

# NIKITHA SEELAM BALAJI

Los Gatos, CA | (669) 208-9792 | nikithasb2001@gmail.com | linkedin.com/in/sbnikitha  
Software Engineering · LLM & Agentic AI · Machine Learning · Data Engineering · Reinforcement Learning & Robotics

## SUMMARY

Software and AI engineer with 3+ years of production experience at Samsung and hands-on graduate research in RL, LLMs, and data systems at SJSU. Built end-to-end ML monitoring platforms, RAG pipelines, multi-agent agentic systems, real-time data pipelines, and curriculum RL environments. Comfortable across the full stack: Python, Java, REST APIs, cloud infrastructure, and AI/ML workflows. Currently pursuing M.S. Data Analytics while contributing to autonomous driving research under Prof. Vishnu Pendyala.

## TECHNICAL SKILLS

**Languages:** Python, Java, SQL, JavaScript, C/C++, Shell Script, PHP, Scala

**AI / ML / RL:** LLMs, RAG, LangChain, FAISS, HuggingFace, Reinforcement Learning, Curriculum Learning, RLHF, Reward Modeling, Policy Gradient, Multi-Agent Systems, Offline RL, Prompt Engineering

**Agentic / LLM Infra:** Multi-agent orchestration, MCP tool calling, supervisor-agent patterns, Gemini API, OpenAI API, FastAPI AI services, human-in-the-loop workflows

**Data Engineering:** Apache Spark, PySpark, Kafka, Airflow, ETL, dbt, Great Expectations, Apache Iceberg, Hadoop, Redshift, BigQuery, Snowflake

**Software Engineering:** OOP, Spring Boot, FastAPI, Django, REST APIs, React, Angular, JUnit, Git, Agile, Maven, CI/CD

**Cloud / DevOps:** AWS (S3, Glue, Lambda, Redshift), Azure, GCP, Docker, Kubernetes, Jenkins, Linux

**Databases:** MySQL, PostgreSQL, MongoDB, Redis, Snowflake, Oracle, Amazon Redshift

**Visualization:** Power BI, Tableau, QuickSight, Figma

## PROFESSIONAL EXPERIENCE

### Software Engineer — Java Full Stack & Data Engineering

May 2022 – Jun 2025  
Bangalore, India

Samsung Electro-Mechanics

- Built and deployed AI/ML monitoring platforms, REST/JSON APIs, and backend services in Java, Spring Boot, and Python, achieving ~99% reduction in production downtime and improving model-serving reliability.
- Designed Airflow-based ETL pipelines and data drift detection for image and tabular ML models, directly improving prediction reliability and reducing retraining costs; work earned a Project-Level Recognition Award.
- Implemented SOA microservices with OAuth 2.0/Keycloak, JUnit testing, Docker/Kubernetes CI/CD (Jenkins), and reusable OOP design patterns, accelerating delivery cycles by ~85%.
- Maintained metadata governance, data validation workflows, and cross-team technical documentation supporting production ML pipelines and distributed systems.

### Full Stack Web Developer

Apr 2021 – May 2022  
Remote

Stargazer Edutech LLP

- Solely built and launched Physicsopedia.com a full-stack educational platform (HTML/CSS/JS, PHP/Django, SQL) serving 350+ monthly active users; designed responsive UI flows and SQL-backed content pipelines.

## RESEARCH & AI/ML PROJECTS

### Curriculum Learning for Autonomous Driving Agent

2025–2026

Research Assistant · Prof. Vishnu Pendyala, SJSU | Python · RL · Curriculum Learning · Reward Modeling · Multi-Agent · LiDAR Simulation

- Designed a 6-stage RL curriculum progressively increasing task complexity from lane-following to dense mixed-traffic navigation; engineered 106-dimensional observation space (ego state, LiDAR, vehicle/pedestrian positions).
- Built multi-component reward functions (collision avoidance, goal completion, forward progress, off-road penalties) and applied checkpoint-based staged transfer, achieving 98%+ success rate in early environments.
- Conducted systematic failure analysis on reward imbalance, policy degradation, and distribution shift across vehicle-only to mixed-agent scenarios.

### LLM-Powered Code Query System for Apache Spark

2026

Independent AI Project | Python · LangChain · FAISS · HuggingFace Embeddings · Gemini API · RAG · PySpark · Scala

- Built a RAG pipeline to query Apache Spark internals: parsed 33+ source directories, indexed 110,000+ code chunks using FAISS and HuggingFace embeddings, and integrated Gemini API for context-grounded code explanations.
- Compared baseline LLM outputs against RAG-enhanced responses to evaluate improvements in architectural reasoning for Catalyst Optimizer, RDD execution flow, and SparkConf modules.

### Agentic Multi-Agent Recruiting Platform with MCP Integration

2026

Distributed Systems Group Project | FastAPI · Kafka · MCP Tool Calling · MySQL · MongoDB · Redis · Docker · AWS

- Built a LinkedIn-style platform with Kafka async workflows, agentic AI layer (supervisor + specialized agents), MCP-style tool calls, and human-in-the-loop review for resume parsing, job matching, and recruiter outreach.

### AKI-Trajectory: Real-Time ICU Deterioration Prediction System

2026

Healthcare ML & Data Engineering Project | Spark · Kafka · Airflow · Azure · AWS · Gradient Boosted Trees · MIMIC-IV · eICU

- Designed Bronze→Silver→Gold→Mart pipelines on Azure/AWS to predict Stage 3 kidney deterioration 48 hours ahead; implemented KDIGO staging logic and trained GBT/Logistic Regression models on MIMIC-IV/eICU datasets.
- Integrated Kafka + Spark Streaming for near real-time inference; structured ICU event traces mirroring offline RL data preparation paradigms for sequential decision-making.

## NYC 311–Airbnb Urban Livability Data Warehouse

2025

Cloud Data Engineering Project | AWS S3 · Glue · Lambda · Airflow · Redshift · Power BI · QuickSight · SQL

- Built scalable ETL pipelines, Redshift star-schema models, and Power BI/QuickSight dashboards to analyze complaint trends, neighborhood livability signals, and short-term rental patterns across NYC boroughs.

## EDUCATION

---

### M.S. Data Analytics — San Jose State University, CA

Aug 2025 – May 2027 (Expected)

CGPA: 3.57 Data Warehousing, Data Visualization, Mathematical Methods for Data Analytics

### B.E. Computer Science Engineering — Jyothy Institute of Technology (VTU), India

2019 – 2023

CGPA: 8.47 (First Class Distinction) | DSA, OOP, DBMS, OS, Computer Networks, Software Engineering

## ACHIEVEMENTS & LEADERSHIP

---

- Project-Level Recognition Award, Samsung Electro-Mechanics innovation in ML monitoring, ETL pipeline design, and data drift detection.
- 1st Place, JIT Hack 2021 (National Hackathon, 300+ participants) independently built Arcade, a full-stack web application.
- Toastmasters: Most Improved Toastmaster & Leader of the Term; developed public speaking, storytelling, and technical communication skills.
- Mentored 30+ junior engineers (Java Tech Team Lead) in coding, debugging, Git, and software fundamentals; served as judge/mentor at national-level student hackathons.