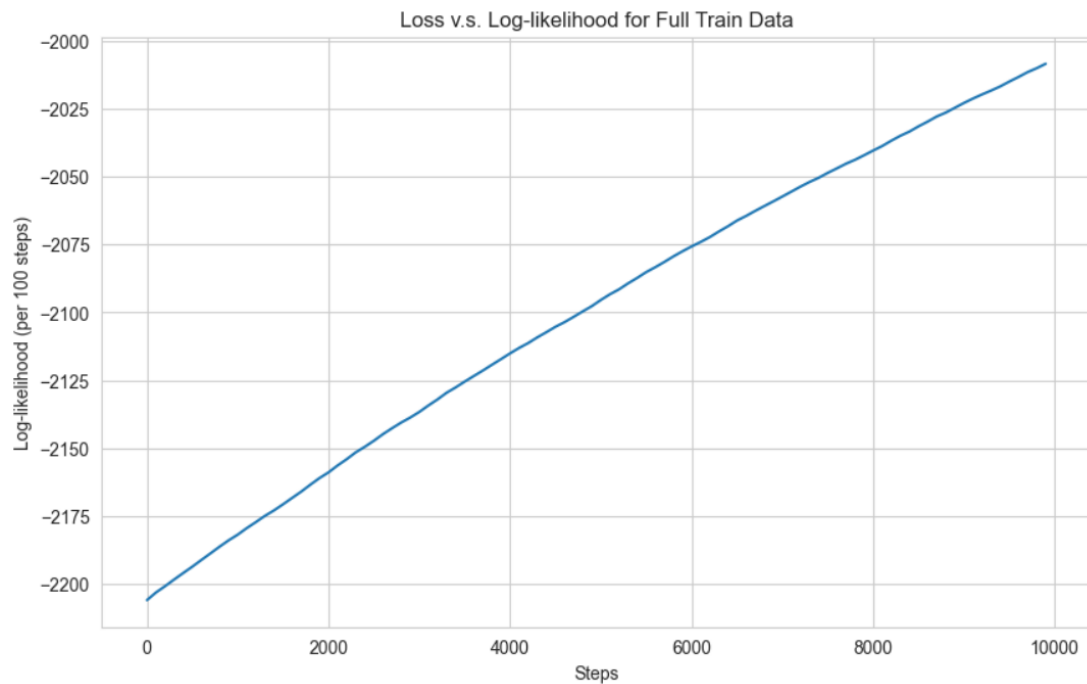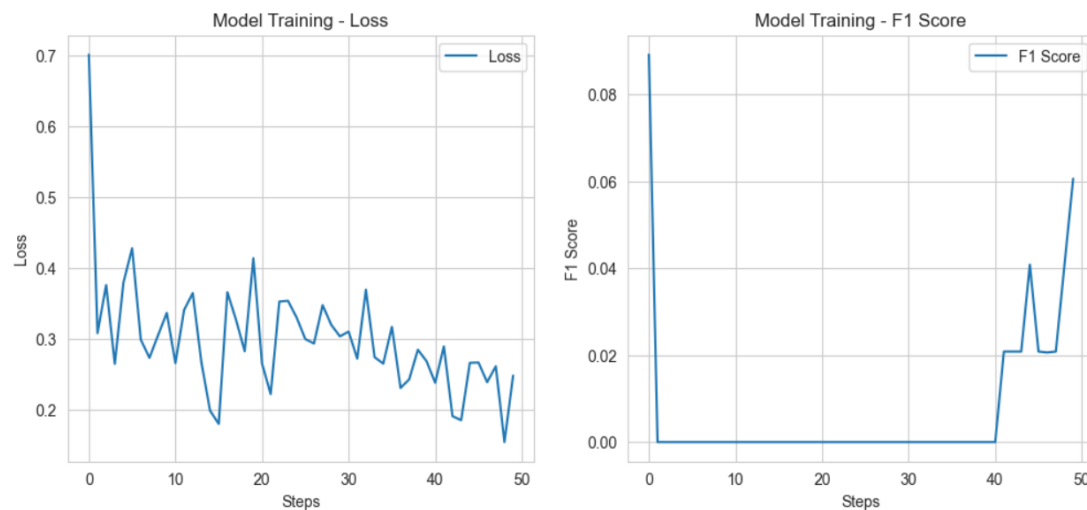**Username on