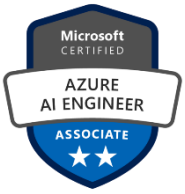
AI/ML Engineer
Lokeshwar Kodipunjula
Email: lokeshwarkodipunjula@gmail.com
Contact: +1 6823901628 | LinkedIn | GitHub



Professional Summary:

- AI/ML Engineer with 4 years of experience designing, developing, and deploying scalable machine learning and NLP solutions across insurance, healthcare, and fintech domains.
- Proven expertise in building supervised, unsupervised, and deep learning models using Python, scikit-learn, XGBoost, TensorFlow, and Keras to solve complex business problems such as fraud detection, patient risk prediction, and document classification.
- Strong hands-on experience with NLP frameworks like spaCy, Transformers (Hugging Face), and NLTK for extracting insights from unstructured text data in clinical, customer service, and financial domains.
- Developed and optimized real-time and batch data pipelines using PySpark, Apache Kafka, and Apache Flink to support high-throughput inference systems and feature engineering at scale.
- Skilled in deploying ML models as RESTful microservices using FastAPI and Docker, orchestrated via Kubernetes (AWS EKS) for production-grade resilience and scalability.
- Experienced in implementing robust MLOps practices using MLflow, Airflow, CI/CD pipelines, and model monitoring stacks (Prometheus, Grafana) to ensure model traceability, reproducibility, and performance.
- Demonstrated ability to integrate model outputs with BI tools (Power BI, Tableau) and backend systems, enabling real-time decision support and actionable insights for business stakeholders.
- Deep understanding of regulatory and compliance frameworks (HIPAA, RBI) with hands-on implementation of security best practices using TLS encryption, HashiCorp Vault, and audit logging.
- Worked in Agile/Scrum teams alongside data scientists, DevOps, product managers, and domain experts to deliver and deploy end-to-end AI solutions.

Technical Skills:

Programming Languages	Python, SQL, R, Java, JavaScript, Scala
ML Libraries	scikit-learn, XGBoost, LightGBM
Deep Learning	TensorFlow, Keras, PyTorch, Autoencoders, CNNs, Transfer Learning, Vision Models, Multi-modal Learning, High-content Imaging Analysis
LLMs, NLP & Model Explainability	spaCy, Transformers (Hugging Face), NLTK, SHAP, LangChain, OpenAI API, Embeddings, Knowledge Graphs, Semantic Layers, RAG pipelines
Data Engineering	PySpark, pandas, NumPy, SQLAlchemy, Apache Kafka, Apache Flink, Databricks
ETL & Workflow	Apache Airflow, LangGraph (Agentic Workflow Design)
Model Tracking & MLOps	MLflow, CI/CD, Git, Bitbucket, LLMOps frameworks
APIs & Web Frameworks	FastAPI, RESTful APIs
Cloud Platforms	AWS (S3, EC2, Lambda, EKS), Azure
Containerization	Docker, Kubernetes (K8s), AWS EKS
Monitoring & Observability	Prometheus, Grafana
BI & Reporting	Power BI, Tableau, Jupyter Notebook
Security & Compliance	HIPAA, RBI guidelines, HashiCorp Vault, TLS encryption
Development Methodologies	Agile, Scrum
Version Control & Tools	Git, Bitbucket, JIRA, Confluence

Professional Experience:

AIG, NY **November 2024 – Till Date**
AI-ML Engineer
Project: SentinelAI
Responsibilities:

- Designed and implemented supervised and unsupervised ML models using Python, scikit-learn, and XGBoost to predict claim fraud likelihood, reducing false positives by 23%.
- Designed and implemented deep learning pipelines using TensorFlow and Keras for document classification and anomaly detection, enabling automated review of scanned insurance claim forms.
- Developed NLP pipelines using spaCy and Transformers (Hugging Face) to perform entity recognition and sentiment analysis from customer service transcripts, improving case triage efficiency.
- Built real-time data pipelines with Apache Kafka and batch ETL workflows using PySpark to preprocess and engineer policy and claim data features for model training and inference.

- Orchestrated training, validation, and deployment workflows in Apache Airflow to ensure continuous and traceable ML lifecycle management across environments.
- Containerized model microservices using Docker and deployed scalable inference endpoints on AWS EKS through Kubernetes, reducing latency in fraud scoring pipelines.
- Leveraged MLflow for model versioning, experiment tracking, and deployment lifecycle control, supporting reproducible results and collaborative model governance across the ML team.

Environment: Python, scikit-learn, TensorFlow, Keras, PySpark, Apache Kafka, Airflow, Docker, Kubernetes, AWS EKS, MLflow, Power BI, Prometheus, Grafana, Agile.

VNS Health, NY **January 2024 – October 2024**
Python Data Scientist
Project: Clinical Insight Engine – NLP & ML for Patient Risk Prediction
Responsibilities:

- Designed and implemented NLP pipelines using spaCy and scikit-learn to extract clinical entities and sentiment from provider notes, enabling integration with structured risk models.
- Built and validated patient risk prediction models using XGBoost, Random Forest, and Logistic Regression on top of feature sets engineered from time-series vitals, lab results, and NLP-derived features.
- Preprocessed large-scale EHR datasets (HL7/C-CDA formats) using pandas, NumPy, and SQLAlchemy, ensuring data consistency and quality for modeling workflows.
- Developed and maintained automated model training pipelines using Airflow and version-controlled experimentation with MLflow, ensuring reproducibility across clinical releases.
- Tuned and evaluated model performance using stratified cross-validation, SHAP for explainability, and AUROC/F1 metrics to support clinical adoption and trust.
- Deployed models into a clinical decision support interface using FastAPI and Docker, ensuring scalable inference in production aligned with HIPAA compliance.
- Collaborated with clinical informaticists and data engineers in Agile sprints, documenting findings and presenting model behavior using Jupyter, Markdown, and Tableau dashboards.

Environment: Python 3.10, Jupyter, pandas, NumPy, scikit-learn, spaCy, XGBoost, SQLAlchemy, PostgreSQL, MLflow, Airflow, Docker, FastAPI, SHAP, Git, Tableau, Agile (Scrum), HIPAA-compliant cloud (Azure).

Axis Bank, India **July 2021 – August 2023**
Jr. ML Engineer
Project: Real-Time Transaction Fraud Detection Engine
Responsibilities:

- Built and deployed supervised ML models using Python, scikit-learn, and XGBoost on large-scale financial data to detect fraudulent activity and enhance real-time risk scoring accuracy.
- Designed streaming and batch data pipelines with Apache Kafka, Apache Flink, and Redis to support high-throughput transaction monitoring and ensure sub-second inference latency.
- Engineered data features with PySpark and SQL by deriving behaviour-based indicators like merchant risk, transaction velocity, and device fingerprinting to improve model interpretability.
- Implemented automated retraining, version control, and drift detection pipelines using MLflow, Airflow, and Docker within a CI/CD framework to maintain continuous model reliability.
- Collaborated with security and compliance teams to align model deployment with RBI governance standards, ensuring data encryption and audit trail consistency across production systems.
- Deployed scalable ML inference services using FastAPI, Docker, and Kubernetes, with Prometheus and Grafana used to monitor model precision, recall, and operational performance in real time.

Environment: Agile, Scrum, AWS (S3, EC2, Lambda, EKS), Docker, Kubernetes, Airflow, CI/CD, JIRA, Bitbucket.

Education:

Master's: University of Texas at Arlington - Arlington, Texas, United States
Bachelor's: Kingston Engineering College - Vellore, Tamil nadu, India

Certification:

- Microsoft Certified: Azure AI Engineer Associate