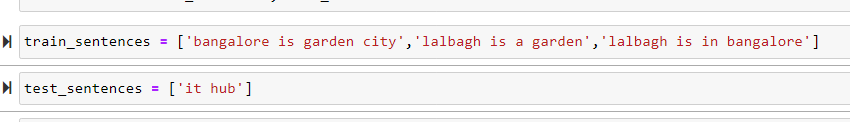
# **InducT GCN**

LINK OF THE PAPER: [[2206.00265] InducT-GCN: Inductive Graph Convolutional Networks for Text Classification (arxiv.org)](https://arxiv.org/abs/2206.00265)

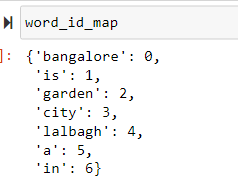
The paper talks about converting Text GCN which is transductive approach to Inductive approach.

**Dated: April/18/2023**

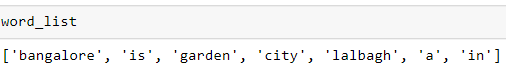
* Mine 1st experiment was based on taking a general small train and testing dataset and visualizing the graph.
* 

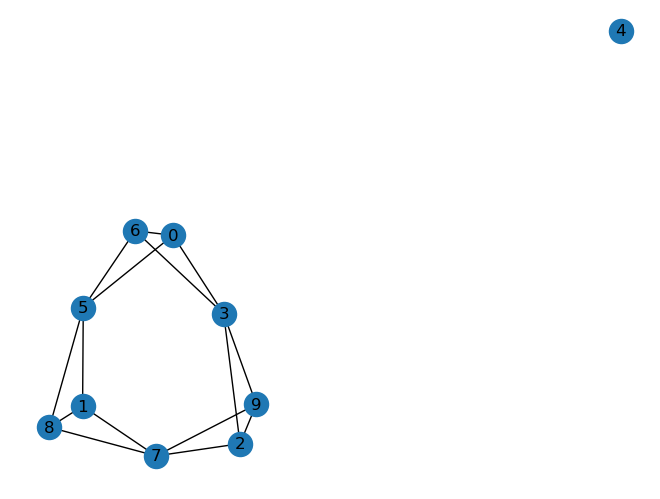
As shown above it was my small dataset.

* Now using the above dataset, vocabulary, word\_id\_map , word frequency is formed.



Vocab of the dataset.

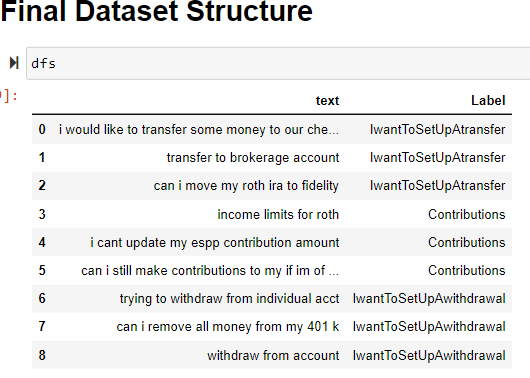




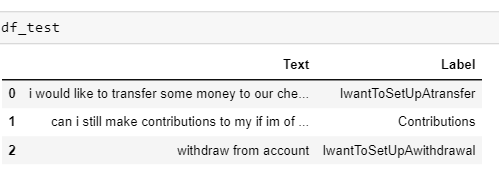
This is how graph looks like, where 0,1,2 is document id, and 3-9 is word id.

Dated: April/21/2023

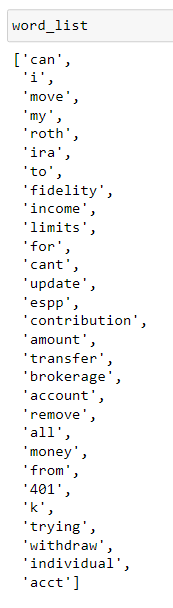
Now the second part of the internship was taking a small Fidelity Dataset and then visualizing the graph.

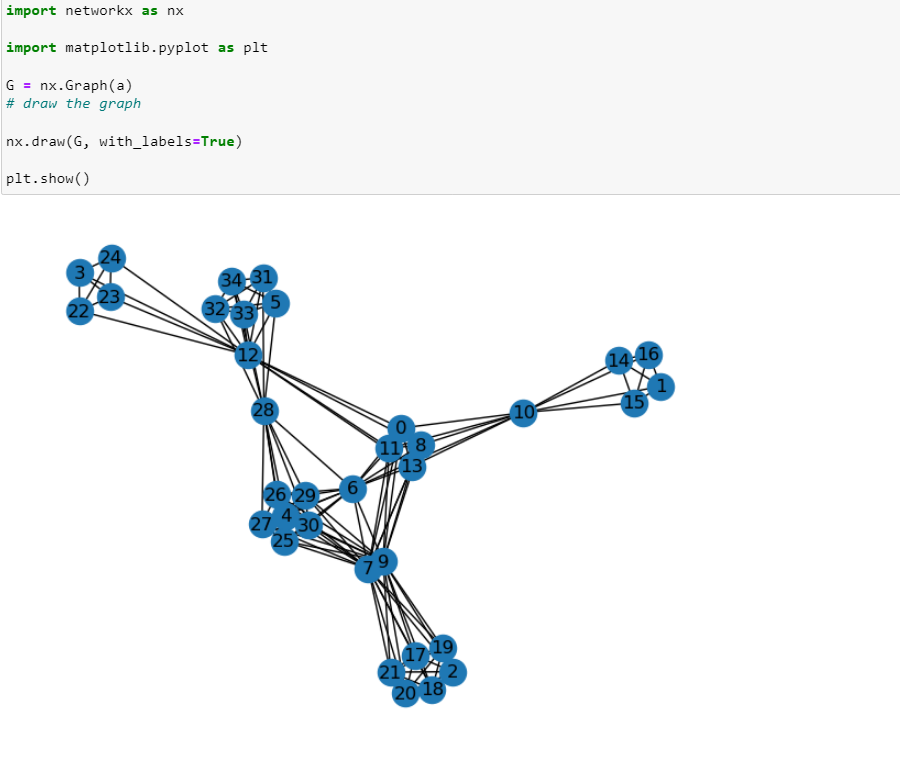


This is how mine small Fidelity Dataset looks like, it has 3 Labels to classify for, I did train test slit on dataset to train InducTGCN to train and predict the test labels.



