

Aishanee Sinha

San Jose, CA | aishanee.sinha@gmail.com | (669) 265-2905 | linkedin/aishanee-sinha | github/aishanee-sinha

Education

MS Applied Data Intelligence San Jose State University (GPA: 3.9/4)	Aug 2024 - May 2026 (Grad)
• Special Scholarship Awardee - exceptional standing among graduate students	
• Graduate Student Assistant Generative AI, Big Data Technologies	
• <i>Relevant Coursework : Machine Learning, Deep Learning, Gen AI, Big Data, Data Engineering, Distributed Systems</i>	
MS Economics Indira Gandhi Institute of Development Research	Jul 2018 – Jul 2020
BE Electronics & Telecommunication Jadavpur University (Gold Medalist)	Aug 2013 – Jul 2017

Work Experience

Business Systems and Data Management Intern, PACCAR Financial Corporation	May 2025 – Aug 2025
• Document Intelligence Agent: Built a Python-based AI tool using NLP to parse complex auction PDF reports and automatically generate structured insights for sales teams, reducing manual analysis by 80%	
• Live Data Integrity Monitoring: Developed a dashboard with live connections to multiple Snowflake sources to autonomously flag data integrity issues and system anomalies in real-time.	
• Strategic ROI Modeling: Analyzed lead conversion data to optimize marketing budget, projecting a 15% increase in Campaign ROI .	
Senior Data Analyst, Citibank	Jan 2022 – Aug 2024
• Led automation of 14+ campaign dashboards , leveraging PySpark and SQL for data orchestration and achieving a 95% reduction in manual workload.	
• Championed Citi Ignite 2023 global ideation challenge as a semifinalist, proposing an ML-driven win-back model to re-engage 1.5M inactive customers at a lower acquisition cost.	
• Architected scalable PySpark data pipelines on Hadoop clusters, integrating portfolio, transaction, and CRM data for unified customer analytics.	
• Designed and deployed customer segmentation models using K-Means, Decision Trees, and Random Forests , boosting campaign targeting efficiency and revenue by 23% .	
• Improved existing anomaly detection models to identify fraudulent campaign activity, saving \$0.8MM .	
Data Analyst, Citibank	Aug 2020 – Dec 2021
• Automated monthly campaign reporting via Python-SQL pipelines , cutting manual effort by 70% .	
• Designed A/B tests to optimize campaign offer strategy and contact frequency, driving 20% higher conversion rates .	

Skills

- **Languages & Core:** Python (Pandas, NumPy), SQL, SAS, RESTful APIs, Webhooks, GraphQL
- **ML & Gen AI:** LLM Fine-Tuning, LangChain, LangGraph, RAG, NLP, VectorDB (Pinecone), vLLMs, PyTorch, Keras, Scikit-learn, XGBoost
- **Integrations & Big Data:** API Orchestration, Snowflake, PySpark, SparkML, dbt, Airflow, ETL/ELT Pipelines, AWS, GCP, Docker, OAuth
- **Tools & Workflows:** Git (GitHub API), Jira API, Slack API, CI/CD Pipelines, Tableau, PowerBI

Academic Project Highlights

Agentic AI-Powered Workforce Assistants

- **Autonomous Workflow Orchestration:** Built Python-based agents using LangChain to integrate Slack, Jira, and Email APIs, autonomously syncing "sources of truth" across engineering workflows.
- **Intelligent Task Management:** Engineered proactive bottleneck detection and automated task extraction, pushing parsed to-do items from natural language directly to Jira to resolve blockers.
- **Agentic User Experience:** Designed a conversational interface featuring RLHF-based feedback loops and proactive suggestions to streamline team operations and productivity.

MultiModal RAG Document Retrieval Chatbot

- Built an end-to-end chatbot using **LangChain**, Pinecone vector embeddings, and hybrid sparse-dense retrieval for complex technical documentation.
- Citation-aware answer generation and query reformulation to ensure continuity in multi-turn technical interactions.

Emotion-Aware Transformer-Based Text-to-Speech (TTS) System

- Developed an end-to-end emotion-conditioned TTS pipeline from scratch using **PyTorch** and a Transformer architecture, implementing custom text-to-mel spectrogram generation and a Griffin-Lim based waveform inversion pipeline.

Energy Demand Forecasting based on Real-time data and Weather forecasts

- Orchestrated **ETL/ELT workflows** and developed an end-to-end data pipeline using **Python, Snowflake, Airflow, dbt, Docker and Power BI** to automate energy consumption forecasting 94% prediction accuracy with Snowflake ML.