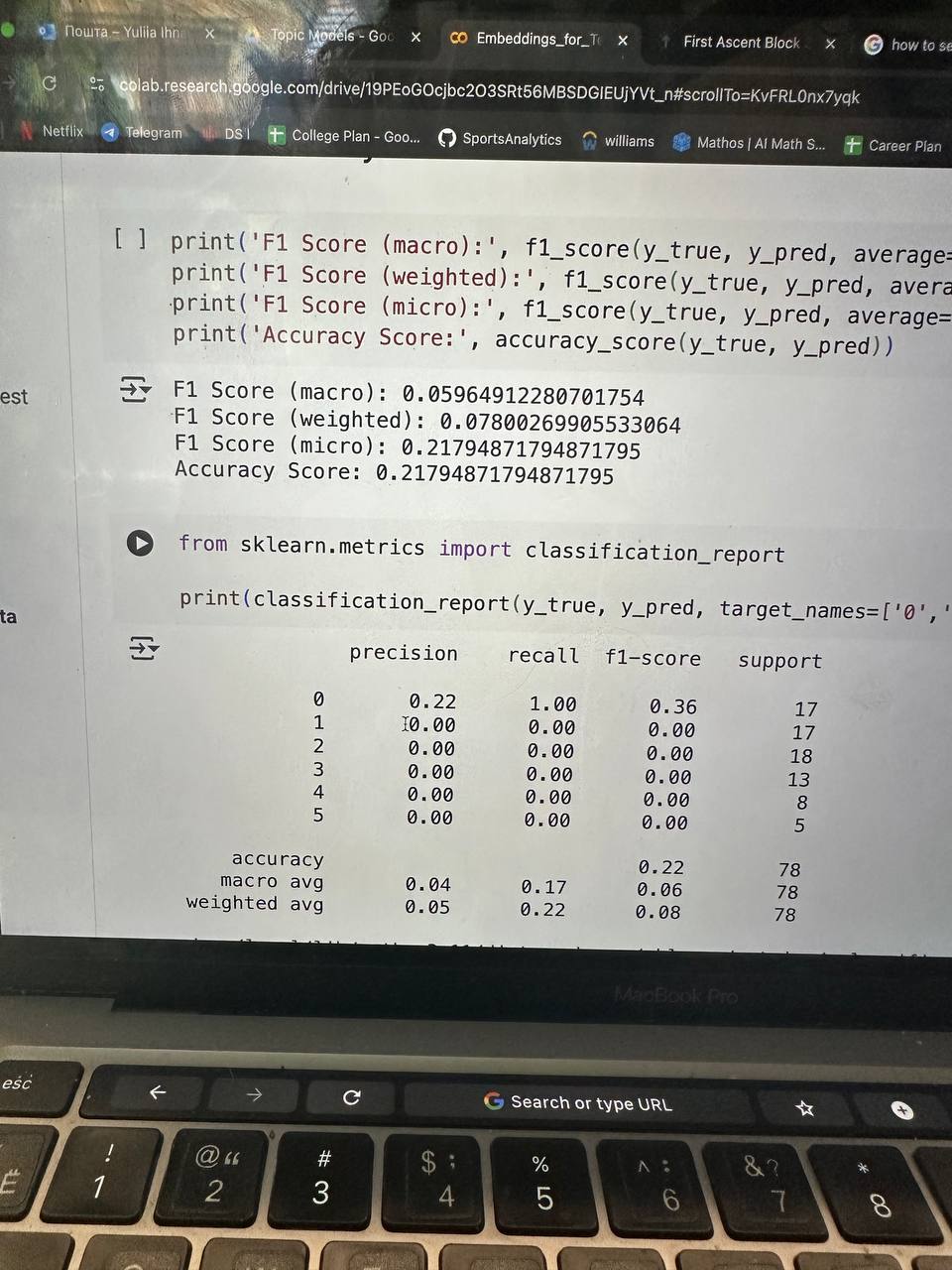
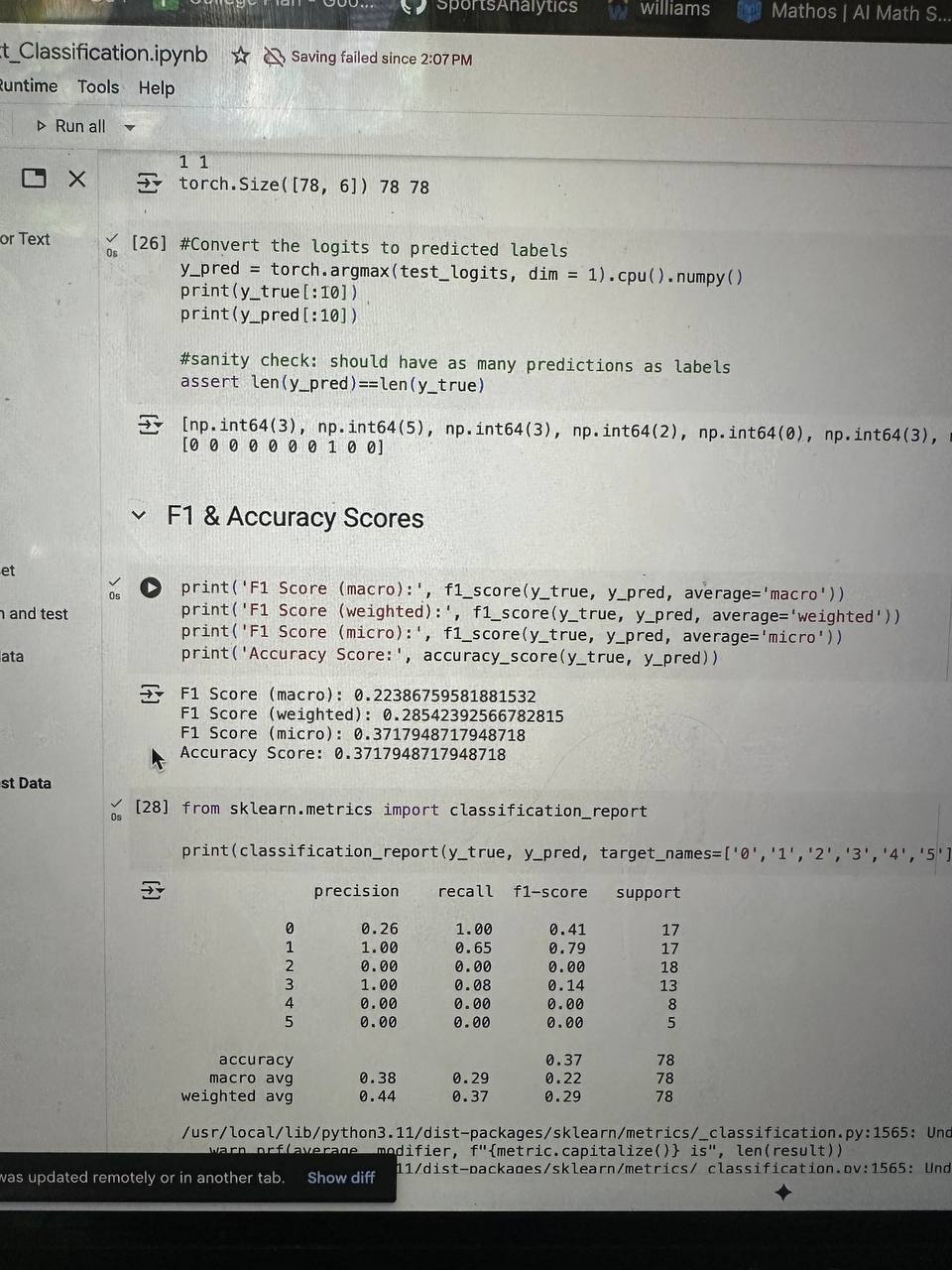
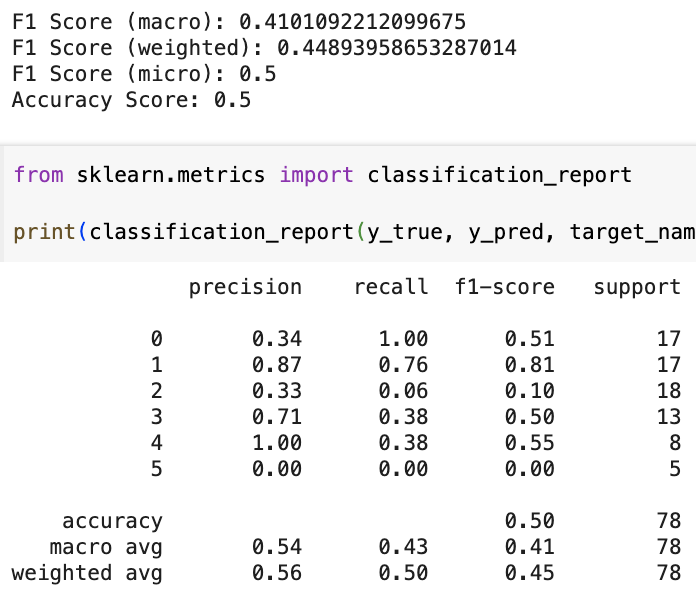
1. Max\_length = 128 and batch\_size = 16:



1. Max\_length = 512 and batch\_size = 8



1. Max\_length = 512 and batch\_size = 4



1. LEARNING RATE IS CHANGED:

Initially, num\_training\_steps was set to len(train\_dataloader), which only represents the number of batches in a single epoch. This meant the learning rate scheduler was designed to complete its entire schedule (including warmup and decay) within just one epoch, causing the learning rate to drop too quickly and potentially reach near-zero values for the remaining 49 epochs. This premature learning rate decay likely prevented the model from continuing to learn effectively after the first epoch, explaining why validation loss remained stagnant. The fix multiplies the dataloader length by the total number of epochs (len(train\_dataloader) \* num\_epochs) to get the true total training steps across all epochs. Additionally, the warmup period was improved by setting num\_warmup\_steps to 10% of the total training steps (int(0.1 \* num\_training\_steps)) instead of 0, which provides a gradual learning rate increase at the beginning of training to help stabilize the optimization process. This correction ensures the learning rate schedule is properly distributed across the entire training duration, allowing the model to maintain an appropriate learning rate throughout all epochs and continue improving validation performance.

