

NEELAM HASWANTH RAJESH

+1 9452411532 | haswanthrajeshn@gmail.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY:

Data Scientist and AI Engineer with 4+ years of experience designing and deploying scalable, data-driven systems across cloud environments. Proven expertise in **statistical modeling, machine learning, and data architecture** to solve complex product and business challenges. Skilled in **Python, SQL, and Spark** for large-scale analytics, complemented by advanced proficiency in **Generative AI** and **LLM-based application development**. Experienced in building **ETL/ELT pipelines** and **data lakehouse architectures** using Azure and AWS platforms. Adept at translating data insights into business strategy, improving decision-making, and driving product innovation through applied AI and data science.

SKILLS:

Data Science & ML: Predictive modeling, hypothesis testing, regression/classification, clustering, A/B testing

Statistical Analysis: Experiment design, causal inference, feature engineering, statistical inference

Data Engineering: Azure Data Factory, Databricks, Redshift, SQL, PySpark, ETL/ELT workflows

Programming: Python, SQL, Spark SQL, PySpark, Scikit-learn, Pandas, NumPy

AI & LLMs: OpenAI API, LangChain, Hugging Face, Embedding Models, RAG Architecture, Vector Databases

Visualization & BI: Power BI, Matplotlib, Streamlit

Cloud Platforms: Azure (ADLS, Blob, Databricks, ADFv2), AWS (S3, Redshift)

Tools & Frameworks: Git, MLflow, Supabase, Vercel, Azure DevOps, CI/CD

GEN AI SKILL SET:

Category	Skills & Tools	Proficiency	Applications
Machine Learning	Regression, Classification, Clustering, Feature Selection	Proficient	Predictive modeling, trend analysis
LLMs & RAG	GPT, LangChain, FAISS, Azure OpenAI, Pinecone	Proficient	Semantic search, intelligent Q&A
Vector Databases	Chroma, FAISS, Pinecone	Proficient	Embedding storage, similarity retrieval
Prompt Engineering	CoT reasoning, few-shot learning, output control	Proficient	Consistent LLM reasoning, domain adaptation
Responsible AI	Hallucination detection, privacy compliance	Proficient	Ethical AI deployment, enterprise safety

PROFESSIONAL EXPERIENCE:

- Azure Data Engineer / AI Engineer — DatafactZ, Northville, MI** *July 2023 – Present*
- Designed and deployed modern **data lakehouse solutions** using **Azure Databricks** and **Delta Lake**, enabling real-time analytics with medallion (Bronze–Silver–Gold) architecture.
 - Developed **predictive and sentiment models** leveraging Hugging Face transformers and PySpark for unstructured data at scale.
 - Built **data ingestion and transformation pipelines** using Azure Data Factory, Databricks, and SQL for data migration across multiple sources.
 - Implemented **LLM-powered automation** for resume rewriting and semantic search using **OpenAI GPT**, **FAISS**, and **sentence-transformers**, improving matching accuracy by 40%.
 - Conducted **statistical analysis and feature engineering** for customer segmentation and performance optimization.
 - Partnered with stakeholders to define key metrics, establish KPIs, and create **Power BI dashboards** that improved reporting efficiency by 30%.
 - Designed **Unity Catalog governance models** to ensure data compliance, lineage, and access control across enterprise datasets.

Key Projects

- **Scalable Data Lakehouse Architecture:** Architected multi-layer Delta pipelines with schema evolution and query optimization; improved data refresh time by 35%.
- **AI-Powered Job Portal:** Engineered AI-driven job matching and resume optimization using OpenAI GPT and Hugging Face embeddings; deployed via React.js/Node.js stack.
- **Product Review Sentiment Analyzer:** Built NLP pipeline on Azure Databricks; automated sentiment trend analysis for 100K+ reviews, supporting product strategy decisions.

- Data Engineer — DatafactZ, Hyderabad, India** *July 2020 – August 2022*
- Developed **data ingestion frameworks** using Azure Data Factory and PySpark, processing terabytes of structured and semi-structured data.
 - Created and optimized **ETL/ELT pipelines** using Databricks with advanced tuning (Z-ordering, AQE, Bloom filters).
 - Designed **metadata-driven automation** for schema inference and error handling, reducing manual maintenance by 50%.
 - Implemented **data quality frameworks** and **governance policies** using Delta Live Tables and Unity Catalog.
 - Collaborated cross-functionally with analysts and ML teams to enable feature pipelines for predictive analytics.
 - Conducted **statistical data profiling and exploratory data analysis (EDA)** to improve data reliability for downstream ML models.

EDUCATION:

- 05/2024 **Master of Science: Computer Science**, The University of Texas at Arlington - Arlington, TX, USA
- 05/2022 **Bachelor of Science: Computer Science**, Amrita Vishwa Vidyapeetham - Coimbatore, TN, India