AKASH AGALAVE

Pune, India

📞 +91-9359372603 ✉ akash.agalave.work@gmail.com 🔗 linkedin.com/in/akash-agalave ⬡ github.com/akashagalave

EDUCATION

Dr. D.Y. Patil University   **Dec 2021 - June 2025**
B.Tech - Computer Science and Engineering | CGPA - 8.45   *Pune, India*

Maharashtra Vidyalaya Barshi   **May 2019 – June 2021**
Higher Secondary Education (Science) | Percentage - 86.85%   *Barshi, India*

EXPERIENCE

Euron   **August 2025**
AI Developer Intern   *Remote, India — Individual project-based*

- Designed and implemented AI applications by developing **machine learning models** for a medical chatbot using domain-specific **data**.
- Implemented **vector search** for fast and relevant medical information retrieval.
- Developed backend services and application **systems**, containerized using Docker to support **scalable deployment**.
- Deployed on **AWS EC2** using images from AWS ECR with CI/CD automation.

CodersCave   **May 2024 - June 2024**
Data Science Intern   *Remote, India — Individual project-based*

- Developed and trained **machine learning models** on text data, achieving 80.7% accuracy.
- Streamlined text preprocessing, tokenization, and feature extraction for training pipeline.
- Evaluated and optimized **model performance**,delivering 0.72 F1-score.

PROJECTS

AutoML Brain – ML Governance & Canary Deployment Platform | GitHub   **Jan 2026 – Feb 2026**

- Built a **cloud-native autonomous ML governance platform** on AWS EKS with reproducible pipelines and stable/canary deployments.
- Achieved **82% P95 latency reduction 2000 ms → 350 ms** through inference optimization and Kubernetes resource tuning.
- Implemented **Prometheus-driven canary promotion** using real-time P95 latency and error-rate validation before 100% traffic shift.
- Delivered an **LLM-powered audit layer** generating automated JSON/PDF reports from ROC-AUC 0.83+% , F1 0.63+% , PSI drift, and SHAP.

FraudGuard – Real-Time Fraud Detection & MLOps Platform | GitHub   **Nov 2025 – Jan 2026**

- Designed a **real-time fraud detection platform** for **highly imbalanced** financial transactions, optimized for low-latency, high-throughput inference under production constraints.
- Improved **model performance** to 94.5% PR-AUC, sustaining 1.2K RPS with P95 < 25 ms and P99 < 32 ms inference-path application latency, measured via Prometheus behind an AWS ALB.
- Built a LightGBM pipeline with advanced feature engineering and imbalance handling, ensuring reproducible training via DVC and MLflow.
- Deployed a **Kubernetes-based microservices system** on AWS EKS using FastAPI, Kafka, observability, and accuracy-triggered CI/CD retraining.

Real-Time YouTube Sentiment Analysis System | GitHub   **Mar 2025 - Apr 2025**

- Engineered a **real-time YouTube sentiment analyzer Chrome extension** using fine-tuned LightGBM with TF-IDF trigram features.
- Attained **86% accuracy** via Optuna Bayesian tuning with expanded search space over 50 trials.
- Resolved **class imbalance** using ADASYN and LightGBM class-weight, boosting model performance.
- Deployed scalable Flask API on AWS EC2 with Docker, ECR, and CI/CD via AWS CodeDeploy for real-time inference.

TECHNICAL SKILLS

Languages: Python, SQL
Technologies: Machine Learning, Deep Learning, NLP, RAG, AI Agents , Hugging Face
Frameworks & Libraries: PyTorch, Langchain, SHAP
Database: SQL,Vector Database
MLOps-Tools: Git, GitHub, DVC, MLflow, Github Actions, DagsHub, Prometheus, Grafana
Deployment & Cloud Services: FastAPI, Docker, AWS,Kubernetes (EKS)