Dataset: Adverse Drug Events from week 9

Machine: 4 core CPU, 30 GiB of RAM

Part 1:

Varying Maxlen (epochs set to 2)
Maxlen = 100

Epoch 1/2
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 25s 41ms/step - accuracy: 0.7548 - loss: 0.4946 - val_accuracy: 0.8325 - val_loss: 0.3745
Epoch 2/2
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 22s 38ms/step - accuracy: 0.8954 - loss: 0.2667 - val_accuracy: 0.8367 - val_loss: 0.3808

Maxlen = 200

Epoch 1/2
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 44s 72ms/step - accuracy: 0.7555 - loss: 0.4917 - val_accuracy: 0.8165 - val_loss: 0.4073
Epoch 2/2
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 80s 68ms/step - accuracy: 0.8905 - loss: 0.2657 - val_accuracy: 0.9252 - val_loss: 0.1834

Maxlen = 225

Epoch 1/2
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 49s 81ms/step - accuracy: 0.7561 - loss: 0.4970 - val_accuracy: 0.9443 - val_loss: 0.1800
Epoch 2/2
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 45s 76ms/step - accuracy: 0.8891 - loss: 0.2764 - val_accuracy: 0.8382 - val_loss: 0.3732

Maxlen = 250 :
Your notebook tried to allocate more memory than is available. It has restarted.

Part 2:

Epoch = 5

Epoch 1/5
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 49s 81ms/step - accuracy: 0.7533 - loss: 0.4952 - val_accuracy: 0.8971 - val_loss: 0.2525

Epoch 2/5
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 45s 76ms/step - accuracy: 0.8928 - loss: 0.2600 - val_accuracy: 0.9345 - val_loss: 0.1801
Epoch 3/5
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 83s 77ms/step - accuracy: 0.9438 - loss: 0.1498 - val_accuracy: 0.9324 - val_loss: 0.1660
Epoch 4/5
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 81s 76ms/step - accuracy: 0.9703 - loss: 0.0829 - val_accuracy: 0.8480 - val_loss: 0.4536
Epoch 5/5
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 45s 77ms/step - accuracy: 0.9777 - loss: 0.0594 - val_accuracy: 0.8863 - val_loss: 0.3657

Since results are deteriorating I will reduce number of epochs to 3

Epoch = 3

Epoch 1/3
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 49s 81ms/step - accuracy: 0.7441 - loss: 0.5082 - val_accuracy: 0.7864 - val_loss: 0.4631
Epoch 2/3
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 79s 76ms/step - accuracy: 0.8902 - loss: 0.2756 - val_accuracy: 0.8663 - val_loss: 0.3112
Epoch 3/3
588/588 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 44s 75ms/step - accuracy: 0.9454 - loss: 0.1461 - val_accuracy: 0.9562 - val_loss: 0.1145

Part 3:

[Best model among all] Glove vectors (maxlen 225, epochs 5, batch size 64):

Epoch 1/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 43s 137ms/step - accuracy: 0.7395 - loss: 0.5073 - val_accuracy: 0.7713 - val_loss: 0.4765
Epoch 2/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 39s 131ms/step - accuracy: 0.8852 - loss: 0.2811 - val_accuracy: 0.8926 - val_loss: 0.2596
Epoch 3/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 38s 130ms/step - accuracy: 0.9393 - loss: 0.1565 - val_accuracy: 0.8867 - val_loss: 0.2832
Epoch 4/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 42s 133ms/step - accuracy: 0.9611 - loss: 0.0989 - val_accuracy: 0.8250 - val_loss: 0.4649
Epoch 5/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 38s 129ms/step - accuracy: 0.9829 - loss: 0.0498 - val_accuracy: 0.9560 - val_loss: 0.1238

Word2Vec from current dataset

Epoch 1/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 23s 74ms/step - accuracy: 0.7265 - loss: 0.5372 - val_accuracy: 0.8348 - val_loss: 0.3608
Epoch 2/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 20s 67ms/step - accuracy: 0.8767 - loss: 0.2858 - val_accuracy: 0.8608 - val_loss: 0.3218
Epoch 3/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 20s 68ms/step - accuracy: 0.9381 - loss: 0.1650 - val_accuracy: 0.8767 - val_loss: 0.3032
Epoch 4/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 20s 68ms/step - accuracy: 0.9712 - loss: 0.0884 - val_accuracy: 0.8276 - val_loss: 0.4660
Epoch 5/5
294/294 ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 20s 66ms/step - accuracy: 0.9764 - loss: 0.0633 - val_accuracy: 0.8869 - val_loss: 0.3579

Observations:

300 maxlen would have been ideal too but I could not find a machine capable of running that computation. However a maxlen of 225 covers a sufficient majority. Intuitively 225 maxlen models should perform better than 200 maxlens but that is not necessarily the case.

For batch lens 32 and 64, 5 epochs end up being enough or more than enough for this datasets. The models begin to overfit after the 4th/5th epoch.

GloveVectors and GoogleNews Vectors performed almost similarly and very good. When forming vectors from the documents itself all values seen in validation accuracy were below 0.9.