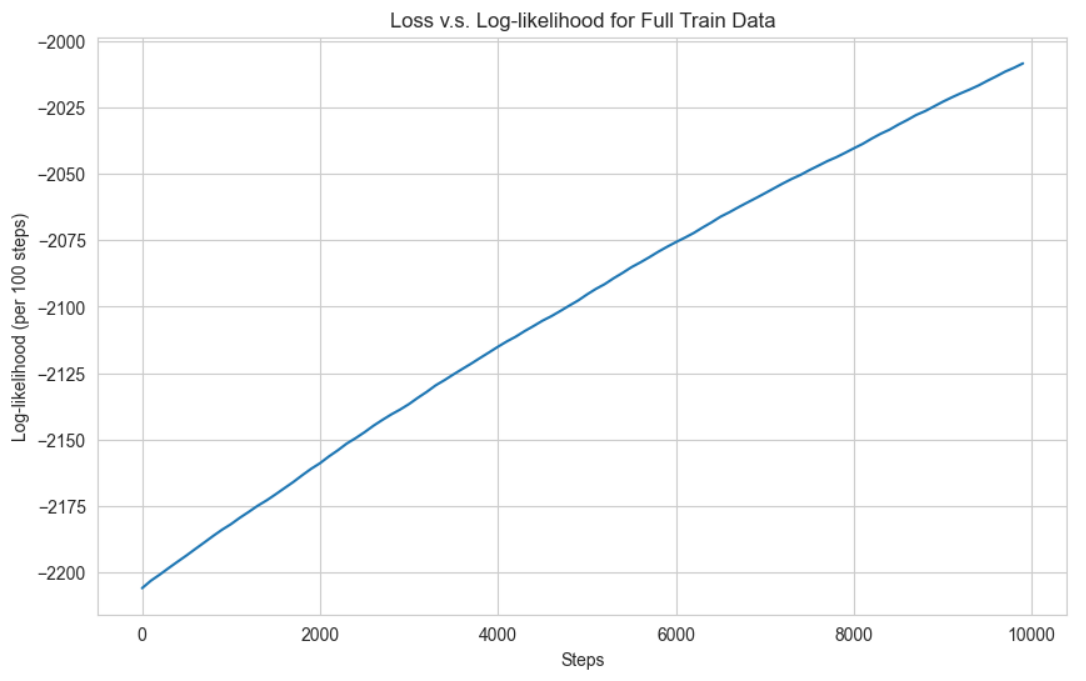
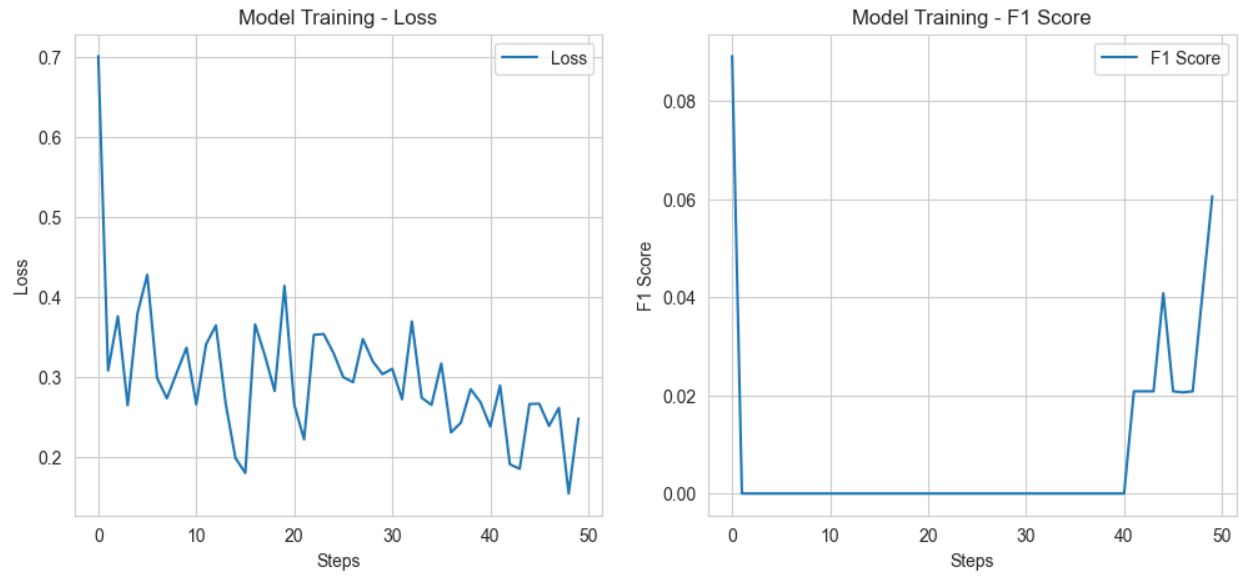
**Username on Kaggle: yanzhuo2001**

