

This document contains all test results for *Argument Mining as Graph Prediction*. The classification of proposition types is reported in section **Propositions** (pp. 1-10). The classification of relations and relation types is reported in section **Relations** (pp. 11-38).

Propositions

Model Training, 23.07.2024

1) CDCP DistilBERT proposition classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

```
'fact': {'precision': 0.5436241610738255, 'recall': 0.6136363636363636, 'f1-score': 0.5765124555160143, 'support': 132},
'policy': {'precision': 0.8198757763975155, 'recall': 0.8627450980392157, 'f1-score': 0.8407643312101911, 'support': 153},
'reference': {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 1},
'testimony': {'precision': 0.9136363636363637, 'recall': 0.8237704918032787, 'f1-score': 0.8663793103448277, 'support': 244},
'value': {'precision': 0.8282828282828283, 'recall': 0.8266129032258065, 'f1-score': 0.8274470232088799, 'support': 496},
'accuracy': 0.804093567251462,
'macro avg': {'precision': 0.8210838258781067, 'recall': 0.8253529713409329, 'f1-score': 0.8222206240559826, 'support': 1026},
'weighted avg': {'precision': 0.8108722598500199, 'recall': 0.804093567251462, 'f1-score': 0.8065758889269157, 'support': 1026}
}
```

Epoch: 5/5 | Training Loss: 0.3701 | Validation Loss: 0.4843
Training accuracy: 92.30%
Validation accuracy: 81.60%

2) CDCP DistilBERT proposition classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

```
'fact': {'precision': 0.6491228070175439, 'recall': 0.5606060606060606, 'f1-score': 0.6016260162601625, 'support': 132},  
'policy': {'precision': 0.9078014184397163, 'recall': 0.8366013071895425, 'f1-score': 0.8707482993197279, 'support': 153},  
'reference': {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 1},  
'testimony': {'precision': 0.8547717842323651, 'recall': 0.8442622950819673, 'f1-score': 0.8494845360824742, 'support': 244},  
'value': {'precision': 0.8279773156899811, 'recall': 0.8830645161290323, 'f1-score': 0.8546341463414633, 'support': 496},  
'accuracy': 0.8255360623781677,  
'macro avg': {'precision': 0.8479346650759213, 'recall': 0.8249068358013206, 'f1-score': 0.8352985996007657, 'support': 1026},  
'weighted avg': {'precision': 0.8234102256164914, 'recall': 0.8255360623781677, 'f1-score': 0.8234033989588201,  
'support': 1026}  
}
```

Epoch: 5/5 | Training Loss: 0.3419 | Validation Loss: 0.5075

Training accuracy: 92.71%

Validation accuracy: 81.39%

3) CDCP DistilBERT proposition classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

```

'fact': {'precision': 0.5355191256830601, 'recall': 0.7424242424242424, 'f1-score': 0.6222222222222222, 'support': 132},

'policy': {'precision': 0.8791946308724832, 'recall': 0.8562091503267973, 'f1-score': 0.8675496688741721, 'support': 153},

'reference': {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 1},

'testimony': {'precision': 0.8787878787878788, 'recall': 0.8319672131147541, 'f1-score': 0.8547368421052632, 'support': 244},

'value': {'precision': 0.8506493506493507, 'recall': 0.7923387096774194, 'f1-score': 0.8204592901878913, 'support': 496},

'accuracy': 0.8050682261208577,

'macro avg': {'precision': 0.8288301971985547, 'recall': 0.8445878631086426, 'f1-score': 0.8329936046779098, 'support': 1026},

'weighted avg': {'precision': 0.8212004127290198, 'recall': 0.8050682261208577, 'f1-score': 0.810304122883002, 'support': 1026}
}

```

Epoch: 5/5 | Training Loss: 0.3514 | Validation Loss: 0.5096

Training accuracy: 92.45%

Validation accuracy: 83.44%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
fact	0.58	\$\pm\$0.063	0.64	\$\pm\$0.094	0.6	\$\pm\$0.023
policy	0.87	\$\pm\$0.045	0.85	\$\pm\$0.014	0.86	\$\pm\$0.016
reference	1	\$\pm\$0.000	1	\$\pm\$0.000	1	\$\pm\$0.000
testimony	0.88	\$\pm\$0.030	0.83	\$\pm\$0.010	0.86	\$\pm\$0.009

value		0.84	± 0.013		0.83	± 0.046		0.83	± 0.018	
+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+
weighted avg		0.82	± 0.007		0.81	± 0.012		0.81	± 0.009	
+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+	+-----+

Model Training, 26.07.2024

1) CDCP BERT proposition classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

```
'fact': {'precision': 0.5459770114942529, 'recall': 0.7196969696969697, 'f1-score': 0.6209150326797387, 'support': 132},
'policy': {'precision': 0.8, 'recall': 0.8888888888888888, 'f1-score': 0.8421052631578948, 'support': 153},
'reference': {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 1},
'testimony': {'precision': 0.8760330578512396, 'recall': 0.8688524590163934, 'f1-score': 0.8724279835390946, 'support': 244},
'value': {'precision': 0.8792710706150342, 'recall': 0.7782258064516129, 'f1-score': 0.8256684491978611, 'support': 496},
'accuracy': 0.8089668615984406,
'macro avg': {'precision': 0.8202562279921054, 'recall': 0.8511328248107729, 'f1-score': 0.832223345714918, 'support': 1026},
'weighted avg': {'precision': 0.8239176244230028, 'recall': 0.8089668615984406, 'f1-score': 0.8130671231603914, 'support': 1026}
}
```

Epoch: 5/5 | Training Loss: 0.3119 | Validation Loss: 0.4881
Training accuracy: 94.67%
Validation accuracy: 82.82%

2) CDCP BERT proposition classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```

Test score: {
  'fact': {'precision': 0.6027397260273972, 'recall': 0.6666666666666666, 'f1-score': 0.6330935251798561, 'support': 132},
  'policy': {'precision': 0.8516129032258064, 'recall': 0.8627450980392157, 'f1-score': 0.8571428571428572, 'support': 153},
  'reference': {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 1},
  'testimony': {'precision': 0.8828451882845189, 'recall': 0.8647540983606558, 'f1-score': 0.8737060041407868, 'support': 244},
  'value': {'precision': 0.8474226804123711, 'recall': 0.8286290322580645, 'f1-score': 0.837920489296636, 'support': 496},
  'accuracy': 0.8216374269005848,
  'macro avg': {'precision': 0.8369240995900187, 'recall': 0.8445589790649205, 'f1-score': 0.8403725751520271, 'support': 1026},
  'weighted avg': {'precision': 0.825140636895832, 'recall': 0.8216374269005848, 'f1-score': 0.8231033432437441, 'support': 1026}
}

