

Narendra Meenuga

AI/ML Engineer

☎ +1 (475) 332-5025

✉ narendra.meenuga97@gmail.com



Professional Summary

ML Engineer with 6+ years of experience building and deploying production-grade AI/ML Platforms. I specialize in developing and deploying Enterprise-Grade domain-specific AI/ML Systems on cloud platforms and edge devices. Proficiency in building data pipelines, preparing structured inputs for ML/AI models. Expertise in ideating, implementing and deploying data-driven products by incorporating software engineering best practices, design patterns, and frameworks.

Technical Skills

Programming Languages	Python, SQL, R, Java
Frameworks & Tools	PyTorch, LangChain, LangGraph, JAX, PaLM, CrewAI, Stable Diffusion, NeMo, NVIDIA Nim, FastAPI, Docling, Ollama, Agentic AI.
Data Visualization	Tableau, Power BI
LLMs	Claude sonnet(on AWS), Llama, T5, BLOOM, Bard, RoBERTa, BERT
AI/Machine Learning	Fine-tuning LLMs, Quantization, Prompt Engineering, RAG,CT-GANs, Clustering, Classification, Ensemble Models(XGboost, AdaBoost and LightGBM), Predictive Modeling, NLTK, BERT, Text Generation, Topic Modeling, Transformers
Cloud Platforms	AWS (S3, EMR, Redshift, SageMaker, Bedrock), Azure (Databricks, Synapse Analytics, AI studio), GCP (BigQuery, Vertex AI)
MLOps	MLflow, Kubeflow, Model Monitoring, Model Deployment
Databases	Oracle, MySQL, PostgreSQL, SQL Server, MongoDB, Cassandra, vector DB(pgvector & Pinecone), GraphQL
Version Control	Git, Bitbucket, Git lab
Big Data Technologies	Apache Spark, Kafka, ETL/ELT, Airflow, Databricks, Pyspark
Other Technical Skills	Docker, Kubernetes, Jenkins, Terraform & CloudFormation, shell Scripting, PowerShell

Professional Experience

ML Engineer

CapitalOne, NY, USA
March 2025 – Present

Responsibilities

- Implemented an automated testing framework for Docker image releases for the CFRM converged offering, standardizing image validation and reducing manual release effort.
- Built an internal Python library that converts Jupyter notebooks into Kubeflow Pipelines and submits runs with a single command, integrating it into the image-testing repo and significantly reducing manual modularization work for data scientists, enabling much faster experimentation.
- Created a highly customized integration-testing script using the Behave framework to validate each pipeline component against different images (single or mutiple runs), catching image and dependency issues early.
- Implemented end-to-end model reproducibility and scalability tests using Kubeflow Pipelines, enabling automated validation of ML workloads across environments.
- Automated image validation pipelines in Jenkins (with CodeGenie integration), integrating build, test, and security checks into a single Mopl workflow.
- Built custom Docker images for SynapseML LightGBM workloads on Spark clusters, optimized for performance, reproducibility, and cluster compatibility.
- Developed Recursive Feature Elimination (RFE) and Hyperparameter Optimization (HPO) solutions for LightGBM on Spark, improving model accuracy while controlling training cost. Contributed these new features to mlkit..

- Configured service accounts and role mappings for Jenkins, onboarding multiple repositories onto the Mopl platform with secure, least-privilege access patterns and performed COS to AWS S3 migration for existing repos during migration.
- Implemented vulnerability-remediation workflows for Python and Java dependencies in SynapseML Docker images, improving security posture of production images.
- Executed continuous vulnerability scanning, patching, and release of SynapseML images as part of the container hardening process.
- Performed ongoing Spark cluster and platform management (upgrades, secrets rotation, tuning), driving operational efficiency and reliability of core ML services.

Tech Stack: Kubeflow, Docker, Kubernetes, AWS, Pyspark, Behave, Python, Vulnerability Management, Shell Scripting.