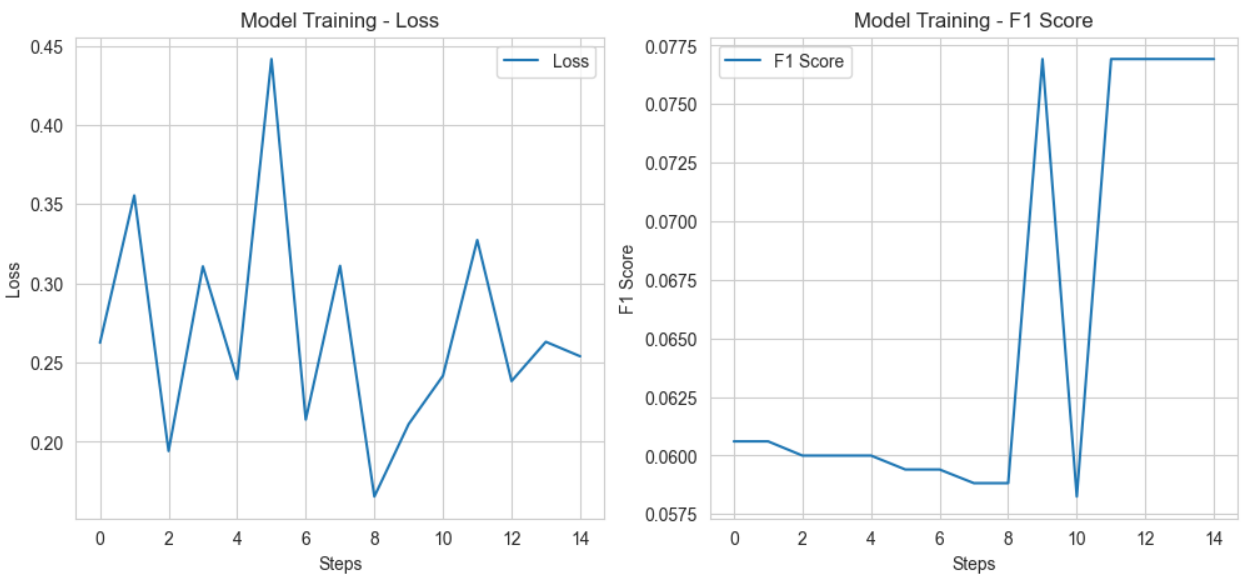
**Part 2.1: Plot log-likelihood**

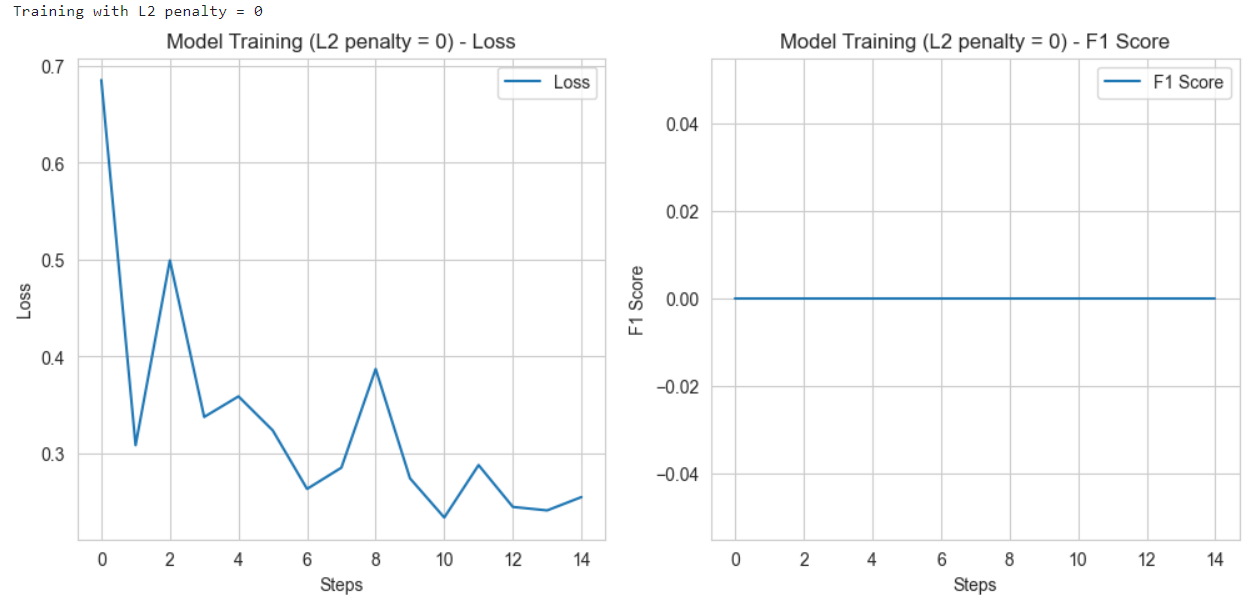


