

HEMANTH KUMAR KOTTAPALLI

ML Engineer

GA | 401-216-9336 | hemanthkk2025@gmail.com

SUMMARY

Machine Learning Engineer with 3+ years of experience delivering **production-grade AI and ML solutions** across healthcare, finance, and supply chain domains. Specialized in **Generative AI, Transformer architectures, and LLM fine-tuning** using Hugging Face, TensorFlow, and PyTorch. Proficient in **MLOps** (Docker, Kubernetes, MLflow, Kubeflow, CI/CD) for **scalable, automated cloud deployments** on AWS SageMaker, GCP AI Platform, and Azure ML. Skilled in **big data processing** (Apache Spark, Dask, Kafka) and **end-to-end AI lifecycle management** from data ingestion to model monitoring. Adept at applying NLP, computer vision, and predictive analytics to drive measurable business outcomes.

SKILLS

Methodologies: SDLC, Agile, Waterfall
Programming: Python, R, Java, C++, SQL
Machine Learning & AI: TensorFlow, PyTorch, Scikit-learn, Keras, XGBoost, LightGBM, Generative AI, LLMs, Hugging Face Transformers, LangChain
Deep Learning: CNNs, RNNs, Transformers, GANs
Data Processing: Pandas, NumPy, Apache Spark, Dask, Apache Beam
Cloud Platforms: AWS (SageMaker, Lambda, EC2, S3), GCP (AI Platform, BigQuery), Azure ML
MLOps & Deployment: Docker, Kubernetes, MLflow, Kubeflow, REST APIs, CI/CD Pipelines
NLP: BERT, GPT-based models, SpaCy, NLTK, Sentiment Analysis
Big Data Technologies: Hadoop, Apache Kafka, Snowflake, Delta Lake
Visualization: Tableau, Power BI, Matplotlib, Seaborn, Excel
Databases: MySQL, PostgreSQL, MongoDB, Redis
Other Tools: Jupyter Notebook, FastAPI, Flask, Airflow, GitHub, GitLab, Bitbucket

EXPERIENCE

McKesson, GA | July 2024 – Current | ML Engineer

- Engineered and operationalized **LLM-augmented predictive analytics pipelines** leveraging Hugging Face Transformers, Scikit-learn, and XGBoost within a hybrid AWS SageMaker–EC2 environment, achieving **15% uplift in demand forecasting precision** and mitigating supply chain disruption risks by **12%**.
- Architected **multi-stage convolutional neural network ensembles** in TensorFlow, incorporating transfer learning and fine-tuned hyperparameter optimization via Bayesian search to enhance radiology image classification accuracy by **18%**.
- Designed and deployed **modular, containerized ML workflows** orchestrated via Kubernetes, with integrated **feature store abstraction layers** enabling scalable, reproducible experimentation across business units.
- Constructed **asynchronous, event-driven RESTful microservices** using Flask and AWS Lambda to deliver sub-200ms inference latency for time-critical clinical decision support applications.
- Developed **Tableau-based real-time MLOps observability dashboards** integrating model drift detection, concept drift alerts, and automated retraining triggers via Airflow.

Infinite Infolab, India | Jan 2021 - July 2023 | ML Engineer

- Designed and productionized **Transformer-based NLP architectures** (BERT, RoBERTa) with domain-adapted tokenizers and knowledge distillation pipelines, realizing a **12% improvement** in sentiment analysis accuracy for high-volume social data streams.
- Implemented **multi-cluster Kubernetes deployments** for ML microservices using Helm charts, enabling elastic scaling of inference workloads under variable traffic loads.
- Leveraged **Apache Spark Structured Streaming** and Dask for fault-tolerant, distributed ETL of **multi-terabyte datasets**, optimizing shuffle-intensive operations via Catalyst optimizer tuning.
- Integrated **MLflow tracking servers with model registry workflows** to establish end-to-end reproducibility, lineage tracking, and automated deployment approvals, reducing model release cycles by **25%**.
- Developed **LSTM and GRU hybrid time-series forecasting frameworks** in Keras, incorporating attention mechanisms for enhanced predictive interpretability in financial datasets.
- Optimized **BigQuery analytical workloads** with partitioned and clustered tables, reducing query runtime by **40%** for iterative model validation processes.

EDUCATION

Master of Science in Business Analytics Mercer University – Atlanta, GA	Dec 2024
Master of Science in Advanced Computer Science University of Leicester – Leicester, UK	Nov 2022
Bachelor's degree in Computer Science Andhra University	April 2017