Kaggle: yanzhuo2001**
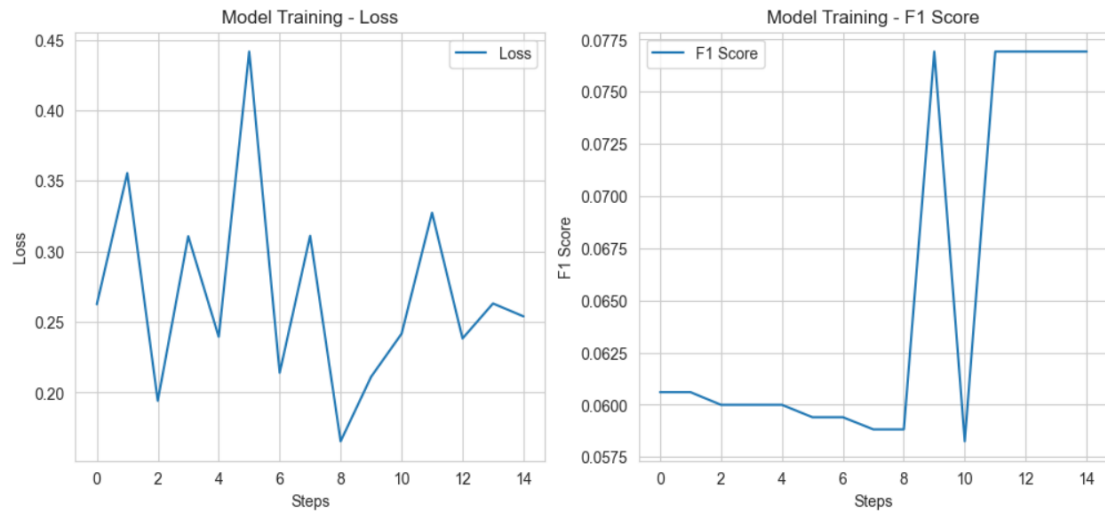
Part 2.1: Plot log-likelihood

