**Naomi Berumen**

**ITAI-3377 Patricia McManus**

**L03 Reflective Journal**

**Reflective Journal: Deploying a Simple AI Model on Edge Impulse**

**Introduction**

This reflective journal captures my journey through setting up, training, deploying, and testing a simple AI model on Edge Impulse. Throughout the process, I encountered various challenges, made important discoveries, and gained a deeper understanding of model deployment in an edge computing environment.

**Setup and Initial Challenges**

At the very start, I realized that setting up a proper development environment is not as straightforward as it seems. Installing Python was an unexpected challenge—I initially had an incorrect version, which led to TensorFlow installation errors. Fixing that alone took a significant amount of time, and I learned just how critical it is to verify dependencies before diving into a project. Additionally, I was using Visual Studio Code for the first time, so navigating through the interface and setting up the correct workspace added another layer of complexity.

One major takeaway here was that having the right tools set up correctly from the beginning can save hours of debugging later.

**Training and Model Conversion**

Training the CNN model on MNIST felt straightforward in theory but not in execution. I initially trained the model for only 5 epochs, which turned out to be insufficient for real-world performance. At the time, I thought the high validation accuracy meant the model was solid, but once deployed, it became clear that the accuracy didn't translate well when tested on real images.

Converting the trained model into TensorFlow Lite (TFLite) was actually one of the easier steps. The conversion process worked without major errors, and saving the `.tflite` file was seamless. However, I later learned that just because a model is successfully converted does not mean it will work perfectly on deployment.

**Deployment and Testing**

Deploying to Edge Impulse was another interesting part of the journey. Uploading the model required setting up an API key, and the interface itself had many settings that required manual adjustments. One of the biggest surprises was that Edge Impulse initially expected a live camera feed, which forced me to adjust my deployment approach.

When testing the model, the classification results were not what I expected. For example:

Testing on "3" → Model thought it was a "2" or "8"(wrong!)

Testing on "0" → Model had low confidence and was confused between "0" and "2".

Testing on "1" → The worst performance; the model essentially guessed randomly.

This was a huge realization moment: A model that seems to work well during training can completely fall apart in real-world scenarios.

**Key Lessons Learned**

* Training Epochs Matter: Five epochs were not enough. More training cycles (15-20 epochs) would improve accuracy significantly.
* Preprocessing Consistency is Critical: The images fed into Edge Impulse need to be preprocessed exactly the same way as the training data. Even slight variations in brightness, noise, or scaling could throw off predictions.
* Debugging is an Expected Part of AI Workflows: Every step in the process required troubleshooting—from software installation to deployment issues. Patience and structured debugging are key.
* Deployment is Not the Same as Training: Just because a model performs well on test data does not mean it will work well once deployed. Real-world testing is the real test.

**Final Thoughts**

Despite all the obstacles, this project was an eye-opening experience. I walked away with a deeper understanding of not just model training, but the full lifecycle of edge AI deployment. The final model did not perform as well as expected, but that in itself was a learning experience—failure is part of the iteration process. If I revisit this, I now know exactly what to fix: more training epochs, an extra convolutional layer, and ensuring preprocessing is consistent across all environments.

At the end of the day, this exercise reinforced one of the most important lessons in AI: machine learning is not magic—it’s a continuous process of trial, error, and improvement.