

Cost Analysis:

In evaluating the overall cost-effectiveness of this application, it is apparent that the primary expense stems from operating the SageMaker endpoint, amounting to approximately \$5 per day. To optimize costs, consideration could be given to utilizing a smaller compute node for the endpoint. It's worth noting, however, that while this adjustment may reduce expenses, there is a potential trade-off in user experience due to prolonged result display times.



Outside of sagemaker the cost of this application is minimal due to the services used.

Frontend:

AWS amplify cost: Free (static webpage hosting)

Backend:

Lambda: Price per ms $\$0.0000000021$ * average run time of 100ms = $\$0.00000021$ per run or \$1 per roughly 4 million invocations.

API Gateway: Free for first 1 Million calls received, then \$3 per million

Auto Scaling:

The only service that would need scaling for this application would be the sagemaker endpoint as Lambda functions are serverless as is the frontend hosting.