Epoch: 5/5 | Training Loss: 0.3174 | Validation Loss: 0.4507
Training accuracy: 95.32%
Validation accuracy: 83.03%

```

3) CDCP BERT proposition classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```

Test score: {
  'fact': {'precision': 0.5986394557823129, 'recall': 0.6666666666666666, 'f1-score': 0.6308243727598566, 'support': 132},
  'policy': {'precision': 0.8253012048192772, 'recall': 0.8954248366013072, 'f1-score': 0.858934169278997, 'support': 153},

```

```

'reference': {'precision': 1.0, 'recall': 1.0, 'f1-score': 1.0, 'support': 1},

'testimony': {'precision': 0.8455284552845529, 'recall': 0.8524590163934426, 'f1-score': 0.8489795918367347, 'support':
244},

'value': {'precision': 0.8583690987124464, 'recall': 0.8064516129032258, 'f1-score': 0.8316008316008315, 'support':
496},

'accuracy': 0.8128654970760234,

'macro avg': {'precision': 0.8255676429197181, 'recall': 0.8442004265129285, 'f1-score': 0.834067793095284, 'support':
1026},

'weighted avg': {'precision': 0.8171067334809152, 'recall': 0.8128654970760234, 'f1-score': 0.8141430584660462,
'support': 1026}
}

```

Epoch: 5/5 | Training Loss: 0.2692 | Validation Loss: 0.5179

Training accuracy: 96.49%

Validation accuracy: 82.62%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
fact	0.58	\$\pm\$0.032	0.68	\$\pm\$0.031	0.63	\$\pm\$0.006
policy	0.83	\$\pm\$0.026	0.88	\$\pm\$0.017	0.85	\$\pm\$0.009
reference	1	\$\pm\$0.000	1	\$\pm\$0.000	1	\$\pm\$0.000
testimony	0.87	\$\pm\$0.020	0.86	\$\pm\$0.009	0.87	\$\pm\$0.014
value	0.86	\$\pm\$0.016	0.8	\$\pm\$0.025	0.83	\$\pm\$0.006
weighted avg	0.82	\$\pm\$0.004	0.81	\$\pm\$0.006	0.82	\$\pm\$0.006

Model Training, 24.07.2024

1) Microtext DistilBERT proposition classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.9166666666666666, 'recall': 0.21568627450980393, 'f1-score': 0.3492063492063492, 'support': 51},
  '1': {'precision': 0.8648648648648649, 'recall': 0.9961089494163424, 'f1-score': 0.925858951175407, 'support': 257},
  'accuracy': 0.8668831168831169,
  'macro avg': {'precision': 0.8907657657657657, 'recall': 0.6058976119630732, 'f1-score': 0.637532650190878, 'support': 308},
  'weighted avg': {'precision': 0.8734424359424361, 'recall': 0.8668831168831169, 'f1-score': 0.8303742670831279, 'support': 308}
}
```

Epoch: 5/5 | Training Loss: 0.3325 | Validation Loss: 0.2690
Training accuracy: 86.88%
Validation accuracy: 87.66%

2) Microtext DistilBERT proposition classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {
  '0': {'precision': 1.0, 'recall': 0.0196078431372549, 'f1-score': 0.038461538461538464, 'support': 51},
  '1': {'precision': 0.8371335504885994, 'recall': 1.0, 'f1-score': 0.9113475177304965, 'support': 257},
  'accuracy': 0.8376623376623377,
  'macro avg': {'precision': 0.9185667752442996, 'recall': 0.5098039215686274, 'f1-score': 0.47490452809601746, 'support': 308},
  'weighted avg': {'precision': 0.8641016963492533, 'recall': 0.8376623376623377, 'f1-score': 0.7668112029814157, 'support': 308}
}
```

```
}
```

```
Epoch: 0005/0005 | Batch 0000/0009 | Loss: 0.3730
Epoch: 5/5 | Training Loss: 0.3489 | Validation Loss: 0.3961
Training accuracy: 83.44%
Validation accuracy: 83.77%
```

3) Microtext DistilBERT proposition classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```
Test score: {
```

```
'0': {'precision': 1.0, 'recall': 0.1568627450980392, 'f1-score': 0.2711864406779661, 'support': 51},
'1': {'precision': 0.8566666666666667, 'recall': 1.0, 'f1-score': 0.9228007181328546, 'support': 257},
'accuracy': 0.8603896103896104,

'macro avg': {'precision': 0.9283333333333333, 'recall': 0.5784313725490196, 'f1-score': 0.5969935794054103, 'support': 308},

'weighted avg': {'precision': 0.8804004329004328, 'recall': 0.8603896103896104, 'f1-score': 0.8149035488140257, 'support': 308}
}
```

```
Epoch: 0005/0005 | Batch 0000/0009 | Loss: 0.3136
Epoch: 5/5 | Training Loss: 0.3222 | Validation Loss: 0.3441
Training accuracy: 88.84%
Validation accuracy: 87.01%
```

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.97	± 0.048	0.13	± 0.101	0.22	± 0.162
1	0.85	± 0.014	1	± 0.002	0.92	± 0.008
weighted avg	0.87	± 0.008	0.85	± 0.015	0.8	± 0.033

+-----+-----+-----+-----+-----+-----+-----+
Model Training, 26.07.2024

1) Microtext BERT proposition classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 51},
  '1': {'precision': 0.8344155844155844, 'recall': 1.0, 'f1-score': 0.9097345132743363, 'support': 257},
  'accuracy': 0.8344155844155844,
  'macro avg': {'precision': 0.4172077922077922, 'recall': 0.5, 'f1-score': 0.45486725663716815, 'support': 308},
  'weighted avg': {'precision': 0.6962493675156013, 'recall': 0.8344155844155844, 'f1-score': 0.7590966555568326,
  'support': 308}
}
```

Epoch: 5/5 | Training Loss: 0.4284 | Validation Loss: 0.4409
Training accuracy: 83.44%
Validation accuracy: 83.77%

2) Microtext BERT proposition classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 51},
  '1': {'precision': 0.8344155844155844, 'recall': 1.0, 'f1-score': 0.9097345132743363, 'support': 257},
  'accuracy': 0.8344155844155844,
  'macro avg': {'precision': 0.4172077922077922, 'recall': 0.5, 'f1-score': 0.45486725663716815, 'support': 308},
  'weighted avg': {'precision': 0.6962493675156013, 'recall': 0.8344155844155844, 'f1-score': 0.7590966555568326,
  'support': 308}
```

