介绍： 本次实验的任务目标是依据用户的评论文本预测用户对商品的评分（文本五分类问题）。在实验中尝试了基础的机器学习模型和lstm模型来完成这个任务

概要设计：

项目的目录在homework文件夹下，其中

dataloader.py是工程的数据处理模块，有get\_features\_data和 get\_lstm\_data函数，分别对应两种模型的数据预处理逻辑。

model.py 模块是模型初始化模块，用于构造各种分类器。

text\_rating\_predict\_bacsi.py 和text\_rating\_predict\_lstm.py 是各种的main函数， 他们具有的统一的流程（生产训练数据， 构造模型， 训练模型，测试模型，保存结果）

部分详细设计：

get\_features\_data：

1. 获取原数据
2. 清理数据
   1. 统一为小写
   2. 停用词
3. 构造特征
   1. TIIDF度量
   2. Word2vector
4. 切分数据

get\_lstm\_data：

1. 序列化
2. Padding

使用说明：

1. 工程构建：

以ML文件夹为root 打开工程

1. 数据准备：

在<https://www.kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones> 中下载数据，解压后，将Amazon\_Unlocked\_Mobile.csv放入dataset文件夹

1. 依赖包下载：

分别运行text\_rating\_predict\_bacsi.py 和text\_rating\_predict\_lstm.py根据编译提示安装缺失的依赖包：如：pip install bs4

1. 运行text\_rating\_predict\_bacsi.py查看基础模型的预测效果。
   1. 在result\_basc.csv
2. 运行text\_rating\_predict\_lstm.py查看LSTM模型的运行结果。
   1. 在result\_lstm.csv