Jilani Shaik

Minneapolis, MN
763-321-1545 | jilani2420@gmail.com | www.linkedin.com/in/jilanishaik/

PRINCIPAL ENGINEER

Passionate Engineer with hands-on experience delivering robust, scalable ML solutions. Deeply specialized in 6+ years in Machine Learning, Generative AI(RAG), LLM with extensive experience across the Azure and Databricks ecosystems. Demonstrated success in translating complex data into actionable business insights by doing statistical analysis using Python and SQL, building and deploying ML pipelines by leveraging deep learning frameworks. Proven leader guiding engineering teams and collaborating with product and business stakeholders on product strategies. Competencies include:

Business Development | Data Analysis | Statistical Techniques | Project Management | Team Leadership and Collaboration | Stakeholder Management | Machine Learning Pipelines | Generative AI Mentoring | Problem Resolution | Communication (Written and Verbal) | Strategic Planning | Risk Management

TECHNICAL SKILLS

Programming Languages: Python | Java and Scala.
ML and AI: Deep Learning (PyTorch | TensorFlow/Keras) | Time Series | Causal Inference and Transformer Architecture.
LLM and Frameworks: LlamaIndex | LangChain | Fine-tuning (LoRA/QLoRA) | Hugging Face | Chroma | Pinecone | and Azure AI Search.
ML Ops and Deployment: ML Pipeline Design and Deployment | Model Evaluation and Monitoring | MLOps | Data Drift Handling | A/B Testing | Docker | Kubernetes | GitHub CI/CD Pipelines.
Data Science and Analytics: Statistics | Feature Engineering | Data Mining/Analysis/Cleaning | Data Visualization (Tableau | Power BI) | Model Selection/Evaluation/Deployment/Monitoring | Optuna.
Cloud Platforms: Azure (Expert) | AWS (Familiar) | GCP (Familiar).
Data Technologies: Databricks | PySpark | SQL | Data Governance | Stream Processing (Kafka) | Delta Lake | Unity Catalog.
Architecture Patterns: Deep understanding and practical application of various patterns (ML System Design | Distributed Training | Real-time/Batch Inference | Data Lakehouse | Microservices).
Tools and Technologies: Jupyter | MLflow | Airflow.

SELECTED ACCOMPLISHMENTS

- Engineered a Generative AI feature using Azure Search, Retrieval Augmented Generation (RAG), and OpenAI(gpt-4o-mini model) improving product satisfaction by 20% through context-aware information to providers during patient visits and coders for coding.
- Deployed a model that reduced the run time of the overall risk and quality chassis, which produced results fast to provider groups during the program launch. Which helped early launch of the In-Office Assessment Program (IOA) in mid of January, instead of late Feb for every year.
- Enabled cross-functional team collaboration to improve project efficiency through knowledge sharing and integrated workflows; developed a flexible cloud-native streaming framework supporting diverse Kafka consumption with dynamic business logic.

PROFESSIONAL EXPERIENCE

Optum **01/2016 – Present**
Principal Engineer
Machine Learning for Healthcare | Data Science Practitioner
Implemented healthcare payer data, risk stratification, and quality gap management to drive impactful data science and machine learning initiatives, uncovering critical insights. Designed and implemented robust data environments, including feature stores, model registries, and scalable pipelines, to facilitate seamless transition of machine learning models from research to production.

- Generated data insights through exploratory data analysis (EDA) and plotting images using various libraries (e.g., klib, plotly, seaborn), and enhanced model performance via feature reduction techniques like PCA, FAMD, and autoencoders.

- As part of data analysis utilized various statistics libraries including scipy stat, PySpark ML Lib statistical functions for big data and stats models, time series to do recommendations for readmission.
- Developed and deployed a predictive model for hospital readmission risk leveraging diverse healthcare claims data, dynamic time warping (DTW) for temporal pattern analysis, and integrated with social determinants of health (SDOH), to provide deep insights into patient risk factors.
- Applied causal inference techniques to understand potential impact of various factors on readmission, providing actionable insights for intervention strategies.
- Managed the end-to-end ML lifecycle, from research to staging and production deployment, including tracking and monitoring by using ML flow.
- Deployed a predictive model in a distributed cluster to identify deployable members for provider groups, measured with A/B tests, and gathered feedback for continuous improvement, facilitating early launch of the In-Office Assessment Program (IOA) annually and enabling 20% more prospective returns each year.
- Enhanced ensemble model accuracy to 83% by refining feature engineering and hyperparameter tuning, resulting in a 15% improvement in identifying critical medication gaps.
- Utilized classification models like LightGBM, XGBoost, Random Forest, Logistic Regression to identify and target high-risk patient population, deriving key insights for preventive care initiatives.
- Converted business needs into ML applications to predict and manage member risk, aiming for measurable ROI through targeted in-person visits.
- Maintained a base line data which is continuously updated after drift identified and fixed by updating incoming data pipelines, including schema evolution and erroneous label data cleanup.