```

}

Epoch: 5/5 | Training Loss: 0.3702 | Validation Loss: 0.4899
Training accuracy: 83.44%
Validation accuracy: 83.77%

3) Microtext BERT proposition classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {
  '0': {'precision': 0.8, 'recall': 0.47058823529411764, 'f1-score': 0.5925925925925927, 'support': 51},
  '1': {'precision': 0.9028776978417267, 'recall': 0.9766536964980544, 'f1-score': 0.9383177570093458, 'support': 257},
  'accuracy': 0.8928571428571429,

  'macro avg': {'precision': 0.8514388489208633, 'recall': 0.723620965896086, 'f1-score': 0.7654551748009693, 'support': 308},

  'weighted avg': {'precision': 0.8858427543679342, 'recall': 0.8928571428571429, 'f1-score': 0.8810710577065718, 'support': 308}
}

Epoch: 5/5 | Training Loss: 0.2974 | Validation Loss: 0.3366
Training accuracy: 93.58%
Validation accuracy: 83.77%

```

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.27	± 0.462	0.16	± 0.272	0.2	± 0.342
1	0.86	± 0.040	0.99	± 0.013	0.92	± 0.017
weighted avg	0.76	± 0.109	0.85	± 0.034	0.8	± 0.070

Relations

[A] undirected graph without replacement

Model Training, 23.07.2024

1) CDCP DistilBERT relation (undirected graph without replacement) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'0': {'precision': 0.9410854312835328, 'recall': 0.9422520661157024, 'f1-score': 0.9416683873632047, 'support': 9680},

'1': {'precision': 0.12106918238993711, 'recall': 0.11882716049382716, 'f1-score': **0.11993769470404984**, 'support': 648},

'accuracy': 0.8905886909372579,

'macro avg': {'precision': 0.5310773068367349, 'recall': 0.5305396133047648, 'f1-score': 0.5308030410336273, 'support': 10328},

'weighted avg': {'precision': 0.8896359222514791, 'recall': 0.8905886909372579, 'f1-score': **0.8901113105968286**,
'support': 10328}
}

Epoch: 5/5 | Training Loss: 0.1077 | Validation Loss: 0.2167

Training accuracy: 98.25%

Validation accuracy: 91.73%

2) CDCP DistilBERT relation (undirected graph without replacement) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

'0': {'precision': 0.9391891891891891, 'recall': 0.9764462809917356, 'f1-score': 0.9574554294975689, 'support': 9680},

'1': {'precision': 0.13636363636363635, 'recall': 0.05555555555555555, 'f1-score': **0.07894736842105263**, 'support': 648},

```

'accuracy': 0.9186676994577847,

'macro avg': {'precision': 0.5377764127764127, 'recall': 0.5160009182736456, 'f1-score': 0.5182013989593107, 'support': 10328},

'weighted avg': {'precision': 0.888818259848469, 'recall': 0.9186676994577847, 'f1-score': 0.9023360236515596, 'support': 10328}
}

```

Epoch: 5/5 | Training Loss: 0.1079 | Validation Loss: 0.1888
Training accuracy: 97.90%
Validation accuracy: 93.18%

3) CDCP DistilBERT relation (undirected graph without replacement) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```

Test score: {

'0': {'precision': 0.9426019098470069, 'recall': 0.9483471074380165, 'f1-score': 0.9454657809361966, 'support': 9680},

'1': {'precision': 0.15110356536502548, 'recall': 0.13734567901234568, 'f1-score': 0.14389652384801938, 'support': 648},

'accuracy': 0.8974632068164213,

'macro avg': {'precision': 0.5468527376060162, 'recall': 0.5428463932251811, 'f1-score': 0.544681152392108, 'support': 10328},

'weighted avg': {'precision': 0.8929416728965495, 'recall': 0.8974632068164213, 'f1-score': 0.895173674178534, 'support': 10328}
}

```

Epoch: 5/5 | Training Loss: 0.1102 | Validation Loss: 0.2284
Training accuracy: 98.12%
Validation accuracy: 92.22%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.9426019098470069	0.0000000000000000	0.9483471074380165	0.0000000000000000	0.9454657809361966	0.0000000000000000
1	0.15110356536502548	0.0000000000000000	0.13734567901234568	0.0000000000000000	0.14389652384801938	0.0000000000000000

0	0.94	± 0.002	0.96	± 0.018	0.95	± 0.008
1	0.14	± 0.015	0.1	± 0.043	0.11	± 0.033
weighted avg	0.89	± 0.002	0.9	± 0.015	0.9	± 0.006

Model Training, 26.07.2024

1) Microtext DistilBERT relation (undirected graph without replacement) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

```
Test score: {
'0': {'precision': 0.7553648068669528, 'recall': 0.8712871287128713, 'f1-score': 0.8091954022988507, 'support': 1010},
'1': {'precision': 0.47580645161290325, 'recall': 0.29280397022332505, 'f1-score': 0.36251920122887865, 'support': 403},
'accuracy': 0.7062986553432413,
'macro avg': {'precision': 0.615585629239928, 'recall': 0.5820455494680982, 'f1-score': 0.5858573017638646, 'support': 1413},
'weighted avg': {'precision': 0.6756323106409217, 'recall': 0.7062986553432413, 'f1-score': 0.6817994298776201, 'support': 1413}
}
```

```
Epoch: 5/5 | Training Loss: 0.4501 | Validation Loss: 0.5468
Training accuracy: 82.29%
Validation accuracy: 71.53%
```

2) Microtext DistilBERT relation (undirected graph without replacement) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {
'0': {'precision': 0.7567340067340067, 'recall': 0.8900990099009901, 'f1-score': 0.8180163785259327, 'support': 1010},
```

```
'1': {'precision': 0.5066666666666667, 'recall': 0.28287841191067, 'f1-score': 0.36305732484076436, 'support': 403},
'accuracy': 0.7169143665958952,
'macro avg': {'precision': 0.6317003367003367, 'recall': 0.5864887109058301, 'f1-score': 0.5905368516833485, 'support': 1413},
'weighted avg': {'precision': 0.6854126068421893, 'recall': 0.7169143665958952, 'f1-score': 0.6882580638513942, 'support': 1413}
}
```

Epoch: 5/5 | Training Loss: 0.4504 | Validation Loss: 0.6021
Training accuracy: 81.61%
Validation accuracy: 70.54%

3) Microtext DistilBERT relation (undirected graph without replacement) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```
Test score: {
'0': {'precision': 0.7881773399014779, 'recall': 0.7920792079207921, 'f1-score': 0.7901234567901234, 'support': 1010},
'1': {'precision': 0.4723618090452261, 'recall': 0.4665012406947891, 'f1-score': 0.4694132334581773, 'support': 403},
'accuracy': 0.6992215145081387,
'macro avg': {'precision': 0.630269574473352, 'recall': 0.6292902243077906, 'f1-score': 0.6297683451241504, 'support': 1413},
'weighted avg': {'precision': 0.6981039790132475, 'recall': 0.6992215145081387, 'f1-score': 0.6986540866536943, 'support': 1413}
}
```

