ATHARV DABHADE

716-520-5275 | atharvda27@gmail.com | www.linkedin.com/in/atharvdabhade | https://atharvd27.github.io/atharv/

SUMMARY

Skilled Machine Learning Engineer with 3+ years of experience in developing and implementing AI-driven solutions using Python. Proven ability to optimize data extraction and analysis processes, leveraging machine learning to solve complex challenges, demonstrating strong analytical and technical expertise across projects. Thrives in dynamic settings, delivering projects with precision. Skilled in technical writing, communication, leadership, and project management, fostering teamwork and success.

EDUCATION

Master of Science: Robotics, 3.48 GPA, University at Buffalo, The State University of New York, Expected May 2024

Bachelor of Technology: Electronics and Communication, 3.2 GPA, Indian Institute of Information Technology, May 2022

SKILLS & TOOLS

Languages & Frameworks: Python, SQL, Tensorflow, Keras, PyTorch, MLflow, Robot Operating System (ROS), MATLAB, Docker
Data Analytics: OpenCV, Numpy, Pandas, Scikit-learn, Scipy, Matplotlib, Seaborn, Tableau, Hypothesis testing, Statistics
Tools: Git, Visual Studio code, Jupyter Notebook, Office Suite (Word, PowerPoint and Excel), Linux, Windows, MacOS

WORK EXPERIENCE

Research Assistant (Machine Learning Engineer), CUBS Lab - University at Buffalo, SUNY, Buffalo, NY

- **Context-Aware Data Extraction & Field Association from Documents:** May 2023 - Present
  - Spearheaded creating and fine-tuning an OCR Pipeline by deploying multimodal networks, incorporating, and LoRA fine-tuning Large Language (LLM) Models achieving 65% accuracy in Context Aware Data Extraction.
  - Improved biographic data digitization efficiency from diverse documents by leveraging advanced NLP techniques and designing post-processing pipelines, resulting in a 15% increase in data accuracy and significantly reduced need for manual verification.
  - Led project management using agile methodologies, authored detailed documentation for ongoing R&D efforts, and containerized the OCR pipeline with Docker to ensure scalable and efficient deployment.

- **GestSpoof, Spatio-Temporal Gesture Learning for Fingerprint Attack Detection** July 2023 – September 2023
  - Collaborated on a study enhancing fingerprint spoof detection with intentional distortions, creating a unique dataset of 3,680 videos from 184 fingers, and establishing baseline numbers due to the dataset's uniqueness.
  - Leveraged spatio-temporal networks and optimized distributed training with PyTorch and timm libraries for a gesture-based dataset, increasing biometric security and achieving a 50% faster training process for fingerprint spoof detection.
  - Co-authored a research paper on the innovative approach, accepted for publication at the Conference on Automatic Face and Gesture Recognition 2023.

- **SCC-PG Scene Text Pipeline:** January 2023 - May 2023
  - Developed a web scraper to extract and categorize over 20,000 images from various sources for OCR applications and compiled a manually labeled testing dataset from these images to validate model performance.
  - Improved OCR model efficacy by 20% and overall system efficiency by 30% through a multi-stage pipeline utilizing EasyOCR and MMOCR, alongside innovative techniques and existing Python libraries.

Research Assistant, Indian Institute of Information Technology, Guwahati, India

- **Extreme Weather Predictions and Localization:** June 2021 - April 2022
  - Supervised a team of 4, consisting of personnel from IIIT and NARSS to research and incorporate novel methods to nowcast extreme weather events cyclones, and sandstorms.
  - Crafted Machine Learning pipeline leveraging Time Series Forecasting and Image Segmentation for weather predictions.
  - Conducted comprehensive studies on temporal and spatial characteristics of extreme weather events, enhancing the capability of this pipeline in the accurate detection and localization of such events.

PROJECT

Customer Segmentation & Retention Analysis: Python, SQL, Tableau

- Leveraged SQL and Python for data retrieval, clustering, and feature engineering, employing XGboost and Neural Networks to attain a 75% accuracy rate in customer segmentation and churn prediction.
- Explored retention strategies through exploratory data analysis with Matplotlib and Seaborn and streamlined workflow with an automated Python script for efficient data extraction, processing, and analysis.

Development of Sales Performance Dashboard: Python, SQL, Tableau

- Leveraged SQL, Python, and Real-Time Analytics for extracting, transforming, and analyzing sales data; developed a Tableau dashboard to visualize key sales metrics and trends, enabling data-driven decision-making.
- Implemented ARIMA for accurate time series forecasting of future sales, subsequently providing insights for business strategy enhancement through Artificial Intelligence techniques.