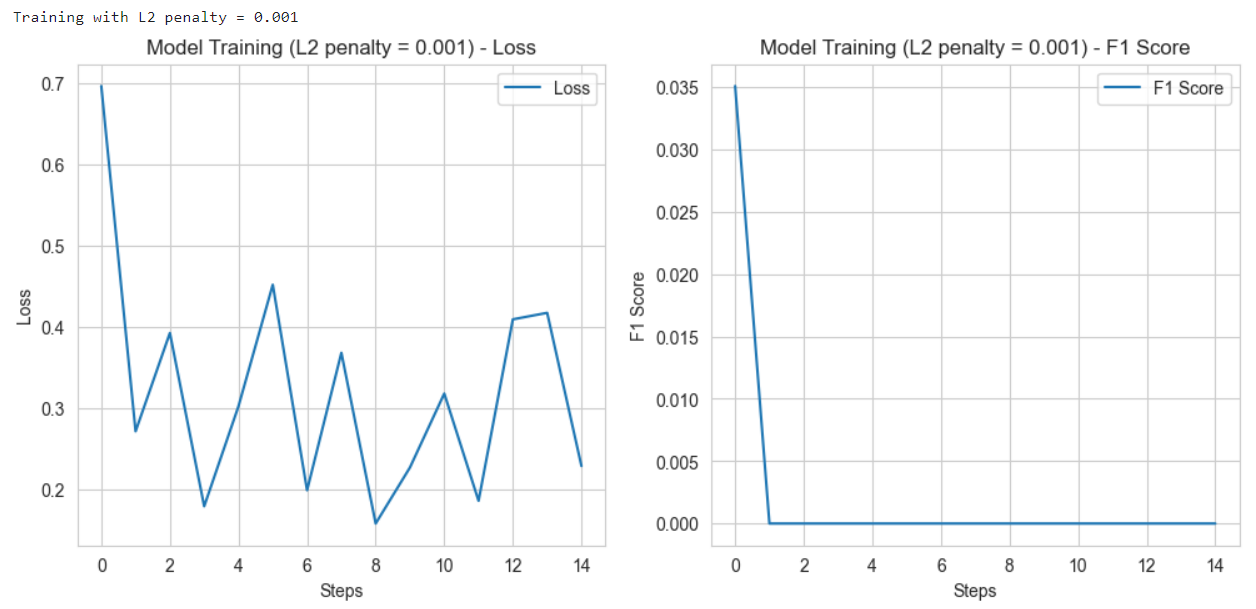
**Part 3.1: 1000 steps training and reporting losses every 20 steps**