Epoch: 5/5 | Training Loss: 0.4537 | Validation Loss: 0.5525
Training accuracy: 82.07%
Validation accuracy: 70.96%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.77	± 0.019	0.85	± 0.052	0.81	± 0.014
1	0.48	± 0.019	0.35	± 0.103	0.4	± 0.062
weighted avg	0.69	± 0.011	0.71	± 0.009	0.69	± 0.009

[B] directed graph without replacement

Model Training, 23.07.2024

1) CDCP DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'0': {'precision': 0.9720244660583576, 'recall': 0.9690123950419832, 'f1-score': 0.9705160935075337, 'support': 10004},

'1': {'precision': 0.1267605633802817, 'recall': 0.13888888888888889, 'f1-score': **0.1325478645066274**, 'support': 324},

'accuracy': 0.9429705654531371,

'macro avg': {'precision': 0.5493925147193196, 'recall': 0.5539506419654361, 'f1-score': 0.5515319790070805, 'support': 10328},

'weighted avg': {'precision': 0.9455076666327479, 'recall': 0.9429705654531371, 'f1-score': **0.9442281668812464**,

'support': 10328}

}

Epoch: 5/5 | Training Loss: 0.1377 | Validation Loss: 0.3103

Training accuracy: 97.67%

Validation accuracy: 90.18%

2) CDCP DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.9736564805057956, 'recall': 0.9236305477808876, 'f1-score': 0.9479839950754078, 'support': 10004},
  '1': {'precision': 0.0883054892601432, 'recall': 0.22839506172839505, 'f1-score': 0.12736660929432014, 'support': 324},
  'accuracy': 0.9018202943454686,
  'macro avg': {'precision': 0.5309809848829694, 'recall': 0.5760128047546413, 'f1-score': 0.5376753021848639, 'support': 10328},
  'weighted avg': {'precision': 0.945882107813736, 'recall': 0.9018202943454686, 'f1-score': 0.9222403822759236, 'support': 10328}
}

Epoch: 5/5 | Training Loss: 0.1533 | Validation Loss: 0.2897
Training accuracy: 96.55%
Validation accuracy: 88.46%
```

3) CDCP DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.9709699791935005, 'recall': 0.9796081567373051, 'f1-score': 0.9752699407871822, 'support': 10004},
  '1': {'precision': 0.13191489361702127, 'recall': 0.09567901234567901, 'f1-score': 0.110912343470483, 'support': 324},
  'accuracy': 0.951878388845856,
  'macro avg': {'precision': 0.5514424364052609, 'recall': 0.5376435845414921, 'f1-score': 0.5430911421288326, 'support': 10328},
  'weighted avg': {'precision': 0.944647956756748, 'recall': 0.951878388845856, 'f1-score': 0.9481541524902601, 'support': 10328}
}

Epoch: 5/5 | Training Loss: 0.1535 | Validation Loss: 0.2627
Training accuracy: 96.79%
Validation accuracy: 90.84%
```


Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.97	± 0.001	0.96	± 0.030	0.96	± 0.015
1	0.12	± 0.024	0.15	± 0.068	0.12	± 0.011
weighted avg	0.95	± 0.001	0.93	± 0.027	0.94	± 0.014

Model Training, 25.07.2024

1) Microtext DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'sup': {'precision': 0.40476190476190477, 'recall': 0.3493150684931507, 'f1-score': **0.37500000000000006**, 'support': 146},

'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 7},

'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 49},

'None': {'precision': 0.8942006269592476, 'recall': 0.9508333333333333, 'f1-score': **0.9216478190630047**, 'support': 1200},

'accuracy': 0.8502139800285307,

'macro avg': {'precision': 0.3247406329302881, 'recall': 0.325037100456621, 'f1-score': 0.3241619547657512, 'support': 1402},

'weighted avg': {'precision': 0.8075149717876856, 'recall': 0.8502139800285307, 'f1-score': **0.8279082616801752**, 'support': 1402}

Epoch: 5/5 | Training Loss: 0.3964 | Validation Loss: 0.4718

Training accuracy: 89.48%
Validation accuracy: 85.45%

2) Microtext DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

'sup': {'precision': 0.49137931034482757, 'recall': 0.3904109589041096, 'f1-score': **0.4351145038167939**, 'support': 146},

'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 7},

'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 49},

'None': {'precision': 0.8989113530326595, 'recall': 0.9633333333333334, 'f1-score': **0.9300080450522928**, 'support': 1200},

'accuracy': 0.8651925820256776,

'macro avg': {'precision': 0.34757266584437174, 'recall': 0.33843607305936074, 'f1-score': 0.3412806372172717, 'support': 1402},

'weighted avg': {'precision': 0.8205670491794125, 'recall': 0.8651925820256776, 'f1-score': **0.8413240881740395**, 'support': 1402}
}

Epoch: 5/5 | Training Loss: 0.4089 | Validation Loss: 0.4161

Training accuracy: 88.83%

Validation accuracy: 85.88%

3) Microtext DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

'sup': {'precision': 0.43023255813953487, 'recall': 0.2534246575342466, 'f1-score': **0.31896551724137934**, 'support': 146},

'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 7},

```

'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 49},
'None': {'precision': 0.8799392097264438, 'recall': 0.965, 'f1-score': 0.9205087440381557, 'support': 1200},
'accuracy': 0.8523537803138374,
'macro avg': {'precision': 0.3275429419664947, 'recall': 0.30460616438356164, 'f1-score': 0.30986856531988377,
'support': 1402},
'weighted avg': {'precision': 0.7979607740086337, 'recall': 0.8523537803138374, 'f1-score': 0.8210980444814753,
'support': 1402}
}

```

Epoch: 5/5 | Training Loss: 0.3871 | Validation Loss: 0.4295
Training accuracy: 89.34%
Validation accuracy: 84.31%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
sup	0.44	± 0.045	0.33	± 0.070	0.38	± 0.058
exa	0	± 0.000	0	± 0.000	0	± 0.000
reb	0	± 0.000	0	± 0.000	0	± 0.000
None	0.89	± 0.010	0.96	± 0.008	0.92	± 0.005
weighted avg	0.81	± 0.011	0.86	± 0.008	0.83	± 0.010

