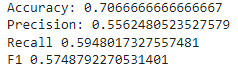
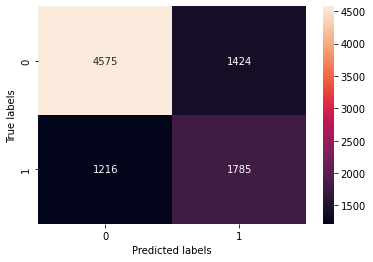
**Experimental Analysis**

The extent of similarity for each pair was denoted in the range of 0-1 by these models. After analysis the most suitable threshold value of 0.65 is taken to classify the each input as similar or dissimilar.

**Case-1: using ‘bert-base-nli-mean-tokens’ model:**





Number of True positives- 1785

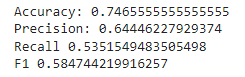
Number of True negatives- 4575

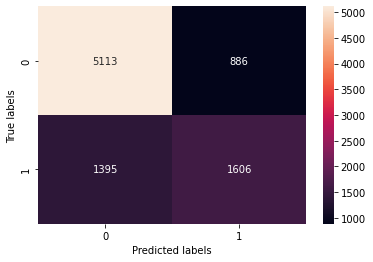
Number of False positives- 1424

Number of False negatives- 1216

**Case-2 :**

1. **‘bert-base-uncased’ pretrained model fine tuned on NLI datasets:**





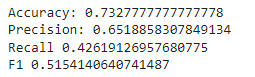
Number of True positives- 1606

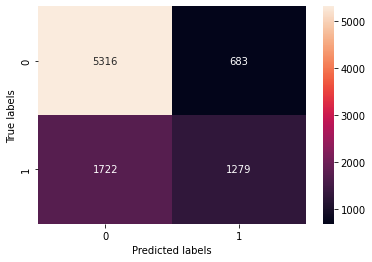
Number of True negatives- 5113

Number of False positives- 886

Number of False negatives- 1395

1. **Pretrained ‘bert-base-uncased’ model with fine tuning:**





Number of True positives- 1279

Number of True negatives- 5316

Number of False positives- 683

Number of False negatives- 1722