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

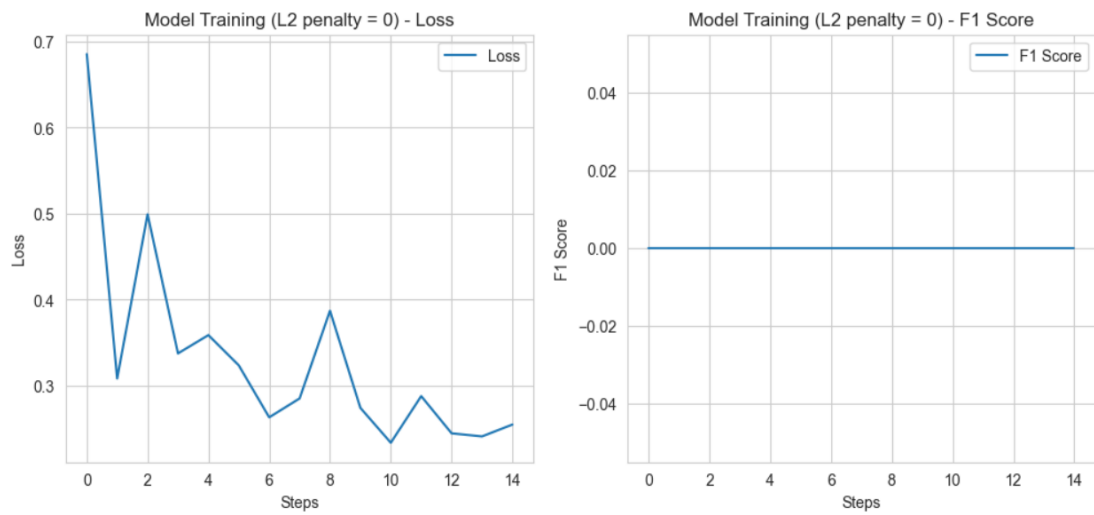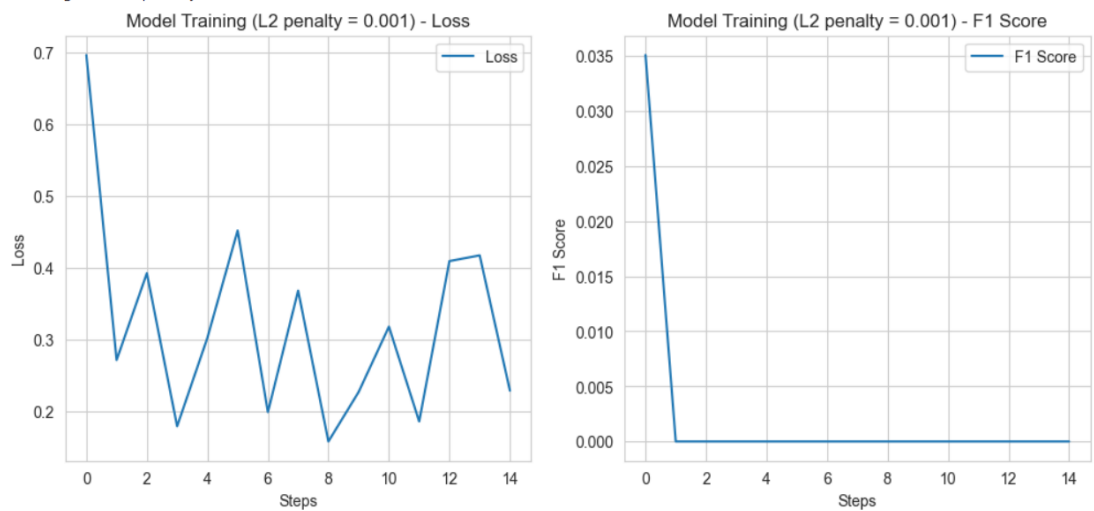**Part 3.2: Train at least 1 epoch and calculate the loss and F1 score**

Part 3.3：Add regularization (L2 penalty)
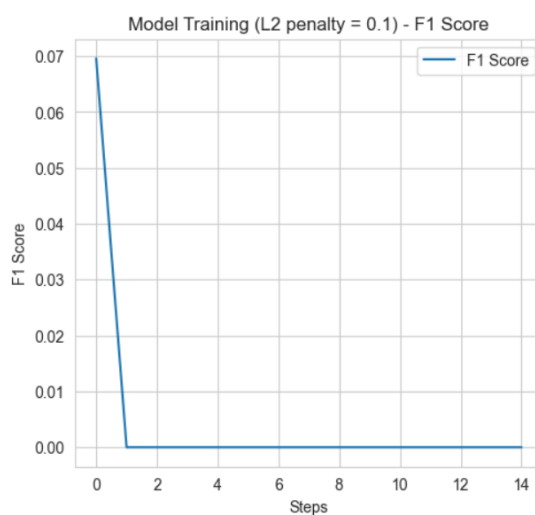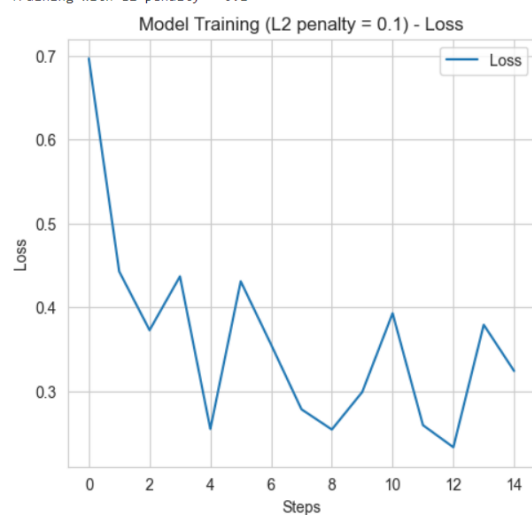
Training with L2 penalty = 0

