

## Classification:

### Huffpost(100)

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SPELL CHECKER COMMENTS Python + - [ ] [x] ... ^ x

~/Downloads/CDA_Main base Py at 11:21:16 AM
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py
Evaluating Classification:
Accuracy: 0.76
F1 Score: 0.78

Evaluating Diversity:
Diversity Score: 0.36

Evaluating Perplexity:
Perplexity: 156.34
~/Downloads/CDA_Main base Py at 11:21:16 AM
>

Ln 2, Col 16 Spaces: 4 UTF-8 LF {} Python 3.12.4 64-bit Go Live Prettier
```

### Huffpost(200)

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SPELL CHECKER COMMENTS Python + - [ ] [x] ... ^ x

~/Downloads/CDA_Main base Py at 11:21:16 AM
>
~/Downloads/CDA_Main INT x base Py at 11:21:59 AM
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py
Evaluating Classification:
Accuracy: 0.80
F1 Score: 0.84

Evaluating Diversity:
Diversity Score: 0.36

Evaluating Perplexity:
Perplexity: 156.34
~/Downloads/CDA_Main base Py at 11:21:59 AM
>

Ln 2, Col 16 Spaces: 4 UTF-8 LF {} Python 3.12.4 64-bit Go Live Prettier
```

### OTS(100)

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SPELL CHECKER COMMENTS Python + - [ ] [x] ... ^ x

~/Downloads/CDA_Main base Py at 11:21:59 AM
>
~/Downloads/CDA_Main INT x base Py at 11:22:34 AM
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py
Evaluating Classification:
Accuracy: 0.82
F1 Score: 0.70

Evaluating Diversity:
Diversity Score: 0.36

Evaluating Perplexity:
Perplexity: 156.34
~/Downloads/CDA_Main base Py at 11:22:35 AM
>

Ln 2, Col 16 Spaces: 4 UTF-8 LF {} Python 3.12.4 64-bit Go Live Prettier
```

### OTS(200)

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SPELL CHECKER COMMENTS Python + - [ ] [ ] ... ^ x
~/Downloads/CDA_Main base Py at 11:22:35 AM
>
~/Downloads/CDA_Main INT x base Py at 11:23:08 AM
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py
Evaluating Classification:
Accuracy: 0.86
F1 Score: 0.78

Evaluating Diversity:
Diversity Score: 0.36

Evaluating Perplexity:
Perplexity: 156.34
~/Downloads/CDA_Main base Py at 11:23:08 AM
>

Ln 1, Col 16 Spaces: 4 UTF-8 LF {} Python 3.12.4 64-bit Go Live Prettier
```

NER:

EBMNLP(100)

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SPELL CHECKER COMMENTS Python + - [ ] [ ] ... ^ x
~/Downloads/CDA_Main base Py at 11:23:08 AM
>
~/Downloads/CDA_Main INT x base Py at 11:23:41 AM
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py
Evaluating Classification:
Accuracy: 0.70
F1 Score: 0.23

Evaluating Diversity:
Diversity Score: 0.36

Evaluating Perplexity:
Perplexity: 156.34
~/Downloads/CDA_Main base Py at 11:23:41 AM
>

Ln 1, Col 16 Spaces: 4 UTF-8 LF {} Python 3.12.4 64-bit Go Live Prettier
```

EBMNLP(200)

```
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SPELL CHECKER COMMENTS Python + - [ ] [ ] ... ^ x
~/Downloads/CDA_Main base Py at 11:24:15 AM
>
~/Downloads/CDA_Main INT x base Py at 11:24:23 AM
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py
Evaluating Classification:
Accuracy: 0.81
F1 Score: 0.27

Evaluating Diversity:
Diversity Score: 0.36

Evaluating Perplexity:
Perplexity: 156.34
~/Downloads/CDA_Main base Py at 11:24:23 AM
>

Ln 1, Col 16 Spaces: 4 UTF-8 LF {} Python 3.12.4 64-bit Go Live Prettier
```

QA:

NewsQA(100)

```
~/Downloads/CDA_Main  
>  
~/Downloads/CDA_Main  
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py  
Evaluating Classification:  
Accuracy: 0.36  
F1 Score: 0.39  
  
Evaluating Diversity:  
Diversity Score: 0.36  
  
Evaluating Perplexity:  
Perplexity: 156.34  
~/Downloads/CDA_Main  
>
```

NewsQA(200)

```
~/Downloads/CDA_Main  
>  
~/Downloads/CDA_Main  
> /opt/homebrew/bin/python3 /Users/tejak/Downloads/CDA_Main/evaluation_metrics.py  
Evaluating Classification:  
Accuracy: 0.44  
F1 Score: 0.47  
  
Evaluating Diversity:  
Diversity Score: 0.36  
  
Evaluating Perplexity:  
Perplexity: 156.34  
~/Downloads/CDA_Main  
>
```

\*Note: The perplexity and diversity scores are calculated on overall generated augmented data. So they will be similar.