Creating EMR Cluster:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

Adding a Step to the EMR Cluster and run the spark application in the Cluster:

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Once the step is added, the initial status will be Pending and it will move to Running in few minutes.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The best model identified by CrossValidator and logs information are saved to the s3 bucket.

A screenshot of a computer

Description automatically generated

docker build ran successfully (Screenshot is captured after fixing the issues in the Docker file)

A computer screen with blue text

Description automatically generated

Spark application ran successfully by executing the docker run command

A black background with many small dots

Description automatically generated

Printing the Validation Dataset

A black background with many small squares

Description automatically generated with medium confidence

Application running successfully with the F1 score printed in the logs (boxed for reference)

A black screen with many small dots

Description automatically generated with medium confidence