Model Training, 26.07.2024

1) Augmented DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.9710764337541, 'recall': 0.976609356257497, 'f1-score': 0.9738350361325692, 'support': 10004},
  '1': {'precision': 0.12359550561797752, 'recall': 0.10185185185185185, 'f1-score': 0.1116751269035533, 'support': 324},
  'accuracy': 0.9491673121611154,
  'macro avg': {'precision': 0.5473359696860388, 'recall': 0.5392306040546744, 'f1-score': 0.5427550815180613, 'support': 10328},
  'weighted avg': {'precision': 0.9444900839558715, 'recall': 0.9491673121611154, 'f1-score': 0.946788191575036, 'support': 10328}
}

Epoch: 5/5 | Training Loss: 0.1116 | Validation Loss: 0.2082
Training accuracy: 97.82%
Validation accuracy: 93.89%
```

2) Augmented DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {
  '0': {'precision': 0.9741736635136571, 'recall': 0.916233506597361, 'f1-score': 0.9443156647607274, 'support': 10004},
  '1': {'precision': 0.088139281828074, 'recall': 0.25, 'f1-score': 0.13032984714400644, 'support': 324},
  'accuracy': 0.8953330751355538,
  'macro avg': {'precision': 0.5311564726708655, 'recall': 0.5831167532986805, 'f1-score': 0.5373227559523669, 'support': 10328},
  'weighted avg': {'precision': 0.9463778521594619, 'recall': 0.8953330751355538, 'f1-score': 0.9187800910864615, 'support': 10328}
}
```

Epoch: 5/5 | Training Loss: 0.1163 | Validation Loss: 0.2220
 Training accuracy: 96.98%
 Validation accuracy: 92.26%

3) Augmented DistilBERT relation (directed graph without replacement) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

'0': {'precision': 0.9710000993147284, 'recall': 0.9773090763694522, 'f1-score': 0.9741443730384098, 'support': 10004},
 '1': {'precision': 0.12355212355212356, 'recall': 0.09876543209876543, 'f1-score': **0.1097770154373928**, 'support': 324},
 'accuracy': 0.9497482571649883,
 'macro avg': {'precision': 0.547276111433426, 'recall': 0.5380372542341088, 'f1-score': 0.5419606942379013, 'support': 10328},
 'weighted avg': {'precision': 0.944414783266405, 'recall': 0.9497482571649883, 'f1-score': **0.9470282785513136**,
 'support': 10328}
 }

Epoch: 5/5 | Training Loss: 0.1225 | Validation Loss: 0.1967
 Training accuracy: 97.10%
 Validation accuracy: 93.97%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.97	± 0.002	0.96	± 0.035	0.96	± 0.017
1	0.11	± 0.020	0.15	± 0.086	0.12	± 0.011
weighted avg	0.95	± 0.001	0.93	± 0.031	0.94	± 0.016

[C] undirected graph window=1

Model Training, 23.07.2024

1) CDCP DistilBERT relation (undirected graph window=1) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

```
'0': {'precision': 0.8172185430463577, 'recall': 0.9060205580029369, 'f1-score': 0.8593314763231197, 'support': 1362},
'1': {'precision': 0.47107438016528924, 'recall': 0.2923076923076923, 'f1-score': 0.36075949367088606, 'support': 390},
'accuracy': 0.769406392694064,
'macro avg': {'precision': 0.6441464616058235, 'recall': 0.5991641251553146, 'f1-score': 0.6100454849970028, 'support': 1752},
'weighted avg': {'precision': 0.7401659040488596, 'recall': 0.769406392694064, 'f1-score': 0.7483479870340951, 'support': 1752}
}
```

Epoch: 5/5 | Training Loss: 0.2322 | Validation Loss: 0.3942

Training accuracy: 96.29%

Validation accuracy: 85.82%

2) CDCP DistilBERT relation (undirected graph window=1) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

```
'0': {'precision': 0.8213538032100488, 'recall': 0.8641703377386197, 'f1-score': 0.8422182468694097, 'support': 1362},
'1': {'precision': 0.4200626959247649, 'recall': 0.3435897435897436, 'f1-score': 0.37799717912552894, 'support': 390},
'accuracy': 0.7482876712328768,
'macro avg': {'precision': 0.6207082495674069, 'recall': 0.6038800406641817, 'f1-score': 0.6101077129974694, 'support': 1752},
'weighted avg': {'precision': 0.7320253033006535, 'recall': 0.7482876712328768, 'f1-score': 0.7388813653510802, 'support': 1752}
}
```

```
}
```

Epoch: 5/5 | Training Loss: 0.2429 | Validation Loss: 0.3752
Training accuracy: 95.60%
Validation accuracy: 84.25%

3) CDCP DistilBERT relation (undirected graph window=1) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

```
'0': {'precision': 0.8155339805825242, 'recall': 0.9251101321585903, 'f1-score': 0.8668730650154798, 'support': 1362},  
'1': {'precision': 0.5072463768115942, 'recall': 0.2692307692307692, 'f1-score': 0.35175879396984927, 'support': 390},  
'accuracy': 0.7791095890410958,  
  
'macro avg': {'precision': 0.6613901786970593, 'recall': 0.5971704506946798, 'f1-score': 0.6093159294926646, 'support': 1752},  
  
'weighted avg': {'precision': 0.7469083153595433, 'recall': 0.7791095890410958, 'f1-score': 0.7522072170087469,  
'support': 1752}  
}
```

Epoch: 5/5 | Training Loss: 0.2589 | Validation Loss: 0.4235
Training accuracy: 94.86%
Validation accuracy: 83.89%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.82	± 0.003	0.9	± 0.031	0.86	± 0.013
1	0.47	± 0.044	0.3	± 0.038	0.36	± 0.013
weighted avg	0.74	± 0.007	0.77	± 0.016	0.75	± 0.007

Model Training, 26.07.2024

1) Microtext DistilBERT relation (undirected graph window=1) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

```
'0': {'precision': 0.6469002695417789, 'recall': 0.7868852459016393, 'f1-score': 0.7100591715976331, 'support': 305},  
'1': {'precision': 0.49612403100775193, 'recall': 0.3282051282051282, 'f1-score': 0.39506172839506176, 'support': 195},  
'accuracy': 0.608,  
'macro avg': {'precision': 0.5715121502747654, 'recall': 0.5575451870533837, 'f1-score': 0.5525604499963475, 'support': 500},  
'weighted avg': {'precision': 0.5880975365135084, 'recall': 0.608, 'f1-score': 0.5872101687486303, 'support': 500}  
}
```

Epoch: 5/5 | Training Loss: 0.6035 | Validation Loss: 0.6332

Training accuracy: 71.89%

Validation accuracy: 64.80%

2) Microtext DistilBERT relation (undirected graph window=1) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

```
'0': {'precision': 0.6181434599156118, 'recall': 0.9606557377049181, 'f1-score': 0.7522464698331194, 'support': 305},  
'1': {'precision': 0.5384615384615384, 'recall': 0.07179487179487179, 'f1-score': 0.12669683257918551, 'support': 195},  
'accuracy': 0.614,  
'macro avg': {'precision': 0.5783024991885751, 'recall': 0.5162253047498949, 'f1-score': 0.43947165120615245, 'support': 500},  
'weighted avg': {'precision': 0.5870675105485231, 'recall': 0.614, 'f1-score': 0.5082821113040852, 'support': 500}  
}
```


Epoch: 5/5 | Training Loss: 0.6047 | Validation Loss: 0.6094
 Training accuracy: 63.94%
 Validation accuracy: 62.40%

3) Microtext DistilBERT relation (undirected graph window=1) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

'0': {'precision': 0.6648351648351648, 'recall': 0.7934426229508197, 'f1-score': 0.7234678624813153, 'support': 305},
 '1': {'precision': 0.5367647058823529, 'recall': 0.37435897435897436, 'f1-score': **0.4410876132930514**, 'support': 195},
 'accuracy': 0.63,
 'macro avg': {'precision': 0.6007999353587589, 'recall': 0.583900798654897, 'f1-score': 0.5822777378871834, 'support': 500},
 'weighted avg': {'precision': 0.6148876858435682, 'recall': 0.63, 'f1-score': **0.6133395652978924**, 'support': 500}
 }

Epoch: 5/5 | Training Loss: 0.5893 | Validation Loss: 0.6180
 Training accuracy: 75.71%
 Validation accuracy: 63.60%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.64	± 0.024	0.85	± 0.098	0.73	± 0.022
1	0.52	± 0.024	0.26	± 0.163	0.32	± 0.170
weighted avg	0.6	± 0.016	0.62	± 0.011	0.57	± 0.055

[D_1] directed graph window=1

Model Training, 23.07.2024

1) CDCP DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'0': {'precision': 0.9164086687306502, 'recall': 0.9505459216441875, 'f1-score': 0.9331651954602774, 'support': 1557},

'1': {'precision': 0.43795620437956206, 'recall': 0.3076923076923077, 'f1-score': **0.3614457831325301**, 'support': 195},

'accuracy': 0.8789954337899544,

'macro avg': {'precision': 0.6771824365551061, 'recall': 0.6291191146682475, 'f1-score': 0.6473054892964037, 'support': 1752},

'weighted avg': {'precision': 0.8631562540340393, 'recall': 0.8789954337899544, 'f1-score': **0.869532041690922**,
'support': 1752}
}

Epoch: 5/5 | Training Loss: 0.2266 | Validation Loss: 0.3674

Training accuracy: 94.38%

Validation accuracy: 85.22%

2) CDCP DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

'0': {'precision': 0.9178433889602053, 'recall': 0.9184328837508028, 'f1-score': 0.9181380417335472, 'support': 1557},

'1': {'precision': 0.34536082474226804, 'recall': 0.3435897435897436, 'f1-score': **0.3444730077120822**, 'support': 195},

'accuracy': 0.8544520547945206,

'macro avg': {'precision': 0.6316021068512367, 'recall': 0.6310113136702732, 'f1-score': 0.6313055247228148, 'support': 1752},

'weighted avg': {'precision': 0.8541252953400582, 'recall': 0.8544520547945206, 'f1-score': **0.8542883376044458**,
'support': 1752}

```
}
```

Epoch: 5/5 | Training Loss: 0.2178 | Validation Loss: 0.3363
Training accuracy: 93.49%
Validation accuracy: 84.13%

3) CDCP DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

```
'0': {'precision': 0.8980891719745223, 'recall': 0.9961464354527938, 'f1-score': 0.9445797807551766, 'support': 1557},  
'1': {'precision': 0.76, 'recall': 0.09743589743589744, 'f1-score': 0.17272727272727276, 'support': 195},  
'accuracy': 0.8961187214611872,  
  
'macro avg': {'precision': 0.8290445859872612, 'recall': 0.5467911664443457, 'f1-score': 0.5586535267412247, 'support': 1752},  
  
'weighted avg': {'precision': 0.8827196579705087, 'recall': 0.8961187214611872, 'f1-score': 0.858671539279468, 'support': 1752}  
}
```

Epoch: 5/5 | Training Loss: 0.2347 | Validation Loss: 0.3301
Training accuracy: 91.39%
Validation accuracy: 89.42%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.91	± 0.011	0.96	± 0.039	0.93	± 0.013
1	0.51	± 0.218	0.25	± 0.133	0.29	± 0.104
weighted avg	0.87	± 0.015	0.88	± 0.021	0.86	± 0.008

Model Training, 26.07.2024

1) Microtext DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

```
Test score: {
'sup': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 71},
'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 5},
'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 22},
'None': {'precision': 0.8, 'recall': 1.0, 'f1-score': 0.8888888888888889, 'support': 392},
'accuracy': 0.8,
'macro avg': {'precision': 0.2, 'recall': 0.25, 'f1-score': 0.2222222222222224, 'support': 490},
'weighted avg': {'precision': 0.64, 'recall': 0.8, 'f1-score': 0.7111111111111111, 'support': 490}
}
```

Epoch: 5/5 | Training Loss: 0.6433 | Validation Loss: 0.6263

Training accuracy: 80.06%

Validation accuracy: 80.00%

2) Microtext DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {
'sup': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 71},
'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 5},
'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 22},
'None': {'precision': 0.8, 'recall': 1.0, 'f1-score': 0.8888888888888889, 'support': 392},
'accuracy': 0.8, 'macro avg': {'precision': 0.2, 'recall': 0.25, 'f1-score': 0.2222222222222224, 'support': 490},
```

```
'weighted avg': {'precision': 0.64, 'recall': 0.8, 'f1-score': 0.7111111111111111, 'support': 490}
}
```

```
Epoch: 0005/0005 | Batch 0010/0014 | Loss: 0.5153
Epoch: 5/5 | Training Loss: 0.6133 | Validation Loss: 0.6511
Training accuracy: 80.06%
Validation accuracy: 80.00%
```

3) Microtext DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

