CS 529 ASSIGNMENT 1

OBSERVATIONS REPORT
Stance classification and Headline-Body Congruity Checking

Name: BUDDI KIRAN CHAITANYA
Roll No: 214161002

DATASET 1: FNC-1 (4 CLASS):

Keras tokenizer has been fitted on the vocabulary in the article bodies and in the headlines. And a 2030 dimension feature vector (30 dimensions features of headline concatenated with 2000 dimension feature vector of article body) has been generated for each headline body pair. The train data has been split in 80:20 ratio and the following is details of the metrics observed on 4 different classifiers (Decision Trees, Random Forest, KNearest, SGDC). The scoring metric is arrived at using FNC-1 baseline criteria (as % of best score possible)

Raw features:

```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.03      0.15      0.05       142
           1       0.01      0.08      0.01        13
           2       0.04      0.40      0.08       217
           3       0.94      0.52      0.67      5411

    accuracy                           0.50      5783
   macro avg       0.25      0.28      0.20      5783
weighted avg       0.88      0.50      0.63      5783

-------------------------------------------------------------
|            |   agree    | disagree  |  discuss  | unrelated |
-------------------------------------------------------------
|   agree    |     21     |     3     |     40    |     78    |
-------------------------------------------------------------
| disagree   |      2     |     1     |      3    |      7    |
-------------------------------------------------------------
|  discuss   |     27     |     6     |     86    |     98    |
-------------------------------------------------------------
| unrelated  |     664    |    168    |   1792    |   2787    |
-------------------------------------------------------------
Score: 825.0 out of 1724.75     (47.83301927815626%)
```

```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.03      0.11      0.05       142
           1       0.00      0.00      0.00        13
           2       0.05      0.41      0.08       217
           3       0.94      0.58      0.71      5411

    accuracy                           0.56      5783
   macro avg       0.25      0.27      0.21      5783
weighted avg       0.88      0.56      0.67      5783

-------------------------------------------------------------
|            |   agree    | disagree  |  discuss  | unrelated |
-------------------------------------------------------------
|   agree    |     15     |     1     |     36    |     90    |
-------------------------------------------------------------
| disagree   |      2     |     0     |      3    |      8    |
-------------------------------------------------------------
|  discuss   |     10     |     0     |     90    |    117    |
-------------------------------------------------------------
| unrelated  |     457    |     8     |   1821    |   3125    |
-------------------------------------------------------------
Score: 899.25 out of 1724.75    (52.137991013190316%)
```

Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = none, dimension reduction = none)

	precision	recall	f1-score	support
0	0.03	0.15	0.05	142
1	0.00	0.00	0.00	13
2	0.04	0.33	0.08	217
3	0.94	0.57	0.71	5411
accuracy			0.55	5783
macro avg	0.25	0.26	0.21	5783
weighted avg	0.88	0.55	0.67	5783

	agree	disagree	discuss	unrelated
agree	22	4	44	72
disagree	3	0	1	9
discuss	29	0	72	116
unrelated	701	75	1542	3093

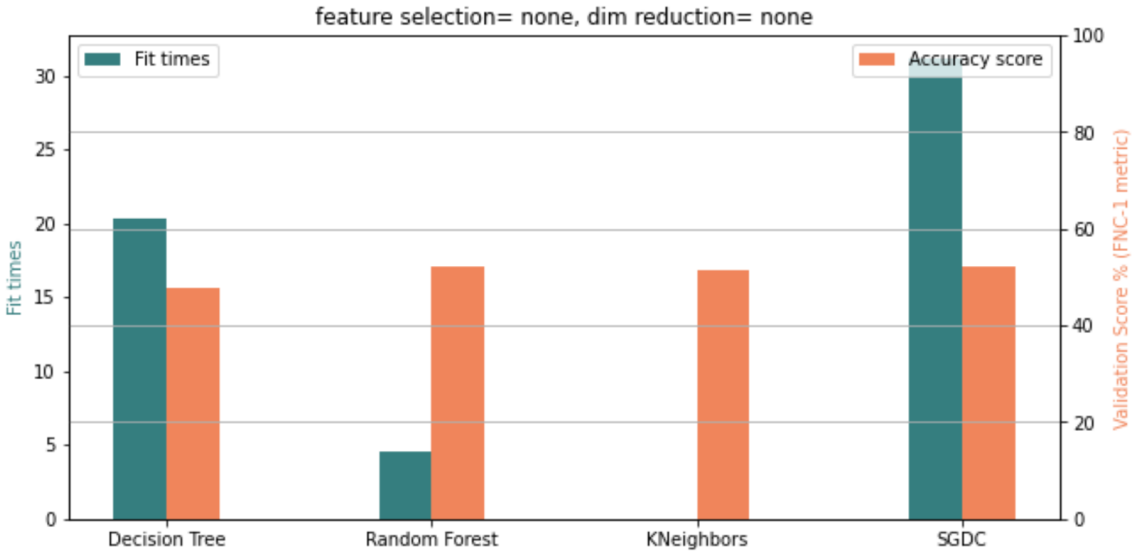
Score: 887.5 out of 1724.75 (51.45673285983476%)

Classification Report & Confusion matrix
(classifier= SGDC, feature selection = none, dimension reduction = none)

	precision	recall	f1-score	support
0	0.03	0.13	0.05	142
1	0.00	0.00	0.00	13
2	0.04	0.30	0.07	217
3	0.94	0.58	0.72	5411
accuracy			0.56	5783
macro avg	0.25	0.26	0.21	5783
weighted avg	0.88	0.56	0.68	5783

	agree	disagree	discuss	unrelated
agree	19	7	44	72
disagree	1	0	4	8
discuss	26	11	66	114
unrelated	531	236	1489	3155

Score: 897.0 out of 1724.75 (52.00753732424989%)



Feature Selection with Chi2 Method (1000-Best)

Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features, dimension reduction = none)

	precision	recall	f1-score	support
0	0.03	0.17	0.06	142
1	0.00	0.08	0.01	13
2	0.05	0.32	0.08	217
3	0.94	0.59	0.72	5411
accuracy			0.57	5783
macro avg	0.26	0.29	0.22	5783
weighted avg	0.88	0.57	0.68	5783

	agree	disagree	discuss	unrelated
agree	24	2	33	83
disagree	2	1	3	7
discuss	16	9	70	122
unrelated	679	208	1349	3175

Score: 905.0 out of 1724.75 (52.47137266270474%)

Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features, dimension reduction = none)

	precision	recall	f1-score	support
0	0.03	0.11	0.05	142
1	0.00	0.00	0.00	13
2	0.04	0.44	0.08	217
3	0.94	0.51	0.66	5411
accuracy			0.50	5783
macro avg	0.25	0.27	0.20	5783
weighted avg	0.88	0.50	0.62	5783

	agree	disagree	discuss	unrelated
agree	16	0	53	73
disagree	1	0	2	10
discuss	21	0	96	100
unrelated	509	3	2135	2764

Score: 822.25 out of 1724.75 (47.6735758805624%)

Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features, dimension reduction = none)

	precision	recall	f1-score	support
0	0.04	0.23	0.07	142
1	0.00	0.00	0.00	13
2	0.03	0.33	0.06	217
3	0.94	0.48	0.64	5411
accuracy			0.47	5783
macro avg	0.25	0.26	0.19	5783
weighted avg	0.88	0.47	0.60	5783

	agree	disagree	discuss	unrelated
agree	32	6	53	51
disagree	0	0	7	6
discuss	28	3	72	114
unrelated	732	73	2004	2602

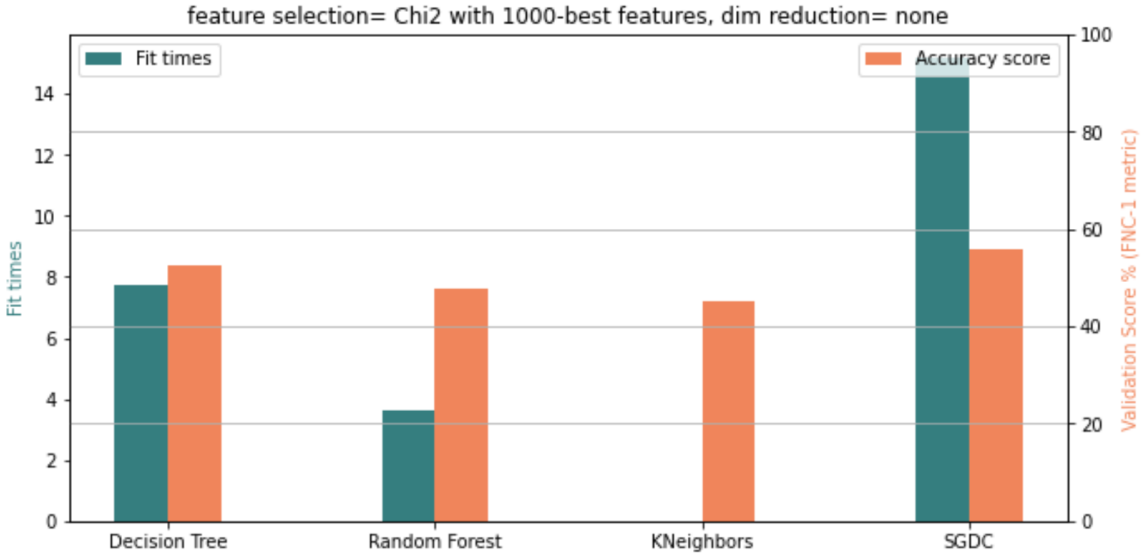
Score: 778.75 out of 1724.75 (45.15147122771416%)

Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features, dimension reduction = none)

	precision	recall	f1-score	support
0	0.03	0.07	0.05	142
1	0.00	0.00	0.00	13
2	0.03	0.23	0.06	217
3	0.94	0.65	0.77	5411
accuracy			0.62	5783
macro avg	0.25	0.24	0.22	5783
weighted avg	0.88	0.62	0.72	5783

	agree	disagree	discuss	unrelated
agree	10	4	43	85
disagree	2	0	4	7
discuss	23	12	49	133
unrelated	254	242	1390	3525

Score: 962.25 out of 1724.75 (55.790694303522255%)



Feature Selection with Chi2 (1000-best) & dimn’ reduction with PCA (300 top PCs)

Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features,
tion = PCA redn to 300 features)

	precision	recall	f1-score	support
0	0.03	0.20	0.04	142
1	0.00	0.08	0.01	13
2	0.04	0.43	0.07	217
3	0.94	0.34	0.50	5411
accuracy			0.34	5783
macro avg	0.25	0.26	0.16	5783
weighted avg	0.88	0.34	0.48	5783

	agree	disagree	discuss	unrelated
agree	28	8	64	42
disagree	1	1	5	6
discuss	35	13	94	75
unrelated	1045	238	2264	1864

Score: 620.5 out of 1724.75 (35.97622843890419%)

Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features,
tion = PCA redn to 300 features)

	precision	recall	f1-score	support
0	0.04	0.16	0.06	142
1	0.00	0.00	0.00	13
2	0.04	0.61	0.08	217
3	0.95	0.38	0.54	5411
accuracy			0.38	5783
macro avg	0.26	0.29	0.17	5783
weighted avg	0.89	0.38	0.51	5783

	agree	disagree	discuss	unrelated
agree	23	1	66	52
disagree	3	0	4	6
discuss	26	0	132	59
unrelated	564	0	2777	2070

Score: 697.5 out of 1724.75 (40.44064357153211%)

Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= PCA redn to 300 features)

	precision	recall	f1-score	support
0	0.04	0.23	0.07	142
1	0.00	0.00	0.00	13
2	0.03	0.35	0.06	217
3	0.94	0.46	0.62	5411
accuracy			0.45	5783
macro avg	0.25	0.26	0.19	5783
weighted avg	0.88	0.45	0.58	5783

	agree	disagree	discuss	unrelated
agree	32	5	57	48
disagree	1	0	6	6
discuss	29	2	77	109
unrelated	727	55	2128	2501

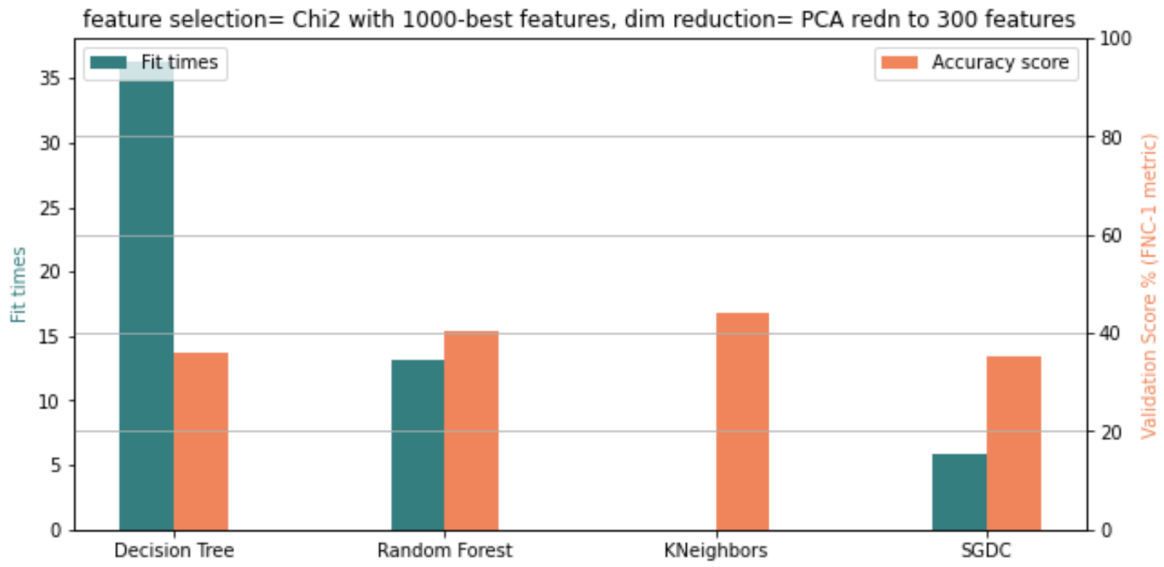
Score: 759.25 out of 1724.75 (44.020872590230 47%)

Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features,
A redn to 300 features)

	precision	recall	f1-score	support
0	0.03	0.36	0.05	142
1	0.00	0.23	0.01	13
2	0.04	0.21	0.06	217
3	0.94	0.35	0.51	5411
accuracy			0.34	5783
macro avg	0.25	0.29	0.16	5783
weighted avg	0.88	0.34	0.48	5783

	agree	disagree	discuss	unrelated
agree	51	26	24	41
disagree	6	3	2	2
discuss	52	39	46	80
unrelated	1652	750	1133	1876

Score: 606.25 out of 1724.75 (35.15002174228149%)



Feature selection (Chi2, 1000-best) and dimension reduction (SVD, 300-top)

```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features,
tion = SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.03      0.25      0.06       142
           1       0.00      0.08      0.01        13
           2       0.04      0.41      0.07       217
           3       0.94      0.35      0.51      5411

    accuracy                           0.35      5783
   macro avg       0.25      0.27      0.16      5783
weighted avg       0.88      0.35      0.48      5783
```

	agree	disagree	discuss	unrelated
agree	36	5	46	55
disagree	4	1	3	5
discuss	42	18	90	67
unrelated	1028	312	2165	1906

```
Score: 633.0 out of 1724.75     (36.700971155239895%)
```

```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features,
tion = SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.04      0.15      0.06       142
           1       0.00      0.00      0.00        13
           2       0.04      0.53      0.07       217
           3       0.94      0.38      0.54      5411

    accuracy                           0.38      5783
   macro avg       0.25      0.26      0.17      5783
weighted avg       0.88      0.38      0.51      5783
```

	agree	disagree	discuss	unrelated
agree	21	1	70	50
disagree	4	0	6	3
discuss	19	0	114	84
unrelated	546	3	2826	2036

```
Score: 669.0 out of 1724.75     (38.788230178286715%)
```

```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.04      0.23      0.07       142
           1       0.00      0.00      0.00        13
           2       0.03      0.35      0.06       217
           3       0.94      0.46      0.61      5411

    accuracy                           0.45      5783
   macro avg       0.25      0.26      0.19      5783
weighted avg       0.88      0.45      0.58      5783
```

	agree	disagree	discuss	unrelated
agree	32	6	53	51
disagree	1	0	6	6
discuss	32	2	76	107
unrelated	735	56	2148	2472

```
Score: 751.0 out of 1724.75     (43.542542397448905%)
```

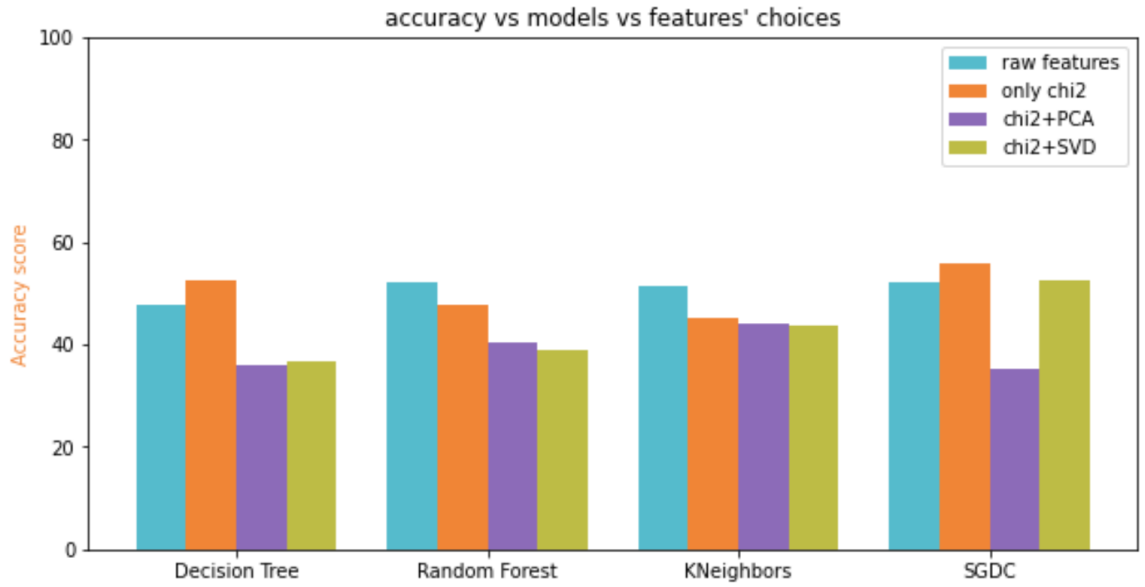
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features, d
D redn to 300 features)

	precision	recall	f1-score	support
0	0.04	0.24	0.07	142
1	0.00	0.00	0.00	13
2	0.03	0.24	0.06	217
3	0.94	0.59	0.72	5411
accuracy			0.57	5783
macro avg	0.25	0.27	0.21	5783
weighted avg	0.88	0.57	0.68	5783

	agree	disagree	discuss	unrelated
agree	34	1	38	69
disagree	5	0	4	4
discuss	35	1	53	128
unrelated	792	6	1427	3186

Score: 904.5 out of 1724.75 (52.44238295405131%)

Comparison between the feature engineering methods



Word Embedding with GloVe vectors:

A shallow network is fed a feature vector of length 810 (extracted from both body and headline) and the network summary is as follows:

