Data Structures, Algorithms, Statistics, Discrete Math, Multivariable Calculus, Linear Algebra

Experience

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**.
- Participated in daily standup, sprint planning, and retrospective meetings. Presented findings to **upper management** and **stakeholders**.

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.
- Automated the **ETL** process of multiple data pipelines from a variety of sources using **KNIME**, **Python**, and **Pandas**, allowing the team to have more reliable and convenient access to current data.
- Performed literature reviews on machine learning in telecommunications, contributing over **10** articles to the department's internal wiki.

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**.
- Added compatibility for all modern devices, creating a seamless experience on desktop, laptop, tablet, and phone screens.
- Consulted with accessibility experts to ensure that lessons were **inclusive** to screen-readers and keyboard-only users.
- Improved a custom, **Python**-based static site generator, allowing for instructors to easily add more text and image-based content to their lessons without developer intervention.

Projects

- **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