Abhinita Sanabada

San Jose, CA | (669) 241-8492 | abhinita.sanabada@sjsu.edu | linkedin.com/in/abhinita-sanabada | github.com/Sanabada

Summary

M.S. Data Science candidate leveraging 9 years of software engineering experience in building robust, data-centric systems. Expertise in architecting and automating scalable data pipelines using Airflow, Spark, and Kafka and managing cloud infrastructure (AWS/Azure, Kubernetes) and high-availability cloud platforms (AWS/Azure, Docker, Kubernetes) and data-intensive applications.. Seeking a 2026 Data Engineering or MLOps or software engineering internship.

Education

San Jose State University M.S., Data Science (2025 – Jan 2027, expected)
Coursework: Machine Learning, AI, Deep Learning, Big Data Analytics, Data Visualization, Cloud Computing,Database Systems, Distributed Systems, Software Engineering, Algorithms

Technical Skills

Languages: Python, SQL, Java
Data / ML: Pandas, NumPy, scikit-learn, Matplotlib, Seaborn, statistical analysis
Back-End: Spring Boot, Spring MVC, Microservices, REST APIs
Data Engineering: Airflow, Spark (PySpark),ETL / ETL, data modeling, testing / validation
Platforms: Hadoop, Kafka, Kubernetes, Docker
Cloud / DevOps: AWS, Azure, Git, CI/CD (Jenkins, GitHub Actions), Splunk, Redis
Databases: MySQL, Oracle, **MongoDB**
Visualization: Tableau, Power BI, Preset

Academic Projects

Real-Time SMS Spam Classification Pipeline

- Architected a real-time, event-driven pipeline to classify SMS messages as 'spam' or 'ham' using a machine learning model.
- The entire system, including Kafka, Zookeeper, and Python-based producers/consumers, was orchestrated using Docker and Docker Compose.
- This project demonstrates a complete, containerized data flow from live ingestion to real-time ML-driven classification.

Scalable Weather Data Ingestion Pipeline

- Built an ML pipeline to predict Architected a scalable data pipeline on HDFS to process raw NOAA-style station feeds, using Apache Spark for batch and streaming ETL.Engineered a real-time ingestion layer using Kafka to handle live data topics. Built a lightweight REST API to serve processed daily/hourly data to downstream applications and a simple GUI (PyQt5) for data validation.
- Models were trained with Spark MLlib (baseline linear/GBTs) and evaluated via time-series splits to avoid leakage.
- I exposed daily/hourly forecasts as Spark jobs writing back to HDFS + a lightweight REST layer, and tracked accuracy with MAE/RMSE for continuous improvements.

Mental Health Risk Prediction (Social & Behavioral Features)

- Designed an AI/ML pipeline to automate the prediction of mental-health risk from social/behavioral signals.
- Trained and compared Logistic Regression and Random Forest with stratified cross-validation, class-imbalance handling, and threshold tuning; evaluated using ROC-AUC/PR-AUC and F1.
- Delivered dashboards (Tableau/Power BI) to explain drivers and model lift, documented assumptions and limits

Professional Experience

Wells Fargo **Senior Software Engineer** Nov 2021–Jan 2024

- Maintained and scaled core banking microservices by building robust REST APIs using Java/Spring, ensuring high availability and transactional integrity. (Use "microservices" to sound more modern/cloud-native).
- Created a Selenium + TestNG automation framework from scratch, reducing regression cycles by ~40%
- Led monthly releases and triaged post-production issues, driving faster defect resolution and stakeholder updates
- Migrated core banking services to Azure and containerized with Docker to improve scalability and performance
- *Tech:* Java (Spring Boot/MVC), SQL, Docker, Azure, Jenkins, JUnit, Unix

SAP Labs **Software Development Engineer 2** Apr 2019–Oct 2021

- Delivered features for an automotive product on SAP HANA Cloud used by global clients
- Built a Kyma/Kubernetes POC that simplified runtime ops and improved traceability across data flows
- Managed post-production support and hotfixes using ServiceNow and Splunk to stabilize releases
- *Tech:* Spring Boot, HANA Cloud, GraphQL, Kyma/Kubernetes, Splunk, REST APIs

J.P. Morgan Chase & Co. **Software Engineer** Jan 2017–Apr 2019

- Built microservices and ingestion pipelines feeding analytics dashboards; improved availability via replication and indexing on Oracle/MongoDB
- Implemented data quality checks and metadata practices to support reliable reporting for internal stakeholders
- Engineered and automated critical ETL data pipelines using Airflow and Spark, supporting backend analytics systems and stakeholder reporting.
- *Tech:* Python, SQL, Oracle, MongoDB, Airflow, Spark, Tableau, Power BI, CI/CD

Capgemini **Software Engineer** Apr 2014–Aug 2016

- Built a multi-database abstraction layer (Core Java) to enhance complex transactional systems (Rapid Cash transfers), improving performance and portability across heterogeneous databases. (Use "abstraction layer" and "portability" to emphasize robust design).
- Delivered JasperReports reporting using Core Java and Oracle SQL