Pranav Pabba

+1-(650)-400-3754 | pranav.pabba@gmail.com | pranav-pabba | ppabba101 | pranavpabba.com

Profile

Sophomore in Computer Science at the University of Michigan; six years of Python with strengths in AI/ML, NLP, and data engineering. Recent work includes AWS Athena/Redshift lineage and dataset ranking for AWS cost optimization and marketing analytics.

Skills

- **General Programming:** Python, SQL, C++, Java, C, JavaScript, TypeScript, Git
- **Artificial Intelligence & Machine Learning:** PyTorch, TensorFlow/Keras, OpenCV, Scikit-learn, NumPy, Pandas
- **Data Engineering & Platforms:** AWS (Athena, Redshift, Glue, Lambda, S3), Apache Airflow
- **Web Technologies:** React, Flask, Docker, WordPress, Figma, HTML/CSS

Experience

- **GoDaddy [🌐]** *May 2025 - August 2025*
  SDE Intern Remote/Tempe, AZ
  - Built a campaign-to-table lineage graph across AWS Athena/Redshift; ingested SFMC, MessageGears, and internal APIs; scanned tens of thousands of campaigns to generate 100,000+ campaign-to-table mappings.
  - Created a table-value attribution model tying campaign impact to underlying datasets; ranked tables by downstream influence to guide maintenance, deprecation, and storage/compute spend.
  - Developed scoring formulas and ML/AI heuristics to quantify value-per-GB/compute, surfacing high-leverage tables and pruning low-yield data paths.
  - Implemented a Model Context Protocol (MCP) server behind an internal data-hydration tool, enabling non-technical ops to define and schedule validated hydration jobs for reliable downstream use.
- **Intentwise [🌐]** *June 2023 - August 2023*
  Python Development Intern Remote
  - Built an NLP feature for ad optimization that generated higher-quality Amazon keywords from product text and reviews, improving relevance and coverage.
  - Implemented scalable data pipelines for large keyword corpora (cleaning, tokenization, dedup, ranking) to support fast candidate generation.
  - Applied ML techniques like TF–IDF and embedding-based similarity to raise the precision/recall of suggestions versus baseline heuristics.
- **Exemplifi [🌐]** *July 2022 - October 2022*
  Web Development Intern Palo Alto, CA
  - Contributed front-end components for an enterprise WordPress launch (www.acronym.com), focusing on responsive, reusable widgets and page performance hygiene.

Projects

- **Japanese Language Assistant** *February 2023 - Current*
  Tools: PyTorch, torchaudio, Selenium, Kaldi, Praat, NumPy, Matplotlib, Seaborn, Pandas
  - Developing an app alongside Japanese teachers that provides immediate, customized language feedback for hundreds of Japanese language students.
  - Implemented automated learner feedback approaches, including phoneme/IPA alignment and spectrogram analysis; exploring Transformer-based scoring.
  - Built scrapers to collect diverse-accent audio and metadata, expanding training/eval coverage.
- **Phenotype prediction using single-cell RNA data** *June 2022 - September 2022*
  Tools: PyTorch, NumPy, SciPy, Pandas, Seaborn, ScanPy, AnnData
  - Constructed predictive models using XGBoosted forests, Gaussian Mixture Models, and CNNs to analyze scRNA data, achieving high accuracy in phenotype classification.
  - Utilized ScanPy and AnnData for efficient preprocessing and analysis of large single-cell RNA sequencing datasets.
  - Combined datasets from diverse studies in a non-biased manner using varied statistical methods, enhancing the robustness of phenotype predictions.

Education

- **University of Michigan, Ann Arbor** *August 2024 - May 2028*
  B.S. in Computer Science Ann Arbor, MI
  - **GPA:** 4.00/4.00 | **Honors:** James B. Angell Scholar, William J. Branstrom Freshman Prize
  - **Notable Classes:** Introduction to Machine Learning, Web Systems, Foundations of Computer Science, Data Structures and Algorithms, Intro to Computer Organization, Linear Algebra, Discrete Math, Differential Equations