**统计学习**

期末的作业

1. 按时间顺序完整的写下本学期统计学习各次课程所讲授的主要内容（一两句话的内容概要）。（20分）
2. 大作业（80分）

* 在如下页面下载数据 <https://data.ed.gov/dataset/college-scorecard-all-data-files-through-6-2020/resources> ，具体方法：进入页面，左侧点击 GO TO RESOURCE
* 下载字段说明 <https://collegescorecard.ed.gov/assets/FullDataDocumentation.pdf>
* 着重关注MERGED20XX-20XX的所有csv文件

作业内容：

1. 根据与Repayment 有关的特征（可以是某一特征，也可以是几个特征的综合），构建每一间机构（学校）的（学生）贷款的违约指数，指数可以是连续型的，也可以是分级的（比如：1, 2, … 9, 10），解释您构建方法的理由
2. 根据csv文件中其他（除了与earning有关的字段）来对您所构建的指数进行预测，可以是分类方法也可以是回归方法
3. 使用多种统计学习方法
4. 可以逐年构建指数，并考虑时间序列，也可以将每年的数据当成单独的样本
5. 讨论影响该指数的主要因素
6. 报告8页以内，需要写清楚目的、方法、讨论
7. 上交code（允许仅调包，软件不限 ）
8. 中文即可，word和latex均可以接受
9. 结果可能不好，但是没有关系，只要有适当的统计学习方法运用与合理的说明即可