Aishvarya Salvi

+1 7164657631 | aishvary@buffalo.edu | www.linkedin.com/in/aishvaryasalvi | Buffalo, New York

Summary

Data-driven professional with 3+ years of experience in analytics, machine learning, and data engineering. Proficient in Python, SQL, Spark, Kafka, and GCP, with expertise in building scalable ML pipelines and recommendation engines. Adept at real-time data processing, model deployment, and interactive dashboard development. Strong foundation in MLOps, CI/CD, and containerization using Docker, Kubernetes, and Jenkins. Passionate about delivering actionable insights and production-ready solutions from complex datasets.

Experience

Software Engineer — Mediaocean 2023 – 2024

- Automated large-scale data validation and analytics workflows by developing Karate API & Selenium frameworks, increasing API regression coverage from 40% to 95% and reducing production defects by 30%.
- Orchestrated data-driven QA pipelines in CI/CD Jenkins, GitHub Actions with Docker, building Power BI dashboards for KPI and coverage tracking—cutting environment setup from 2 hours to 10 minutes, improving stability by 50%, and reducing MTTR by 40%.

Associate Software Engineer — Mediaocean 2021 – 2023

- Built SQL-based validation scripts and analytics reports to ensure data integrity across releases, improving coverage from 60% to 95% and reducing escaped defects by 40%.
- Partnered with product & engineering teams in Agile sprints to analyze requirements, monitor 200+ defect trends, and apply frontend/API inspection tools—cutting bug resolution time by 30% and sustaining a 97% build success rate.

Projects

E-Commerce Product Recommendation System *Python, ALS*

- Engineered a full-stack ALS-based recommendation engine on 2.7M+ RetailRocket events with optimized sparse matrix processing and real-time predictions, deploying a FastAPI backend + Streamlit frontend fully containerized with Docker Compose for scalable, production-ready delivery.

Real-Time Analytics and Churn Prediction Pipeline *Kafka, Spark*

- Built a real-time streaming pipeline processing 10K+ Kafka events per min with Spark Structured Streaming, storing data in PostgreSQL and delivering Logistic Regression churn predictions with 85% accuracy via FastAPI in less than 200ms, and developed a Streamlit dashboard with 15-second auto-refresh.

Heart Attack Risk Prediction Pipeline *Python, MLOps*

- Designed and deployed a machine learning pipeline for heart attack risk prediction, performing data wrangling, feature engineering, and exploratory analysis on 20+ clinical and lifestyle variables, implementing Logistic Regression, Random Forest, and XGBoost models, managing 7+ ML experiments with MLflow/DagsHub, and deploying a production-ready model via FastAPI and Streamlit on GCP.

Education

University at Buffalo - Master of Science in Data Science Dec 2025

Pune University - Bachelor of Engineering in Computer Science May 2021

Technical Skills

Programming Languages: Python, SQL, Java, R, C/C++

Frameworks & Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Streamlit, FastAPI, Selenium WebDriver, Karate API, TF-IDF, Apache Spark, A/B Testing

Databases & Data Storage: PostgreSQL, Oracle Database, SQLite, Snowflake

Developer Tools & Platforms: Git, GitHub, Docker, Docker Compose, Kubernetes, Jenkins, GitHub Actions, VS Code, PyCharm, IntelliJ, Eclipse, JIRA, Power BI, Tableau, Browser Developer Tools, AWS, GCP, Azure

Certifications

- AWS Cloud Practitioner Essentials
- Foundations of Data Science
- Jenkins Essential Training
- Python for Data Science, AI & Development
- Programming Foundations: APIs and Web Services
- Oracle Database