# Batch-Processing Data Architecture for Sentiment Analysis ML Application

**Abstract:**
This project develops a batch-processing data pipeline for sentiment analysis using the Amazon reviews dataset, leveraging BigQuery, PySpark, and MongoDB. It employs Docker for containerization, Airflow for orchestration, and Terraform for automated infrastructure provisioning. Microservices handle data ingestion and preprocessing, ensuring efficient and scalable data processing. The system is fully reproducible, deployed in the cloud, and integrates CI/CD for seamless updates. This architecture enables efficient data handling, automation, and scalability, making it a robust solution for large-scale sentiment analysis.