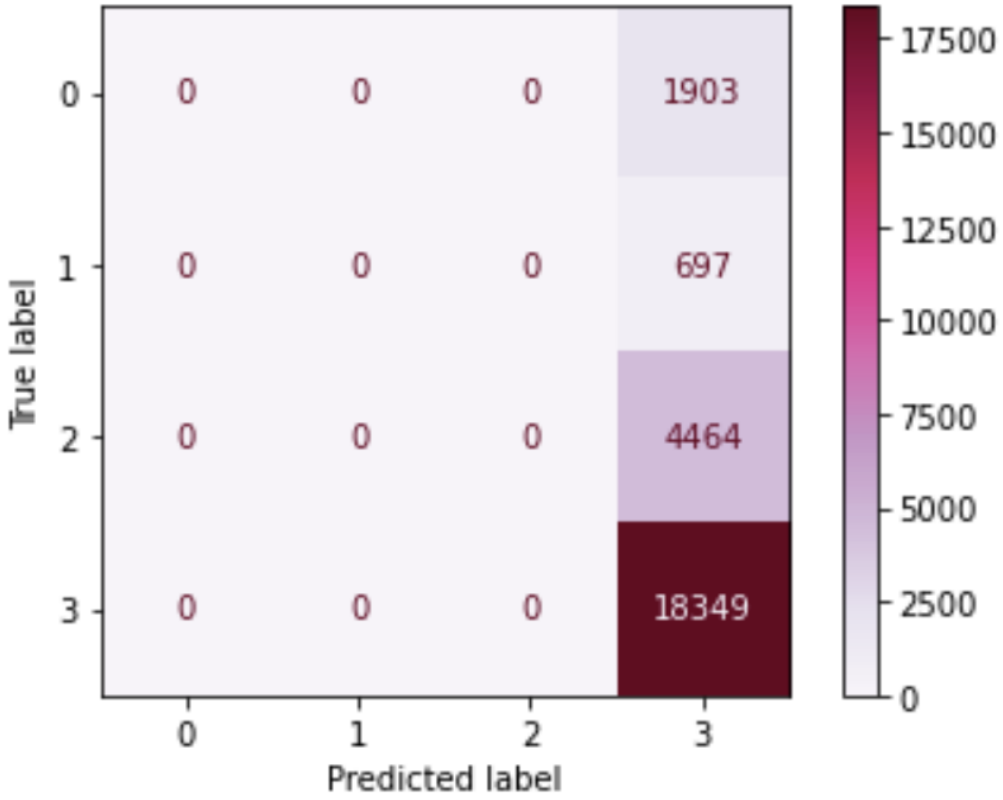
```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 815, 300)          1500300
_________________________________________________________________
lstm_1 (LSTM)                (None, 1)                 1208
_________________________________________________________________
dense_1 (Dense)              (None, 4)                 8
=================================================================
Total params: 1,501,516
Trainable params: 1,216
Non-trainable params: 1,500,300
_________________________________________________________________
```

```
model.fit(train_features,y_train, epochs=1, verbose=1)
```

```
Train on 22073 samples
```

```
accuracy_embedding
```

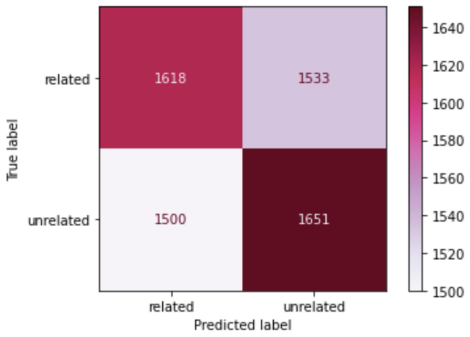
```
0.7220320308503522
```



Dataset 2: NELA (2-CLASS)
Performance metric used: Accuracy score (0-1)

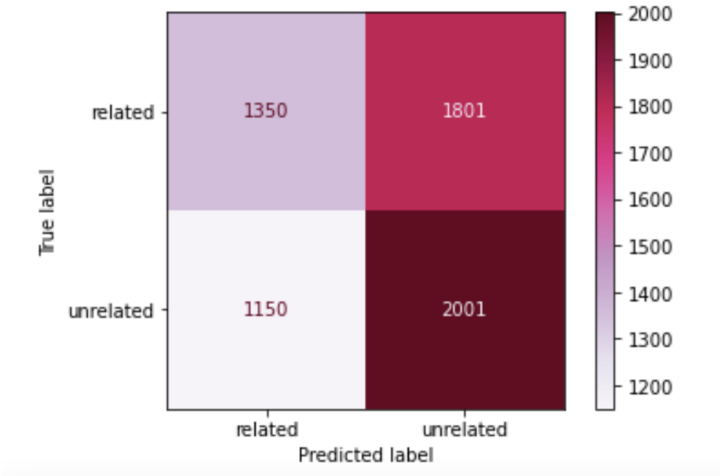
Raw Features (2030 features):

```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.52      0.52      0.52      3151
           1       0.52      0.51      0.52      3151

    accuracy                           0.52      6302
   macro avg       0.52      0.52      0.52      6302
weighted avg       0.52      0.52      0.52      6302

[[1618 1533]
 [1500 1651]]
```



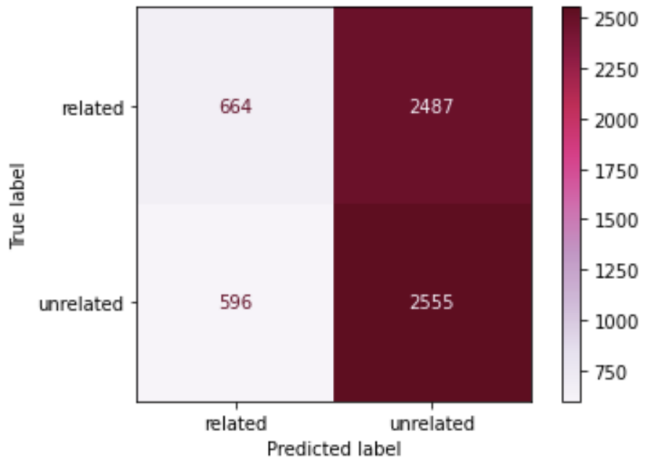
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = none, dimension reduction = none)

	precision	recall	f1-score	support
0	0.53	0.64	0.58	3151
1	0.54	0.43	0.48	3151
accuracy			0.53	6302
macro avg	0.53	0.53	0.53	6302
weighted avg	0.53	0.53	0.53	6302



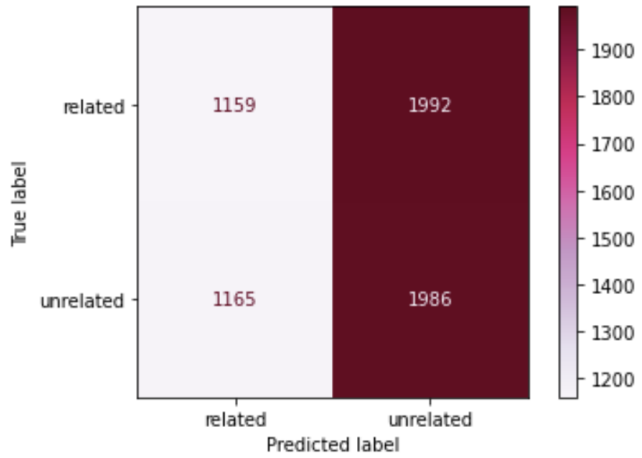
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = none, dimension reduction = none)

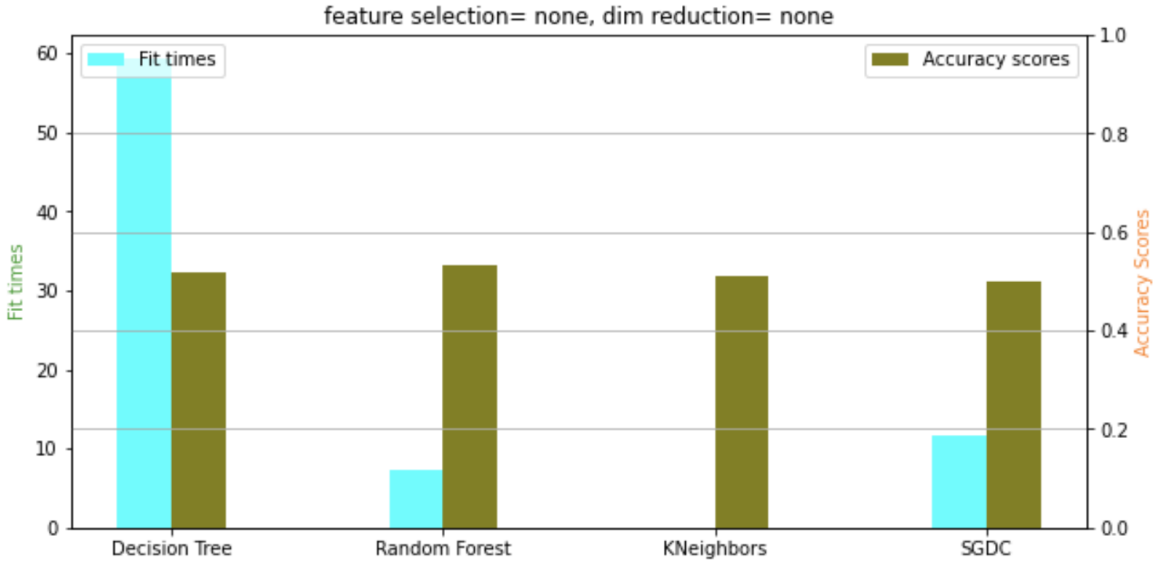
	precision	recall	f1-score	support
0	0.51	0.81	0.62	3151
1	0.53	0.21	0.30	3151
accuracy			0.51	6302
macro avg	0.52	0.51	0.46	6302
weighted avg	0.52	0.51	0.46	6302



