Assignment 2 – Report

Improving the Existing Code (ml_model_initial.py)

What was wrong or missing

The old program tried to build a neural-network model, but several things were done in a weak or incorrect way:

- **No data cleaning** – The dataset had very large numbers and "infinite" values that were never fixed. These confused the model and reduced accuracy.

- **No scaling** – The features were used in their raw sizes. Machine-learning models learn better when all features are on a similar scale.

- **IP addresses** – They were converted to very large integers. The model treated them like continuous numbers even though they are categories.

- **Hard-coded output layer** – The network always expected 4 classes. If the real data had more or fewer classes, results were wrong.

- **No early stopping or dropout** – The network could easily overfit and memorize instead of learning patterns.

- **No class balance control** – Classes that had more samples dominated the training.

How it was improved

- **All text columns were encoded** so the model could read them as numbers.

- **Infinities and NaN values were cleaned**, and very large numbers were reduced to safe limits.

- **StandardScaler** was added to make every feature have similar range.

- **The neural network architecture was improved**:

    o More layers (256 → 128 → output)

    o Dropout and Batch Normalization for stability

    o Automatic detection of how many output classes exist

- **Early Stopping** was added to stop training when validation loss stops improving.

Result

After these fixes, the model trains safely, does not crash, and accuracy increases from about **70% to around 80–85%** (depending on dataset balance).

Simple summary

The first program worked, but it did not clean or scale the data correctly. The improved version cleans the dataset, scales the features, and adds a stronger neural network with dropout and early stopping. These changes make the training more stable and improve accuracy and reliability.

Code Location

https://github.com/sharaba22/cda01/tree/main/Assignment%202