Gen AI Clinical Decision Support | Data Science Practitioner
Created a Generative AI feature using Azure, Retrieval Augmented Generation (RAG), and Open AI (gpt-4o-mini model), leveraging LlamaIndex for advanced indexing and retrieval strategies and storing vector embeddings in Pinecone for efficient semantic search, to provide real-time, context-aware, and domain-specific answers to provider questions during member visits.

- Indexed member charts submitted by providers to Azure AI Search as Vector data further enhancing search capabilities by exploring alternative vector databases like Chroma and FAISS for specific use cases.
- Lead the development of service using fastapi.
- Extracted medical codes and suspect conditions from member charts using Azure AI Document Intelligence NLP libraries and transformers, and prototyped fine-tuning techniques (e.g., LoRA, QLoRA) on smaller domain-specific models to optimize performance for specialized healthcare terminology.
- Provided technical leadership for data infrastructure, strategically aligning it with the evolving needs of data science and Generative AI applications, ensuring data accessibility and quality for advanced analytical exploration, often designing modular components following microservices principles for enhanced agility and maintainability.

Batch and Real-time Data Processing | Lead Engineer

- Built efficient ETL/ELT pipelines, capable of handling high-velocity, volume, and variety data for feature engineering and model training.
- Designed and built a fast system using Spark and Kafka to process high-volume and high-velocity data into a Delta Lake and Cosmos DB (NoSQL), enabling real-time data analysis and insights for efficiently generating 500K customized PDFs with member clinical profile.
- Implemented a flexible cloud-native streaming framework supporting diverse Kafka consumption with dynamic business logic.
- Implemented a scalable Medallion architecture on Delta Lake within Azure Databricks.
- Led the migration of big data systems from on-prem to Azure Databricks, utilizing Terraform for automated infrastructure provisioning and management, ensuring consistent and reproducible deployments.
- Lead the micro service implementation to manage member data for other systems to consume.
- Optimized Spark cloud-native data pipelines by analyzing health care population data(member, provider, summarized claims) and strategically tuning read and write patterns, achieving over 50% reduction in the processing time, which reduces cluster usage resulted in savings.
- Mentored 5 engineers, which makes project progress faster, which reduced 30% build to deploy time.

CES Technology Services **06/2015 – 12/2015**
Senior Technical Lead
Provided technical guidance for projects, focusing on system design and data infrastructure to support analytical capabilities.

- Designed and implemented REST micro services using spring.

Bank of America **08/2014 – 05/2015**
Sr. Analyst
Managed sourcing and integration of critical data from various business lines into Netezza, supporting downstream applications and risk factor analysis.

- Designed and implemented Oozie workflows for complex data pipelines, ensuring data integrity and automated processing for market data from ingestion to dissemination
- Led the migration of Hadoop infrastructure.

Adroitent **02/2014 – 08/2014**
Sr. Technical Lead
Contributed to system design considerations for future search integration within medical research content, that streamlined nightly content pull processes, improving data availability efficiency.

- Designed core components to convert repository content into optimized Solr documents, enhancing search capabilities and content discoverability for insights.

ReThink IT Services **06/2011 – 02/2014**
Technical Project Lead
Developed and deployed a big data analytics platform, applying advanced statistical and machine learning techniques to derive actionable insights.

- Designed and implemented robust data infrastructure leveraging Hadoop, NoSQL, Solr, and distributed computing for large-scale data processing.
- Engineered comprehensive data ingestion and transformation pipelines, optimizing data for efficient exploration, feature engineering, and strategic decision-making.

ADDITIONAL RELEVANT EXPERIENCE

United Health Group
Sr. Software Engineer
Supported critical business functions through application development and maintenance.

CTE
Software Engineer
Developed reusable software (File Upload, Email Utility, Notifications) to optimize project timelines.

EDUCATION and CERTIFICATIONS

Master of Computer Applications
Osmania University - Hyderabad, India

Bachelor of Mathematics
Acharya Nagarjuna University – Guntur, India

PG Program in Data Science and AI with Deep Learning
University of Texas at Austin