Classification Report & Confusion matrix
(classifier= SGDC, feature selection = none, dimension reduction = none)

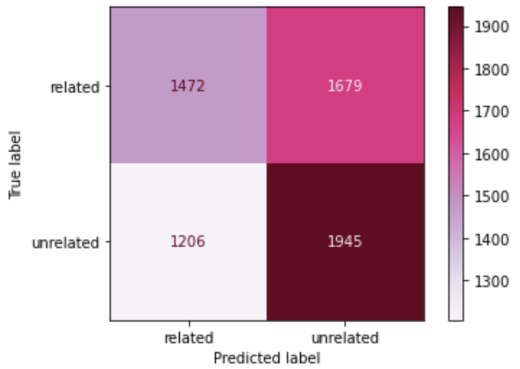
	precision	recall	f1-score	support
0	0.50	0.63	0.56	3151
1	0.50	0.37	0.42	3151
accuracy			0.50	6302
macro avg	0.50	0.50	0.49	6302
weighted avg	0.50	0.50	0.49	6302



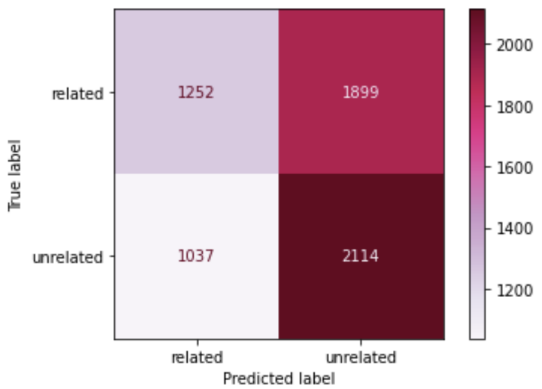


Feature Selection with Chi2 (1000-best features)

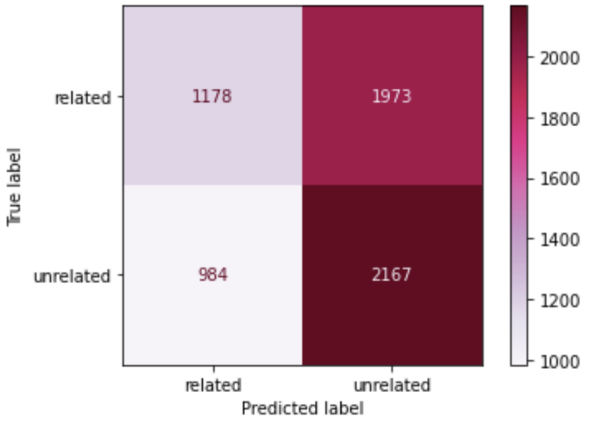
```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features,
tion = none)

              precision    recall  f1-score   support

           0       0.54      0.62      0.57      3151
           1       0.55      0.47      0.51      3151

    accuracy                           0.54      6302
   macro avg       0.54      0.54      0.54      6302
weighted avg       0.54      0.54      0.54      6302

[[1472 1679]
 [1206 1945]]

<Figure size 432x288 with 0 Axes>
```



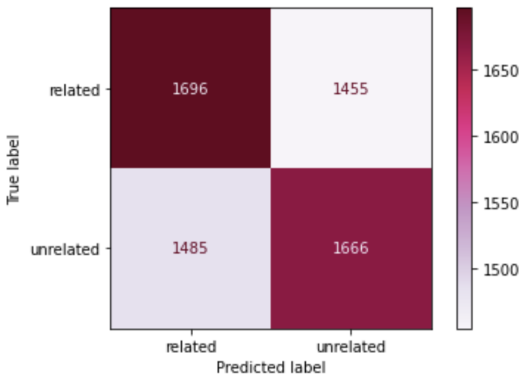

```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best
tion = none)

              precision    recall  f1-score   support

           0       0.53      0.67      0.59      3151
           1       0.55      0.40      0.46      3151

    accuracy                           0.53      6302
   macro avg       0.54      0.53      0.53      6302
weighted avg       0.54      0.53      0.53      6302

[[1252 1899]
 [1037 2114]]

<Figure size 432x288 with 0 Axes>
```

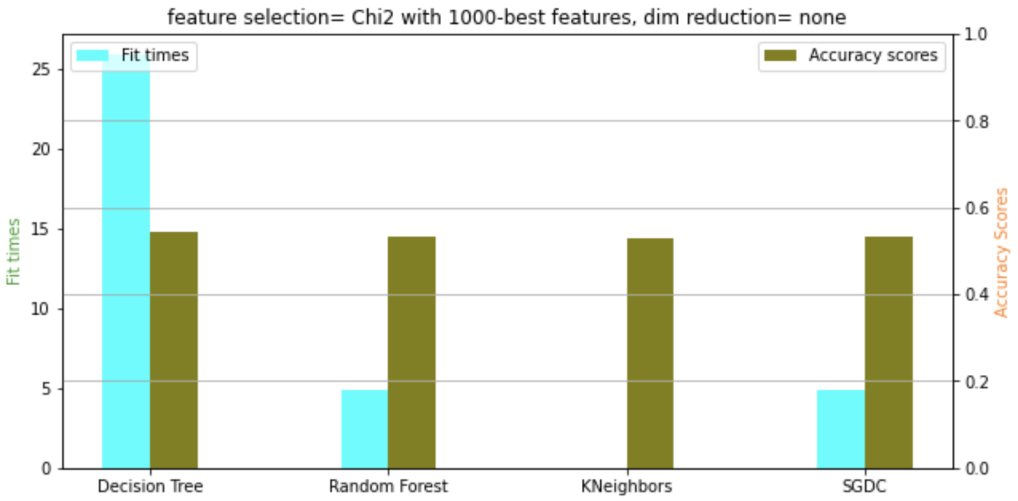


```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features
= none)

              precision    recall  f1-score   support

           0       0.52      0.69      0.59      3151
           1       0.54      0.37      0.44      3151

    accuracy                           0.53      6302
   macro avg       0.53      0.53      0.52      6302
weighted avg       0.53      0.53      0.52      6302

[[1178 1973]
 [ 984 2167]]

<Figure size 432x288 with 0 Axes>
```



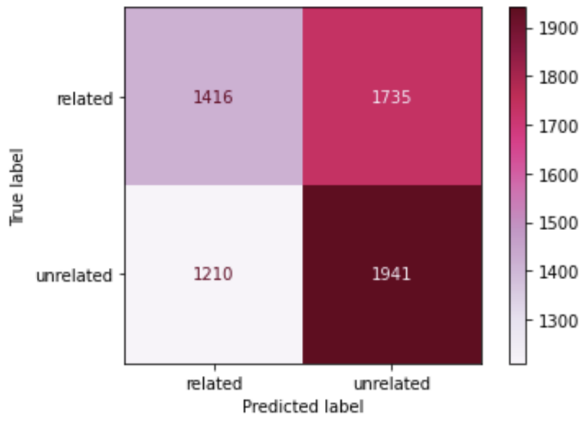

```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features,
ne)

              precision    recall  f1-score   support

           0       0.53      0.53      0.53      3151
           1       0.53      0.54      0.54      3151

    accuracy                           0.53      6302
   macro avg       0.53      0.53      0.53      6302
weighted avg       0.53      0.53      0.53      6302

[[1696 1455]
 [1485 1666]]

<Figure size 432x288 with 0 Axes>
```



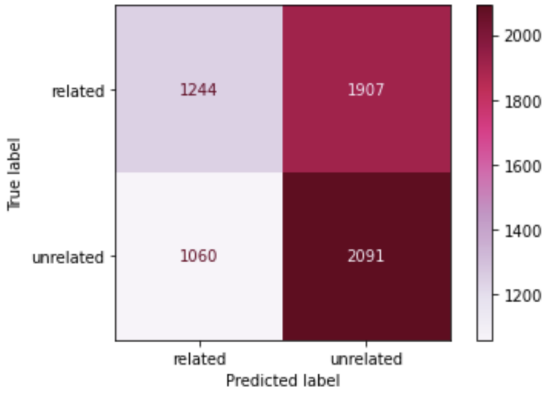


Feature Selection with Chi2 and dimension reduction using PCA (to 300 features)

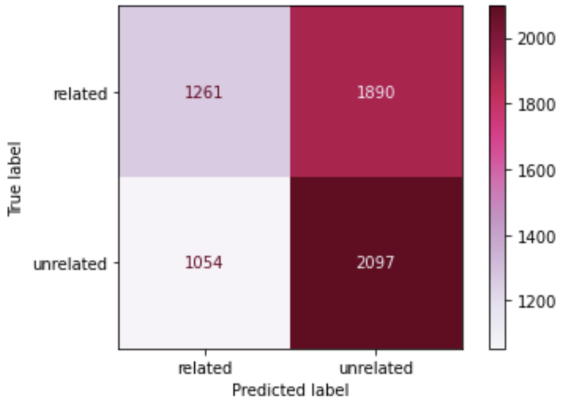
```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features,
tion = PCA redn to 300 features)

              precision    recall  f1-score   support

           0       0.53      0.62      0.57      3151
           1       0.54      0.45      0.49      3151

    accuracy                           0.53      6302
   macro avg       0.53      0.53      0.53      6302
weighted avg       0.53      0.53      0.53      6302

[[1416 1735]
 [1210 1941]]

<Figure size 432x288 with 0 Axes>
```



