Adeyemi Kolawole

Professor McManus

ITAI 3377

11 February 2025

Test and Validate the Model

Run Inference on the Simulated Edge Device

Testing the deployed model is crucial. Edge Impulse allows us to run inference and check performance metrics like accuracy and inference speed.

Observations and Adjustments

- **Accuracy**: The model reached about 98% accuracy, meaning it correctly identified most handwritten digits.
- **Latency**: The model's response time was quick enough for real-time processing, making it well-suited for edge deployment.
- **Edge Device Performance**: The model ran smoothly with minimal lag, showing that even a small, low-power device can handle AI computations effectively.

Reflective Journal

Challenges Faced

Diving into edge AI was both exciting and frustrating. Converting a TensorFlow model into TensorFlow Lite wasn't as simple as flipping a switch. Understanding how to optimize the model for a small device required learning about quantization and performance tuning. Another challenge was dealing with authentication issues when uploading the model to Edge Impulse—getting the API key configuration right took longer than expected!

Learning Outcomes

This project taught me the importance of efficiency in AI. Training a model is one thing, but making sure it runs smoothly on a small device is another challenge entirely. I also gained a deeper appreciation for how edge computing can make AI more responsive by eliminating cloud

dependency. Seeing the model process data in real time without needing an internet connection was truly mind-blowing!

Application of Knowledge

This experience gave me a clearer understanding of how AI can be used in real-world applications like smart cameras, IoT devices, and automated monitoring systems. Deploying AI on edge devices opens the door to countless possibilities—from AI-powered security cameras that don't need internet access to smart appliances that can learn user preferences locally.

Conclusion

This project was a whirlwind of learning, problem-solving, and excitement. Seeing an AI model go from training to deployment on an edge device was incredibly rewarding. It also highlighted areas for improvement, like optimizing model size even further and reducing inference time. Moving forward, I'm eager to explore more about how AI can be integrated into everyday devices, making them smarter and more efficient. The possibilities are endless!