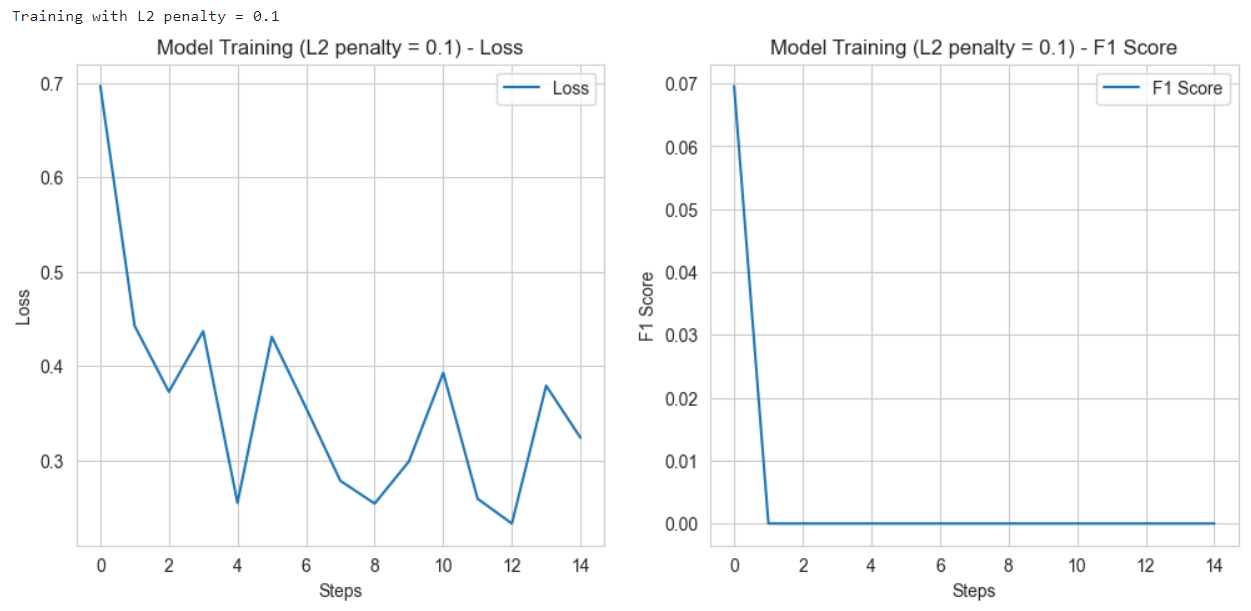
****

**Part 3.2: Train at least 1 epoch and calculate the loss and F1 score**