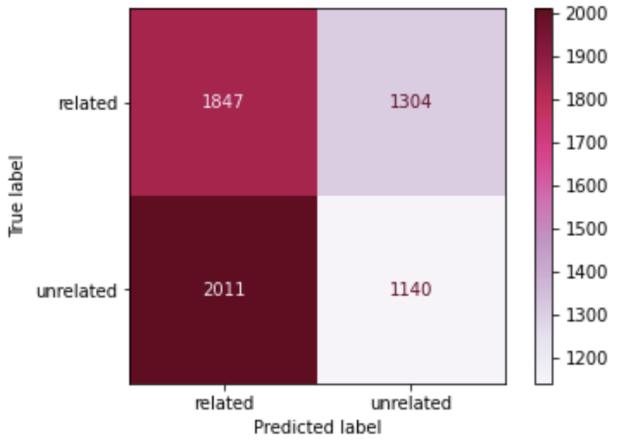
```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features,
tion = PCA redn to 300 features)

              precision    recall  f1-score   support

           0       0.52      0.66      0.58      3151
           1       0.54      0.39      0.46      3151

    accuracy                           0.53      6302
   macro avg       0.53      0.53      0.52      6302
weighted avg       0.53      0.53      0.52      6302

[[1244 1907]
 [1060 2091]]

<Figure size 432x288 with 0 Axes>
```

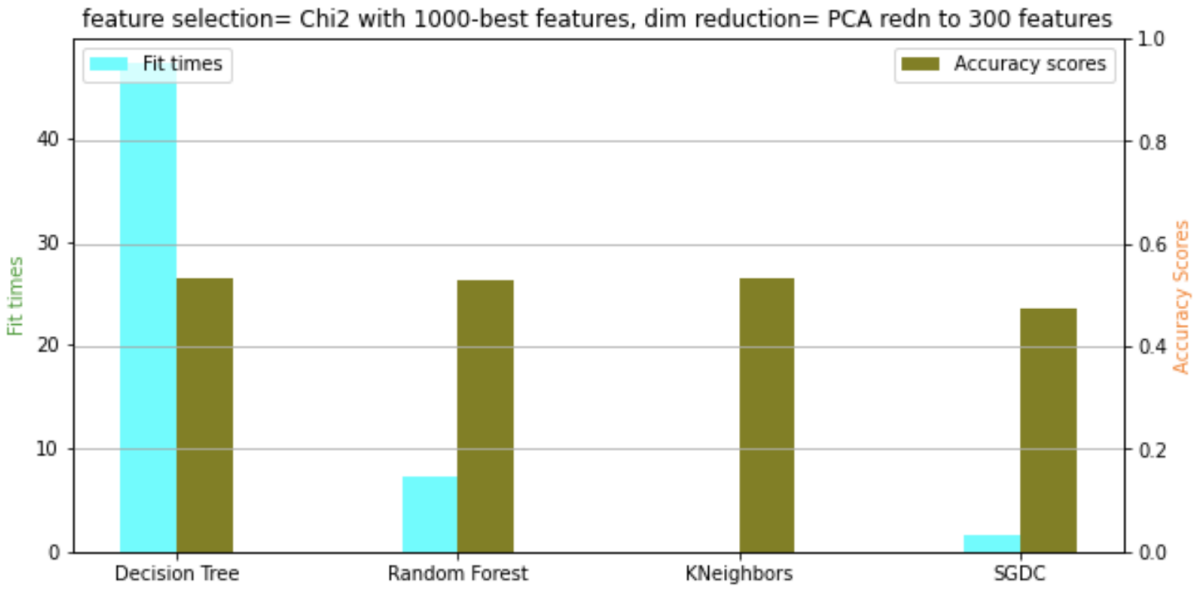



```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= PCA redn to 300 features)

              precision    recall  f1-score   support

           0       0.53      0.67      0.59      3151
           1       0.54      0.40      0.46      3151

    accuracy                           0.53      6302
   macro avg       0.54      0.53      0.52      6302
weighted avg       0.54      0.53      0.52      6302

[[1261 1890]
 [1054 2097]]

<Figure size 432x288 with 0 Axes>
```



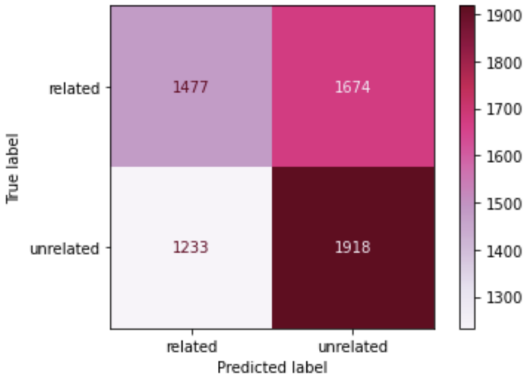
```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features,
A redn to 300 features)

              precision    recall  f1-score   support

           0       0.47      0.36      0.41      3151
           1       0.48      0.59      0.53      3151

    accuracy                           0.47      6302
   macro avg       0.47      0.47      0.47      6302
weighted avg       0.47      0.47      0.47      6302

[[1847 1304]
 [2011 1140]]

<Figure size 432x288 with 0 Axes>
```



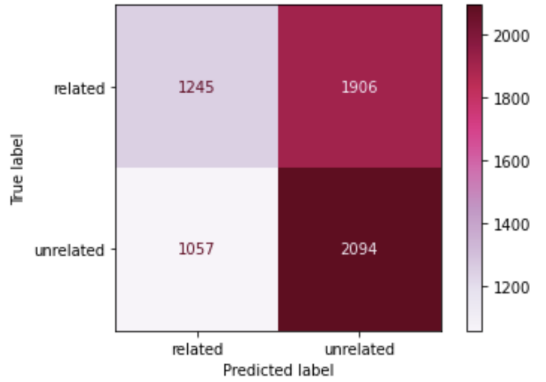


Feature selection with Chi2 (1000-best) and then dimension reduction using SVD (300 top SVs)

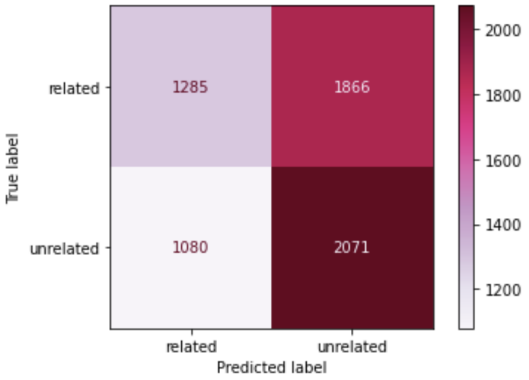
```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features,
tion = SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.53      0.61      0.57      3151
           1       0.55      0.47      0.50      3151

    accuracy                           0.54      6302
   macro avg       0.54      0.54      0.54      6302
weighted avg       0.54      0.54      0.54      6302

[[1477 1674]
 [1233 1918]]

<Figure size 432x288 with 0 Axes>
```



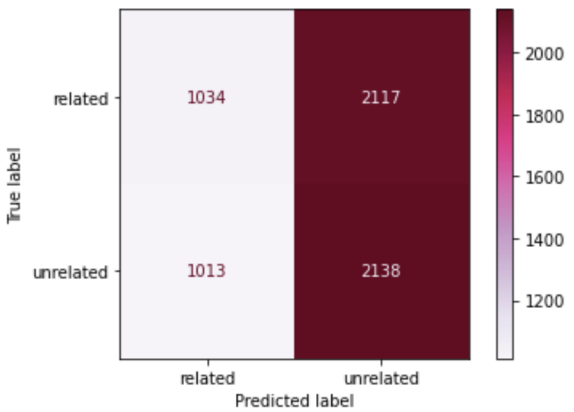

```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features,
tion = SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.52      0.66      0.59      3151
           1       0.54      0.40      0.46      3151

    accuracy                           0.53      6302
   macro avg       0.53      0.53      0.52      6302
weighted avg       0.53      0.53      0.52      6302

[[1245 1906]
 [1057 2094]]

<Figure size 432x288 with 0 Axes>
```

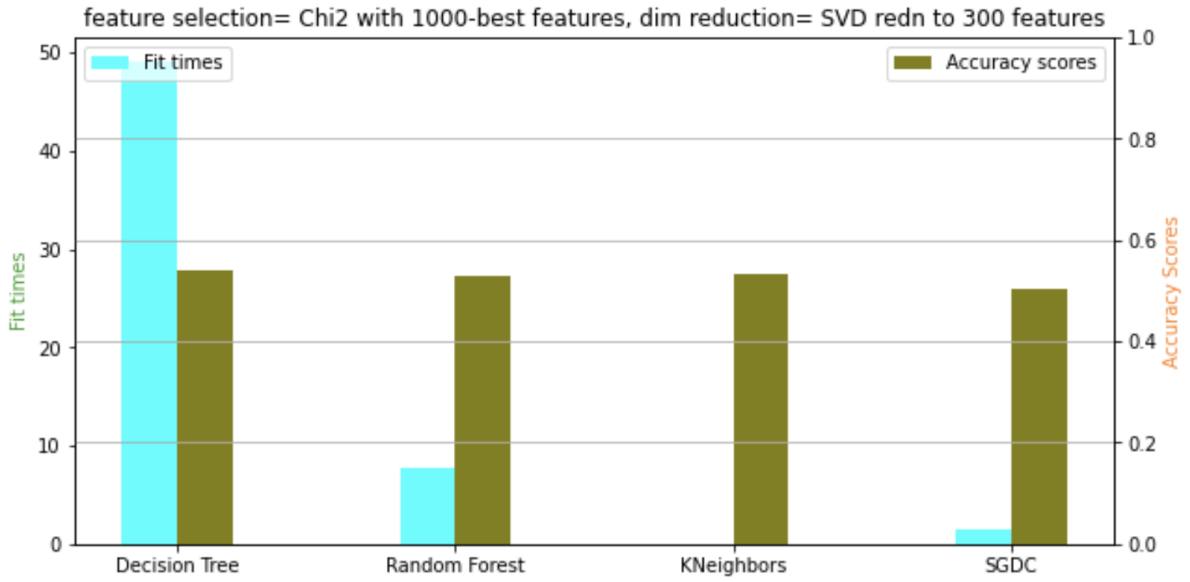


```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.53      0.66      0.58      3151
           1       0.54      0.41      0.47      3151

    accuracy                           0.53      6302
   macro avg       0.53      0.53      0.53      6302
weighted avg       0.53      0.53      0.53      6302

[[1285 1866]
 [1080 2071]]

<Figure size 432x288 with 0 Axes>
```

