Executive summary
Sentiment analysis evaluation for the Dublin City Council - [Colab notebook](#)
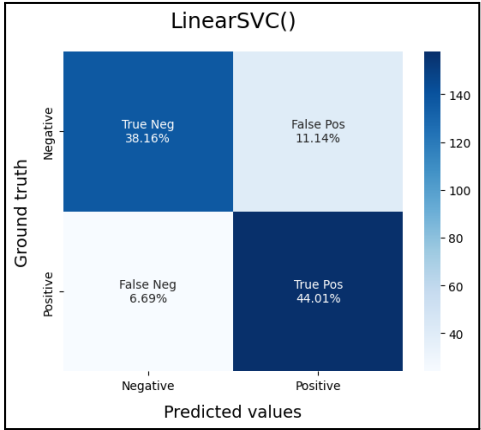<u>Baseline Model:</u>
Trained with the whole dataset (1.6M samples).

S140_test dataset evaluation for a SVC model and a Logistic Regression model:

```
              precision    recall  f1-score   support

           0       0.85      0.77      0.81       177
           1       0.80      0.87      0.83       182

    accuracy                           0.82       359
   macro avg       0.82      0.82      0.82       359
weighted avg       0.82      0.82      0.82       359

CPU times: user 54.5 s, sys: 609 ms, total: 55.1 s
Wall time: 57.5 s
```
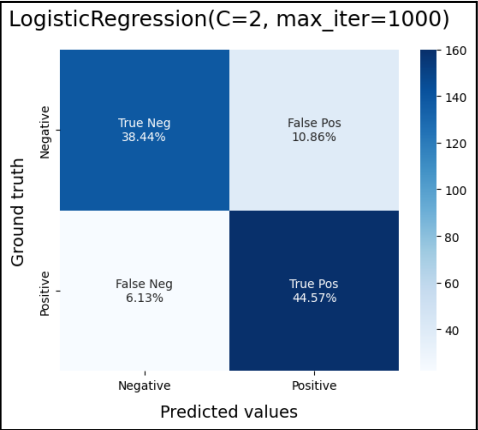
```
              precision    recall  f1-score   support

           0       0.86      0.78      0.82       177
           1       0.80      0.88      0.84       182

    accuracy                           0.83       359
   macro avg       0.83      0.83      0.83       359
weighted avg       0.83      0.83      0.83       359

CPU times: user 5min 47s, sys: 2min 36s, total: 8min 24s
Wall time: 5min 59s
```
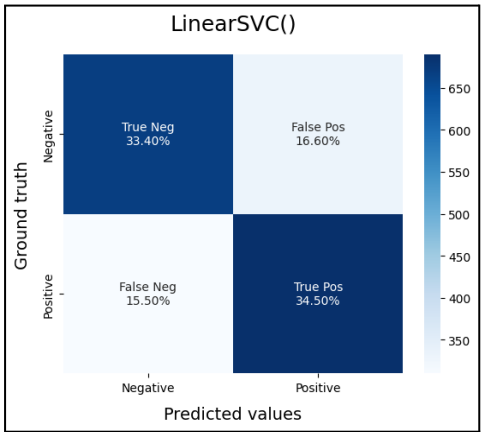

Dublin_test dataset evaluation:

```
              precision    recall  f1-score   support

           0       0.68      0.67      0.68      1000
           1       0.68      0.69      0.68      1000

    accuracy                           0.68      2000
   macro avg       0.68      0.68      0.68      2000
weighted avg       0.68      0.68      0.68      2000

CPU times: user 41.4 s, sys: 434 ms, total: 41.9 s
Wall time: 41.9 s
```
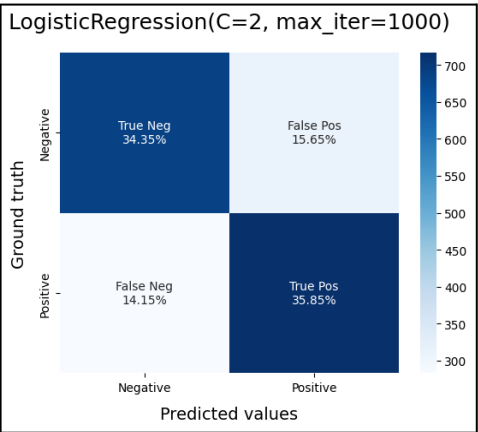
```
              precision    recall  f1-score   support

           0       0.71      0.69      0.70      1000
           1       0.70      0.72      0.71      1000

    accuracy                           0.70      2000
   macro avg       0.70      0.70      0.70      2000
weighted avg       0.70      0.70      0.70      2000

CPU times: user 4min 51s, sys: 2min 30s, total: 7min 21s
Wall time: 4min 55s
```
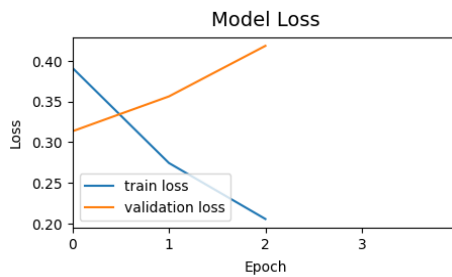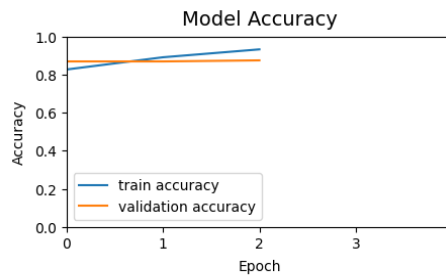
<u>Deep Learning Model:</u>
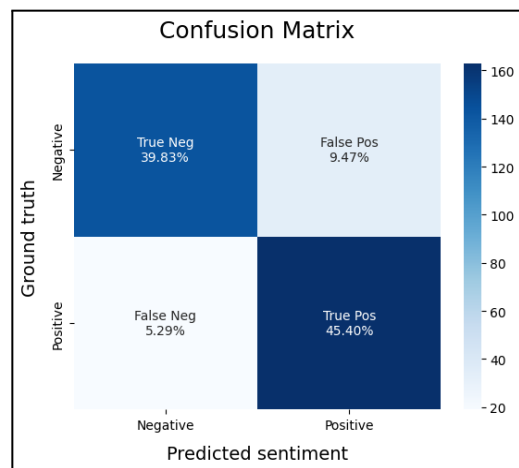*Trained with a subset (160k samples)*

Model Accuracy

Model Loss

Hyperparameters:
- *Batch size: 16*
- *Learning rate: 2e-5*
- *Dropout: 0,3*
- *Number of epochs: 3*

S140_test dataset:

	precision	recall	f1-score	support
negative	0.88	0.81	0.84	177
positive	0.83	0.90	0.86	182
accuracy			0.85	359
macro avg	0.86	0.85	0.85	359
weighted avg	0.85	0.85	0.85	359

Confusion Matrix

Dublin_test dataset:

	precision	recall	f1-score	support
negative	0.70	0.57	0.63	1000
positive	0.64	0.75	0.69	1000
accuracy			0.66	2000
macro avg	0.67	0.66	0.66	2000
weighted avg	0.67	0.66	0.66	2000

Confusion Matrix