**项目背景：**通过NLP模型，对学生摘要文本进行打分，分别预测其内容与词汇分数

**整体思路：**

1. 通过SKLEARN获取停用词表，通过去停用词、去特殊符号等方式，精炼文本关键信息
2. 基于TRANSFORMERS库，使用DEBERTA\_V3\_LARGE模型，将EMBEDDING层和encoder 18层冻结，提高模型稳定性，防止过拟合
3. 基于TRASNFORMERS库的TRAINER函数，对模型进行训练
4. 构建如文本长度、文本与问题重合度、文本拼写错误数等，结合DEBERTA模型预测结果作为特征，使用LIGHTGBM模型进行二阶段训练
5. 将DEBERTA预估结果与LIGHTGBM结果进行ENSEMBLE加权，得到最终结果

项目成绩：成功获得KAGGLE COMMONLIT STUDENT SUMMARIES比赛XX牌