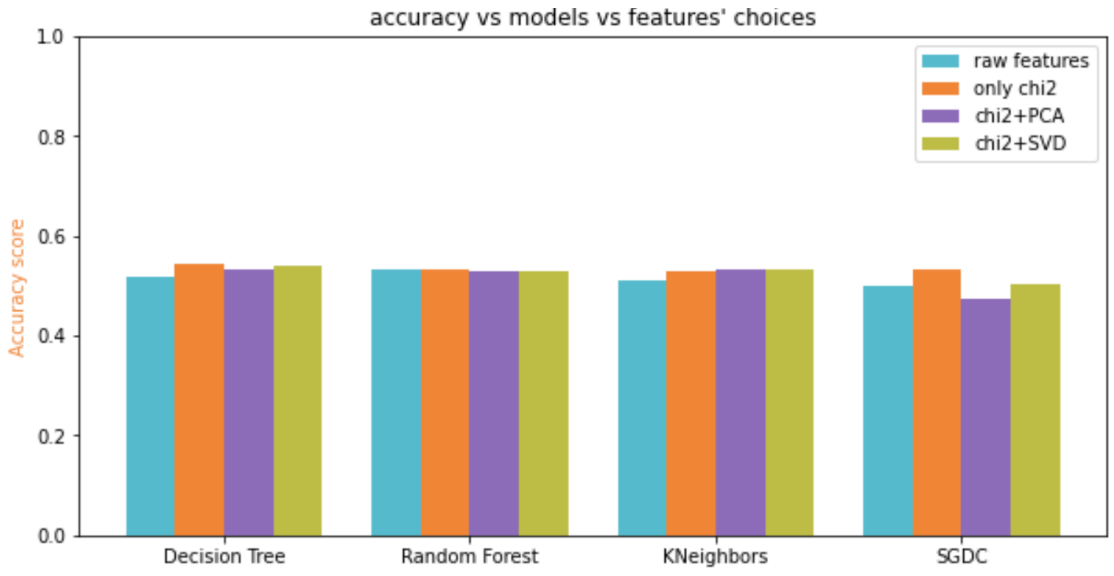



```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features,
D redn to 300 features)

              precision    recall  f1-score   support

           0       0.50      0.68      0.58      3151
           1       0.51      0.33      0.40      3151

    accuracy                           0.50      6302
   macro avg       0.50      0.50      0.49      6302
weighted avg       0.50      0.50      0.49      6302

[[1034 2117]
 [1013 2138]]

<Figure size 432x288 with 0 Axes>
```





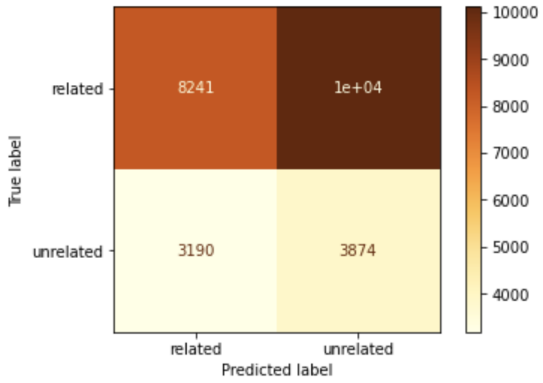
Comparison of Feature engineering approaches over NELA



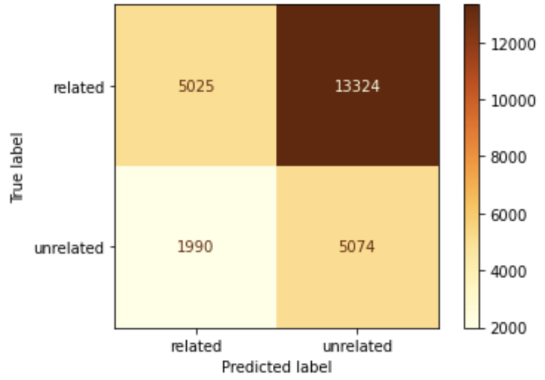
Dataset 3 : FNC (2-Class):

Raw Features (2030 features from headline & body)

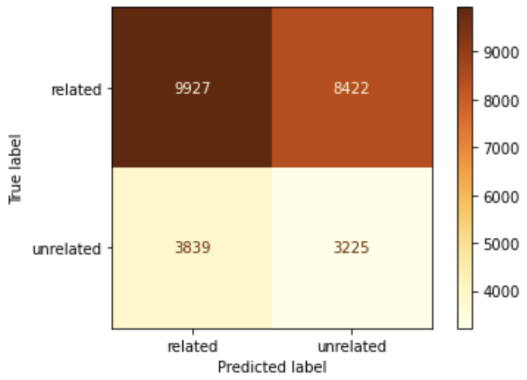
```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.28      0.55      0.37      7064
           1       0.72      0.45      0.55     18349

    accuracy                           0.48     25413
   macro avg       0.50      0.50      0.46     25413
weighted avg       0.60      0.48      0.50     25413

[[ 8241 10108]
 [ 3190  3874]]

<Figure size 432x288 with 0 Axes>
```



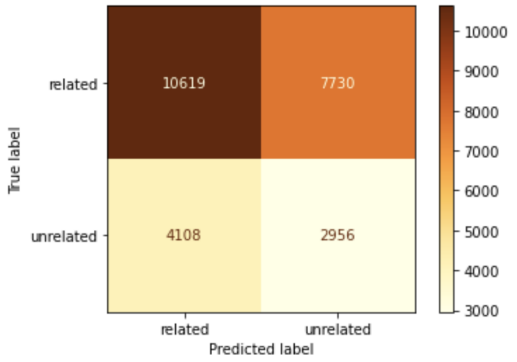

```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.28      0.72      0.40      7064
           1       0.72      0.27      0.40     18349

    accuracy                           0.40     25413
   macro avg       0.50      0.50      0.40     25413
weighted avg       0.59      0.40      0.40     25413

[[ 5025 13324]
 [ 1990  5074]]

<Figure size 432x288 with 0 Axes>
```

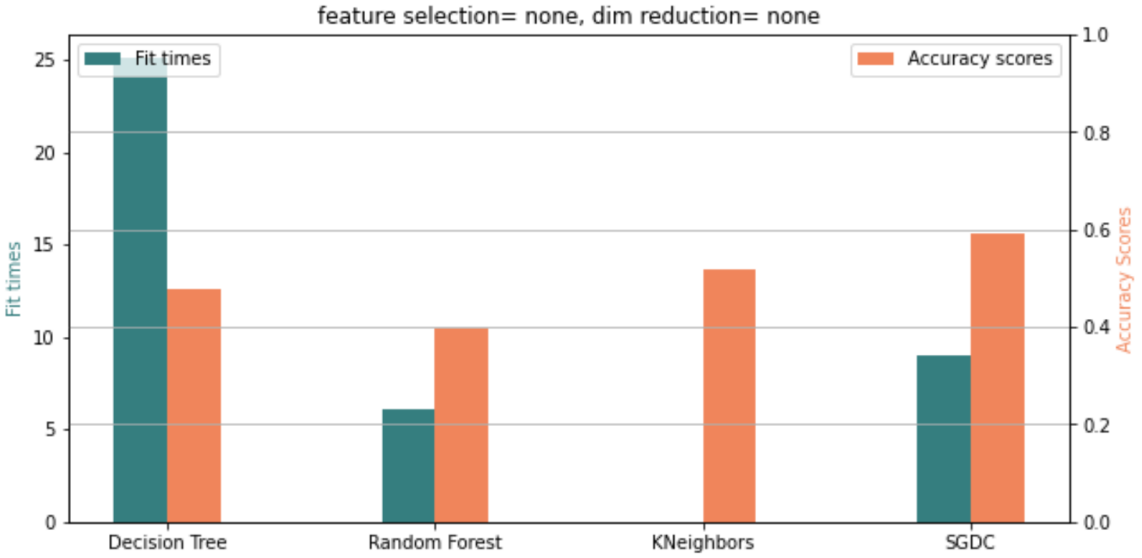


```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.28      0.46      0.34      7064
           1       0.72      0.54      0.62     18349

    accuracy                           0.52     25413
   macro avg       0.50      0.50      0.48     25413
weighted avg       0.60      0.52      0.54     25413

[[9927 8422]
 [3839 3225]]

<Figure size 432x288 with 0 Axes>
```



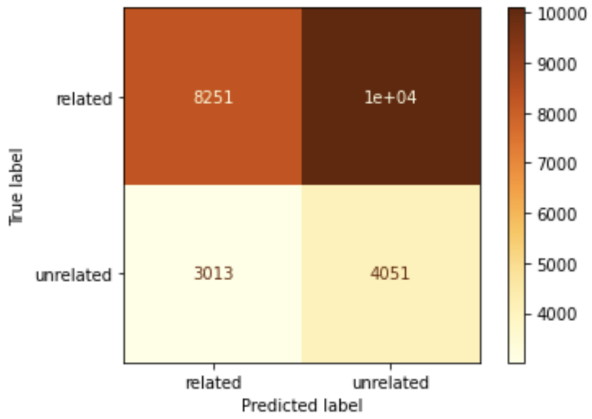

```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = none, dimension reduction = none)

              precision    recall  f1-score   support

           0       0.28      0.42      0.33      7064
           1       0.72      0.58      0.64     18349

    accuracy                           0.53     25413
   macro avg       0.50      0.50      0.49     25413
weighted avg       0.60      0.53      0.56     25413

[[10619  7730]
 [ 4108  2956]]

<Figure size 432x288 with 0 Axes>
```



