题目：根据图书书评预测图书评分
成员：陈楚东 2006500018
张颂研 2006500022
Kaggle 链接：https://www.kaggle.com/competitions/goodreads-books-reviews-290312
数据集：数据集包括训练集，测试集和示例提交文件

- **goodreads_train.csv** - 训练集
- **goodreads_test.csv** - 测试集
- **goodreads_sample_submission.csv** - 示例提交文件

以下是数据集的列 -

user_id - 用户的 ID

book_id - 书的ID

review_id - 评论编号

评分 - 评分从 0 到 5

review_text - 审阅文本

date_added - 添加日期

date_updated - 更新日期

read_at - 阅读于

started_at - 开始于

n_votes票 - 票数

n_comments - 没有评论

预想的特征包括，review_text,n_votes 与 n_comments.
其他的特征就不在算法中加入.
预期能完成基于以上三个特征进行训练，能达到较为准确的图书评分，将测试集的结果写入示例提交文件，并上传到 kaggle 获得 score。

打算用 BERT 来完成评分预测，使用 Hugging face 提供的 transformers 库，该库提供了 Python 中使用 BERT 和其他预训练模型的简单接口，使模型的应用变得相对容易。基于输入文本获取模型产生的能够捕获上下文信息的向量表示，然后将此向量作为输入特征传递给回归器来预测评分。