Training with L2 penalty = 0.001
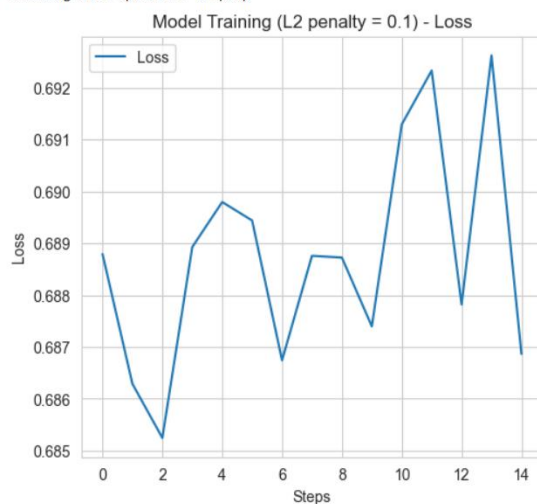
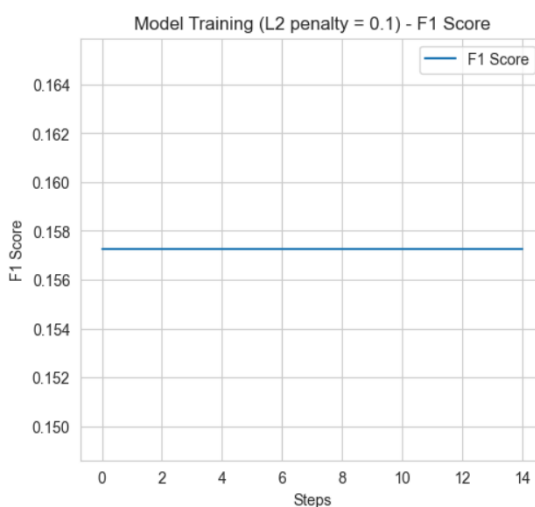Model Training (L2 penalty = 0.1) - Loss

Model Training (L2 penalty = 0.1) - F1 Score

Part 3.4: Use different optimizers

Training with optimizer RMSprop

Model Training (L2 penalty = 0.1) - Loss

Model Training (L2 penalty = 0.1) - F1 Score
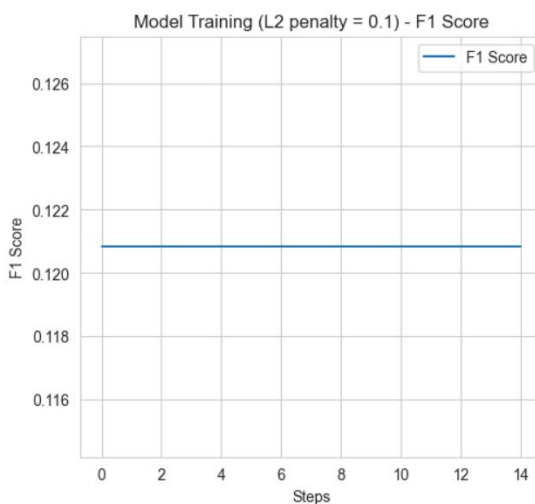
Training with optimizer AdamW