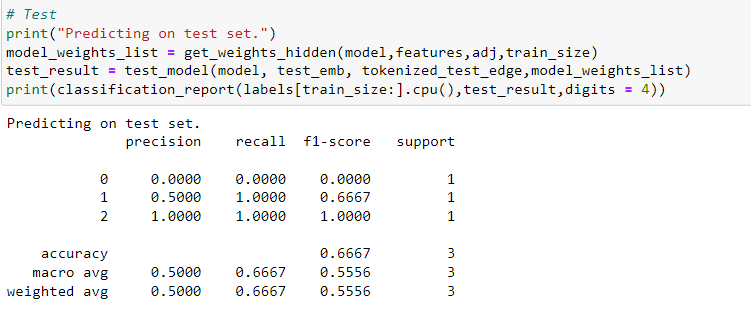
This is how test Dataset split looks like.

This is vocab of the small Fidelity Dataset.



This is how graph of the dataset looks like, with 0-5 document node and remaining word node.

Let's Look into prediction result for small test sentences.

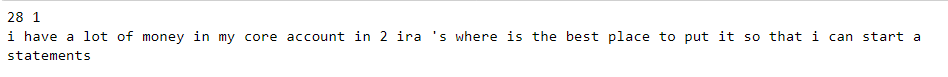


As the dataset was small so model didn’t perform that well.

# **Dated: April/22/2023**

Now looking into dataset provided by fidelity.

* There are 27727 rows of text and label associated with it provided.
* The total number of labels is 193.

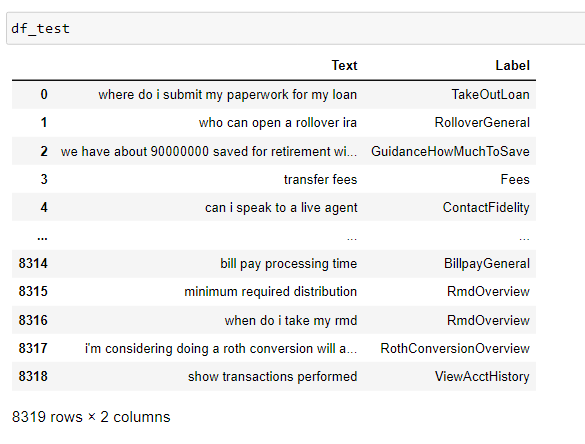


The maximum length of a sentence is 28, minimum length of sentence is 1.

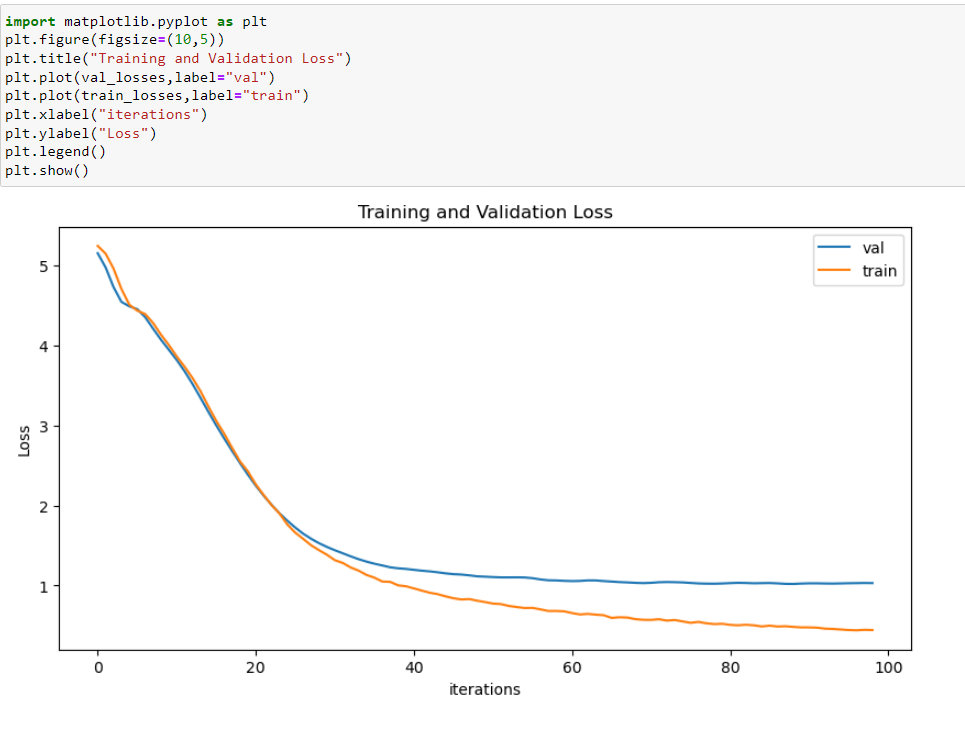
There are few more sentence with token length of 1 as shown below.



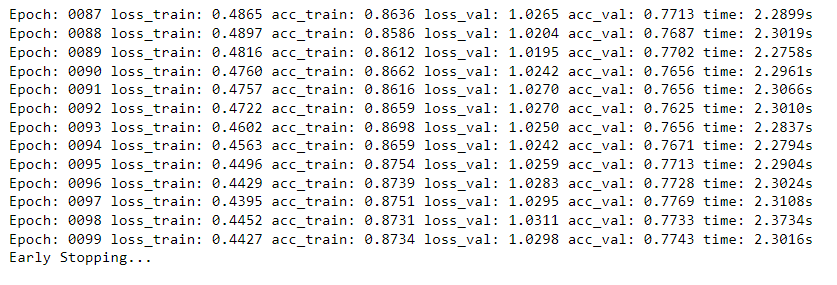
* Now dividing the dataset into train and test, mine test dataset looks like:



Training has vocab length of 4448. At the time of training the model divided the train dataset into train and validation dataset, Noted the graph for train and validation.



Looking into the last ten rows for validation accuracy it looks like below.



I haven’t constructed the graph, as it was showing me an out of memory error.