Generative AI Engineer

Symbotic Inc, MA, USA

Feb 2024 – March 2025

Responsibilities

- Engineered robust ML risk management frameworks using **XGBoost** and scikit-learn, implementing Kubernetes-optimized distributed computing pipelines that improved model throughput by 40%.
- Established enterprise-wide model governance through comprehensive documentation protocols and CI/CD-integrated testing harnesses, reducing production incidents by 32%.
- Designed automated validation workflows leveraging Python-based statistical testing libraries, achieving 99.7% benchmark compliance across risk prediction models and optimized model performance through feature engineering and **hyperparameter tuning**, deploying containerized solutions via Kubernetes that reduced inference latency by 58%.
- Implemented risk mitigation strategies using ensemble techniques and **SHAP analysis**, enhancing model stability metrics by 27% across credit risk portfolios.
- Developed distributed data processing pipelines using Dask and Spark, enabling real-time feature generation for high-volume (1M+ TPS) fraud detection systems.
- Evaluated **Autogen**, **CrewAI**, and **LangGraph** agentic frameworks through rigorous testing to select optimal solutions for future projects.
- Built a production-scale knowledge base combining LightRAG and LangGraph for hybrid graph + vector retrieval, using Claude 3.5 Sonnet (with Haiku for cost-sensitive paths) as the reasoning backbone. Consistently hit top-percentile scores on our internal GraphRAG benchmarks, beating pure vector RAG by 35–48% on tricky multi-hop questions. Designed and shipped a multi-modal ingestion pipeline that actually works reliably.
- Docling for best-in-class parsing of PDFs, DOCX, and HTML with proper layout and table preservation, automatic knowledge graph extraction into Neo4j, embeddings stored in PostgreSQL(pgvector), and raw documents safely archived in S3.
- Performed LLM evaluations DeepEval and G-Eval.

Tech Stack: Pytorch, Scikit-learn, LangGraph, CrewAI, Autogen, DeepEval, XgBoost, Spark, MoE, RLHF, Generative AI, vLLM, Docling, Vector Databases.

AI/ML Engineer

Synergy Technologies, IN

Jan 2021 – Dec 2022

Responsibilities

- Boosted customer engagement by 30% with a state-of-the-art **NLP sentiment analysis** platform delivering targeted, **A/B-tested** recommendations; built using Hugging Face transformers (**BERT**, **RoBERTa**) with spaCy/PyTorch and deployed on Kubernetes via automated CI/CD pipelines.
- Shortened the model development and deployment lifecycle by 20%; implemented CI/CD automation on Azure Cloud platform to enable cross-functional teams to rapidly deploy reinforcement learning-based anomaly detection systems.
- Utilized **MLOps** tools like **MLflow**, **Kubeflow** integrated with **Azure Databricks** for model packaging, deployment, monitoring, and lifecycle management of Machine Learning models.
- Leveraged supervised, unsupervised, and reinforcement learning algorithms to develop predictive models, recommendation systems, and anomaly detection solutions, using libraries like scikit-learn, TensorFlow, and PyTorch.

- Collaborated closely with cross-functional teams to gather requirements, prototype solutions, and deploy AI systems into production, aligning with business objectives and user needs.

Tech Stack: Scikit-learn, Hugging Face Transformers, PyTorch, Azure, Databricks, SQL, Python, Pandas, NumPy, NLTK, SpaCy, Docker, Kubernetes, CI/CD Pipelines

Data Scientist/ML engineer

Cognizant Technology Solutions, IN

Jan 2019 – Dec 2020

Responsibilities

- Boosted predictive model accuracy by 22% for customer churn prediction using XGBoost, LightGBM, and hyperparameter tuning (Bayesian optimization) while engineering high-impact features from Spark-processed datasets.
- Developed predictive models using supervised learning algorithms such as linear regression, logistic regression, decision trees, random forests, and gradient boosting machines (GBMs), achieving accuracy and performance improvements through feature engineering and model optimization.
- Employed unsupervised learning techniques, including clustering, dimensionality reduction, and anomaly detection, to identify patterns and relationships within data, uncovering hidden insights and opportunities for optimization.
- Employed cross-validation strategies such as k-fold cross-validation to assess model performance robustly, evaluating metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve.
- Built scalable data pipelines and workflows using technologies like Apache Spark and Apache Airflow, orchestrating data processing tasks and ensuring reliability and efficiency in data processing and analysis.
- Designed and Implemented data processing workflows, optimizing performance, and ensuring data consistency and reliability Using Apache Spark and Apache Kafka for processing large-scale data streams and managing distributed data processing pipelines.

Tech Stack: TensorFlow, PyTorch, Keras, Scikit-Learn, Apache Spark, Apache Kafka, AWS, SQL, Python, Pandas, NumPy

Education

Master of Science in Computer Science

University of Bridgeport, Bridgeport, CT

Bachelors in Electronics and Communication Engineering

Lovely Professional University, Punjab, IN

Certifications

- Azure AI Engineer Associate - Certified
- AWS Certified Cloud Practitioner - Certified
- Certified Ethical Hacker (CEH) - Certified