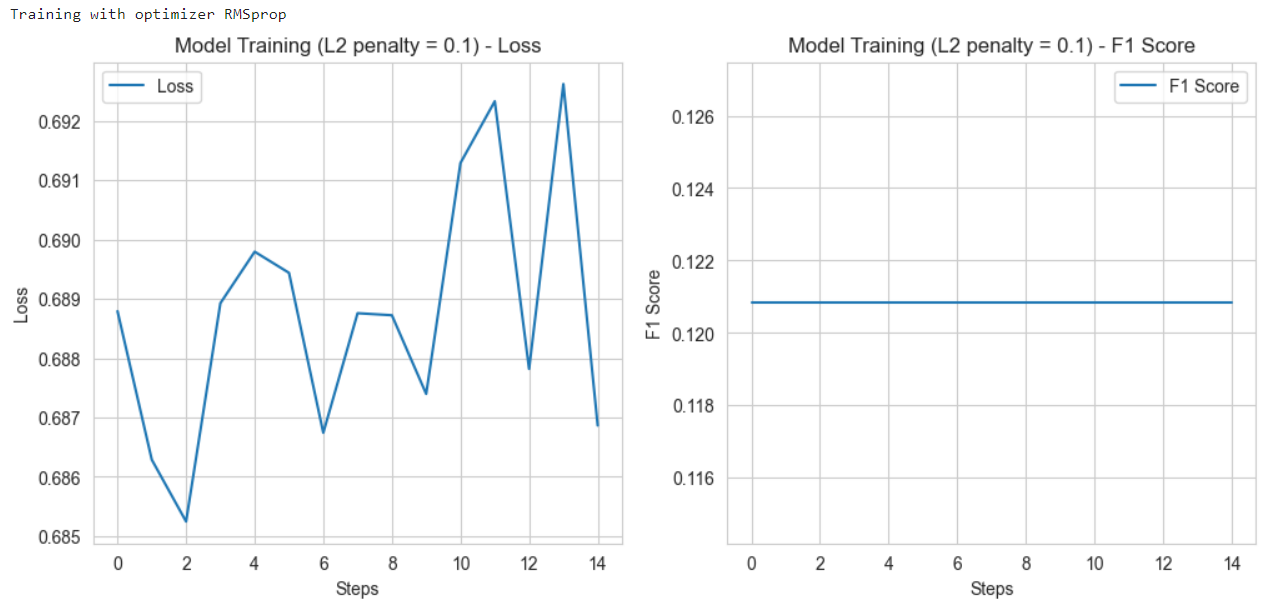
****

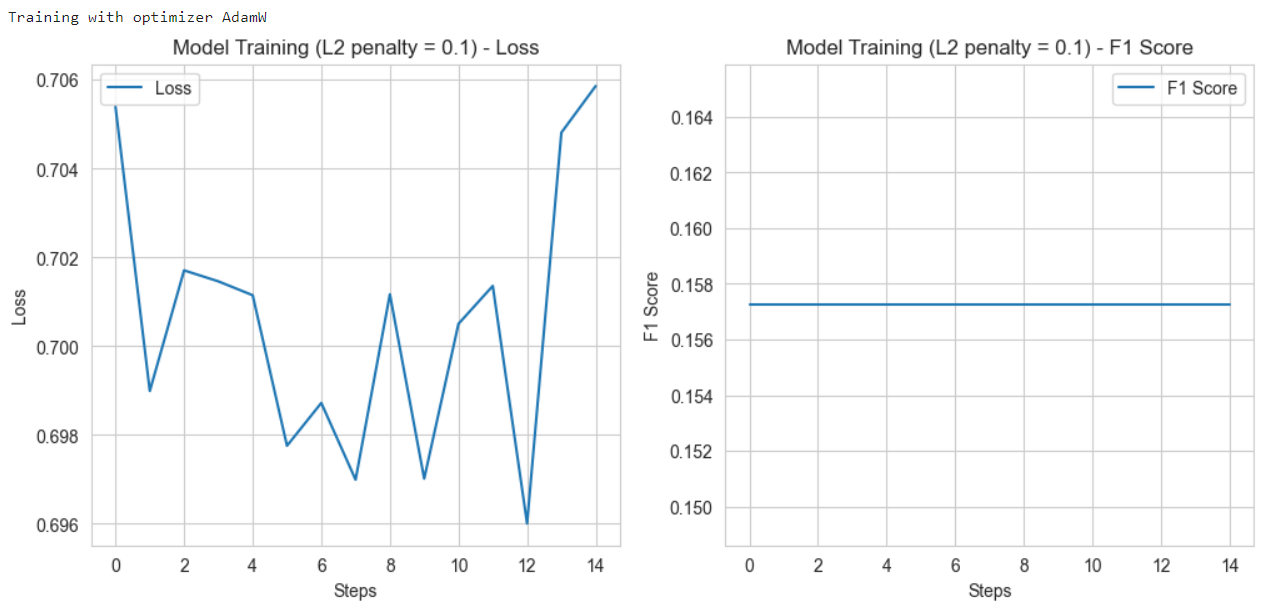
**Part 3.3：Add regularization (L2 penalty)**