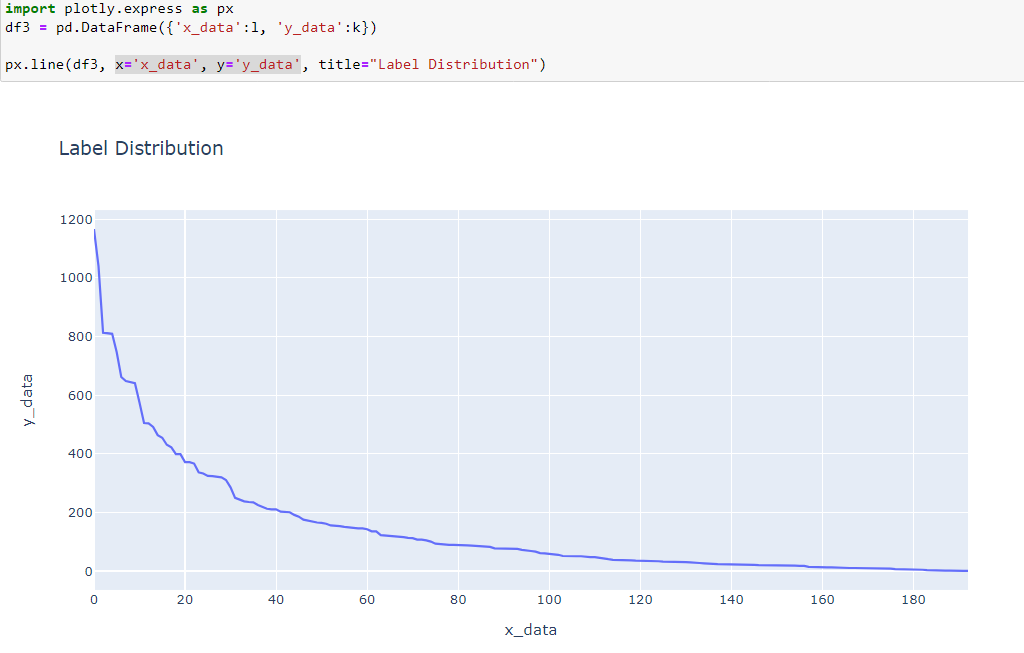
Now looking into the model prediction on test dataset per labels wise.

precision recall f1-score support  
  
 0 1.0000 0.4286 0.6000 7  
 1 0.5918 0.6304 0.6105 46  
 2 0.8478 0.9070 0.8764 129  
 3 0.4615 0.7500 0.5714 32  
 4 0.3276 0.4222 0.3689 45  
 5 0.8919 0.7674 0.8250 43  
 6 1.0000 0.4286 0.6000 7  
 7 0.8333 0.8333 0.8333 6  
 8 0.8333 0.5000 0.6250 10  
 9 0.8358 0.8750 0.8550 64  
 10 0.8750 0.4667 0.6087 45  
 11 1.0000 1.0000 1.0000 4  
 12 0.7213 0.8000 0.7586 110  
 13 0.7436 0.7073 0.7250 41  
 14 1.0000 0.7500 0.8571 4  
 15 1.0000 0.2500 0.4000 4  
 16 1.0000 0.1667 0.2857 6  
 17 0.8755 0.9259 0.9000 243  
 18 1.0000 0.3333 0.5000 6  
 19 0.8750 0.6364 0.7368 11  
 20 0.9262 0.9417 0.9339 120  
 21 0.8117 0.9191 0.8621 136  
 22 0.8571 0.6000 0.7059 60  
 23 1.0000 0.6111 0.7586 18  
 24 1.0000 0.5000 0.6667 2  
 25 0.6667 0.2857 0.4000 7  
 26 0.9231 1.0000 0.9600 12  
 27 0.6000 0.5455 0.5714 11  
 28 0.8750 0.7000 0.7778 10  
 29 0.8333 0.4167 0.5556 12  
 30 0.8163 0.8511 0.8333 47  
 31 0.7838 0.6591 0.7160 44  
 32 0.8968 0.9205 0.9085 151  
 33 0.3333 0.1111 0.1667 9  
 34 0.8348 0.9391 0.8839 312  
 35 0.8000 0.4706 0.5926 17  
 36 0.9459 0.6863 0.7955 51  
 37 0.9143 0.6038 0.7273 53  
 38 0.9730 0.7200 0.8276 50  
 39 1.0000 1.0000 1.0000 2  
 40 1.0000 0.8333 0.9091 6  
 41 0.0000 0.0000 0.0000 1  
 42 0.8182 0.8438 0.8308 32  
 43 0.9000 0.5714 0.6990 63  
 44 0.8438 0.7500 0.7941 36  
 45 0.7500 0.3000 0.4286 10  
 46 0.7647 0.7222 0.7429 36  
 47 0.9302 0.7143 0.8081 56  
 48 1.0000 0.3333 0.5000 3  
 49 0.0000 0.0000 0.0000 1  
 50 1.0000 0.5000 0.6667 10  
 51 0.9091 0.7143 0.8000 14  
 52 0.8737 0.8646 0.8691 192  
 53 1.0000 0.6190 0.7647 21  
 54 0.9231 0.8000 0.8571 15  
 55 0.8056 0.9134 0.8561 127  
 56 0.7419 0.8214 0.7797 28  
 57 0.6824 0.8169 0.7436 71  
 58 1.0000 1.0000 1.0000 6  
 59 1.0000 0.1429 0.2500 7  
 60 0.9167 0.7333 0.8148 15  
 61 0.6842 0.7429 0.7123 70  
 62 0.5000 0.5294 0.5143 34  
 63 1.0000 0.6154 0.7619 13  
 64 0.6667 0.5185 0.5833 27  
 65 0.7097 0.7938 0.7494 194  
 66 0.7154 0.7750 0.7440 120  
 67 1.0000 1.0000 1.0000 3  
 68 0.0000 0.0000 0.0000 1  
 69 0.7500 0.8969 0.8169 97  
 70 0.8704 0.8393 0.8545 112  
 71 0.0000 0.0000 0.0000 1  
 72 0.7447 0.7143 0.7292 49  
 73 0.9500 0.8837 0.9157 86  
 74 0.7000 0.6364 0.6667 22  
 75 0.8000 0.6957 0.7442 23  
 76 0.6571 0.7188 0.6866 224  
 77 0.8811 0.9065 0.8936 139  
 78 1.0000 0.5714 0.7273 7  
 79 0.4545 0.6250 0.5263 8  
 80 0.6250 0.8333 0.7143 18  
 81 0.5833 0.5833 0.5833 12  
 82 0.2667 0.6154 0.3721 13  
 83 0.7500 0.7500 0.7500 4  
 84 1.0000 0.7500 0.8571 4  
 85 0.8442 0.8784 0.8609 148  
 86 0.7778 0.2333 0.3590 60  
 87 0.6000 0.3750 0.4615 8  
 88 0.7500 0.6000 0.6667 10  
 89 0.7128 0.7204 0.7166 93  
 90 0.7018 0.8943 0.7864 350  
 91 0.7099 0.7623 0.7352 244  
 92 1.0000 0.8750 0.9333 16  
 93 1.0000 1.0000 1.0000 15  
 94 0.8548 0.8413 0.8480 63  
 95 0.7143 0.2174 0.3333 23  
 96 0.8750 0.6901 0.7717 71  
 97 1.0000 0.6667 0.8000 6  
 98 0.7037 0.9500 0.8085 20  
 99 1.0000 0.7222 0.8387 36  
 100 0.7049 0.5733 0.6324 75  
 101 1.0000 0.3333 0.5000 6  
 102 0.7778 0.6364 0.7000 11  
 103 1.0000 0.7273 0.8421 11  
 104 0.6061 0.8000 0.6897 25  
 105 1.0000 1.0000 1.0000 3  
 106 0.0000 0.0000 0.0000 1  
 107 1.0000 0.6667 0.8000 3  
 108 0.7763 0.8082 0.7919 73  
 109 0.9286 0.4815 0.6341 27  
 110 0.0000 0.0000 0.0000 1  
 111 1.0000 1.0000 1.0000 3  
 112 1.0000 0.5000 0.6667 4  
 113 0.0000 0.0000 0.0000 4  
 114 0.9167 0.8462 0.8800 26  
 115 0.8966 0.9630 0.9286 27  
 116 0.0000 0.0000 0.0000 3  
 117 1.0000 0.4286 0.6000 7  
 118 0.8889 0.6154 0.7273 26  
 119 0.5000 0.2963 0.3721 27  
 120 1.0000 0.4667 0.6364 15  
 121 1.0000 0.5000 0.6667 2  
 122 1.0000 0.5000 0.6667 10  
 123 0.8000 0.6400 0.7111 25  
 124 0.7027 0.5532 0.6190 47  
 125 1.0000 0.6250 0.7692 8  
 126 0.8000 0.4800 0.6000 25  
 127 0.5000 0.6667 0.5714 3  
 128 0.8214 0.6216 0.7077 37  
 129 0.6531 0.7273 0.6882 44  
 130 0.9643 0.8308 0.8926 65  
 131 0.6500 0.5200 0.5778 100  
 132 0.6250 0.3061 0.4110 49  
 133 0.8922 0.8613 0.8765 173  
 134 0.7938 0.9506 0.8652 243  
 135 0.6594 0.9381 0.7745 97  
 136 0.0000 0.0000 0.0000 6  
 137 1.0000 0.2143 0.3529 14  
 138 1.0000 0.7500 0.8571 4  
 139 1.0000 0.2857 0.4444 7  
 140 0.0000 0.0000 0.0000 1  
 141 0.9643 0.5192 0.6750 52  
 142 1.0000 0.7500 0.8571 8  
 143 0.8333 0.7143 0.7692 7  
 144 1.0000 0.9000 0.9474 10  
 145 0.9737 0.6379 0.7708 58  
 146 1.0000 0.8000 0.8889 5  
 147 1.0000 0.6667 0.8000 3  
 148 0.6875 0.5946 0.6377 37  
 149 0.9375 0.5769 0.7143 26  
 150 0.8222 0.9801 0.8943 151  
 151 0.9375 0.7317 0.8219 41  
 152 1.0000 0.3000 0.4615 10  
 153 0.7308 0.9048 0.8085 21  
 154 0.7609 0.7292 0.7447 96  
 155 0.7008 0.9296 0.7991 199  
 156 1.0000 0.5455 0.7059 11  
 157 1.0000 1.0000 1.0000 2  
 158 0.8571 0.2609 0.4000 23  
 159 0.5333 0.6957 0.6038 23  
 160 0.8750 0.6364 0.7368 44  
 161 0.8333 0.7143 0.7692 7  
 162 0.6667 0.4444 0.5333 9  
 163 0.5556 0.2632 0.3571 19  
 164 0.6950 0.9326 0.7965 193  
 165 0.9000 0.3333 0.4865 27  
 166 0.7500 0.4000 0.5217 15  
 167 0.7812 0.7143 0.7463 35  
 168 0.7308 0.6230 0.6726 61  
 169 1.0000 0.5000 0.6667 6  
 170 0.5217 0.4286 0.4706 28  
 171 0.5652 0.4194 0.4815 31  
 172 0.7581 0.8393 0.7966 112  
 173 1.0000 0.6667 0.8000 3  
 174 1.0000 1.0000 1.0000 2  
 175 1.0000 1.0000 1.0000 1  
 176 0.0000 0.0000 0.0000 3  
 177 0.4382 0.4021 0.4194 97  
 178 1.0000 0.2000 0.3333 5  
 179 1.0000 0.8000 0.8889 10  
 180 0.3000 0.3333 0.3158 18  
 181 0.7000 0.6087 0.6512 23  
 182 0.5882 1.0000 0.7407 30  
 183 0.9565 0.9565 0.9565 23  
 184 1.0000 0.1429 0.2500 7  
 185 0.0000 0.0000 0.0000 4  
 186 0.7727 0.6733 0.7196 101  
 187 0.0000 0.0000 0.0000 1  
 188 0.4351 0.8382 0.5729 68  
 189 0.4894 0.6765 0.5679 34  
  
