Purushothaman Natarajan

Cuddalore, TN, India | purushothamann@icloud.com | linkedin.com/in/purushothamann

github.com/Purushothaman-natarajan

Summary

Experienced Data Scientist with 4+ years specializing in Computer Vision and NLP, delivering high-impact AI solutions at Blackstraw AI, Amazon, and DRDO. Proven track record designing and deploying scalable, explainable ML pipelines (LIME, SHAP) that improved decision transparency by 40% for critical defense applications. Hands-on expertise in building vector similarity search, segmentation, OCR, and privacy-compliant asset detection pipelines for real-world production. Currently researching large-scale vector search, visual-language models, and self-supervised learning.

Skills

Programming Languages: Python, SQL, C++ (Basics), JavaScript (Basics)

Technologies & Tools: PyTorch, TensorFlow, Scikit-learn, Keras, OpenCV, NLTK, Pandas, Matplotlib, Seaborn, Spark, Kubernetes, Docker, FastAPI, Streamlit, Git, CUDA

ML/AI: BERT, GPT, Claude, LLaMA, Mistral, DALL·E 3, Stable Diffusion, Qwen, CNN, RNN, LSTM, Transformers, GANs, Autoencoders, Prompt Engineering, Fine-tuning, RAG (Retrieval-Augmented Generation), LoRA, LangChain, spaCy, Reinforcement Learning (PPO, DPO, GRPO)

Agents & Frameworks: DSPy, OpenAI, LangGraph, AutoGen, CrewAI

Classical ML Models: Linear/Logistic Regression, Decision Trees, Random Forest, SVM, XGBoost, LightGBM, K-Means, DBSCAN, PCA

Databases: MySQL, PostgreSQL, LMDB, Neo4j

VectorDBs: FAISS, Pinecone, Milvus, USearch

Cloud Platforms: AWS, GCP, Azure

Education

SRM University, PhD in Computer Science (withdrew prior to completion) Feb 2024 – Nov 2024

- **Coursework:** eXplainable AI, Artificial Intelligence, Comparison of Learning Algorithms, and Computational Theory.

BITS, Pilani, M.Tech in Data Science Sept 2022 – Sept 2024

- **Coursework:** Data Science, Applied Machine Learning, Deep Learning, Natural Language Processing, Information Retrieval, Artificial and Computational Intelligence.

Anna University, B.E in Mechanical Engineering Aug 2015 – Nov 2020

- **Coursework:** Kinematics of Machinery, Thermodynamics, Manufacturing Technology, Internal Combustion Engines, Design and Development.

Experience

Data Scientist (Computer Vision & NLP), Blackstraw AI – Chennai, IN Nov 2024 – Present

- Developed an agentic AI framework for intelligent coding assistance that retrieves product images via SERP and uses LLMs to extract product-specific details aligned with user needs.
- Designed and deployed scalable ML pipelines for image detection using YOLO, large-scale vector search-based image classification with SimCLR, fuzzy histogram based clustering, and Mahalanobis distance for anomaly detection.
- Built text detection and recognition models using custom OCR architectures (DeepLabV3Plus, SegFormer, YOLO11N/L, ViT-STR, MGP-STR, FFT-Former); integrated CharCNN with attention and super-resolution, reducing manual task effort by 40% in production.
- Developed in-house layout understanding models for client-specific documents, leveraging line detection to enhance OCR accuracy, reading order consistency, and structured data extraction from complex layouts.

- Built asset detection and privacy compliance pipelines using object detection and embedding-based similarity mapping to locate and redact sensitive content.
- Implemented NLP-driven privacy compliance models (Flair, spaCy) to identify and extract personally identifiable or confidential information across unstructured documents.
- Designed an end-to-end agentic automation framework using Crew AI for seamless codebase migration from Oracle to MySQL and from Java to Python.

Research Fellow (Machine Learning & eXplainable AI), SRMIST – Chennai, IN Sept 2023 – Oct 2024

- Selected as a **funded PhD candidate under DRDO sponsorship**, with multiple published research papers in explainable AI and underwater SONAR image analysis.
- Developed algorithms and reliable AI/ML models for underwater sonar image detection, enhancing submarines underwater surveillance and exploration capabilities.
- Integrated Explainable AI into a framework, ensuring transparency and trust in decision-making processes within the critical defense domain and yielding reliable, interpretable AI models.
- Delivered a self-explainable AI model using LIME, SHAP and Grad-CAM for underwater SONAR image detection and classification, tested and utilized by NPOL, DRDO, and the Defense Ministry of India.

Machine Learning Associate, Amazon – Chennai, IN Jul 2022 – May 2023

- Ensured data quality for Alexa, Ring, and Halo; processed text with NLTK and spaCy; standardized pipelines for cleaner labeling.
- Developed performance dashboards for actionable insights and process improvements.

Customer Support Executive, Amazon – Coimbatore, IN Aug 2021 – Nov 2021

- Managed customer and seller queries, refunds, and returns, achieving top performer status for two consecutive months.

Publications

Synth-SONAR: Sonar Image Synthesis with Enhanced Diversity and Realism via Dual Diffusion Models and GPT Prompting Sept 2024

Purushothaman Natarajan, Kamal Basha, Athira Nambiar

VALE: A Multimodal Visual and Language Explanation Framework for Image Classifiers using eXplainable AI and Language Models June 2024

Purushothaman Natarajan, Athira Nambiar

Underwater SONAR Image Classification and Analysis using LIME-based Explainable Artificial Intelligence Dec 2023

Purushothaman Natarajan, Athira Nambiar

Projects

Open-Source OCR Models `github.com/Purushothaman-natarajan/doctane`

- Built an open-source OCR repository (doctane) combining state-of-the-art research for robust text extraction and document understanding.
- Tools: Python, Transformers, NLTK, Streamlit, Pytorch.

XAI for AID Scene Classification on Remote Sensing `github.com/Purushothaman-natarajan/eXplainable-AI-for-Image-Classification-on-Remote-Sensing`

- Designed a transfer learning-based scene classifier with LIME and Grad-CAM for visual explanations.
- Tools: Python, TensorFlow, Scikit-learn, LIME, Grad-CAM, Gradio.

Additional Experience and Awards

Business Analyst, IIFL (Oct 2019 – May 2021): Managed daily client transactions worth $10M, liaised between clients and the trading desk, and provided strategic market insights.

Instructor, BrightNext Academy (2023-2024): Taught Machine Learning and Deep Learning courses to over 100 students.

Freelancer (Upwork and LinkedIn) (2022-2024): Successfully Delivered AI and ML projects valued between $10K to $20K as an independent contributor.

Third Prize, Innovation & Design on Remote Sensing Data by Hack2Skill: Recognized for designing a synchronized research and production-ready dashboard for an explainable image classifier, competing against 100+ entries.

Licenses & Certifications

Udacity: Deep Learning, Computer Vision, Generative AI Nanodegree's.

DeepLearning.AI: Retrieval Augmented Generation (RAG), Reliable AI using Guardrails.

Codecademy: BI Dashboards with Power BI & Tableau, Business Intelligence Data Analyst, SQL for Marketers and Product Managers.

LinkedIn: Advanced SQL, Artificial Intelligence Foundations: ANN, CNN, R-CNN, RNN, LSTM, GNN & Transformers, Advanced NLP with Python for Machine Learning, GANs and Diffusion Models with TensorFlow and PyTorch, Transfer Learning Using PyTorch, Deep Learning for Computer Vision Applications.