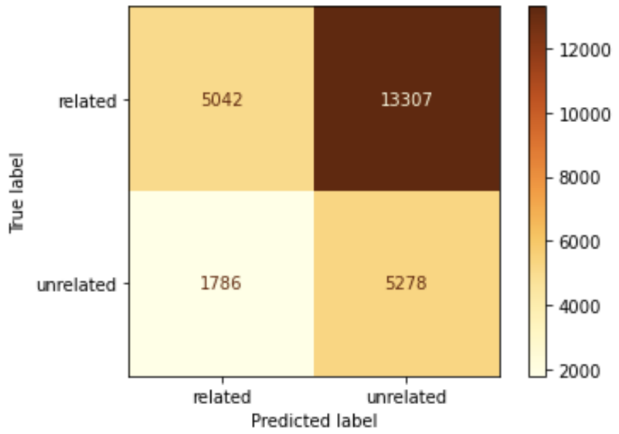


Feature Selection with Chi2 (1000-best)

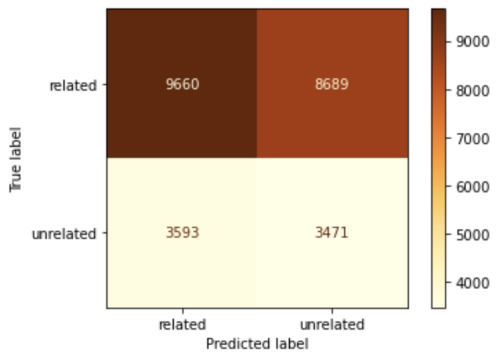
```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best
tion = none)

              precision    recall  f1-score   support

           0       0.29      0.57      0.38      7064
           1       0.73      0.45      0.56     18349

    accuracy                           0.48     25413
   macro avg       0.51      0.51      0.47     25413
weighted avg       0.61      0.48      0.51     25413

[[ 8251 10098]
 [ 3013  4051]]

<Figure size 432x288 with 0 Axes>
```



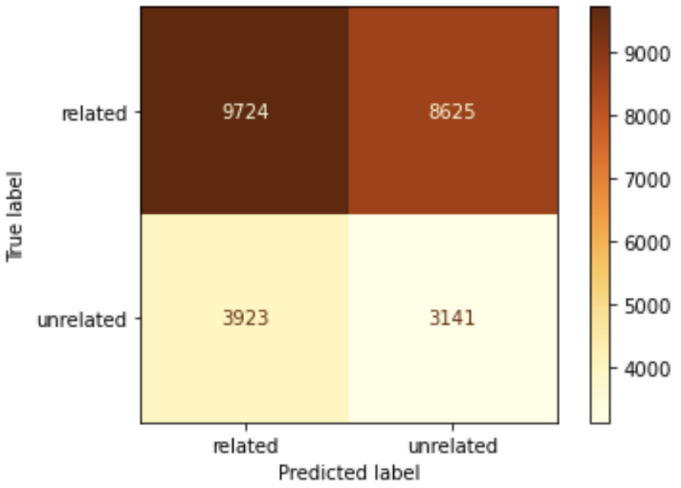
```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best f
tion = none)

              precision    recall  f1-score   support

           0       0.28      0.75      0.41      7064
           1       0.74      0.27      0.40     18349

    accuracy                           0.41     25413
   macro avg       0.51      0.51      0.41     25413
weighted avg       0.61      0.41      0.40     25413

[[ 5042 13307]
 [ 1786  5278]]

<Figure size 432x288 with 0 Axes>
```

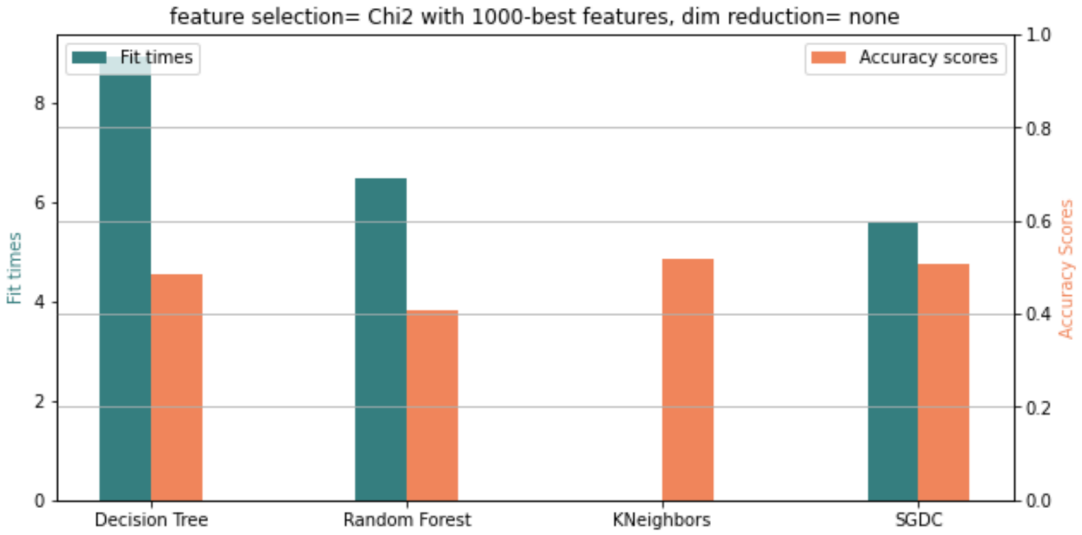



```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= none)

              precision    recall  f1-score   support

           0       0.29      0.49      0.36      7064
           1       0.73      0.53      0.61     18349

    accuracy                           0.52     25413
   macro avg       0.51      0.51      0.49     25413
weighted avg       0.61      0.52      0.54     25413

[[9660 8689]
 [3593 3471]]

<Figure size 432x288 with 0 Axes>
```



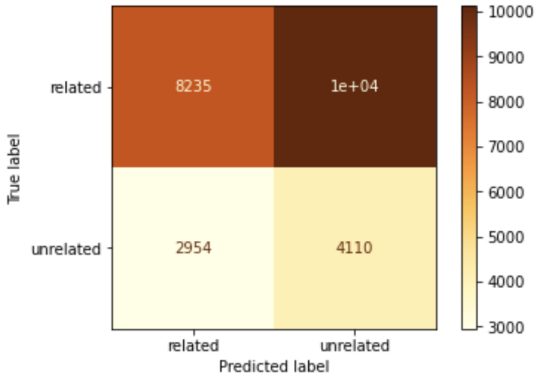
```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features,
ne)

              precision    recall  f1-score   support

           0       0.27      0.44      0.33      7064
           1       0.71      0.53      0.61     18349

    accuracy                           0.51     25413
   macro avg       0.49      0.49      0.47     25413
weighted avg       0.59      0.51      0.53     25413

[[9724 8625]
 [3923 3141]]

<Figure size 432x288 with 0 Axes>
```



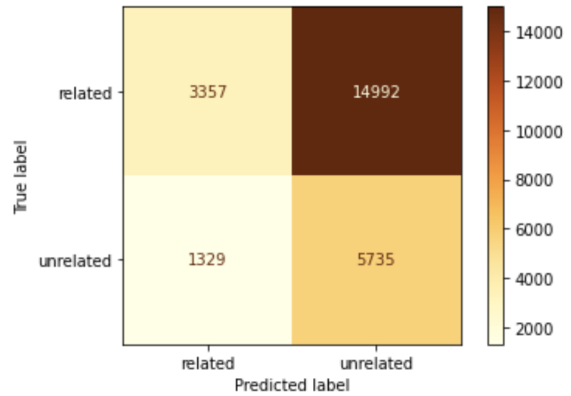


with feature selection (Chi2, 1000-best) and dimension reduction (PCA, top 300)

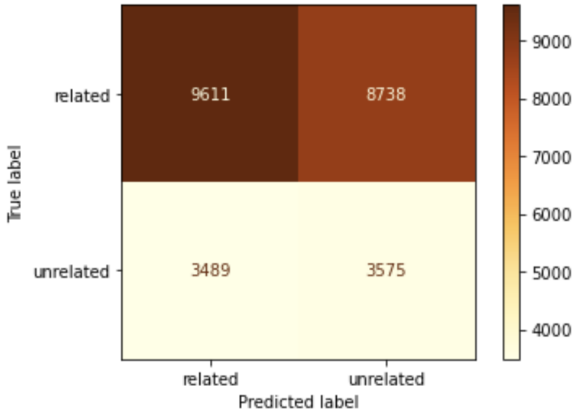
```
Classification Report & Confusion matrix
(classifier= Decision tree, feature selection = Chi2 with 1000-best features,
tion = PCA redn to 300 features)

              precision    recall  f1-score   support

           0       0.29      0.58      0.39      7064
           1       0.74      0.45      0.56     18349

    accuracy                           0.49     25413
   macro avg       0.51      0.52      0.47     25413
weighted avg       0.61      0.49      0.51     25413

[[ 8235 10114]
 [ 2954  4110]]

<Figure size 432x288 with 0 Axes>
```



