

ABSTRACT

This research evaluates the performance of a livestock disease detection application for pinkeye using two machine learning models, MobileNet and VGG16, deployed on two cloud platforms, Google App Engine and Google Cloud Run. Testing was conducted using the Locust framework to assess metrics such as response time, total number of requests, request failures, and CPU usage. The test results show that for the MobileNet model, Google App Engine had an average response time of 118.53 ms without any request failures, while Cloud Run demonstrated a faster average response time of 88.94 ms. Both platforms fall under the "Very Good" category according to TIPHON standards. For the VGG16 model, Google App Engine exhibited a very high average response time of 25,804.68 ms, placing it in the "Poor" category according to TIPHON standards, while Cloud Run showed an average response time of 5,372.74 ms with some request failures. A comparison between the models indicates that MobileNet is more efficient than VGG16 on both platforms. CPU usage for MobileNet on Cloud Run is more efficient, with an average of 17.42% compared to 24.62% on App Engine. Cost evaluation shows that Cloud Run is more economical, with a total cost of Rp. 992,079.30 over 17 days, compared to App Engine, which reached Rp. 2,235,753.06. Additionally, a comparison with AWS App Runner shows that Google Cloud Run remains superior in terms of latency and cost, with an average latency of 18.00 ms for Cloud Run compared to 20.16 ms on App Runner, though the difference is not significant. The MobileNet model on Cloud Run is a better choice for applications prioritizing low cost and fast performance. However, for applications requiring more complex models like VGG16, the availability of instances and auto-scaling on Cloud Run should be considered to avoid request failures.

Keywords: *Google Cloud Run, Google App Engine, Machine Learning, Deployment, Locust, Docker, MobileNet, VGG16*