Manichandra Reddy Bethi

Location: Pennsylvania | **Mail:** bethimanichandrareddy@gmail.com | **Ph.:** +1 (913)-428-9647 | LinkedIn

PROFESSIONAL SUMMARY:

- Data Scientist with around of 5 years of experience delivering scalable data and AI solutions across banking, healthcare, and advanced AI-driven platforms.
- Proven expertise in building production-grade data architectures and machine learning systems in regulated, high-impact environments.
- Strong ability to transform complex, high-volume datasets into reliable data assets that support risk management, compliance, and business decision-making.
- Experienced in end-to-end ownership of data and AI lifecycles, from ingestion and engineering to deployment, monitoring, and optimization.
- Recognized for collaborating effectively with cross-functional teams to translate business needs into high-performance, automation-driven technical solutions.

TECHNICAL SKILLS:

Programming & Scripting:	Python (NumPy, Pandas, scikit-learn, PySpark), R (GLM, Survival Models, caret), SQL
Machine Learning & AI:	Gradient Boosting (XGBoost, LightGBM, CatBoost), Random Forest, GLM, Survival Analysis, Churn Prediction, Predictive Modeling, Anomaly Detection
Deep Learning:	TensorFlow, PyTorch, BERT, spaCy (NLP), LSTM, Autoencoders
Big Data & ETL:	Apache Airflow, PySpark, Databricks, Hadoop Ecosystem
Cloud Platforms:	Azure (Data Lakehouse, Databricks), AWS (SageMaker, S3, EC2), Snowflake
Risk & Compliance Analytics:	Credit Risk Modeling (PD, LGD, EAD), Basel III/IV, IFRS 9, Regulatory Reporting (SEC, OCC, IRDAI), Stress Testing, Capital Adequacy
Fraud & Network Analytics:	Fraud Detection Models, Graph Analytics (Neo4j), AML/KYC Compliance
APIs & Market Data:	Bloomberg API, Refinitiv API, Guidewire Integration
Explainable AI (XAI):	SHAP, LIME, ELI5
Visualization & BI Tools:	Tableau, Power BI, Streamlit
NLP & Text Analytics:	BERT, spaCy, Sentiment Analysis, Claims Notes Categorization
Version Control & DevOps:	Git, CI/CD Pipelines, Docker (for model deployment)
Automation & Reporting:	Regulatory Reporting Automation, GPT-based Summarization, Retrieval- Augmented Generation (RAG)

PROFESSIONAL EXPERIENCE:

Minutentag- USA **Jan 2025- Present**

Data Scientist

- Built a production-grade image generation pipeline using Stable Diffusion, GPT-4, and ControlNet, achieving a 40% reduction in image generation latency while demonstrating strong end-to-end AI/ML engineering capabilities.
- Architected and orchestrated a modular Generative AI platform with agent-based prompt engineering and dynamic model selection (SDXL, ControlNet), significantly streamlining production workflows through structured data modeling and scalable design.
- Improved overall system efficiency by integrating vector search and long-term memory using Milvus and Pinecone, enabling real-time contextual retrieval and advanced agent-based reasoning across multimodal tasks.
- Designed and implemented human-in-the-loop feedback mechanisms and optimized prompt engineering strategies, resulting in a 92% improvement in model alignment accuracy and higher output reliability.
- Containerized the complete Generative AI stack and deployed scalable microservices on AWS SageMaker and EKS, highlighting strong expertise in Docker-based containerization, cloud deployment, and MLOps best practices.
- Developed automated MLOps pipelines using Apache Airflow and AWS Batch to support daily retraining cycles, ensuring continuous model monitoring, versioning, and performance improvement.
- Created and deployed image filtering and classification models to clean and curate datasets for multimodal learning, applying robust data preprocessing and quality-control techniques.
- Integrated social media APIs including Facebook, Instagram, and Twitter to automate content publishing workflows, expanding platform reach and enabling seamless cross-channel content delivery.

Morgan Stanley – USA **Jan 2024- Dec 2024**

Data Engineer

- Developed end-to-end credit risk and portfolio analytics solutions using Python (Pandas, NumPy), transforming large-scale banking data into actionable insights for lending and risk teams.
- Built and validated regulatory-grade PD, LGD, and EAD models aligned with Basel III and IFRS 9, strengthening credit risk measurement and audit readiness.
- Designed high-performance predictive models using XGBoost to identify default risk and early warning indicators, supporting proactive credit and collections strategies.

- Conducted customer churn and behavior analysis using Random Forest, enabling data-driven retention strategies across retail and corporate banking portfolios.
- Implemented survival and time-to-event analysis using R (GLM) to evaluate loan prepayment risk and customer lifetime exposure.
- Engineered scalable data processing pipelines using PySpark to handle high-volume transactional, loan, and customer datasets with improved performance and reliability.
- Automated and scheduled ETL workflows with Apache Airflow, ensuring timely availability of clean, compliant financial and regulatory data for analytics use cases
- Deployed secure analytics and machine learning workloads on Azure Data Lakehouse, ensuring scalability, data governance, and regulatory compliance.
- Delivered executive-ready risk and performance dashboards using Tableau, enabling senior stakeholders to monitor portfolio trends, model outputs, and regulatory metrics efficiently.

Accenture – India **Aug 2020- Dec 2022**

Data Engineer

- Designed and maintained scalable healthcare data pipelines using Python and SQL to ingest, transform, and standardize clinical, claims, and patient operational datasets.
- Built robust ETL workflows using Apache Airflow to automate ingestion of EHR, lab, and claims data, ensuring timely and reliable data availability for downstream analytics.
- Processed large-scale healthcare datasets using PySpark, optimizing performance for high-volume patient records and historical medical data.
- Implemented data lake architecture on Azure Data Lakehouse, enabling secure, compliant storage of structured and semi-structured healthcare data.
- Enforced healthcare data quality, validation, and reconciliation rules using Python, improving accuracy of patient-level and provider-level reporting.
- Integrated external healthcare and insurance data sources using SQL, supporting population health analytics and cost-of-care analysis.
- Supported fraud, waste, and abuse detection initiatives by preparing clean, high-integrity datasets for anomaly detection models.
- Collaborated with data science and clinical analytics teams to optimize feature-ready datasets for predictive modeling and risk stratification use cases.
- Implemented data governance, access controls, and audit trails aligned with healthcare compliance and privacy requirements within Azure environments.
- Developed executive and operational dashboards using Power BI, enabling healthcare leaders to track utilization, patient outcomes, and operational efficiency metrics.

EDUCATION:

Master of Information Systems Technology – Wilmington University, Delaware, USA

CERTIFICATION:

AWS data Engineer associate – Amazon

Certified with Python, SQL, Prompt engineering, ML, GenAI, model training, N8N, Image classification crash courses - Udemy