 accuracy 0.7655 8319  
 macro avg 0.7675 0.6178 0.6585 8319  
weighted avg 0.7791 0.7655 0.7578 8319

189 labels because I combined 4 labels as 1(Other). Over model F1 score was 0.7578 weighted with accuracy of 0.7655.

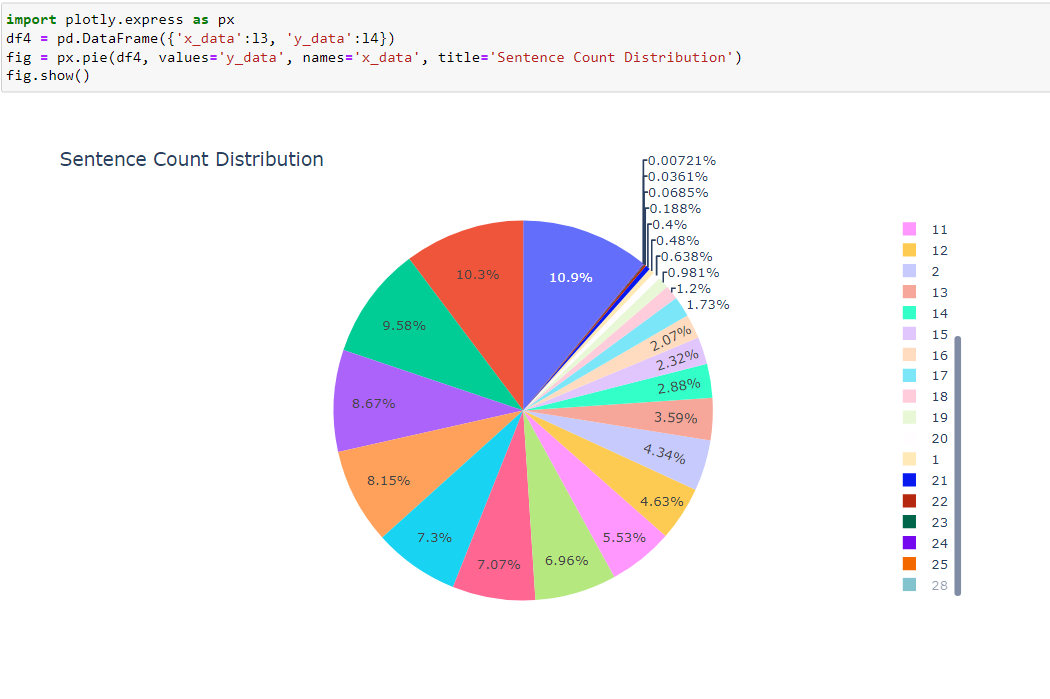
**Dated: April/24/2023**

Dataset Analysis of labels:



With x-axis as label number and y axis as number of instances of that label.

