Saikiranmansa Sunnam

Machine Learning Engineer

ssaikiranmansa2023@gmail.com 7164150744 Buffalo Linkedin Portfolio Hugging Face

SUMMARY

Machine Learning Engineer with **1+ years of industry experience** building and scaling intelligent systems across diverse domains including gig platforms, real estate, and educational technology. Proficient in deploying advanced ML models using PyTorch, TensorFlow, and Scikit-learn with seamless backend integration via FastAPI and Docker. Demonstrated expertise in NLP, computer vision, and anomaly detection, backed by hands-on MLOps experience with AWS. Recent graduate with an **M.Sc. in Computer Science** from the University at Buffalo, combining rigorous academic foundations with hands-on experience in production-level ML systems.

PROFESSIONAL EXPERIENCE

Machine Learning Engineer, *Rivach LLP* 04/2022 – 05/2023 | India

- **Designed** and implemented a geo-personalized gig recommendation engine using spatial clustering and collaborative filtering, resulting in a **42% improvement in task-user match accuracy** across the Brekrr platform.
- **Engineered** a robust identity verification system using Siamese neural networks and OpenCV, **reducing fake profile creation by 35%** and enhancing overall user trust.
- **Developed** predictive maintenance algorithms leveraging XGBoost and Random Forest, achieving **60% forecast accuracy** and decreasing equipment downtime by **28%** within the Landmanager ecosystem.
- **Spearheaded** the deployment of a content-based recommender system powered by matrix factorization techniques, leading to a **50% increase in user engagement** on Guiding Young Minds.
- **Implemented** Transformer-based NLP moderation pipelines for real-time chat analysis and policy enforcement, **cutting manual moderation overhead by 60%** while ensuring regulatory compliance.
- **Optimized** ML deployment pipelines using FastAPI and Docker within a microservices architecture, maintaining **sub-200ms latency** for high-frequency API interactions.
- **Automated** scalable inference systems using AWS Lambda and Firebase Functions, **reducing cloud compute expenses by 20%** and improving deployment agility.
- **Established** comprehensive monitoring solutions via MLflow and Prometheus, delivering **99.9% model uptime** and real-time visibility into production model health.
- **Collaborated** with cross-functional stakeholders to translate business requirements into ML-driven features, **accelerating project delivery by 25%** and improving end-user satisfaction.

TECHNICAL PROJECTS

Domain-Specific QA System Using DeepSeek and RAG

- **Built a production-grade question-answering system** leveraging **DeepSeek** and **Retrieval-Augmented Generation (RAG)**, combining **FAISS**-based dense vector retrieval and a language model to deliver **30%+ improvement in response accuracy**.
- **Processed and indexed over 1 million domain-specific documents** using **FAISS** and **Elasticsearch**, enabling efficient, low-latency (**<250ms**) query retrieval at scale.
- **Enhanced semantic search quality** by integrating **Sentence-BERT** embeddings, boosting **BLEU** and **ROUGE** scores by **15–20%**, and improving relevance in context-sensitive queries.
- **Benchmarked QA performance** using **BLEU**, **ROUGE**, and **Exact Match**, achieving **30%+ lift over baseline retrieval and generation models**, validating end-to-end model effectiveness.
- **Launched a scalable QA** pipeline on **AWS** using **FastAPI** and Docker, reducing document retrieval time by 50%.

Advanced Anomaly Detection and Text Classification Using Deep Learning

- **Engineered an anomaly detection framework** using three autoencoder variants to detect patterns in a time-series dataset of **5,315 records**, achieving a maximum **R^2 of 0.9916** and detecting **25–74 anomalies** depending on the model.
- **Developed a Transformer-based text classifier** using **PyTorch** and fine-tuned it on the **AG News dataset**, increasing classification accuracy from **90.08% to 90.53%** via **L2 regularization** and **dropout**.
- **Preprocessed and visualized 10,000+ tokens**, leveraging **tokenization**, **normalization**, and **data visualization** with **Matplotlib** to identify trends and feature importance for classification.
- **Optimized training using advanced regularization techniques**, fine-tuning dropout rates and learning rates, which resulted in a **0.45% improvement in accuracy** and **12% reduction in overfitting**.
- **Conducted comprehensive model evaluation** using **R^2**, **precision**, **recall**, **F1-score**, **confusion matrix**, and improved **ROC AUC from 0.82 to 0.91**, validating model performance and generalization.

EDUCATION

State University of New York at Buffalo, *Master of Science in Computer Science* 12/2024 | Buffalo, NY

Courses: Machine learning, Deep learning, Computer Vision & Image Processing, Operating Systems, Algorithms Analysis and Design, Data Intensive Computing, Computer Security, Data Mining and Query Language, Software Engineering

SKILLS

Core ML & AI: Generative AI, LLMs (LLaMA, GPT, BERT), RAG, NLP, Computer Vision, Anomaly Detection
Frameworks & Libraries: PyTorch, TensorFlow, Scikit-learn, Hugging Face, Keras
Programming: Python, SQL, R, Java, C
Data & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly
DevOps & Cloud: Docker, FastAPI, REST APIs, AWS (EC2, S3, Lambda, SageMaker, DynamoDB, CloudWatch), CI/CD
Big Data & Databases: Hadoop, Spark, MySQL, Oracle Database
Tools & Collaboration: Git, Jupyter Notebooks, Data Warehousing, Testing, Code Reviews, Documentation

PUBLICATIONS

Granite classification using machine learning and edge computing
Published in **F1000Research**: Link