MENTRAJU MEESALA

AI/ML Engineer

USA| (901) 206-2519 | raju.meesala0218@gmail.com | LinkedIn | https://github.com/RajuMeesala11

SUMMARY

AI/ML Engineer with 6+ years of progressive experience spanning data engineering, machine learning, and enterprise-scale deep learning solutions. Adept at building and deploying NLP, computer vision, and predictive models that enhance business operations, reduce costs, and improve accuracy. Skilled in designing end-to-end MLOps workflows using tools like MLflow, SageMaker Pipelines, Docker, and Airflow to enable scalable, reliable machine learning systems. Experienced in multi-cloud deployments (AWS, GCP, Azure) and integrating AI models into production systems through REST APIs. Passionate about delivering measurable business impact through automation, data-driven insights, and real-time analytics dashboards using Power BI and Tableau. Thrives in Agile/Scrum environments with a focus on collaboration, performance, and continuous improvement.

SKILLS

Programming & Query Languages: Python, SQL, Java, R

Data Engineering & Warehousing: Airflow (ETL), NumPy, Pandas, Data Modeling, Data Quality

ML & Deep-Learning Frameworks: TensorFlow, PyTorch, scikit-learn, Keras, CNNs, RNNs, LSTM, GANs

GenAI & LLM Stack: Hugging Face Transformers, Accelerate, LoRA/PEFT, RLHF (TRL/PPO), LangChain, LlamaIndex

Experiment Tracking & MLOps: Weights & Biases, MLflow, SageMaker Pipelines, Kubeflow, Ray

Model Serving & Observability: NVIDIA Triton, TorchServe, KServe, Prometheus, Grafana

Cloud & DevOps: AWS (EC2, Lambda, S3), Azure, GCP; Docker, Kubernetes, Terraform, Git, CI/CD

Visualization & BI: Power BI, Tableau, Matplotlib, Seaborn

Tools & IDEs: JupyterLab, PyCharm, VS Code, Eclipse

Methodologies: Agile/Scrum, SDLC, Waterfall, A/B Testing

EDUCATION

Master of Science, Artificial Intelligence, 3.4 GPA **March 2025**

Depaul University, Chicago, IL

EXPERIENCE

ServiceNow, USA | Sep 2024 – Current | AI/ML Engineer

- Fine-tuned BERT models on 1.2M labeled ITSM tickets, improving auto-triage F1-score by 37% over baseline; deployed via AWS SageMaker Pipelines, contributing to a 22% reduction in average ticket resolution time (from 3.1 to 2.4 hours).
- Built an MLOps workflow using MLflow, Lambda, and Docker to trigger model retraining on drift signals, reducing manual intervention in model updates by ~90%.
- Integrated the model with ServiceNow REST APIs, enabling inference on 100k+ tickets per day with median latency under 200 ms.
- Developed Power BI dashboards for real-time SLA risk monitoring, supporting operational changes that contributed to $1.8M in annual support cost savings.

Renault Nissan Technologies and Business Centre | Oct 19 – Nov 22 | ML Engineer L1 | Project – BOM STS

- Designed and optimized CNN, RNN/LSTM models to improve predictive accuracy by 28% on BOM datasets; explored GANs for generating synthetic BOM data to augment training sets.
- Refactored ETL workflows processing 500k+ BOM records using Airflow orchestration and Pandas optimizations, reducing data load time by 40%.
- Contributed to Agile/Scrum practices that helped shorten sprint cycles by ~22%, accelerating release cadence.
- Executed hyperparameter sweeps with PyTorch/TensorFlow, enhancing model performance by 20% over initial benchmarks.
- Led feature engineering and k-fold validation workflows, resulting in a 25% improvement in model accuracy over initial baselines.

- Created Matplotlib visualizations that improved stakeholder understanding of model insights and predictions.
- Established data-validation checklists that helped reduce downstream error rates and improved data quality.

Capgemini India Private Limited, India | Nov 16 – Oct 19 | Data Engineer | Project – Global Payments CMF and BV

- Engineered SQL and Airflow pipelines that supported reliable data processing across 50+ global payment systems, achieving high data reliability and availability.
- Standardized data-cleaning protocols and quality checks, improving audit readiness and reducing reliance on spreadsheet-driven processes.
- Developed Tableau dashboards for real-time monitoring of financial operations, enabling faster decision-making and reducing decision latency.
- Built fraud-detection models using Logistic Regression, Decision Trees, and Random Forests, improving anomaly detection rates over prior rule-based approaches.
- Performed A/B testing and hypothesis analysis on payment flows, providing insights that informed region-specific strategy shifts.

ACADEMIC PROJECTS

Capstone Project: Microclimate Impact Analysis of Data Centers

- Orchestrated a large-scale data center impact study, analyzing energy consumption, emissions, and operational efficiency.
- Scraped and curated data from over 3,000 organizations, creating a robust dataset for predictive modeling.
- Engineered ML models to forecast energy needs and recommend operational strategies, driving a 20% reduction in energy costs.
- Leveraged Python, Scrapy, Pandas, NumPy, Scikit-learn, and Matplotlib to transform raw data into actionable insights.

RAG Application for Medical Domain

- Formulated predictive models to analyze clinical data and forecast patient treatment outcomes, enabling data-driven decision making in healthcare.
- Integrated reinforcement learning within a retrieval-augmented generation framework to optimize diagnosis suggestions, achieving a 32% improvement in clinical decision accuracy.

Reinforcement Learning for Highway Environment, Snake and Atari

- Refined reward structures and hyperparameters across multiple environments, accelerating agent convergence by 30%.
- Enhanced policy iteration processes through iterative experimentation, boosting overall agent performance by 25%

Pneumonia Detection

- Developed a CNN-based pneumonia detection system using OpenCV, employing advanced normalization and pooling techniques to boost classification accuracy by 30%.
- Constructed an end-to-end preprocessing and feature extraction pipeline that reduced image processing time by 25%, fortifying model robustness.

AI Campus Assistant Agent

- Designed and deployed a modular multi-agent system using LangChain, OpenAI LLMs, and FAISS for personalized student guidance via natural language interface.
- Implemented event extraction, shortest-path campus navigation (A* algorithm), and a real-time recommendation engine using RAG pipelines and metadata filtering.

CERTIFICATION

- AWS Certified AI Practitioner
- AI/ML Engineer – Great Learnings
- ML Engineer – Stanford Online (Andre NG's Program)