* Will try adding sentence length distribution, for data augmentation part.
* Tried data augmentation + referred GAT code.

GAT Implementation details: 1. [Pytorch Geometric Tutorial (antoniolonga.github.io)](https://antoniolonga.github.io/Pytorch_geometric_tutorials/posts/post3.html)

2. [Graph Attention Networks (GAT) (labml.ai)](https://nn.labml.ai/graphs/gat/index.html)

3. [GitHub - AntonioLonga/PytorchGeometricTutorial: Pytorch Geometric Tutorials](https://github.com/AntonioLonga/PytorchGeometricTutorial)

4. [Graph: GCN and GAT - My Computational Genomic Playground (zqfang.github.io)](https://zqfang.github.io/2020-12-12-ml-gcn-gat/)

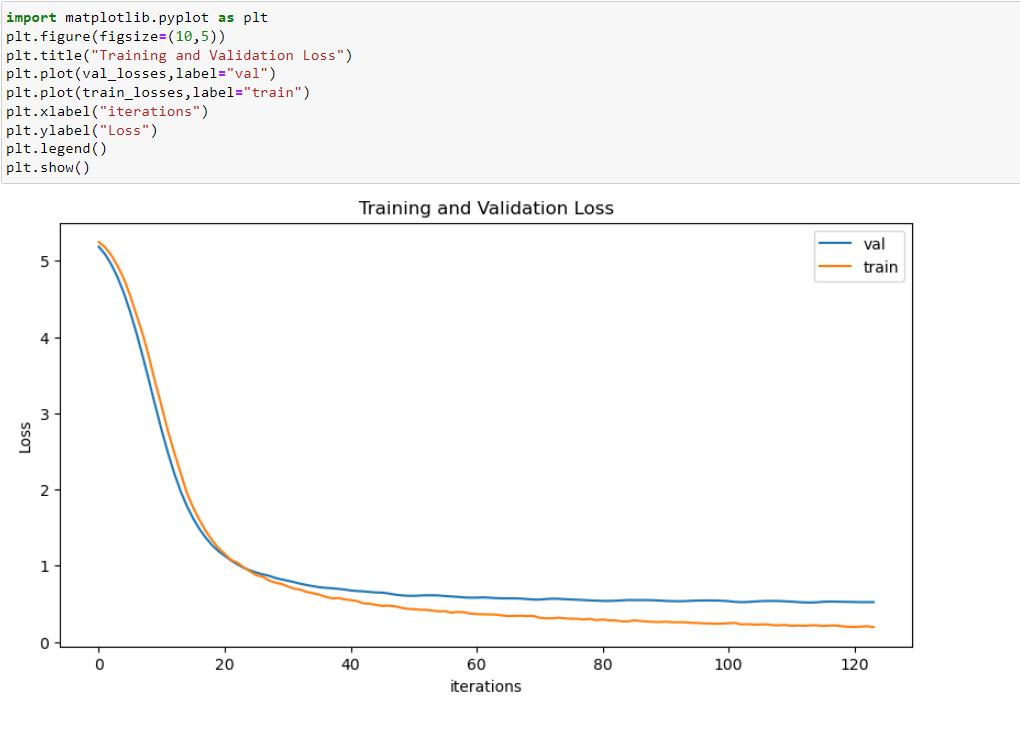
5. [Graph Attention Networks Under the Hood | by Giuseppe Futia | Towards Data Science](https://towardsdatascience.com/graph-attention-networks-under-the-hood-3bd70dc7a87)

6. [GitHub - gordicaleksa/pytorch-GAT: My implementation of the original GAT paper (Veličković et al.). I've additionally included the playground.py file for visualizing the Cora dataset, GAT embeddings, an attention mechanism, and entropy histograms. I've supported both Cora (transductive) and PPI (inductive) examples!](https://github.com/gordicaleksa/pytorch-GAT)

7. [GitHub - rohithteja/pytorch-geometric: GCN, GAT implementation using pytorch geometric on the cora dataset](https://github.com/rohithteja/pytorch-geometric)

