Vishak Bharadwaj S

Machine Learning Engineer III

vishak.svec@gmail.com
+91 948 362 8282
linkedin.com/in/vishakbharadwaj
Bengaluru, India

PROFILE

ML Engineer with 6+ years building data-driven systems across recommendation, deep learning, GenAI, and MLOps — from signal pipelines and ranking models to LLM-powered annotation and production monitoring at scale, delivering measurable engagement lifts and production-grade systems serving 100M+ users. Experienced in model explainability, drift monitoring, and the full ML lifecycle. Strong foundations in computer vision, NLP, and distributed ML infrastructure.

EXPERIENCE

Machine Learning Engineer II → III

Jun 2022 — Present

Glance · Bengaluru, India

- **Hybrid Recommendation Engine:** Designed and deployed a dual-track ranking pipeline that dynamically routes users based on data density — owning the Samsung channel (40M users) within a 150M-user platform. Powered the "dense" track with Gemini-enriched content embeddings, a Two-Tower retrieval model, and an LGBM ranker, while gracefully handling cold-start "sparse" users using Wilson's lower bound popularity and recency signals. Achieved a **40% lift in interactions** — notable for a locked-in lock screen product where users never actively opt in.
- **Signal Extraction & Data Pipeline:** Iterated extensively on high-volume, noisy event logs (150M+ user signals) to isolate clean, high-intent engagement signals, ensuring downstream ML models were trained on accurate behavioral data. 30-min hourly batch pipelines; sub-250ms online serving latency.
- **Infrastructure (Vertex AI → GKE):** Built online and offline serving pipelines on Vertex AI with Vertex Feature Store; migrated to GKE + Argo CD for cost efficiency; wrote Golang prediction services and model controllers; instrumented with OpenTelemetry and Grafana.
- **Experimentation (Alchemist):** Contributed to an internal A/B testing platform; derived minimum sample sizes from inference/confidence equations to ensure statistical significance before shipping ranking changes.
- **AI Annotation Setup:** POC for annotation workflows; GenAI-based image metadata tagging with LLMs and prompt engineering for category classification; drove annotation cleanup and cost reduction.

Machine Learning Engineer I

Nov 2020 — Jun 2022

Censius AI · Bangalore, India

- **Explainability Module:** Used SHAP and LIME to explain model predictions and provide insight into why models produce the outputs they do; logged and monitored using MLflow, Prometheus, WhyLogs and Grafana.
- **Drift Prediction Module:** Created APIs for continuous monitoring of model performance and production data; monitored data and concept drift with custom code and WhyLogs; built data quality, drift and performance monitors on Prefect / Airflow jobs.
- **Deployment & Infrastructure:** Containerized and deployed using Docker, GitHub Actions and AWS ECS; orchestrated workloads on Kubernetes.
- **ML Lifecycle:** Worked across the full ML lifecycle — from model onboarding and deployment to post-production monitoring — enabling clients to detect model degradation early and act on it.

Machine Learning Intern → Jr Machine Learning Engineer

Jul 2018 — Oct 2020

Omni-Eye / The Valley Edutech · Bangalore, Karnataka

- Built models for security systems for the **Omni-Eye** platform.
- **Eye In the Sky — Real-time Image Processing:** Stacked deep learning models to process images for object detection, facial recognition and plate detection. MTCNN network for facial detection; finetuned PyTorch models for facial recognition. OCR and data pipelines for plate recognition. Tracked all experiments with MLflow.
- **Image Search & Clustering:** Created, trained and finetuned a CNN autoencoder that converts unlabelled images into feature vectors; inserted into KNN and LSH (Locality Sensitive Hashing) for fast similarity search and into unsupervised clustering algorithms to group image data.
- **Student Platform & Instruction:** Developed a portal for tracking student progress (Flask, MongoDB) with GitHub commit-tracking APIs; instructed Python, ML and Deep Learning cohorts with a code-first, project-oriented approach.

NOTABLE PROJECTS

- **ResNet50 on ImageNet-1k from Scratch** ERAv4 · AWS EC2 — No pretrained weights; trained on full ImageNet-1k on EC2; 75%+ top-1 accuracy; ~10,000 people globally. HuggingFace demo.
- **YouSum — AI YouTube Summarizer** Chrome Extension — Streaming YouTube summaries via Claude & ChatGPT APIs; 5 detail levels, background generation, persistent storage.
- **YOLO Object Detection** Andrew Ng · C4 — Real-time detection for autonomous driving; bounding box prediction, IoU and non-max suppression from scratch.
- **Face Recognition with FaceNet** Andrew Ng · C4 — One-shot face verification using the FaceNet architecture and triplet loss.
- **Poetry Analysis Studio** Flask · Gemini AI — Poem analysis and generation (Haiku, Sonnet, Limerick, Free Verse) with detailed literary analysis via Google Gemini.
- **Neural Machine Translation with Attention** Andrew Ng · C5 — Seq2seq model with an attention mechanism, learning to focus on relevant input positions at each decoding step.
- **Rossmann Store Sales Prediction** Kaggle · Top 0.4% — Deep embedding network; 10% RMSPE; **11th out of 3,000+ teams**.
- **BlueBook for Bulldozers** Kaggle · Top 0.4% — Random Forest Regressor; RMSLE 0.2214; **2nd out of 476 teams**.

DOMAINS

Machine Learning · Recommendation Systems · Deep Learning · MLOps · Computer Vision · NLP · Model Monitoring

STACK

Python · Go · PyTorch · Scikit-learn · Pandas · NumPy · PySpark · Vertex AI · GKE · OpenTelemetry · Grafana · SHAP / LIME

EDUCATION

Bachelor of Engineering

B M S College of Engineering · Bangalore

2012 — 2016

CERTIFICATIONS

Deep Learning Specialization

Andrew Ng · deeplearning.ai · 5 courses

- Neural Networks & Deep Learning
- Hyperparameter Tuning, Regularization & Optimization
- Structuring Machine Learning Projects
- Convolutional Neural Networks
- Sequence Models

END2 — Extensive NLP via Deep Models

The School of AI

- Transformers, attention, BERT, GPT-1/2/3
- PyTorch for NLP, embeddings, language modelling
- Retrieval-augmented generation techniques

ERAv4 — Extensive & Reimagined AI Program

The School of AI

- LLM pretraining & instruction tuning from scratch
- QAT, RLHF, Vision-Language Models (CLIP)
- Multi-GPU CNN training on ImageNet