Jayantha Nanduri

+1 (617) 992 4852 Email Linkedin Portfolio Github

Professional Summary

Machine learning engineer with 3+ years of experience streamlining workflows in Finance and Healthcare. Proficient in implementing end-to-end ML systems with expertise in LLMs, API development, and leveraging technologies such as LangChain, PyTorch, Apache Spark, Next.js, FastAPI, Docker, and AWS.

Experience

Data Science Co-op | *Boehringer Ingelheim* **Aug 2024 - Jan 2025**

Python, Streamlit, LangChain, LLMs, RAG, Next.Js, PostgreSQL, PowerBI, Docker, OpenShift, AWS *Athens, GA, USA*

- Enhanced retrieval precision by 50% by developing a RAG chatbot using OpenAI's text-ada-001 embedding model and FaissDB vector database, enabling efficient querying and analysis of large unstructured documents.
- Boosted summarization accuracy by 30% by designing a map-reduce workflow and refining prompts to split documents into batches, circumventing LLM context window constraints.
- Increased workflow efficiency by developing a JIRA-inspired asset-tracking application using Next.js, PostgreSQL, and NFC tags, enabling seamless digital monitoring of physical assets and automating manufacturing workflows.
- Accelerated development speeds by 20% by containerizing application code and database into separate services using Docker, improving modularity and streamlining workflows.
- Improved stakeholder decision-making by integrating a PowerBI dashboard into the application, delivering real-time insights into operational efficiency and KPIs such as production cycle times and resource utilization.

Machine Learning Engineer| *Jocata Financial Advisory & Technology* **Jan 2020 - Dec 2022**

Python, Django, Redis, Apache Spark, RandomForests, Object Detection, PyTorch, OpenCV, Docker, A/B Testing, AWS *Hyderabad, TS, India*

- Enabled digital disbursal of unsecured business loans (INR 2L–20L) and overdrafts up to INR 2 crore by designing and deploying SME DNA for ICICI bank, leveraging GSTN-driven underwriting.
- Improved loan-default prediction accuracy by 40% and enhanced financial risk insights by conducting A/B testing comparing Random Forest, XGBoost, and LightGBM ensembles against a logistic regression baseline.
- Optimized data processing times by building a scalable ETL pipeline using Apache Spark to ingest and process tax filing data from government APIs, enabling efficient storage in PostgreSQL and supporting real-time credit risk analytics.
- Reduced API response latency by 50% by integrating a Redis caching layer in Django, consistently handling 10,000+ daily requests.
- Improved text extraction by 70%, accurately detecting different languages in images, filtering out relevant text by training a custom Faster R-CNN model using PyTorch, MLFlow.
- Decreased Storage Footprint by 25% through Efficient Image Preprocessing using AWS Lambda and AWS Batch to perform real-time image resizing, compression, and format conversion.

Education

Northeastern University | Master of Science in Computer Science | GPA: 3.85 ***Jan 2023 - Feb 2025***

Graduate Teaching Assistant - Discrete Mathematics, Algorithms, Unsupervised Data Mining

Coursework - MLOps, Data Mining Techniques, Web Development, Parallel Data Processing, LLMs, Machine Learning, NLP

Technical Skills

Languages	Python, Java, SQL, Scala, Golang, TypeScript
API Frameworks & Libraries	NextJS, FastAPI, Django, Streamlit, Express.js, React.js, GraphQL
ML Frameworks	NumPy, Pandas, PyTorch, Scikit-learn, LangChain, HuggingFace
Development & Database	PostgreSQL, MySQL, MongoDB, Hadoop, Kafka, Redis, Airflow, Splunk
Cloud & DevOps	Git, Jenkins, Docker, DVC, OpenShift, Kubernetes, AWS

Projects

Distributed Graph Analysis | *PageRank, Python, Java, PySpark, Map Reduce, AWS*| Github

- Implemented PageRank in MapReduce and Spark for analyzing extensive web graphs, achieving a 40% reduction in processing time and improving scalability and fault tolerance.
- Employed lineage information to enhance debugging, performance optimization, and result validation.

Story Generation | *LLMs, LangChain, Llama, DistilGPT, Python, HugginFace*| Github

- Fine-tuned DistilGPT model from Hugging Face, targeted at generating coherent stories from a dataset of 300,000 entries sourced from Facebook AI research.
- Achieved a Perplexity of 22.6 and BLEU score of 0.85, complemented by evaluations using Rouge, and human assessments.