

Instituto Tecnológico y de Estudios Superiores de Monterrey



**Tecnológico
de Monterrey**

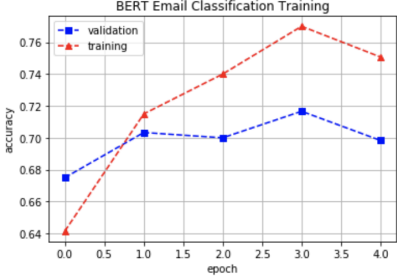
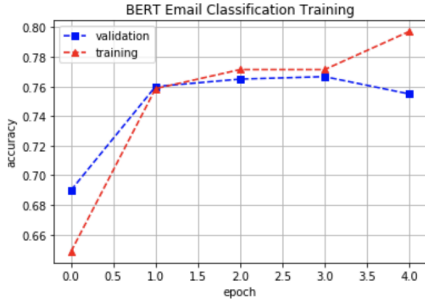
**Inteligencia artificial avanzada para la ciencia de datos I
(Gpo 101)**

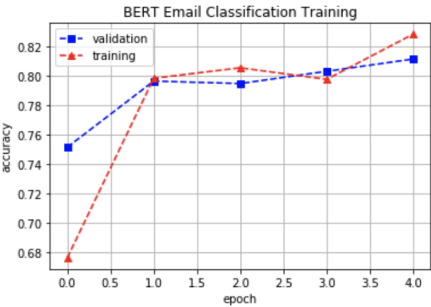
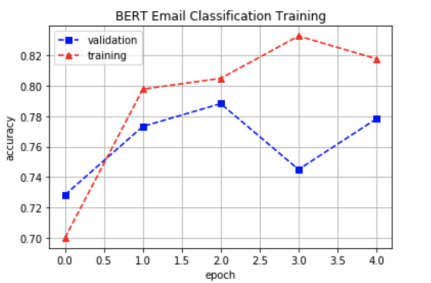
Equipo 4

Tabla comparativa

Integrantes:

Facundo Colasurdo Caldironi A01198015

| | BERT | |
|---|---|--|
| <p>(a) Nsamp = 1000, maxtokens = 50, maxtokenlen = 20</p> |  | <p>Epoch 1/5 1400/1400 [=====] - 11s 8ms/sample - loss: 0.6131 - acc: 0.6679 - val_loss: 0.5569 - val_acc: 0.6967</p> <p>Epoch 2/5 1400/1400 [=====] - 8s 6ms/sample - loss: 0.5333 - acc: 0.7329 - val_loss: 0.5435 - val_acc: 0.6967</p> <p>Epoch 3/5 1400/1400 [=====] - 8s 6ms/sample - loss: 0.4946 - acc: 0.7614 - val_loss: 0.5723 - val_acc: 0.7000</p> <p>Epoch 4/5 1400/1400 [=====] - 9s 6ms/sample - loss: 0.4714 - acc: 0.7743 - val_loss: 0.5813 - val_acc: 0.6950</p> <p>Epoch 5/5 1400/1400 [=====] - 9s 6ms/sample - loss: 0.4300 - acc: 0.8043 - val_loss: 0.5935 - val_acc: 0.7017</p> |
| <p>(b) Nsamp = 1000, maxtokens = 100, maxtokenlen = 100</p> |  | <p>Epoch 1/5 1400/1400 [=====] - 16s 11ms/sample - loss: 0.5905 - acc: 0.6743 - val_loss: 0.5169 - val_acc: 0.7367</p> <p>Epoch 2/5 1400/1400 [=====] - 13s 10ms/sample - loss: 0.4948 - acc: 0.7621 - val_loss: 0.4780 - val_acc: 0.7550</p> <p>Epoch 3/5 1400/1400 [=====] - 14s 10ms/sample - loss: 0.4614 - acc: 0.7807 - val_loss: 0.4694 - val_acc: 0.7717</p> <p>Epoch 4/5 1400/1400 [=====] - 14s 10ms/sample - loss: 0.4390 - acc: 0.7771 - val_loss: 0.5380 - val_acc: 0.7367</p> <p>Epoch 5/5 1400/1400 [=====] - 15s 11ms/sample - loss: 0.4471 - acc: 0.7807 - val_loss: 0.5000 - val_acc: 0.7517</p> |

