各位同学，本课程的大作业是成绩评定的重要组成部分。按照我们已经讲过的原则，课程大作业\*\*可以自选题目\*\*，鼓励大家把课上学到的东西与自己的研究课题结合起来。除此之外，我们也提供\*\*5个\*\*公共选题，大家也可以从其中选择。

对大作业的要求是：

1、在7月5日之前，把完成情况做成PPT，找助教汇报（我也会尽力参加）；

2、5个公共选题（已经发布至课程网站），每个选题配备一名助教，请找相应的助教检查。

3、在汇报之前，将所有代码压缩打包，上次到课程网站相应的作业中（所有代码著作权归作者所有，本课程仅用于核实检查）；

4、成绩评定，以所使用技术和代码质量为评定指标，最终准确率仅做参考，即不以结果论成败。

5、至7月5日，未完成的项目，也可以提交检查，我们重点检查“做了什么”，而不是“是否做完”。

6、\*\*所有自选课题\*\*请联系【陈汉亭助教（微信号: woshihtaaa）】，由陈汉亭助教统一分配时间和其他助教进行检查。

5个公共选题如下：

课程大作业选题之一： 图像处理

Plant Seedlings Classification

Determine the species of a seedling from an image

【任务说明】

1、这是一个Kaggle上的比赛项目，相关信息参考如下：

<https://www.kaggle.com/c/plant-seedlings-classification>

2、主要依赖的技术是对图片进行分类；

【数据集】

数据集及相关说明请参考：

<https://www.kaggle.com/c/plant-seedlings-classification/data>

【提交时间】

本作业提交时间：7月5日

【提交内容】

1、全部源代码压缩后提交至课程网站；

2、制作大作业汇报PPT；

3、联系助教进行现场作业检查；

【联系助教】

陈汉亭 微信号: woshihtaaa

课程大作业选题之二： 程序 自然语言处理

CodRep: Machine Learning on Source Code Competition

A machine learning competition on source code data, with talent and taste.

【任务说明】

1、这是一个由瑞典皇家理工学院发起的比赛项目，相关信息参考如下：

<https://github.com/KTH/CodRep-competition>

2、主要依赖的技术是自然语言处理相关的方法；

【数据集】

数据集及相关说明请参考：

<https://github.com/KTH/CodRep-competition>

【提交时间】

本作业提交时间：7月5日

【提交内容】

1、全部源代码压缩后提交至课程网站；

2、制作大作业汇报PPT；

3、联系助教进行现场作业检查；

【联系助教】

刘芳 微信号: lf94816

课程大作业选题之三： 预测客户的还款能力

Home Credit Default Risk

Can you predict how capable each applicant is of repaying a loan?

【任务说明】

1、这是一个Kaggle上的比赛项目，相关信息参考如下：

<https://www.kaggle.com/c/home-credit-default-risk>

2、主要依赖的技术是基于Feature的分类方法；

【数据集】

数据集及相关说明请参考：

<https://www.kaggle.com/c/home-credit-default-risk/data>

【提交时间】

本作业提交时间：7月5日

【提交内容】

1、全部源代码压缩后提交至课程网站；

2、制作大作业汇报PPT；

3、联系助教进行现场作业检查；

【联系助教】

马连韬 微信号: massltime

课程大作业选题之四： 阅读理解

百度 PaddlePaddle AI 大赛——智能问答

【任务说明】

1、这是一个由百度发起的比赛项目，相关信息参考如下：

<https://www.kesci.com/apps/home/competition/5ad56e667238515d80b53704>

2、主要依赖的技术是基于深度学习的自然语言处理技术；

【数据集】

数据集及相关说明请参考：

<https://www.kesci.com/apps/home/competition/5ad56e667238515d80b53704>

【提交时间】

本作业提交时间：7月5日

【提交内容】

1、全部源代码压缩后提交至课程网站；

2、制作大作业汇报PPT；

3、联系助教进行现场作业检查；

【联系助教】

刘雨轩 微信号: sylimu

课程大作业选题之五： 恶意评论分类识别挑战

Toxic Comment Classification Challenge

Identify and classify toxic online comments

【任务说明】

1、这是一个Kaggle上的比赛项目，相关信息参考如下：

<https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge>

2、主要依赖的技术是对自然语言语句进行分类；

【数据集】

数据集及相关说明请参考：

<https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data>

【提交时间】

本作业提交时间：7月5日

【提交内容】

1、全部源代码压缩后提交至课程网站；

2、制作大作业汇报PPT；

3、联系助教进行现场作业检查；

【联系助教】

贾然 微信号: pkueecsjr