

Dataset: <https://www.kaggle.com/kazanova/sentiment140>

Task: Classification of twitter message: whether it is positive or negative comment

Approach: word to vector embedding, with convolutional neural network for classification

Output:

Use 100,000 twitter messages (out of original 1.4M records), with 64,000 messages for training, 16,000 messages for validation and 16,000 messages for testing

Score 73.6% accuracy

```
Epoch 4/20
64000/64000 [=====] - 281s 4ms/step - loss: 0.3616 - acc: 0.8393 - val_loss: 0.5197 - val_acc: 0.7606
Epoch 5/20
64000/64000 [=====] - 281s 4ms/step - loss: 0.2973 - acc: 0.8721 - val_loss: 0.5945 - val_acc: 0.7557
Epoch 6/20
64000/64000 [=====] - 281s 4ms/step - loss: 0.2443 - acc: 0.8980 - val_loss: 0.6613 - val_acc: 0.7449
Epoch 7/20
64000/64000 [=====] - 281s 4ms/step - loss: 0.2002 - acc: 0.9178 - val_loss: 0.7364 - val_acc: 0.7409
20000/20000 [=====] - 12s 581us/step
Test set
Loss: 0.755
Accuracy: 0.736
```