Model Training (L2 penalty = 0.1) - Loss

Model Training (L2 penalty = 0.1) - F1 Score
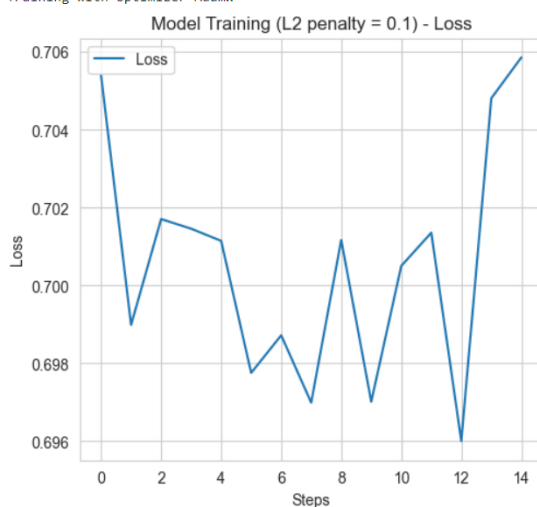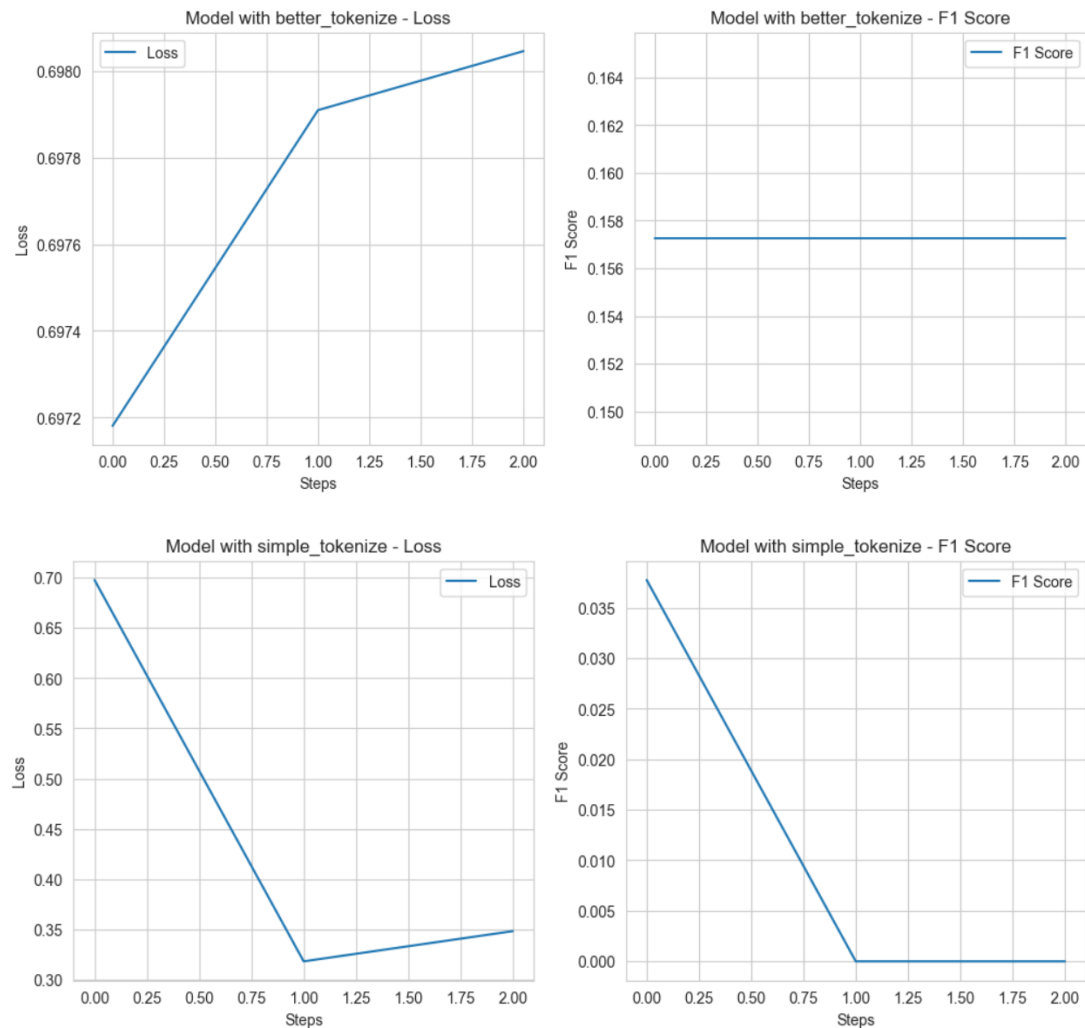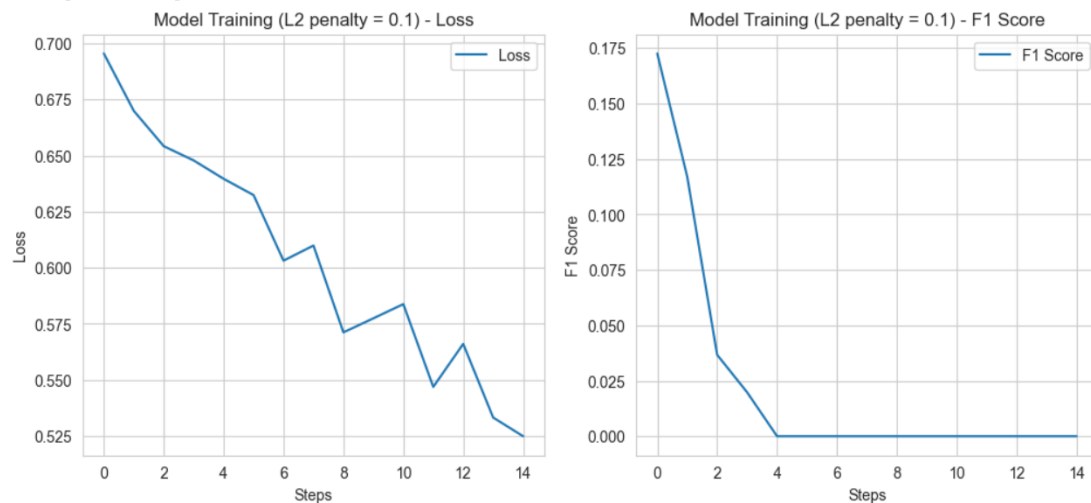
Part 3.5: Comparison of different word segmentation methods

```
Epoch 0, Step 0, Loss: 0.6971805691719055, F1 Score: 0.1572
Epoch 0, Step 50, Loss: 0.6979093351635742, F1 Score: 0.1572