**Dated: April/28/2023**

**Model performance when we do weighted sampling on fidelity dataset.**



Testing Score

precision recall f1-score support  
  
 0 1.0000 0.9773 0.9885 44  
 1 0.6500 0.6047 0.6265 43  
 2 0.8537 0.8537 0.8537 41  
 3 0.8039 0.9318 0.8632 44  
 4 0.4651 0.4762 0.4706 42  
 5 0.8947 0.8500 0.8718 40  
 6 0.9608 1.0000 0.9800 49  
 7 0.9615 1.0000 0.9804 50  
 8 0.9773 1.0000 0.9885 43  
 9 0.8085 0.9268 0.8636 41  
 10 0.8108 0.6818 0.7407 44  
 11 1.0000 1.0000 1.0000 41  
 12 0.7674 0.6875 0.7253 48  
 13 0.6981 0.8222 0.7551 45  
 14 1.0000 1.0000 1.0000 47  
 15 1.0000 1.0000 1.0000 44  
 16 0.8600 1.0000 0.9247 43  
 17 0.8636 0.8085 0.8352 47  
 18 0.9167 1.0000 0.9565 44  
 19 0.9020 1.0000 0.9485 46  
 20 0.8611 0.7561 0.8052 41  
 21 0.8269 0.9348 0.8776 46  
 22 0.9189 0.7727 0.8395 44  
 23 0.9020 0.9583 0.9293 48  
 24 1.0000 1.0000 1.0000 50  
 25 1.0000 1.0000 1.0000 53  
 26 1.0000 1.0000 1.0000 42  
 27 0.9762 0.8723 0.9213 47  
 28 0.9762 1.0000 0.9880 41  
 29 0.9756 1.0000 0.9877 40  
 30 0.9512 0.9070 0.9286 43  
 31 0.9444 0.8095 0.8718 42  
 32 0.9583 0.9020 0.9293 51  
 33 0.8421 1.0000 0.9143 48  
 34 0.9333 0.6222 0.7467 45  
 35 0.9211 0.9459 0.9333 37  
 36 0.9697 0.8000 0.8767 40  
 37 0.8611 0.7750 0.8158 40  
 38 0.8182 0.8780 0.8471 41  
 39 1.0000 1.0000 1.0000 42  
 40 1.0000 1.0000 1.0000 44  
 41 0.8864 1.0000 0.9398 39  
 42 0.9333 0.8235 0.8750 51  
 43 0.9655 0.6087 0.7467 46  
 44 0.8889 0.9091 0.8989 44  
 45 1.0000 0.9574 0.9783 47  
 46 0.8611 0.7561 0.8052 41  
 47 0.8140 0.7778 0.7955 45  
 48 0.9762 1.0000 0.9880 41  
 49 1.0000 1.0000 1.0000 41  
 50 0.9750 0.8864 0.9286 44  
 51 0.9429 0.7857 0.8571 42  
 52 0.9459 0.7609 0.8434 46  
 53 0.9722 0.9211 0.9459 38  
 54 1.0000 0.9778 0.9888 45  
 55 0.9333 0.6222 0.7467 45  
 56 0.9423 0.9800 0.9608 50  
 57 0.7255 0.8409 0.7789 44  
 58 0.9429 0.8919 0.9167 37  
 59 1.0000 0.9545 0.9767 44  
 60 0.9556 0.9556 0.9556 45  
 61 0.8710 0.6136 0.7200 44  
 62 0.8919 0.8049 0.8462 41  
 63 0.9474 0.7500 0.8372 48  
 64 0.8800 0.8800 0.8800 50  
 65 0.9500 0.4750 0.6333 40  
 66 0.8056 0.7250 0.7632 40  
 67 1.0000 1.0000 1.0000 43  
 68 1.0000 1.0000 1.0000 42  
 69 0.6491 0.8409 0.7327 44  
 70 0.9524 0.8333 0.8889 48  
 71 1.0000 1.0000 1.0000 51  
 72 0.8684 0.8919 0.8800 37  
 73 0.9706 0.8462 0.9041 39  
 74 0.9268 0.8636 0.8941 44  
 75 0.9556 0.9149 0.9348 47  
 76 0.8095 0.3864 0.5231 44  
 77 1.0000 0.7317 0.8451 41  
 78 0.8980 1.0000 0.9462 44  
 79 0.8235 0.9545 0.8842 44  
 80 0.8627 1.0000 0.9263 44  
 81 0.8444 0.8636 0.8539 44  
 82 0.8696 0.8889 0.8791 45  
 83 0.9375 1.0000 0.9677 45  
 84 0.9778 1.0000 0.9888 44  
 85 0.9118 0.7750 0.8378 40  
 86 0.9200 0.5000 0.6479 46  
 87 0.8163 0.8511 0.8333 47  
 88 0.8605 0.9024 0.8810 41  
 89 0.6939 0.7083 0.7010 48  
 90 0.6190 0.8125 0.7027 48  
 91 0.5660 0.6122 0.5882 49  
 92 1.0000 1.0000 1.0000 46  
 93 1.0000 1.0000 1.0000 37  
 94 0.9444 0.7727 0.8500 44  
 95 0.5882 0.5128 0.5479 39  
 96 0.9000 0.7826 0.8372 46  
 97 0.9512 1.0000 0.9750 39  
 98 0.7500 0.9750 0.8478 40  
 99 1.0000 0.9615 0.9804 52  
 100 0.7436 0.6170 0.6744 47  
 101 0.9762 1.0000 0.9880 41  
 102 0.9200 0.9787 0.9485 47  
 103 1.0000 1.0000 1.0000 42  
 104 0.7869 1.0000 0.8807 48  
 105 1.0000 1.0000 1.0000 41  
 106 1.0000 1.0000 1.0000 44  
 107 1.0000 1.0000 1.0000 42  
 108 0.9697 0.8000 0.8767 40  
 109 1.0000 0.8837 0.9383 43  
 110 1.0000 1.0000 1.0000 45  
 111 1.0000 1.0000 1.0000 45  
 112 1.0000 1.0000 1.0000 43  
 113 0.9535 1.0000 0.9762 41  
 114 0.8305 1.0000 0.9074 49  
 115 1.0000 0.9535 0.9762 43  
 116 0.9348 1.0000 0.9663 43  
 117 0.9348 1.0000 0.9663 43  
 118 0.8431 0.8776 0.8600 49  
 119 0.6667 0.7895 0.7229 38  
 120 0.9737 0.8605 0.9136 43  
 121 0.9778 1.0000 0.9888 44  
 122 0.9730 0.9730 0.9730 37  
 123 0.9038 1.0000 0.9495 47  
 124 0.6531 0.8205 0.7273 39  
 125 1.0000 1.0000 1.0000 43  
 126 0.7885 0.8542 0.8200 48  
 127 0.9492 1.0000 0.9739 56  
 128 0.9302 0.8333 0.8791 48  
 129 0.6739 0.7750 0.7209 40  
 130 0.8718 0.8293 0.8500 41  
 131 0.6111 0.4889 0.5432 45  
 132 0.7000 0.5490 0.6154 51  
 133 0.9355 0.7073 0.8056 41  
 134 0.7750 0.6458 0.7045 48  
 135 0.8056 0.6591 0.7250 44  
 136 0.8511 1.0000 0.9195 40  
 137 0.7544 0.9773 0.8515 44  
 138 1.0000 1.0000 1.0000 42  
 139 0.9500 1.0000 0.9744 38  
 140 0.9796 1.0000 0.9897 48  
 141 0.9688 0.8378 0.8986 37  
 142 1.0000 0.9545 0.9767 44  
 143 0.9259 1.0000 0.9615 50  
 144 0.9200 1.0000 0.9583 46  
 145 0.9444 0.8718 0.9067 39  
 146 1.0000 1.0000 1.0000 40  
 147 1.0000 1.0000 1.0000 46  
 148 0.7069 0.8723 0.7810 47  
 149 0.9286 0.9070 0.9176 43  
 150 0.7885 0.9762 0.8723 42  
 151 0.9512 0.8125 0.8764 48  
 152 0.9038 0.9792 0.9400 48  
 153 0.9286 0.9070 0.9176 43  
 154 0.7667 0.5610 0.6479 41  
 155 0.5493 0.9070 0.6842 43  
 156 1.0000 1.0000 1.0000 47  
 157 1.0000 1.0000 1.0000 37  
 158 0.6875 0.8980 0.7788 49  
 159 0.8400 0.8936 0.8660 47  
 160 0.8889 0.8163 0.8511 49  
 161 1.0000 0.9286 0.9630 42  
 162 0.7586 1.0000 0.8627 44  
 163 0.6167 0.8043 0.6981 46  
 164 0.6579 0.5556 0.6024 45  
 165 0.9677 0.6977 0.8108 43  
 166 0.9302 0.9302 0.9302 43  
 167 0.6923 0.9474 0.8000 38  
 168 0.8378 0.7561 0.7949 41  
 169 1.0000 0.9535 0.9762 43  
 170 0.5714 0.8182 0.6729 44  
 171 0.7586 0.5946 0.6667 37  
 172 0.7609 0.7609 0.7609 46  
 173 0.9500 1.0000 0.9744 38  
 174 1.0000 1.0000 1.0000 46  
 175 1.0000 1.0000 1.0000 42  
 176 0.9783 1.0000 0.9890 45  
 177 0.4062 0.3333 0.3662 39  
 178 0.8261 1.0000 0.9048 38  
 179 0.8542 0.9762 0.9111 42  
 180 0.5490 0.8000 0.6512 35  
 181 0.8913 0.8723 0.8817 47  
 182 0.7193 0.9535 0.8200 43  
 183 1.0000 0.9000 0.9474 40  
 184 0.9773 0.9773 0.9773 44  
 185 0.9512 0.9070 0.9286 43  
 186 0.7500 0.6136 0.6750 44  
 187 1.0000 1.0000 1.0000 41  
 188 0.5342 0.8478 0.6555 46  
 189 0.6957 0.6957 0.6957 46  
  