****

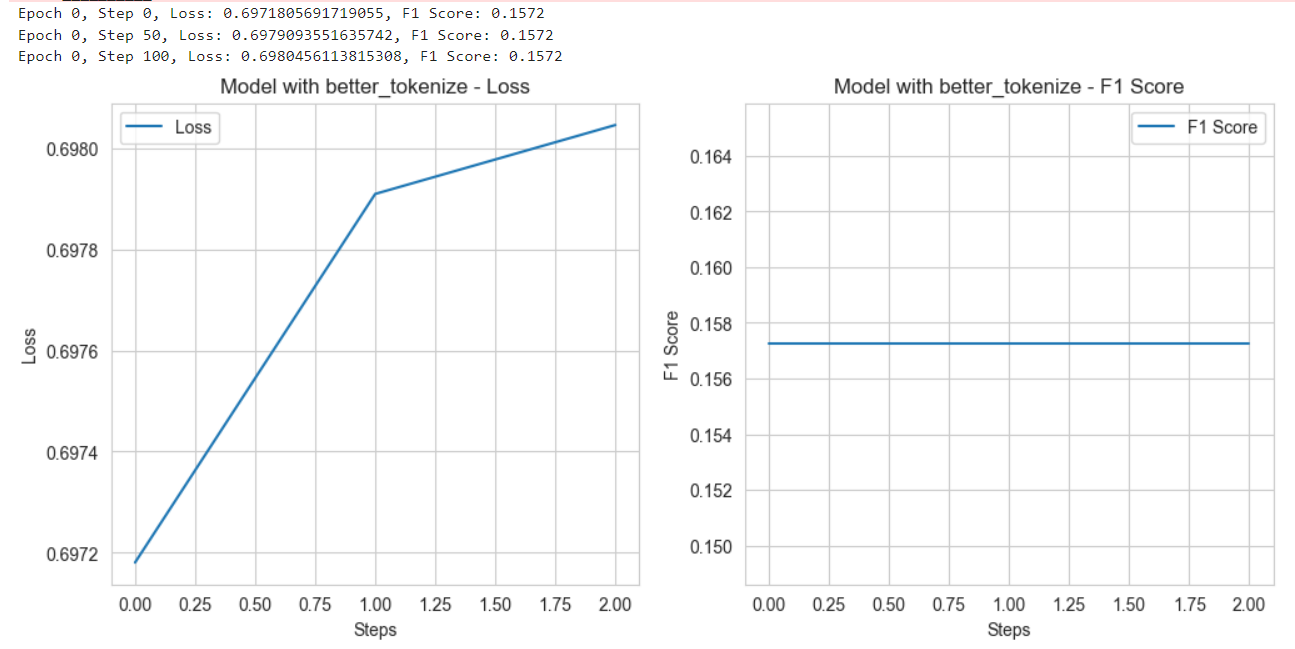
****

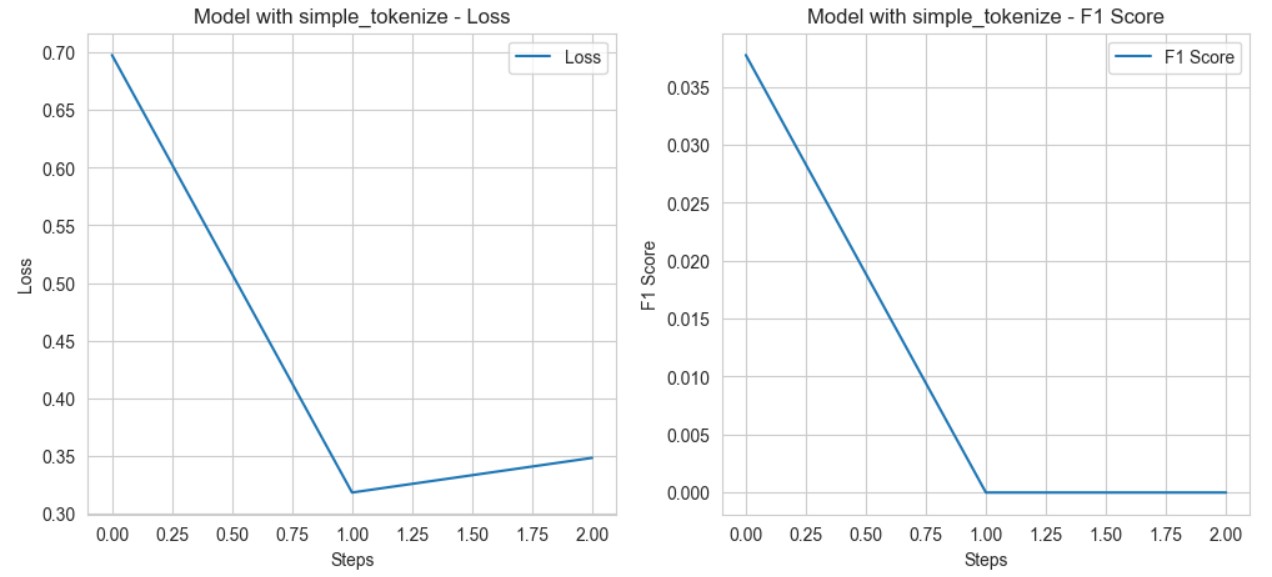
**Part 3.4: Use different optimizers**

