

# DIANA MININE GUDINHO

+1 716-939-5682 | [dianagudinho16@gmail.com](mailto:dianagudinho16@gmail.com) | [LinkedIn](#) | [GitHub](#) | New York

## SUMMARY

Data Scientist with 4+ years of experience designing AI solutions that drive business impact across healthcare, supply chain, and retail. Skilled in deploying LLMs, RAG, and multi-agent systems to improve decision-making and automate knowledge-intensive workflows. Experienced in building end-to-end pipelines and cloud-native deployments that deliver scalable, production-ready AI while translating complex data into actionable strategies.

## EDUCATION

<b>University at Buffalo</b>	<b>New York, USA</b>
<i>Master of Science in Data Science (CGPA: 3.7/4.0)</i>	<i>Aug 2023 – Feb 2025</i>
<b>Visvesvaraya Technological University</b>	<b>Karnataka, IN</b>
<i>Bachelor of Engineering in Information Science (CGPA: 3.5/4.0)</i>	<i>Aug 2016 – Aug 2020</i>

## WORK EXPERIENCE

<b>Research Assistant   University at Buffalo</b>	<b>Feb 2025 – Present</b>
• Improved diagnostic QA accuracy by <b>22%</b> on <b>10K+ MRI</b> scans by combining embedding-driven retrieval with <b>fine-tuned LLMs</b> to surface relevant prior cases for radiologists.	
• Prototyped <b>multi-agent GenAI workflows</b> that coordinated retrieval, reasoning, and summarization tasks, reducing validation by <b>30%</b> .	
• Optimized model training with distributed <b>PyTorch</b> and GPU scheduling, cutting training cycles by <b>25%</b> .	
<b>Graduate Student Assistant   University at Buffalo</b>	<b>Aug 2024 – Dec 2024</b>
• Developed a customer intelligence solution using <b>LLM-based text generation</b> and <b>prompt engineering</b> to analyze unstructured campaign data, driving an <b>18% increase</b> in marketing <b>ROI</b> .	
• Designed a <b>RAG pipeline</b> to integrate structured and unstructured records, enabling <b>semantic search</b> and faster reporting by <b>45%</b> .	
<b>Data Analyst   Tata Consultancy Services Ltd.</b>	<b>Oct 2020 – June 2023</b>
• Built <b>demand forecasting</b> models using <b>XGBoost</b> and <b>LSTM</b> on <b>50K+ SKUs</b> , improving stock accuracy, which cut procurement spend by <b>\$470K</b> annually.	
• Designed <b>pricing and margin</b> models on transactional data, uncovering low-performing SKUs and guiding adjustments that lifted product <b>profitability</b> by <b>12%</b> .	
• Engineered pipelines on <b>AWS Glue, Redshift</b> , and <b>S3</b> , processing <b>5M+</b> daily records to give executives near real-time visibility.	
• Applied <b>clustering</b> and <b>anomaly detection</b> on <b>100K+</b> distribution routes to uncover inefficiencies, <b>reducing delays</b> by <b>17%</b> and improving fleet utilization.	
• Conducted <b>regression</b> and <b>root cause analysis</b> on supplier and claims data, improving compliance report accuracy by <b>25%</b> while also lowering financial risk exposure.	
• Built <b>Tableau</b> dashboards for fulfillment and cost <b>KPIs</b> ; cut refresh time by <b>40%</b> and improved stakeholder adoption by <b>30%</b> .	
• Automated retraining workflows and <b>CI/CD</b> pipelines, cutting release cycles by <b>30%</b> and improving production model reliability.	
• Defined KPIs with <b>finance, logistics</b> , and <b>product</b> teams, embedding metrics like <b>SLA</b> compliance and <b>forecast accuracy</b> into daily decisions.	
<b>Full-Stack Software Engineering Intern   QSpiders</b>	<b>Jul 2019 – Aug 2019</b>
• Optimized <b>PostgreSQL</b> queries and indexing for <b>8K+</b> sessions, cutting response times by <b>45%</b> and stabilizing peak-hour performance.	
• Built automated <b>Excel</b> templates for weekly summaries, streamlining issue tracking for managers.	

## SKILLS

• <b>Programming:</b> Python (Pandas, NumPy, Scikit-learn, PyTorch, TensorFlow, Keras), SQL, R, Scala
• <b>Machine Learning &amp; GenAI:</b> Regression, Classification, Gradient Boosted Trees, Random Forests, Neural Networks, Clustering, Anomaly Detection, Recommender Systems, Forecasting, NLP, LLMs
• <b>Statistical Analysis:</b> Hypothesis Testing, Causal Inference, Bayesian Methods, Time-Series Modeling
• <b>Pipelines &amp; Deployment:</b> PySpark, Airflow, dbt, Kafka, AWS (S3, Glue, Redshift, SageMaker), GCP BigQuery, Snowflake, Jenkins
• <b>Business Intelligence &amp; Visualization:</b> Tableau, Power BI (DAX), Looker, Matplotlib, Seaborn

## PROJECTS

• <b>AI-Powered Skill Gap Analyzer:</b> Deployed a containerized microservice with Ollama and NLP models to identify skill gaps and recommend personalized learning paths, driving a 35% increase in feature adoption.
• <b>RAG Compliance Copilot:</b> Built a retrieval-augmented assistant to auto-summarize and flag risks in regulatory documents, cutting manual review time by 35% and improving issue detection accuracy.
• <b>Customer Retention Insights:</b> Integrated classifiers with LLM-driven narrative generation to produce human-readable churn insights, improving adoption of retention strategies by 15%.
• <b>E-Commerce Product Catalog Enrichment (LLM Ensemble):</b> Fine-tuned LLaMA + GPT ensemble for attribute extraction, boosting catalog completeness by 28%.