 accuracy 0.8745 8319  
 macro avg 0.8824 0.8741 0.8728 8319  
weighted avg 0.8826 0.8745 0.8731 8319

This is how test dataset was looking when we did weight random sampler:

RateOfReturn 56  
CoPayDefinition 53  
Medicare 52  
HdhpDefinition 51  
Definition529 51  
ContactFidelity 51  
RetirementPlanningAndGuidance 51  
SepIraDefinition 50  
AftertaxContribution 50  
FidelityHours 50  
CircuitBreaker 50  
FullView 50  
PortfolioReturns 49  
TaxWithholding 49  
AddressLegalVsMailing 49  
IwantToSetUpAwithdrawal 49  
PensionDt 49  
TaxFormAvailability 49  
BankAcctAdd 48  
StatementsDidntReceive 48  
IwantToSetUpAdeposit 48  
ContributionSummaryMain 48  
SpsVesting 48  
FsaTypes 48  
HardshipWithdrawals 48  
CheckReissue 48  
QualifiedExpenses 48  
RolloverGeneral 48  
NameChange 48  
IwantToSetUpAtransfer 48  
RebalancePortfolio 48  
SecureAct 48  
CodeCantReceive 47  
ProtectAgainstMarketVolatility 47  
BillpayGeneral 47  
WhenCanIretire 47  
MobilePhoneNumber 47  
BeneMinorDefinition 47  
MfDefinition 47  
TaxBracket 47  
SpsGeneralInfo 47  
HomeLoansFidelity 47  
IraDefinition 47  
TaxFormDidntReceive 47  
DivorceGeneral 47  
CaresWithdrawals 46  
DeliveryPreferences 46  
SocialSecurityWhenToClaim 46  
LetterOfAcceptance 46  
SpecialTaxNotice 46  
ExistingLoans 46  
CanIopenA529 46  
InformationLoans 46  
WithdrawalTaxesPenalties 46  
UpdatePersonalInfo 46  
TaxesRetAcct 46  
UsernameNoLongerSaved 46  
YtdTaxActivity 46  
LocateOffice 46  
FaxNumber 45  
UsernameStopSaving 45  
IhaveAquestionAboutStateCodes 45  
HsaPayFrom 45  
OutPocketMaxDefinition 45  
ResearchInvestments 45  
Contributions 45  
DripUpdate 45  
ForeignStatusUpdate 45  
BankInfoChange 45  
Fees 45  
Other 45  
Taxforms 45  
NetUnrealizedAppreciation 44  
HsaChangeJobs 44  
FindProvider 44  
DirectDepositSetup 44  
BuyingHouse 44  
TaxesHsa 44  
Wi457bEligibility 44  
WiPlanEnroll 44  
RothConversionOverview 44  
EmergencyFund 44  
ChangingEmployer 44  
FidelityMailingAddress 44  
HowDoImakeAstockOrMutualFundTrade 44  
HsaFundFromIra 44  
AccessibilityIssue 44  
AcctNumberLocate 44  
LoanPayment 44  
DeMinimisDefinition 44  
SelectPcp 44  
HideAcct 44  
AutoTransfer 44  
PretaxContribution 44  
IhaveAquestionAboutStudentLoans 44  
RothIraRules 44  
UnderstandingTaxForms 44  
Forms 44  
HsaDefinition 44  
BenePerStirpes 44  
HsaInvest 44  
GuidanceHowMuchToSave 44  
PasswordRequirements 43  
PhoneNumberForeign 43  
AnnualIncrease 43  
AccountInvestmentOptions 43  
SpsGrants 43  
UbtiDefinition 43  
OpenOrderUpdateOrCancel 43  
Qdro 43  
PensionLetter 43  
PpaOverview 43  
StatementsTiming 43  
TakeOutLoan 43  
PersonalInfoUsed 43  
TermsOfWithdrawal 43  
ConsolidateAccts 43  
TradeConfirmationView 43  
GoalboosterEligibility 43  
Wi401kDefinition 43  
BeneTypes 43  
WiInvestmentsGeneral 43  
GoalboosterNextSteps 42  
Taxes529 42  
ContactAnAdvisor 42  
UsernameOverlap 42  
EstablishmentDate 42  
MyvoiceOverview 42  
CobraQuestion 42  
SpsPlanDocuments 42  
WashSaleDefinition 42  
NonretirementTaxes 42  
SavedPasswordAcrossFidSites 42  
CumulativeReturn 42  
AcctPositionsView 42  
CostBasisGeneral 41  
ElectiveContribution 41  
FraudConcerns 41  
IraTraditionalVsRoth 41  
StatementsViewOnline 41  
NavDefinition 41  
MobileApp 41  
CollegePlanning 41  
WithdrawalDueToDisability 41  
DocumentTrouble 41  
PasswordStopSaving 41  
AcctClosure 41  
TroubleWithWebsite 41  
CardsGeneral 41  
AverageAnnualReturn 41  
RmdOverview 41  
EftDefinition 41  
ReimbursementGeneral 41  
HowDoIupdateMyBeneficiaries 41  
AuthorizeSomeoneOnMyAcct 41  
MaxedOutRetContributions 40  
AcctRoutingNum 40  
Wi403bDefinition 40  
CompanyMatchContribution 40  
OpenAcct 40  
Correction 40  
SpecialSavings 40  
Coronavirus 40  
GetAccountBalances 40  
RothIraDefinition 40  
FundAvailability 40  
IneedToResetMyPassword 40  
RegisterLogin 40  
PwsGeneral 39  
SomeonePassedAway 39  
LoginRememberMe 39  
LocateAcct 39  
DeductibleDefinition 39  
ViewAcctHistory 39  
HelpFilingTaxes 39  
PotentialTaxForms 38  
ExportAcctInfoToSoftware 38  
ViewPiAndWiAccts 38  
TradingActivityTaxes 38  
UsernameIvr 38  
SeasonalAddress 38  
LifeInsurance 37  
CoreInterestRates 37  
SecureMessageSend 37  
TaxDeadlineDay 37  
FidelityTaxId 37  
ProfitSharing 37  
HealthPlanQuestion 37  
UniversalTracker 37  
WhatKindOfAcctDoIhave 35  
Name: Label, dtype: int64