Epoch 0, Step 100, Loss: 0.6980456113815308, F1 Score: 0.1572
```
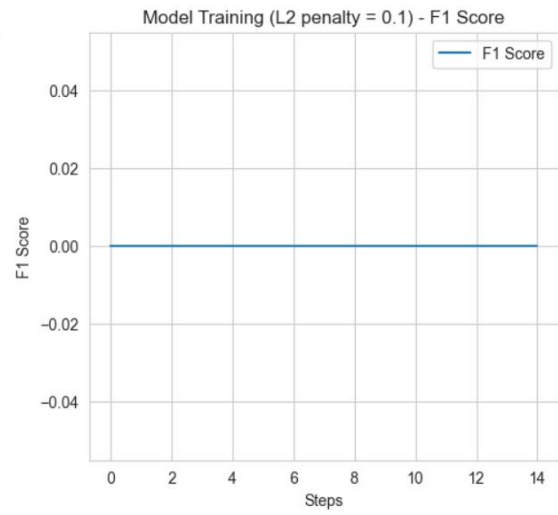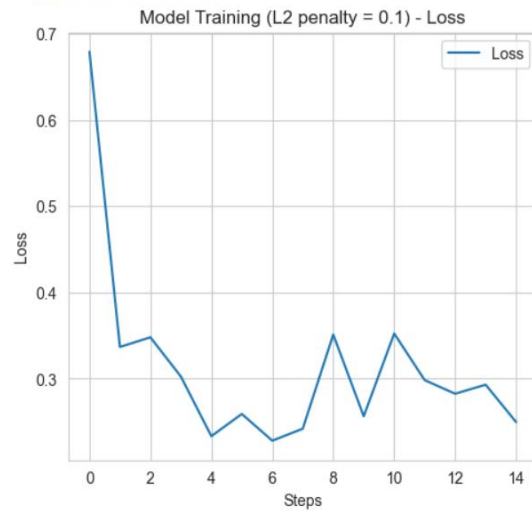

Part 3.6: The effect of learning rate

```
Training with learning rate 0.001
```

Model Training (L2 penalty = 0.1) - Loss

Model Training (L2 penalty = 0.1) - F1 Score

Model Training (L2 penalty = 0.1) - Loss

Model Training (L2 penalty = 0.1) - F1 Score