NISHCHAL MARUR

nishchal-mn.com | nmarur21@umd.edu | linkedin.com/in/nishchal-mn | github.com/NishchalMN | +1 (240) 438-1916 | College Park, MD

EXPERIENCE

Machine Learning Engineer II, Entrupy Inc, Bangalore **Aug 2021 - Aug 2024**

- Built and deployed an end-to-end luxury goods authentication pipeline, achieving **96% TPR** at **5% FPR** and processing **10K+** items monthly in production with Ray Serve for distributed inference.
- Developed a 3D document unwarping system trained entirely on synthetic data, achieving **0.84 SSIM** on real-world documents and improving OCR accuracy by **23%**, outperforming SOTA methods.
- Optimized on-device CoreML inference resulting in a **2x faster** workflow over manual capture, using quantization and dynamic overlays for real-time auto-capture.
- Improved macro fingerprinting pipeline for item tracking and **return fraud** detection using patch embedding similarity and histogram features, enhancing TPR by **15%** while reducing false positives.

AI Intern, Connyct, New York **Jun 2025 - Aug 2025**

- Built a hybrid recommendation system using sentence transformers and Elasticsearch vector search for content-based filtering, achieving **Precision@5** of **0.80** with interaction tracking and dynamic popularity boosting.
- Implemented multi-factor ranking pipeline with user preference modeling and Redis caching, supporting **50+ RPS** with A/B testing framework to continuously improve CTR and recommendation quality.

Software Development Intern, IBM, Bangalore **Jan 2021 - Jul 2021**

- Optimized batch prediction pipelines for large file scoring in **Kubernetes** pods with **Go** concurrency and chunked downloads, reducing inference latency by **15ms** in IBM Watson Cloud deployments.
- Benchmarked TensorFlow, PyTorch, Scikit-Learn and ONNX runtimes to evaluate performance and migration feasibility for an internal architecture shift.

Software Engineering Intern, SLK Software, Bangalore **May 2020 - Jul 2020**

- Saved developers **10+ hours** per week on log retrieval and debugging efforts by building a centralized ELK Stack log aggregation system using Filebeat and Node.js **APIs** across **5+** components.

Machine Learning Intern, PathPartner Technology, Bangalore **May 2019 - Jul 2019**

- Developed a real-time CNN-based gaze tracking system using transposed convolutions and Gaussian heatmap regression, achieving a mean error rate of **1.3 px** across diverse lighting and occlusion scenarios.

TECHNICAL SKILLS

Languages: Python, C++, SQL, Go, Scala, Node.js
Frameworks & Tools: Docker, Kubernetes, AWS, Azure, Apache Spark, Airflow, Ray, Blender, Elasticsearch, Redis
Machine Learning Tools: Pytorch, Tensorflow, HuggingFace, Transformers, OpenCV, Open3D, Flask, Pandas, Keras, Langchain, LangGraph, Scikit-Learn, WandB, MLFlow

EDUCATION

Master of Science (M.S) in Machine Learning **Aug 2024 - May 2026**
University of Maryland, College Park *GPA: 3.7/4*
Courses: Deep Learning, Optimization, Multimodal Foundational Models, Computer Vision, Robotics, NLP, MLOps

Bachelor of Technology (B.Tech) in Computer Science **Aug 2017 - May 2021**
PES University, Bangalore *GPA: 3.6/4*
Courses: Data Structures, Algorithms, Data Science, Data Analytics, Machine Learning, Operating Systems, Cloud Computing

PROJECTS

CAFBrain: Multimodal LLM Platform for Capital Area Food Bank (Top 3) | Project Link

- Built an agentic **RAG** system using **Langchain**, **LangGraph**, **LLMs**, **OCR**, and prompt engineering to generate grant proposals, blogs, and reports from PDFs, images, and videos, reducing content creation time from **hours to under a minute**.
- Indexed **1000+** documents into **FAISS** vector DB and implemented content querying and refinement with human-in-the-loop feedback, enabling transformation of multimodal inputs into ready-to-download files.

FedMedVision: Privacy-Preserving Federated Learning Platform | Project Link

- Developed a federated learning system simulating multi-hospital training on classified X-ray images across **4+** client nodes, improving global **F1 score** by **12-15%** by aggregating class-skewed updates without sharing raw data.
- Tracked global model performance across training rounds using MLflow, with automated evaluation pipelines, and used Docker to containerize and deploy client nodes in distributed hospital environments.

Scalable DBaaS for Rideshare application | Project Link

- Built a fault-tolerant Database-as-a-Service on AWS EC2 using RabbitMQ RPC queues and a custom orchestrator to manage read/write routing and multi-node replication.
- Enabled auto-scaling and leader election for slave containers, simulating **1000+** users and supporting fault recovery based on traffic.