****

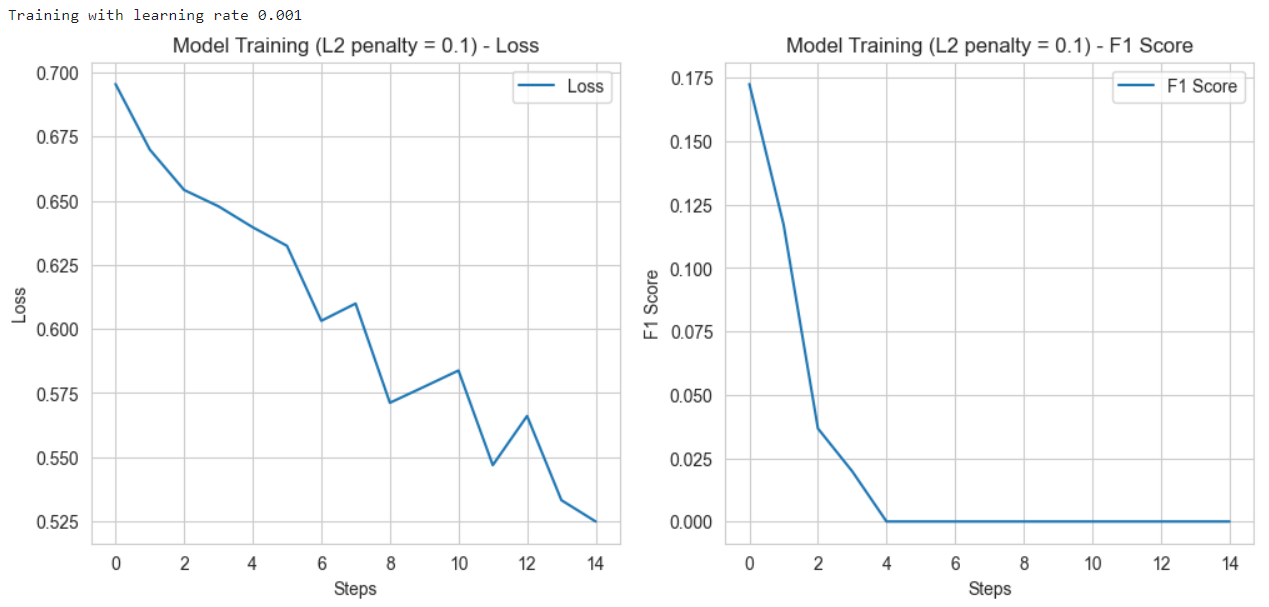
****

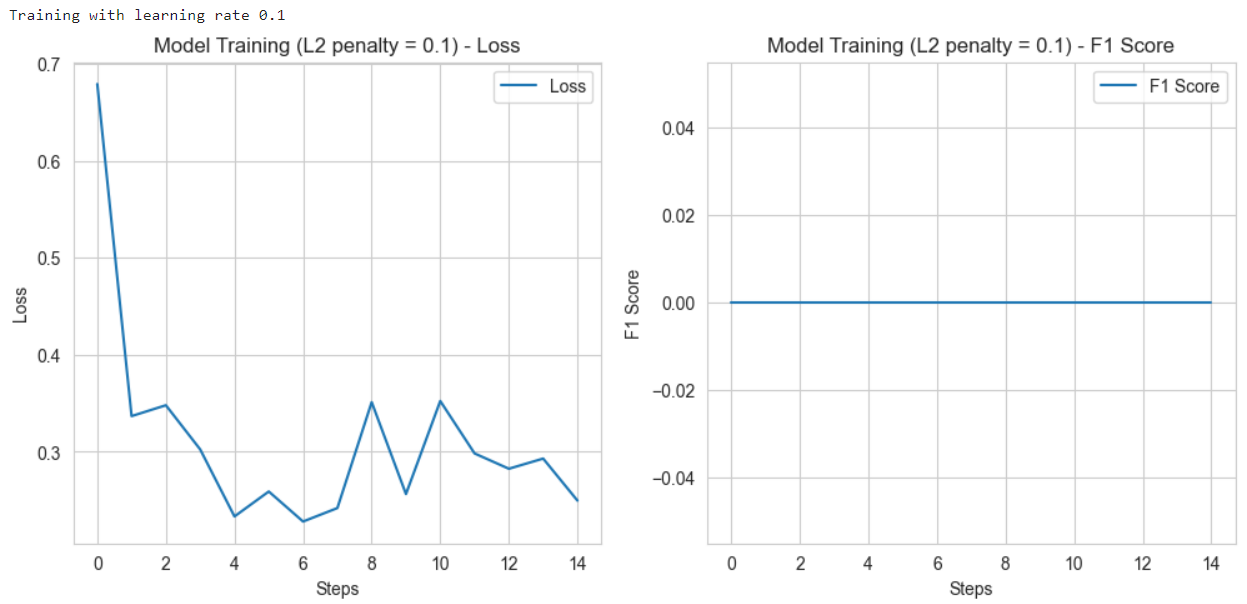
**Part 3.5: Comparison of different word