P1 大家好，我是弗如远甚小组的郁博文，今天我来汇报一下我们组迭代三的情况。

P3 这是我们的分工情况，我们选择的方向是方向二，使用机器学习和深度学习优化情感分析结果

P5 第一部分是sentiCR，我们首先按照PPT要求复现了SentiCR

P6 首先，SentiCR的源代码运行于python2环境，对于xlsx和列表的处理已经不适用于python3环境；在适配python3环境后，换用了openpyxl进行xlsx文件的处理,，右1是在oracle.xlsx（原文数据集）上的运行结果。原作者会在数据集上进行100次10折交叉验证，此处截取了8次10折交叉验证的结果，可以看到在GBT模型上的正确率确实可以达到83%，符合原文所说的GBT83.2%的正确率

sentiCR的作者将中性和积极的情绪归类为非消极，使得原本三分类的情感分析问题变成了二分类问题；因此得以使用1:1的欠采样平衡二类的数据数量；但是助教提供的是一个三分类的数据集，需要对sentiCR进行修改；首先在数据采样方面，我们不需要再进行欠采样，因为根据数据集提供者所言(senti4SD作者)，该数据集是建立在金标准之上的。同时我们也不需要进行10折交叉验证

可以看到就一次实验而言，sentiCR中GBT模型在新的数据集上的结果仍然较为不错。

我们同时还验证了不同的模型在该数据集上的运行情况

P7 通过观察数据集我们发现，数据文本中还存在部分@ + 人名的现象，人名中可能带有影响情感的词语，因此设计函数，移除该部分。

同时，经过观察，我们发现SGD、RF、BT、SVC这四个模型的性能较好，同时观察到他们在三个类的性能各存在差异；同时，额外尝试xgboost效果也不错（平均准确率大概在83%左右）；因此我们设计了集成模型，采取投票方式，集成了以上四个模型

在优化的过程中我们曾考虑使用数据增广，但是在几次实验下来，发现增广后的数据集并未使结果获得提升，甚至还存在略有下降的情况。这可能是我们数据增广方式不恰当或者破坏了论文中所说的"金标准"的原因

我们可以很明显地发现模型在消极一类上的性能较差，三项指标基本处于垫底状态；而在积极一类上的表现最好，召回率显示极大部分的积极一类都被正确的区分出来。而消极和中性的互相区分则很差，很明显，两者存在很大可能的互相归类错误的情况（中性划为消极，消极划为中性).联想到作者在处理有消极词的文本是会进行not翻转的操作，即将否定词附近的动词和副词加前缀NOT的操作，我们不禁怀疑此举的有效性，因此我们将process\_text中handle\_negation操作注释掉（在采用MIX模型后），运行实验。得到以下结果，我们发现总体正确率并未产生大幅波动，因此难以证明该操作对提高消极分类正确率的可用性。

P9 在查阅论文时，发现senti4SD使用了2000万条stackoverflow评论预训练了模型，在数据集上可以达到recall, precision和f1score上皆可以达到87%的水平，因此选取了senti4SD作为研究对象. 由于数据集使用csv文件而csv文件的分隔符作者采用的是‘；’

P12 在决定使用语言预训练模型时，我们最初考虑使用BERT。同时，为了优化实验结果，我们没有使用BERTForSequentialClassification，而是自行包装了一个类，具体来说，我们在BERTModel的基础上，额外增加了这些层（读PPT）

P13 训练参数如下

P14 这是结果对比，可以看到明显Bert偏高