ATHARVA YEOLA

+1 8582203508 | ayeola@ucsd.edu | in/atharvayeola | /atharvayeola /atharvayeola

EDUCATION

University of California San Diego (UCSD) **September 2023 – Present**
Master of Science (M.S) – Machine Learning & Data Science *San Diego, USA*

Pune Institute of Computer Technology (PICT) **July 2019 – June 2023**
Bachelor of Engineering (B.E.) – Electronics & Telecommunication – ***GPA: 3.9/4.0*** *Pune, India*

TECHNICAL SKILLS

Programming Languages & Software Development – Python, R, C++, Matlab, Javascript, Git, Linux, CI/CD
Data Science & Machine Learning – PyTorch, TensorFlow, Keras, Scikit-Learn, LLMs, Linear Algebra, Statistics
Data management & Databases – SQL, MySQL, PostgreSQL, MongoDB, NoSQL, ETL
Cloud & Big Data – AWS, Google Cloud, Docker, PySpark, Hadoop
Specialized Tools & Libraries – Tableau, NLTK, OpenCV, Scipy, CUDA, MLFlow

WORK EXPERIENCE

Data Scientist **July 2024 – Present**
Qualcomm Institute - UC San Diego Division of Calit2 *San Diego, USA*

- Co-developed "Waldo" – an **open-source** NLP tool using RoBERTa architecture to detect Adverse Events (AE) in 360k+ reddit posts, achieving **97.9% accuracy** and **outperforming GPT-3.5 turbo** by 57% in F1-score (95.1% vs. 38%).
- Leveraged synthetic data augmentation and stratified k-fold validation to address class imbalance (6.6% AE rate), enabling Waldo to **detect more than 28k AEs** across diverse sub-reddits (e.g., 12.7% AE rate in r/Marijuana).

Machine Learning Researcher **March 2024 – June 2024**
University of California San Diego *San Diego, USA*

- Improved RNA cell separation accuracy by 15% with a novel thresholding pipeline. The 40% **reduction in time-to-insight** aided 3+ FDA-cleared diagnostic tools, improving patient outcomes in clinical trials.
- Engineered a spectral clustering algorithm with CuPy to resolve overlapping RNA instances, **reducing post-processing time by 4x** and achieving 70% IoU on 300 annotated cells.

Research Intern - Computer Vision **December 2022 – August 2023**
Indian Institute of Technology (IIT) Patna *Remote*

- Created a new dataset by utilizing advanced **image augmentation** techniques to generate high-quality synthetic data, thereby expanding the dataset size to 100,000 images (**30% increase**), improving robustness of the model.
- Tailored a **Vision Transformer** architecture to address edge-case scenarios in traffic sign detection, beating the **state-of-the-art** F1 score by 9%; *paper accepted at CVMI 2024 and* ***published in IEEE Xplore***.

Data Scientist - Recommender Systems **August 2022 – April 2023**
RhythmFlows Solutions Pvt. Ltd *Pune, India*

- Led **A/B testing** comparing content-based and collaborative filtering, achieving a **33% boost** in relevant recommendations. **Deployed** the model on **AWS Sagemaker** with continuous updates for improved performance.
- Reduced document processing time by 7 seconds per document by utilizing **Pytesseract OCR** for electronic receipt scanning, and revamped the data analysis pipeline, **minimizing human effort by 30%**.

Data Scientist - Product **January 2022 – June 2022**
Atomic Loops *Pune, India*

- Spearheaded a team of three in training a **YOLOv5** food detection model achieving **98% precision** which was then integrated with a **food train system**.
- Collaborated with the data platform team to construct preprocessing pipelines, facilitating a notable **20% acceleration** in **project delivery** rate, thereby **enchancing** overall **business efficiency** by **33%**.

PROJECTS

End-to-End Text Summarizer | *Transformers, NLP, GitHub Actions* **July 2024 – July 2024**

- Built a text summarizer for concise summaries, managing the entire workflow from configuration to deployment.
- Automated CI/CD with AWS and GitHub Actions, handling Docker, EC2, ECR, and runner setup.

Retail Vision Enhancement | *YOLOv8, SuperGlue, Docker* **February 2024 – February 2024**

- Implemented YOLOv8 for object detection, achieving over 90% accuracy in labeling on-shelf retail products.
- Introduced SuperGlue for precise product identification, and created a Dockerfile to ensure cross-platform reproducibility.