segmentation methods**

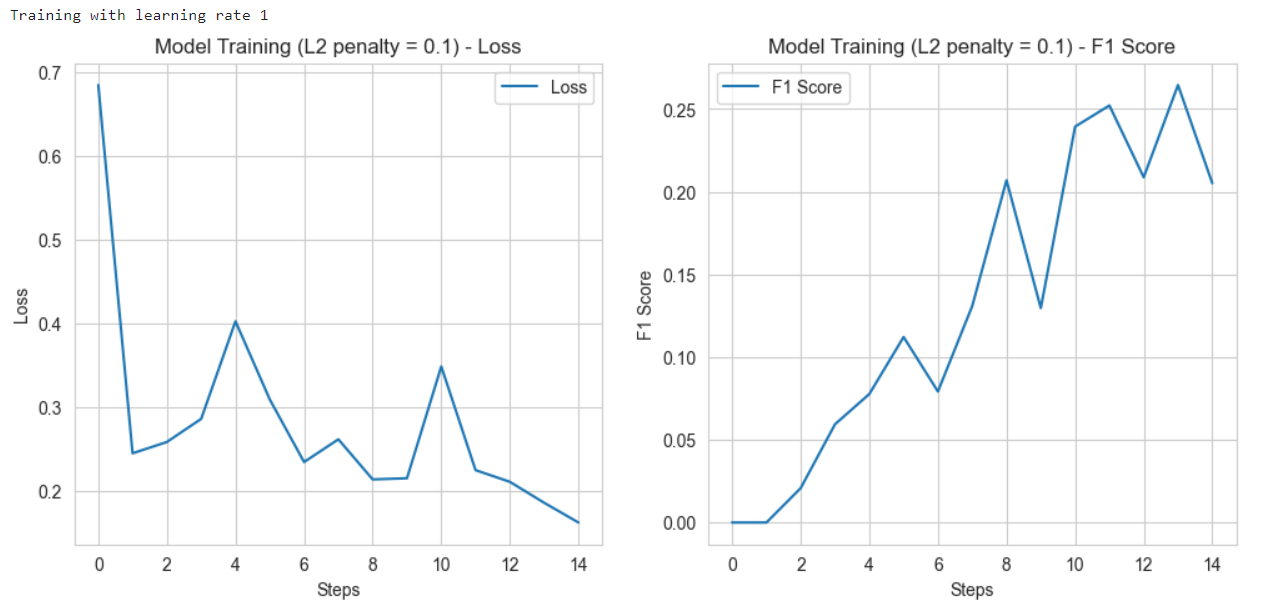
****

****

**Part 3.6: The effect of learning rate**

****

****

****