FROM bitnami/spark:3.4.0

# Switch to root for package installation

USER root

# Install Python pip and necessary Python dependencies

RUN install\_packages python3-pip && \

pip3 install --no-cache-dir boto3 pandas

# Add required Hadoop dependencies for S3

RUN install\_packages wget && \

wget https://repo1.maven.org/maven2/org/apache/hadoop/hadoop-aws/3.3.6/hadoop-aws-3.3.6.jar -P /opt/bitnami/spark/jars/ && \

wget https://repo1.maven.org/maven2/com/amazonaws/aws-java-sdk-bundle/1.12.569/aws-java-sdk-bundle-1.12.569.jar -P /opt/bitnami/spark/jars/

# Set the working directory

WORKDIR /app

# Copy your Spark application

COPY wine\_quality\_prediction.py /app/predict.py

# Entry point for the container

ENTRYPOINT ["spark-submit", "/app/predict.py"]