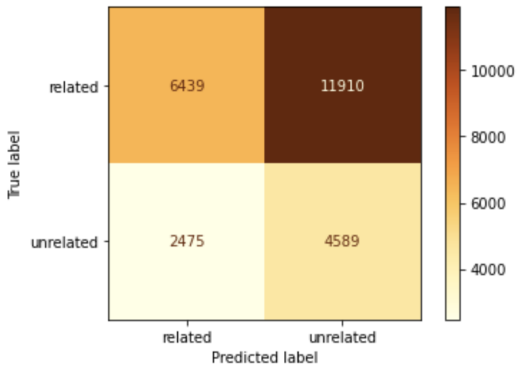

```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features,
tion = PCA redn to 300 features)

              precision    recall  f1-score   support

           0       0.28      0.81      0.41      7064
           1       0.72      0.18      0.29     18349

    accuracy                           0.36     25413
   macro avg       0.50      0.50      0.35     25413
weighted avg       0.59      0.36      0.33     25413

[[ 3357 14992]
 [ 1329  5735]]

<Figure size 432x288 with 0 Axes>
```

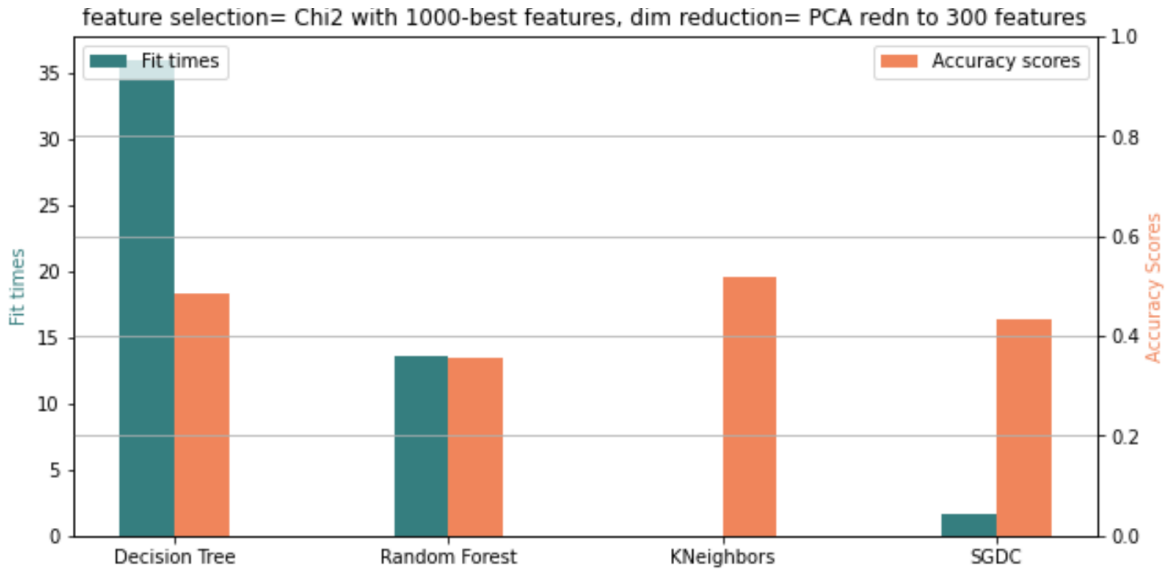


```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= PCA redn to 300 features)

              precision    recall  f1-score   support

           0       0.29      0.51      0.37      7064
           1       0.73      0.52      0.61     18349

    accuracy                           0.52     25413
   macro avg       0.51      0.51      0.49     25413
weighted avg       0.61      0.52      0.54     25413

[[9611 8738]
 [3489 3575]]

<Figure size 432x288 with 0 Axes>
```



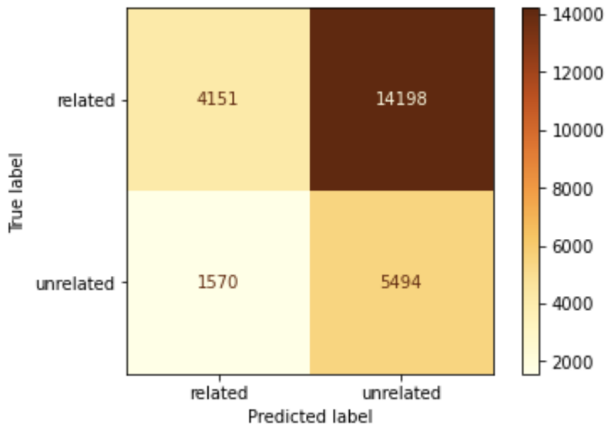

```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features,
A redn to 300 features)

              precision    recall  f1-score   support

           0       0.28      0.65      0.39      7064
           1       0.72      0.35      0.47     18349

    accuracy                           0.43     25413
   macro avg       0.50      0.50      0.43     25413
weighted avg       0.60      0.43      0.45     25413

[[ 6439 11910]
 [ 2475  4589]]

<Figure size 432x288 with 0 Axes>
```



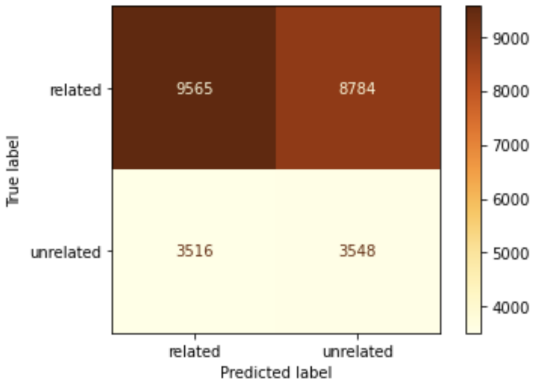


With Feature selection (Chi2 1000-best) and SVD (top-300)

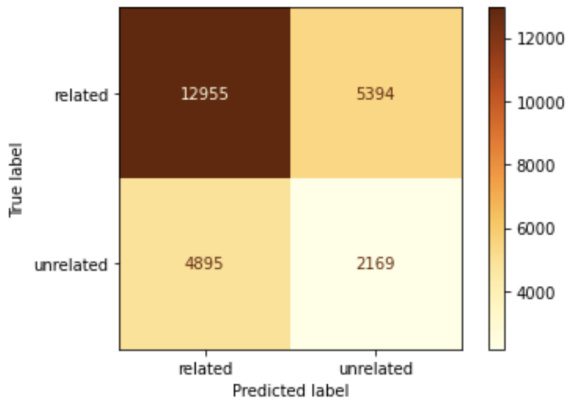
```
Classification Report & Confusion matrix
(classifier= Random Forest, feature selection = Chi2 with 1000-best features,
tion = SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.28      0.78      0.41      7064
           1       0.73      0.23      0.34     18349

    accuracy                           0.38     25413
   macro avg       0.50      0.50      0.38     25413
weighted avg       0.60      0.38      0.36     25413

[[ 4151 14198]
 [ 1570  5494]]

<Figure size 432x288 with 0 Axes>
```

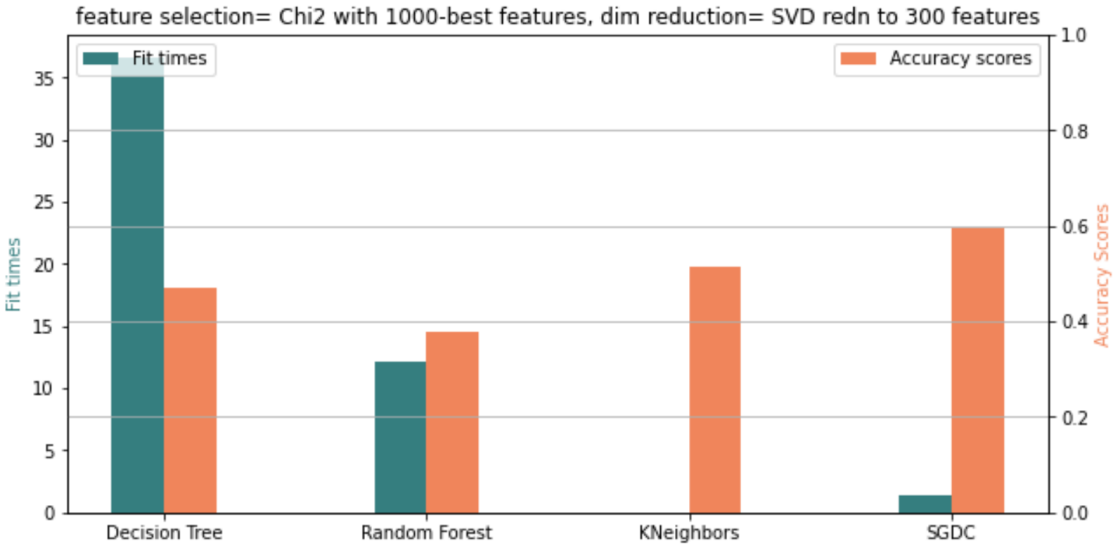


```
Classification Report & Confusion matrix
(classifier= KNeighbor, feature selection = Chi2 with 1000-best features,
= SVD redn to 300 features)

              precision    recall  f1-score   support

           0       0.29      0.50      0.37      7064
           1       0.73      0.52      0.61     18349

    accuracy                           0.52     25413
   macro avg       0.51      0.51      0.49     25413
weighted avg       0.61      0.52      0.54     25413

[[9565 8784]
 [3516 3548]]

<Figure size 432x288 with 0 Axes>
```




```
Classification Report & Confusion matrix
(classifier= SGDC, feature selection = Chi2 with 1000-best features, (
D redn to 300 features)

              precision    recall  f1-score   support

           0       0.29      0.31      0.30      7064
           1       0.73      0.71      0.72     18349

    accuracy                           0.60     25413
   macro avg       0.51      0.51      0.51     25413
weighted avg       0.60      0.60      0.60     25413

[[12955  5394]
 [ 4895  2169]]

<Figure size 432x288 with 0 Axes>
```





Comparison of feature Engineering choices over FNC dataset: