GOWTHAMI CHIKKA

ML Engineer |Data Engineer
(405) 430-9264|chikka.gowthami99@gmail.com|LinkedIn

PROFESSIONAL SUMMARY

AI/ML Engineer with 4+ years of experience building and deploying machine learning solutions at scale across financial, healthcare, and IT domains. Skilled in developing **recommendation systems, search ranking algorithms, and personalization models** using deep learning, NLP, and optimization techniques. Experienced in tackling **cold-start and sparse data problems** by applying foundation models, large language models, and generative AI. Strong foundation in **algorithms, data structures, numerical optimization, and distributed computing**, with hands-on expertise in Python and working knowledge of Java and C++. Proficient in end-to-end ML lifecycle: data preprocessing, feature engineering, model training, hyperparameter tuning, and real-time inference. Adept at scaling solutions on **AWS (SageMaker, EC2, S3, Redshift, Glue), Azure ML, and Databricks**, and deploying production-grade ML models via Docker, Kubernetes, and REST APIs. Collaborates closely with product managers, scientists, and engineers to design experiments, deliver business insights, and drive roadmap decisions, while staying current with research and contributing to publishable work in machine learning and personalization, including research on model selection frameworks built from 300+ academic papers at the University of Oklahoma.

TECHNICAL SKILLS

Programming & Scripting: Python, PySpark, SQL, Pandas, NumPy, Java, R,C++ (working knowledge)
Databases & Querying: PostgreSQL, NoSQL (MongoDB, Cosmos DB), Snowflake, Redshift
Machine Learning & AI: Scikit-learn, TensorFlow, NLP (spaCy, NLTK), Sentiment Analysis, Recommendation Systems, Personalization, Search Ranking, Cold-Start Modeling, Online Learning, Numerical Optimization,Text Classification, Anomaly Detection, Time-Series Modeling, Hyperparameter Tuning (GridSearchCV, Random Search, Cross-validation), Fraud Detection Models, LLMs, Generative AI,
Cloud Platforms & Services: AWS (SageMaker, EC2, S3, Glue, Redshift), Azure (Data Factory, Synapse Analytics), Databricks
Data Engineering & Pipelines: Apache Airflow, ETL/ELT Workflows, Data Ingestion, Data Preprocessing, Data Transformation, MLOps Pipelines, MLflow, GitHub Actions
Deployment & Microservices: Docker, Kubernetes, REST APIs, Microservices Integration
Visualization & Reporting: Tableau, Power BI, A/B Testing, KPI Reporting, Executive Dashboards
Project & Collaboration Tools: Jira, Confluence, Git, Agile/Scrum
Compliance & Domain Knowledge: FINRA, SOX Regulations (Financial Sector), HIPAA, Risk Analytics, Customer Support Automation (IT Sector), Fraud Detection, Algorithms, Data Structures, Distributed Computing

EXPERIENCE

Machine Learning Engineer | Northern Trust, USA **Feb 2025 - Present**

- Developed and deployed ML models for **real-time fraud detection** in payment transactions using Python, PySpark, Scikit-learn, and TensorFlow, **reducing false positives by 22% and cutting investigation time by 18%**.
- Built automated data pipelines with AWS Glue, S3, Redshift, and Apache Airflow to process **1TB+ of daily financial data**, accelerating model training cycles by **40%**.
- Implemented anomaly detection algorithms with NLP and time-series modeling to flag irregular trading patterns, preventing potential losses of **$3M+ annually**.
- Integrated ML models into a microservices architecture (Docker, Kubernetes, REST APIs), enabling **99.9% uptime** and secure, scalable deployment.
- Collaborated with risk teams to deliver executive dashboards in Tableau and Power BI, providing **real-time compliance visibility** across **100+ fraud KPIs** aligned with FINRA & SOX.
- Designed recommendation-style ranking models for fraud likelihood scoring, leveraging optimization methods to prioritize high-risk transactions (parallel to personalization/search ranking systems).
- Collaborated with risk and compliance teams to run A/B tests, analyze user/system behavior, and address cold-start challenges when new transaction types emerged.

AI Data Engineer| Streebo Inc, India **Sep 2019 - Dec 2022**

- Designed and implemented end-to-end ML pipelines to automate support ticket classification, **boosting resolution accuracy by 30%** and reducing average handling time by **25%**.
- Developed sentiment analysis and text classification models deployed on AWS SageMaker, improving customer satisfaction scores by **20%**.
- Built data preprocessing workflows (SQL, Pandas, NumPy) to process **millions of raw logs daily**, improving model readiness and reducing manual data prep time by **60%**.
- Integrated ML models into RESTful APIs and microservices (Docker, Kubernetes), enabling adoption across **10+**

enterprise apps with seamless performance.

- Built early-stage personalization models to recommend resolution templates based on historical customer service interactions, addressing sparse-data cold-start cases.
- Applied ranking algorithms and online learning techniques to continuously improve recommendation accuracy as new support queries appeared.
- Tuned models with GridSearchCV, Random Search, and cross-validation, improving predictive accuracy by **12%** over baseline.
- Set up Airflow + MLflow pipelines to monitor model drift and trigger automated retraining, cutting downtime from weeks to **under 24 hours**.
- Conducted A/B testing with business teams, producing **Tableau/Power BI dashboards adopted by 50+ stakeholders** to drive customer service decisions.

EDUCATION

Masters in Data Science and Analytics| University of Oklahoma, Norman, Oklahoma **Aug 2023 – Dec 2024**

Conducted research on ML model selection frameworks, synthesizing insights from 300+ academic papers, preparing manuscript for publication in data science/ML venues.

Bachelors in Mechanical Engineering| G Pulla Reddy Engineering College, Andhra Pradesh **Sep 2016 – Sep 2020**

PROJECTS

Healthcare AI :Fibroid Detection Pipeline

Built a scalable ML pipeline on Azure to predict fibroid occurrence from ultrasound data; improved recall by 15% while ensuring HIPAA compliance.

- Designed and trained a deep learning model (PyTorch) using MRI imaging + patient EHR data to predict fibroid growth.
- Built preprocessing pipelines in Python (Pandas, NumPy) for structured/unstructured data cleaning and feature engineering.
- Applied ranking and optimization techniques to prioritize treatment recommendations, aligning with personalization workflows
- Deployed end-to-end pipeline on Azure ML, storing medical images in Azure Blob Storage with CI/CD automation via Azure DevOps for continuous retraining and monitoring.
- Delivered explainable AI insights (Grad-CAM visualizations), supporting clinicians in treatment planning and improving decision transparency.

Financial Risk Analytics : Fraud Detection

Designed and deployed real-time fraud detection pipelines on AWS, cutting false positives by 22% and ensuring SOX/FINRA compliance.

- Built a credit risk scoring system with logistic regression, XGBoost, and ensemble models using historical financial datasets.
- Developed scalable ETL workflows (Python + SQL) for data ingestion, cleaning, and transformation from multiple sources.
- Implemented on AWS SageMaker, with real-time scoring endpoints, retraining automation, and monitoring via AWS CloudWatch.
- Designed scoring logic analogous to recommendation ranking systems, prioritizing high-risk transactions in real time transferable to content personalization and search ranking.
- Improved risk detection accuracy by 15% while ensuring compliance with Basel III and financial governance standards.

RAG + LLM Agent :Enterprise Knowledge Retrieval

Built a domain-tuned RAG agent with LangChain + LLaMA 3, enabling natural-language enterprise search and achieving 40% faster knowledge retrieval. Collaborated with OU Systems Realization Lab on machine learning model selection research, designing structured frameworks for dataset–model alignment; work positioned for conference publication. explored foundation models for enterprise retrieval and personalization at scale.

- Developed a domain-tuned RAG agent with LangChain + LLaMA 3, delivering personalized search and retrieval recommendations.
- Created custom embeddings and fine-tuned LLaMA for domain-specific documents, improving retrieval precision over standard LLMs.
- Applied ranking optimization methods to surface the most relevant documents, addressing cold-start and sparse query scenarios.
- Deployed with FastAPI + Docker, orchestrated using Kubernetes for scalable multi-user access.
- Benchmarked retrieval quality improvements like recommendation system evaluation (precision, recall), achieving 40% faster content discovery