Vaishnavi Mocherla

San Jose, CA | +1 (530) 407-3384 | vmocherla25@gmail.com | LinkedIn | Github

Education

San Jose State University — San Jose, CA
M.S. in Applied Data Science — *Aug 2023 – May 2025*

Experience

Informatica — California, USA
Machine Learning Engineer — *May 2024 – Dec 2024*

- Developed an **AI-powered analytics agent** using **Azure OpenAI** and **Python**, enabling **natural language to SQL** query translation, reducing analyst workload by **60%**, and improving query accuracy by **20%** through advanced prompt tuning and schema linking.
- Built scalable **Databricks** pipelines with **PySpark + SQL** for real-time and batch model inference, optimizing processing speeds by **35%** and ensuring production-grade reliability through **containerized CI/CD workflows**.
- Configured and managed **EMR Clusters** on **AWS** to run large-scale Spark jobs for ML data preprocessing, achieving **30% faster** job completion and significant cost savings via **spot instance optimization**.
- Integrated **MLflow** and **Airflow** for automated training, model versioning, and deployment, supporting continuous delivery and reproducibility of models in a cloud-native environment.
- Implemented enterprise-grade data governance, enforcing **row-level security** and **data masking** for a **1TB+ Azure Data Lake** migration to ensure compliance during AI model deployments.
- Created interactive **Power BI dashboards** to monitor ML pipeline performance and LLM usage metrics, providing actionable insights for marketing and operations teams.

Kanerika Software Pvt Ltd — Hyderabad, India
Machine Learning Engineer I — *Aug 2022 – Jun 2023*

- Designed and deployed a **BERT-based NLP system** to assess resume–JD fit and predict candidate progression, automating resume screening for **100K+ resumes** and reducing manual review workload by **60%**.
- Developed a robust **custom parser** handling 10+ document formats (**PDF, DOCX, scanned text**), increasing parsing accuracy by **70%** and enabling reliable ML feature extraction for downstream models.
- Built a real-time prediction pipeline using **Airflow**, **Spark**, and **Python**, storing results in an **Azure Data Lake** and visualizing model confidence and fit scores through **Power BI** dashboards, improving screening turnaround by **30%**.
- Implemented **feature engineering** and prompt-tuning techniques to improve model classification accuracy, with continuous feedback loops integrated for iterative model improvement.
- Presented working prototype to stakeholders, securing approval for beta deployment and integration into ATS workflows.

Projects

Multi-Agent LLM System for Autonomous SDLC | *LangChain, Semantic Kernel, Pinecone, Llama 3, Snowflake*

- Built a production-ready LLM system with **6 specialized agents** (PM, SE, DE, QA, Deployer, Docs) simulating the entire SDLC, from requirements to deployment.
- Orchestrated agents using **LangChain**, **Semantic Kernel**, and **Pinecone**-powered **RAG** pipelines for long-horizon memory and context retrieval.
- Achieved **less than 2s latency**, 88% QA validation accuracy, and 40% hallucination reduction; automated CI/CD with GitHub for reliable releases.

Real-Time Anomaly Detection Pipeline for IoT Sensor Data | *PySpark, Structured Streaming, Airflow, MLflow, Databricks*

- Built a real-time streaming pipeline using **PySpark Structured Streaming** on **Databricks** to process IoT sensor data (temperature, vibration) and detect anomalies with **Isolation Forest**, achieving **95%+ detection precision**.
- Orchestrated workflows with **Airflow** and tracked model performance with **MLflow**, enabling automated retraining, drift monitoring, and alerting dashboards for live anomaly tracking.

Technical Skills

Languages: Python, SQL, R (Basic)

ML Frameworks: PyTorch, TensorFlow, scikit-learn, spaCy, Transformers, Hugging Face

NLP & LLMs: BERT, LSTM, Gemma, LangChain, QLoRA, Semantic Kernel, Prompt Engineering, RAG Pipelines, Pinecone, Vector Search

MLOps & Orchestration: MLflow, Airflow, Docker, Kubernetes, CI/CD

Cloud Platforms: Azure, AWS (EMR, EC2, S3), GCP, Databricks

Databases: PostgreSQL, MongoDB, SQL Server, HDFS

Big Data Processing: PySpark, Spark, Databricks

Visualization: Power BI, Tableau, Matplotlib, Seaborn