| <p>(c) Nsamp = 1000, maxtokens = 200, maxtokenlen = 200</p> |  <table border="1"><caption>BERT Email Classification Training Data for (c)</caption><thead><tr><th>epoch</th><th>validation</th><th>training</th></tr></thead><tbody><tr><td>0.0</td><td>0.75</td><td>0.68</td></tr><tr><td>1.0</td><td>0.79</td><td>0.79</td></tr><tr><td>2.0</td><td>0.79</td><td>0.80</td></tr><tr><td>3.0</td><td>0.80</td><td>0.79</td></tr><tr><td>4.0</td><td>0.81</td><td>0.82</td></tr></tbody></table> | epoch | validation | training | 0.0 | 0.75 | 0.68 | 1.0 | 0.79 | 0.79 | 2.0 | 0.79 | 0.80 | 3.0 | 0.80 | 0.79 | 4.0 | 0.81 | 0.82 | <p>Epoch 1/5 1400/1400 [=====] - 40s 28ms/sample - loss: 0.5882 - acc: 0.6929 - val_loss: 0.4868 - val_acc: 0.7817</p> <p>Epoch 2/5 1400/1400 [=====] - 35s 25ms/sample - loss: 0.4738 - acc: 0.7893 - val_loss: 0.4668 - val_acc: 0.8017</p> <p>Epoch 3/5 1400/1400 [=====] - 34s 24ms/sample - loss: 0.4286 - acc: 0.8057 - val_loss: 0.4485 - val_acc: 0.8033</p> <p>Epoch 4/5 1400/1400 [=====] - 35s 25ms/sample - loss: 0.4098 - acc: 0.8171 - val_loss: 0.4359 - val_acc: 0.7967</p> <p>Epoch 5/5 1400/1400 [=====] - 35s 25ms/sample - loss: 0.3973 - acc: 0.8307 - val_loss: 0.4910 - val_acc: 0.7883</p> |
|---|--|----------|------------|----------|-----|------|------|-----|------|------|-----|------|------|-----|------|------|-----|------|------|--|
| epoch | validation | training | | | | | | | | | | | | | | | | | | |
| 0.0 | 0.75 | 0.68 | | | | | | | | | | | | | | | | | | |
| 1.0 | 0.79 | 0.79 | | | | | | | | | | | | | | | | | | |
| 2.0 | 0.79 | 0.80 | | | | | | | | | | | | | | | | | | |
| 3.0 | 0.80 | 0.79 | | | | | | | | | | | | | | | | | | |
| 4.0 | 0.81 | 0.82 | | | | | | | | | | | | | | | | | | |
| <p>(d) Nsamp = 1000, maxtokens = 230, maxtokenlen = 200</p> |  <table border="1"><caption>BERT Email Classification Training Data for (d)</caption><thead><tr><th>epoch</th><th>validation</th><th>training</th></tr></thead><tbody><tr><td>0.0</td><td>0.73</td><td>0.70</td></tr><tr><td>1.0</td><td>0.77</td><td>0.80</td></tr><tr><td>2.0</td><td>0.79</td><td>0.81</td></tr><tr><td>3.0</td><td>0.74</td><td>0.83</td></tr><tr><td>4.0</td><td>0.78</td><td>0.81</td></tr></tbody></table> | epoch | validation | training | 0.0 | 0.73 | 0.70 | 1.0 | 0.77 | 0.80 | 2.0 | 0.79 | 0.81 | 3.0 | 0.74 | 0.83 | 4.0 | 0.78 | 0.81 | <p>Epoch 1/5 1400/1400 [=====] - 46s 33ms/sample - loss: 0.5695 - acc: 0.7000 - val_loss: 0.5276 - val_acc: 0.7283</p> <p>Epoch 2/5 1400/1400 [=====] - 41s 29ms/sample - loss: 0.4442 - acc: 0.7979 - val_loss: 0.4737 - val_acc: 0.7733</p> <p>Epoch 3/5 1400/1400 [=====] - 39s 28ms/sample - loss: 0.4143 - acc: 0.8050 - val_loss: 0.4497 - val_acc: 0.7883</p> <p>Epoch 4/5 1400/1400 [=====] - 40s 29ms/sample - loss: 0.3925 - acc: 0.8329 - val_loss: 0.5251 - val_acc: 0.7450</p> <p>Epoch 5/5 1400/1400 [=====] - 39s 28ms/sample - loss: 0.4081 - acc: 0.8179 - val_loss: 0.4606 - val_acc: 0.7783</p> |
| epoch | validation | training | | | | | | | | | | | | | | | | | | |
| 0.0 | 0.73 | 0.70 | | | | | | | | | | | | | | | | | | |
| 1.0 | 0.77 | 0.80 | | | | | | | | | | | | | | | | | | |
| 2.0 | 0.79 | 0.81 | | | | | | | | | | | | | | | | | | |
| 3.0 | 0.74 | 0.83 | | | | | | | | | | | | | | | | | | |
| 4.0 | 0.78 | 0.81 | | | | | | | | | | | | | | | | | | |

BOW

| | LR | SVM | RF | XGBoost |
|----|--|--|--|--|
| A) | The logistic regression accuracy score is:: 0.6866666666666666 6 | The SVC Classifier testing accuracy score is:: 0.655 | The RF testing accuracy score is:: 0.6916666666666667 | The Gradient Boosting testing accuracy score is:: 0.6566666666666666 6 |
| B) | The logistic regression accuracy score is:: 0.7533333333333333 3 | The SVC Classifier testing accuracy score is:: 0.7283333333333334 | The RF testing accuracy score is:: 0.745 | The Gradient Boosting testing accuracy score is:: 0.7366666666666666 7 |
| C) | The logistic regression accuracy score is:: 0.79 | The SVC Classifier testing accuracy score is:: 0.79 | The RF testing accuracy score is:: 0.77 | The Gradient Boosting testing accuracy score is:: 0.7616666666666666 7 |

TF-JDF

| | LR | SVM | RF | XGBoost |
|----|--|--|--|--|
| A) | he logistic regression accuracy score is:: 0.7166666666666666 7 | The SVC Classifier testing accuracy score is:: 0.7083333333333334 | The RF testing accuracy score is:: 0.6783333333333333 | The Gradient Boosting testing accuracy score is:: 0.6433333333333333 3 |
| B) | The logistic regression accuracy score is:: 0.7766666666666666 6 | The SVC Classifier testing accuracy score is:: 0.77 | The RF testing accuracy score is:: 0.7333333333333333 | The Gradient Boosting testing accuracy score is:: 0.7166666666666666 7 |
| C) | The logistic regression accuracy score is:: 0.7933333333333333 3 | The SVC Classifier testing accuracy score is:: 0.8116666666666666 | The RF testing accuracy score is:: 0.7483333333333333 | The Gradient Boosting testing accuracy score is:: 0.76 |