**GAT Inductive Learning Documentations**

**GAT Score for Fidelity Dataset: 83.04%**

**Some Short Text Classification Dataset:**

1. [Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification (aclanthology.org)](https://aclanthology.org/D19-1488.pdf)
2. [2206.00265.pdf (arxiv.org)](https://arxiv.org/pdf/2206.00265.pdf)
3. [Text Level Graph Neural Network for Text Classification (aclanthology.org)](https://aclanthology.org/D19-1345.pdf)
4. [Contrastive Learning with Heterogeneous Graph Attention Networks on Short Text Classification | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/9892257?casa_token=PbqGIz0pP6QAAAAA:WWqSzUlpwFXv-SWmFgb0UsoWH33fJIJHhAioeZ4uvaxRTIB1e1KiWc1sZhsFbaATZdNLoY3-Fg)
5. [arXiv:2211.16878v2 [cs.CL] 6 Dec 2022](https://arxiv.org/pdf/2211.16878.pdf) **(V.V. Imp)**
6. [GitHub - FKarl/short-text-classification: This repository contains code to reproduce the results in our paper "Transformers are Short Text Classifiers: A Study of Inductive Short Text Classifiers on Benchmarks and Real-world Datasets".](https://github.com/FKarl/short-text-classification)
7. [2022.emnlp-main.735.pdf (aclanthology.org)](https://aclanthology.org/2022.emnlp-main.735.pdf)
8. [2111.00180.pdf (arxiv.org)](https://arxiv.org/pdf/2111.00180.pdf)
9. [IEEE Xplore Full-Text PDF:](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9892257)
10. [Text Level Graph Neural Network for Text Classification (aclanthology.org)](https://aclanthology.org/D19-1345.pdf)
11. [2109.03777.pdf (arxiv.org)](https://arxiv.org/pdf/2109.03777.pdf)
12. [Deep Attention Diffusion Graph Neural Networks for Text Classification (aclanthology.org)](https://aclanthology.org/2021.emnlp-main.642.pdf)
13. [Donwloadable Corpora (unitn.it)](http://disi.unitn.it/moschitti/corpora.htm)
14. [Data (cornell.edu)](https://www.cs.cornell.edu/people/pabo/movie-review-data/)
15. [Home Page for 20 Newsgroups Data Set (qwone.com)](http://qwone.com/~jason/20Newsgroups/)
16. [ConTextING: Granting Document-Wise Contextual Embeddings to Graph Neural Networks for Inductive Text Classification (aclanthology.org)](https://aclanthology.org/2022.coling-1.100.pdf)

**SCORE STATISTICS:**

**TABLE:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **GAT Score**  **(Accuracy)** | **SOTA Score**  **(Accuracy)** | **Parameter** | **Best Graph Based Model** |
| **NICE-45**  **AvgLength:3.25** | **77.19** | **72.79**  **UsingVanillaBERT** | **Epochs: 80**  **Attention Heads:10**  **Hidden dimension: 256** | **47.19** |
| **NICE-2**  **AvgLen : 3.25** | **99.15** | **99.76**  **Roberta** | **Epochs: 30**  **Attention heads: 10**  **Hidden dimension:256** | **94.76** |
| **TagMyNews** | **68.04** | **62.7**  **(HGAT)** | **Attention heads:2**  **Hidden dimension:256**  **Epoch:100**  **It can be also further improved.** | **62.7**  **(HGAT)** |
| **Proprietary Dataset**  **Avg Length:**  **7.817**  **Classes: 193** | **83.26** | **81.06**  **(InducT GCN)** | **Attention Heads: 12**  **Hidden Dimension: 256**  **Epochs: 500** | **InduT GCN 81.06**  **BERT (80)** |
| **MR-DATASET**  **Avg Length: 20.34** | **85.12** | **90.21(DeBERTa)** | **Attention Heads: 2**  **Hidden Dimension: 256**  **Epochs: 100**  **Further It can also be improved.** | **89.7**  **Roberta GCN**  **(Transductive)**  **Inductive best: 89.43**  **(**ConTextING-RoBERTa**)** |
| **STOPS-41 DATASET**  **Avg Len: 5.25** | **88.70** | **89.73(DEBERTA)** | **Attention Heads: 10**  **Hidden Dimension: 256**  **Epochs:80**  **Further it can also be improved.** | **86.11** |
| **TREC**  **Avg Length: 10.7** | **95.84** | **99.4(BERT)** | **Attention Head:12**  **Hidden Dimension: 256**  **Epochs:100** | **97.99(DADGNN)**  **(Inductive)**  **Transductive (91.42)**  **Text GCN** |
| **SST-2** |  |  |  |  |
| **STOPS-2** |  |  |  |  |
| **Twitter**  **Avg Len: 12** | **97.14** | **99.97**  **(ALBERT)** | **Attention Head :7**  **HiddenDim:256**  **Epochs :100** | **98.16**  **(DADGAN)** |
| **R8**  **Avg Length: 65.72** | **97.57** | **98.53**  TextGCN-CNN-serial-SB | **Attention Head:2**  **Hidden Dimension:256**  **Epochs:300**  **It can be further improved if the attention head is increased even to 4.** | **98.53**  Text GCN-CNN-serial-SB  (Transductive)  (Inductive) 98.28  C-BERT (ESGNN + BERT) |