我目前的SOTA tabular model是一個名為ExcelFormer的模型，我想要將這個模型拆解成幾個部分，並且比較GNN插在哪個地方會讓整體模型的表現提升，在初始化這個模型前，原本的程式碼中有兩個資料前處理的部分，分別是對讀入的表格式資料中的類別型資料經由條件機率等算法轉為數值型資料，因此一個含有類別型資料與數值型資料的表格式資料集，在經過這個轉換後，會變成一個全為數值型資料的資料集。而第二個部分則是mutual\_info\_sort，這個前處理會使用sklearn裡面的mutual\_info來計算該資料集內每個feature column的重要性並且進行排序

這兩個前處理結束後，就會進入excelformer的部分，這個模型會被我列成如下格式:

1.前處理(包含CatToNum與mutual info sort)

2. excelformer encoder(進行基本的編碼以便適應接下來的注意力機制)

3. data augmentation (mix-up)

4. excelformer conv layer (注意力機制以及門控機制)，這個part會有N層

5. excelformer decoder

我會想要將GNN插在不同的模塊中間 並且比較GNN插在哪個地方會使得表現比較好，且我要與原本的模型比較，因此我會有7個模型:

模型1: 原本的ExcelFormer模型(不加GNN)

模型2: 原始的GNN模型

模型3:

1.前處理(包含CatToNum與mutual info sort)

在這裡加入GNN

2. excelformer encoder(進行基本的編碼以便適應接下來的注意力機制)

3. data augmentation (mix-up)

4. excelformer conv layer (注意力機制以及門控機制)，這個part會有N層

5. excelformer decoder

模型4:

1.前處理(包含CatToNum與mutual info sort)

2. excelformer encoder(進行基本的編碼以便適應接下來的注意力機制)

在這裡加入GNN

3. data augmentation (mix-up)

4. excelformer conv layer (注意力機制以及門控機制)，這個part會有N層

5. excelformer decoder

模型5:

1.前處理(包含CatToNum與mutual info sort)

2. excelformer encoder(進行基本的編碼以便適應接下來的注意力機制)

3. data augmentation (mix-up)

在這裡加入GNN

4. excelformer conv layer (注意力機制以及門控機制)，這個part會有N層

5. excelformer decoder

模型6:

1.前處理(包含CatToNum與mutual info sort)

2. excelformer encoder(進行基本的編碼以便適應接下來的注意力機制)

3. data augmentation (mix-up)

4. excelformer conv layer (注意力機制以及門控機制)，這個part會有N層

在這裡加入GNN

5. excelformer decoder

模型7:

1.前處理(包含CatToNum與mutual info sort)

2. excelformer encoder(進行基本的編碼以便適應接下來的注意力機制)

3. data augmentation (mix-up)

4. excelformer conv layer (注意力機制以及門控機制)，這個part會有N層

5. excelformer decoder

在這裡加入GNN

這七個模型會在正常的train test split ratio(0.64:0.16:0.2) 與few shot情境下(0.05:0.15:0.8)對於以下10個種類的資料集(每種類資料集皆有不只一個表格式資料集)進行平均表現排名:

1.small dataset+binclass+numerical類型欄位較多的資料集

2. small dataset+binclass+categorical類型欄位較多的資料集

3. small dataset+binclass+numerical與categorical類型欄位占比差不多的資料集

4.small dataset+regression+numerical類型欄位較多的資料集

5. small dataset+regression+categorical類型欄位較多的資料集

6. small dataset+regression+numerical與categorical類型欄位占比差不多的資料集

7.large dataset+binclass+numerical類型欄位較多的資料集

8. large dataset+multiclass+ numerical類型欄位較多的資料集

9. large dataset+regression+ numerical類型欄位較多的資料集

10. large dataset+regression+ categorical類型欄位較多的資料集