# Mounika Geriki

Stony Brook, NY | mounika.geriki@stonybrook.edu | 934-949-8567 | LinkedIn | Github

## Education

**Stony Brook University**                                                                                                    Stony Brook, NY
*Master of Science, Data Science | GPA: 3.55*                                                                    Jan 2024 - Dec 2025
- **Coursework**: Statistical Computing & Learning, Data Analysis, Big Data Analysis, Data Management, LLM

## Skills

**Programming:** Python, R, SQL, MySQL, Oracle, PostgreSQL, OCAML, C, C++, HTML, CSS
**Data Engineering:** ETL, PySpark, Snowflake, REST/SOAP APIs, Google Cloud(BigQuery, VertexAI, Composer)
**Visualization:** Power BI, Tableau, Looker Studio, MS Excel, Matplotlib
**ML & AI:** Scikit-learn, TensorFlow, LightGBM, OpenCV, Generative AI, LLMs (BERT, RAG, LoRA)
**Cloud & DevOps:** AWS, GCP, Docker, Linux
**Tools:** Git, Jupyter, Agile, Troubleshooting

## Professional Experience

**Stony Brook University** | *Graduate Research Assistant, LLM*                    *Jan 2025 - Present* | Stony Brook, NY
- Conducting research on domain-specific **LLMs** (MaterialsBERT, BatteryBERT, PolymerLLM) for automated extraction of battery and polymer properties from scientific literature.
- Created a curated, benchmarked dataset of battery properties and validated performance across **multiple LLMs (GPT-4, Claude, DeepSeek)**.
- Implemented **prompt-engineering** and **fine-tuning** strategies **(zero-shot, LoRA, prompt-tuning)** to adapt PolymerLLM to BatteryExtract LLM, improving cross-domain knowledge transfer in materials science.

**Tata Consultancy Services** | *Data Analyst*                                                    *Dec 2022 - Jan 2024* | Bengaluru, India
- Engineered scalable **ETL** pipelines using **Python, SQL**, and **SOAP/REST APIs** that processed **50+ GB** of daily banking data, improving stability and reducing processing time by **45%**.
- Directed 4 analysts to deliver **8+ SQL automation solutions** and **20+ Power BI/Tableau dashboards** on loans, deposits, and acquisition costs, improving reporting efficiency by **35%**.
- Supported a **Random Forest churn model** that helped retention teams reduce attrition in high-value customer segments.
- Integrated APIs across digital banking, payments, and service platforms, expanding reach to **200K+** customers with **99.8%** transaction reliability.

**Tata Consultancy Services** | *Associate Business Analyst*                            *Sep 2021 - Nov 2022* | Bengaluru, India
- Led end-to-end migration of **50+ Oracle** dashboards to **Snowflake**, ensuring **100%** accuracy and resolving critical financial discrepancies without business disruption.
- Automated transformations with **SQL stored procedures** in Snowflake, reducing manual data preparation time by **65%** and improving analytical capabilities.
- Designed **20+ interactive Tableau dashboards** (LOD expressions, parameters, calculated fields), transformed financial data into actionable insights, and enabled actionable financial insights.
- Consolidated **5+ data sources through joins and blending**, improving reporting continuity and reducing dashboard latency by 40% for **200+** daily users.

**VI Solutions** | *Machine Learning Intern*                                                        *Sep 2020 - Jan 2021* | Bengaluru, India
- Created a **KNN** blood pressure prediction model (87% accuracy, 22→8 features) and applied **K-Means** for medical image segmentation, accelerating processing by 30% and raising **OCR precision from 76% to 92%.**
- Engineered a real-time vital signs monitoring system by integrating **Python ML** models with **LabVIEW**, achieving less than 500ms latency for immediate clinical feedback.

## Projects

**TravelWise-NYC: Adaptive RAG Travel Guide Chatbot** | *Python, Flask, GCP, FAISS, LangChain*     *Mar 2025 - May 2025*
- Architected an **agentic RAG** pipeline with **adaptive query routing** across curated NYC docs and real-time web search.
- Implemented **Google AI Studio embeddings + FAISS** vector store, achieving **91.4% response rate** with minimal errors.
- Crafted an **interactive Flask web app** with Markdown rendering, suggested queries, and real-time streaming responses.
- Established an evaluation framework **(accuracy, rejection/error rate, latency)**, ensuring reliable outputs.

**NYC Taxi Insights** | *Google Cloud, BigQuery, Looker Studio*                                        *Oct 2024 - Dec 2024*
- Orchestrated an ETL pipeline on **GCP composer**, processing **9+** months of NYC Taxi data with 100% data consistency.
- Enhanced performance by **35%** using **star schema** transformations in **BigQuery**.
- Designed real-time dashboards on **Looker Studio**, providing actionable insights for trip duration predictions.
- Delivered **85%** accuracy in trip duration predictions and **0.012 MSE** in demand forecasting using **BERT** and **LightGBM**.
- Deployed APIs on Render, enabling real-time predictions and improving system responsiveness by **50%**.

**Statistical Ensemble Package Development** | *R*                                                          *Sep 2023 - Dec 2023*
- Developed an **R** package with **5+** models **(linear, logistic, ridge, lasso, SVM)** handling **10K+** variables in high-dimensional data.
- Boosted efficiency by **40%** with **top-K feature selection** and improved predictive accuracy by **25%**, achieving **85%+** across various models.