```
'sup': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 71},
'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 5},
'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 22},
'None': {'precision': 0.8, 'recall': 1.0, 'f1-score': 0.8888888888888889, 'support': 392},
'accuracy': 0.8,
'macro avg': {'precision': 0.2, 'recall': 0.25, 'f1-score': 0.22222222222222224, 'support': 490},
'weighted avg': {'precision': 0.64, 'recall': 0.8, 'f1-score': 0.7111111111111111, 'support': 490}
}
```

```
Epoch: 5/5 | Training Loss: 0.6300 | Validation Loss: 0.6257
Training accuracy: 80.06%
Validation accuracy: 80.00%
```

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
sup	0	± 0.000	0	± 0.000	0	± 0.000
exa	0	± 0.000	0	± 0.000	0	± 0.000

reb	0	± 0.000	0	± 0.000	0	± 0.000
None	0.8	± 0.000	1	± 0.000	0.89	± 0.000
weighted avg	0.64	± 0.000	0.8	± 0.000	0.71	± 0.000

Model Training, 26.07.2024

1) Augmented DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'0': {'precision': 0.9020070838252656, 'recall': 0.9813744380218369, 'f1-score': 0.9400184558597355, 'support': 1557},

'1': {'precision': 0.5, 'recall': 0.14871794871794872, 'f1-score': **0.2292490118577075**, 'support': 195},

'accuracy': 0.8886986301369864,

'macro avg': {'precision': 0.7010035419126328, 'recall': 0.5650461933698928, 'f1-score': 0.5846337338587215, 'support': 1752},

'weighted avg': {'precision': 0.8572631447008782, 'recall': 0.8886986301369864, 'f1-score': **0.8609088430855373**, 'support': 1752}

Epoch: 5/5 | Training Loss: 0.2330 | Validation Loss: 0.3181

Training accuracy: 93.42%

Validation accuracy: 88.73%

2) Augmented DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

'0': {'precision': 0.9097560975609756, 'recall': 0.9582530507385999, 'f1-score': 0.9333750390991555, 'support': 1557},

```
'1': {'precision': 0.41964285714285715, 'recall': 0.24102564102564103, 'f1-score': 0.30618892508143325, 'support': 195},

'accuracy': 0.8784246575342466,

'macro avg': {'precision': 0.6646994773519164, 'recall': 0.5996393458821204, 'f1-score': 0.6197819820902943, 'support': 1752},

'weighted avg': {'precision': 0.8552058225144384, 'recall': 0.8784246575342466, 'f1-score': 0.8635683654499228, 'support': 1752}
}

Epoch: 5/5 | Training Loss: 0.2331 | Validation Loss: 0.2870
Training accuracy: 94.05%
Validation accuracy: 89.82%
```

3) Augmented DistilBERT relation (directed graph window=1) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```
Test score: {

'0': {'precision': 0.9088729016786571, 'recall': 0.9736673089274245, 'f1-score': 0.9401550387596899, 'support': 1557},

'1': {'precision': 0.5119047619047619, 'recall': 0.2205128205128205, 'f1-score': 0.3082437275985663, 'support': 195},

'accuracy': 0.8898401826484018,

'macro avg': {'precision': 0.7103888317917095, 'recall': 0.5970900647201225, 'f1-score': 0.624199383179128, 'support': 1752},

'weighted avg': {'precision': 0.8646898039298504, 'recall': 0.8898401826484018, 'f1-score': 0.8698224441955237, 'support': 1752}
}

Epoch: 5/5 | Training Loss: 0.2406 | Validation Loss: 0.3439
Training accuracy: 93.29%
Validation accuracy: 88.95%
```

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.91	± 0.004	0.97	± 0.012	0.94	± 0.004
1	0.48	± 0.050	0.2	± 0.048	0.28	± 0.045
weighted avg	0.86	± 0.005	0.89	± 0.006	0.86	± 0.005

[D_2] directed graph window=3

Model Training, 27.07.2024

1) CDCP DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'0': {'precision': 0.9451728247914184, 'recall': 0.9628460417678485, 'f1-score': 0.95392758330326, 'support': 4118.0},

'1': {'precision': 0.24630541871921183, 'recall': 0.17857142857142858, 'f1-score': **0.20703933747412007**, 'support': 280.0},

'accuracy': 0.9129149613460664,

'macro avg': {'precision': 0.5957391217553151, 'recall': 0.5707087351696385, 'f1-score': 0.58048346038869, 'support': 4398.0},

'weighted avg': {'precision': 0.9006792200392089, 'recall': 0.9129149613460664, 'f1-score': **0.9063767172659342**, 'support': 4398.0}

Epoch: 5/5 | Training Loss: 0.1559 | Validation Loss: 0.2988

Training accuracy: 97.07%

Validation accuracy: 88.60%

2) CDCP DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split


```
Test score: {  
  
  '0': {'precision': 0.943449575871819, 'recall': 0.9723166585721223, 'f1-score': 0.9576656302320019, 'support': 4118.0},  
  
  '1': {'precision': 0.2597402597402597, 'recall': 0.14285714285714285, 'f1-score': 0.18433179723502302, 'support':  
280.0},  
  
  'accuracy': 0.9195088676671214,  
  
  'macro avg': {'precision': 0.6015949178060394, 'recall': 0.5575869007146326, 'f1-score': 0.5709987137335124, 'support':  
4398.0},  
  
  'weighted avg': {'precision': 0.8999210154996415, 'recall': 0.9195088676671214, 'f1-score': 0.9084310978902206,  
  'support': 4398.0}  
}  
  
Epoch: 5/5 | Training Loss: 0.1586 | Validation Loss: 0.3230  
Training accuracy: 96.59%  
Validation accuracy: 89.30%
```

3) CDCP DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```
Test score: {  
  
  '0': {'precision': 0.9400826446280992, 'recall': 0.9944147644487615, 'f1-score': 0.96648572102903, 'support': 4118.0},  
  
  '1': {'precision': 0.4523809523809524, 'recall': 0.06785714285714285, 'f1-score': 0.11801242236024843, 'support':  
280.0},  
  
  'accuracy': 0.935425193269668,  
  
  'macro avg': {'precision': 0.6962317985045258, 'recall': 0.5311359536529522, 'f1-score': 0.5422490716946392, 'support':  
4398.0},  
  
  'weighted avg': {'precision': 0.9090329689052249, 'recall': 0.935425193269668, 'f1-score': 0.9124674118823135,  
  'support': 4398.0}  
}  
  
