OpenX Intern Recruitment Task

Jakub Michnicki

April 26, 2024

1 Comparison of the performances of the web app solution and the TensorFlow solution.

This is a report of a Locust performance test conducted on a web app and TensorFlow solutions which are deployed in the Kubernetes environment. The maximum amount of simulated users is set to 100. I wanted to set the timer to 15 minutes for this simulation but linux inactivity screen lock is set to 5 minutes so the time was shortened to the one mentioned in the instruction. I have created a bar chart that consists of:

- Average time of a request
- Median of requests
- Minimal and maximal time of a request

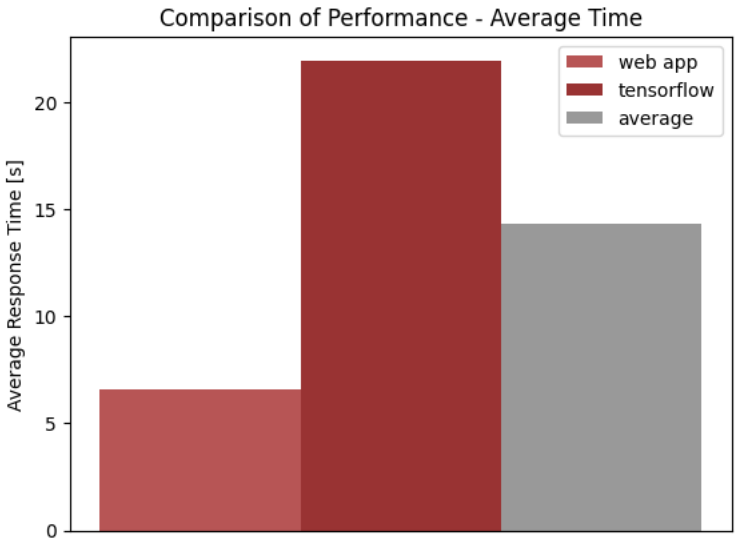

Figure 1: Average time of a request

Implemented web app solution performs much better on average compared to the TensorFlow approach. This might be caused by the fact, that every time a request towards TensorFlow solution is sent, the model is being loaded which is not an optimal solution. To solve this the model should be loaded and passed to the function responsible for the calculations and that is a thing I would potentially change in this project.

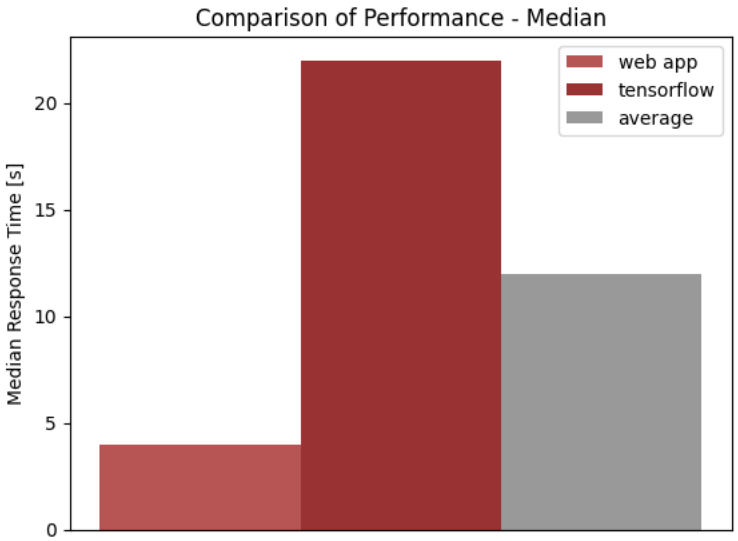

Figure 2: Median of requests

Using median value comaprison is also important in some cases since it is robust against outliers. Here, only the web app time has changed and it performs even faster in this comparison.

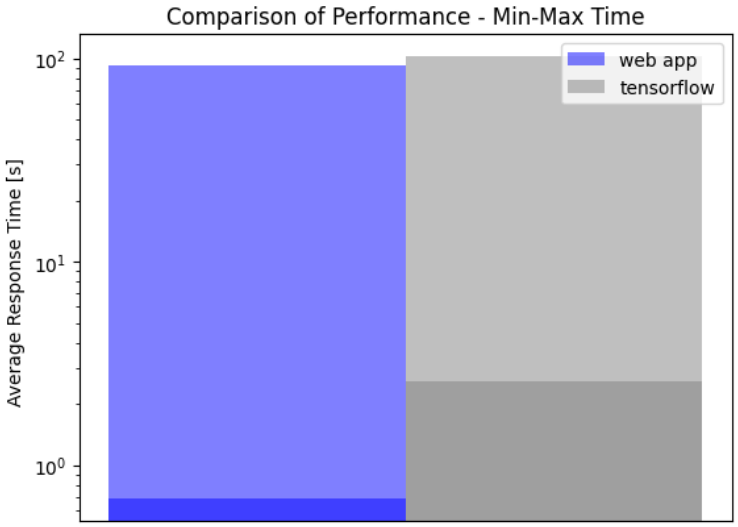

Figure 3: Shortest and longest request time

The darker shade of a color represents the shortest request time, the lighter one - the longest. There is a logarithmic scale applied to the y-axis so that these values are visible since the difference between the min-max values are big. The shortest time needed for the request to go through is still way lower on the web app approach. The minimum for web app is 0.7s and for TensorFlow it is 2.6s.

2 Conclusion

Web app approach seems to be performing better than the TensorFlow model but considering the model is not that complicated since it only measures one variable depending on another, its performance might be even slower in a more complicated scenario. We should still remember AI should be used in more complicated situations where calculating something 'by hand' is not so easy anymore though.