Epoch: 5/5 | Training Loss: 0.1829 | Validation Loss: 0.3218
```

Training accuracy: 93.43%
 Validation accuracy: 91.13%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.94	± 0.003	0.98	± 0.016	0.96	± 0.006
1	0.32	± 0.115	0.13	± 0.057	0.17	± 0.046
weighted avg	0.9	± 0.005	0.92	± 0.012	0.91	± 0.003

1) Microtext DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

'sup': {'precision': 0.37272727272727274, 'recall': 0.33884297520661155, 'f1-score': **0.35497835497835495**, 'support': 121.0},

'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 7.0},

'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': **0.0**, 'support': 39.0},

'None': {'precision': 0.8864970645792564, 'recall': 0.938860103626943, 'f1-score': **0.9119275289380976**, 'support': 965.0},

'accuracy': 0.8365724381625441,

'macro avg': {'precision': 0.31480608432663226, 'recall': 0.31942576970838865, 'f1-score': 0.3167264709791131, 'support': 1132.0},

'weighted avg': {'precision': 0.7955562432146488, 'recall': 0.8365724381625441, 'f1-score': **0.8153378501569304**, 'support': 1132.0}
 }

2) Microtext DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

Test score: {

```
'sup': {'precision': 0.41935483870967744, 'recall': 0.4297520661157025, 'f1-score': 0.4244897959183674, 'support': 121.0},  
'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 7.0},  
'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 39.0},  
'None': {'precision': 0.8958333333333334, 'recall': 0.9357512953367876, 'f1-score': 0.9153573238722758, 'support': 965.0},  
'accuracy': 0.8436395759717314,  
'macro avg': {'precision': 0.3287970430107527, 'recall': 0.3413758403631225, 'f1-score': 0.3349617799476608, 'support': 1132.0},  
'weighted avg': {'precision': 0.8084992068467648, 'recall': 0.8436395759717314, 'f1-score': 0.8256917692958202, 'support': 1132.0}  
}
```

Epoch: 5/5 | Training Loss: 0.4250 | Validation Loss: 0.4688

Training accuracy: 88.51%

Validation accuracy: 84.07%

3) Microtext DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

Test score: {

```
'sup': {'precision': 0.6, 'recall': 0.09917355371900827, 'f1-score': 0.1702127659574468, 'support': 121.0},  
'exa': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 7.0},  
'reb': {'precision': 0.0, 'recall': 0.0, 'f1-score': 0.0, 'support': 39.0},  
'None': {'precision': 0.8633093525179856, 'recall': 0.9948186528497409, 'f1-score': 0.9244102070293692, 'support': 965.0},
```

```

'accuracy': 0.8586572438162544,

'macro avg': {'precision': 0.3658273381294964, 'recall': 0.2734980516421873, 'f1-score': 0.273655743246704, 'support':
1132.0},

'weighted avg': {'precision': 0.800082619416834, 'recall': 0.8586572438162544, 'f1-score': 0.8062293237316187,
'support': 1132.0}
}

Epoch: 5/5 | Training Loss: 0.4737 | Validation Loss: 0.4697
Training accuracy: 85.75%
Validation accuracy: 85.66%

```

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
sup	0.46	\$\pm\$0.120	0.29	\$\pm\$0.171	0.32	\$\pm\$0.131
exa	0	\$\pm\$0.000	0	\$\pm\$0.000	0	\$\pm\$0.000
reb	0	\$\pm\$0.000	0	\$\pm\$0.000	0	\$\pm\$0.000
None	0.88	\$\pm\$0.017	0.96	\$\pm\$0.033	0.92	\$\pm\$0.006
weighted avg	0.8	\$\pm\$0.007	0.85	\$\pm\$0.011	0.82	\$\pm\$0.010

1) Augmented DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 100, 70:10:20 split

Test score: {

```

'0': {'precision': 0.9463667820069204, 'recall': 0.9298203011170472, 'f1-score': 0.9380205781479667, 'support':
4118.0},

'1': {'precision': 0.17897727272727273, 'recall': 0.225, 'f1-score': 0.19936708860759494, 'support': 280.0},

```

```
'accuracy': 0.8849477035015917,  
  
'macro avg': {'precision': 0.5626720273670965, 'recall': 0.5774101505585236, 'f1-score': 0.5686938333777808, 'support':  
4398.0},  
  
'weighted avg': {'precision': 0.8975106968322271, 'recall': 0.8849477035015917, 'f1-score': 0.8909939803600394,  
'support': 4398.0}  
}  
  
Epoch: 5/5 | Training Loss: 0.1420 | Validation Loss: 0.2528  
Training accuracy: 96.57%  
Validation accuracy: 90.86%
```

2) Augmented DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 90, 70:10:20 split

```
Test score: {  
  
'0': {'precision': 0.945715676728335, 'recall': 0.9434191355026712, 'f1-score': 0.9445660102115244, 'support': 4118.0},  
  
'1': {'precision': 0.19655172413793104, 'recall': 0.20357142857142857, 'f1-score': 0.2, 'support': 280.0},  
  
'accuracy': 0.8963165075034106,  
  
'macro avg': {'precision': 0.571133700433133, 'recall': 0.5734952820370499, 'f1-score': 0.5722830051057622, 'support':  
4398.0},  
  
'weighted avg': {'precision': 0.8980199271318563, 'recall': 0.8963165075034106, 'f1-score': 0.8971629900070617,  
'support': 4398.0}  
}  
  
Epoch: 5/5 | Training Loss: 0.1451 | Validation Loss: 0.2287  
Training accuracy: 96.46%  
Validation accuracy: 92.25%
```

3) Augmented DistilBERT relation (directed graph window=3) classification, batch_size=128, epochs=5, random_seed: 80, 70:10:20 split

```
Individual F1 score: {
```

```

'0': {'precision': 0.944015444015444, 'recall': 0.9499757163671685, 'f1-score': 0.9469862018881626, 'support': 4118.0},
'1': {'precision': 0.1889763779527559, 'recall': 0.17142857142857143, 'f1-score': 0.1797752808988764, 'support': 280.0},
'accuracy': 0.9004092769440655,
'macro avg': {'precision': 0.5664959109840999, 'recall': 0.56070214389787, 'f1-score': 0.5633807413935195, 'support': 4398.0},
'weighted avg': {'precision': 0.8959456535430582, 'recall': 0.9004092769440655, 'f1-score': 0.898141486590982, 'support': 4398.0}
}

```

Epoch: 5/5 | Training Loss: 0.1560 | Validation Loss: 0.2177

Training accuracy: 95.82%

Validation accuracy: 92.52%

Mean results from all training and test rounds with standard deviations:

Category	Precision Mean	Precision SD	Recall Mean	Recall SD	F1 Mean	F1 SD
0	0.95	± 0.001	0.94	± 0.010	0.94	± 0.005
1	0.19	± 0.009	0.2	± 0.027	0.19	± 0.011
weighted avg	0.9	± 0.001	0.89	± 0.008	0.9	± 0.004