**《数据挖掘》期末项目说明**

**1.1 实践形式**

本次期末项目通过小组合作方式完成，以组为单位提交结果，每个小组人数在5人以内，最多不超过5人。同学们请自行组队并指定一名组长，组长需于2022年5月3日（周二）20:00之前把成员组成发邮件给助教李哲（zheli20@fudan.edu.cn）。若超时，助教会随机分配未组队的同学（或可选择独立成队）。

**1.2 项目要求**

1）项目选题：自行选取一个数据集，可以从公开渠道获取或者使用爬虫自行爬取。对数据进行预处理、描述性分析、建模分析，具体分析任务自行选定。

2）现场展示：每个小组对其实践项目成果向全班同学进行现场演示和介绍（时间暂定在**2022年6月23日**），并接受老师和同学们的提问。（每个小组展示和提问时间另行通知，预计10分钟左右）

3）项目报告：以书面报告（**PDF格式**）的形式详细阐述本小组完成项目课题的基本思路、数据处理方式、使用的模型、实现工具/语言、实验效果、性能评价和小组成员的分工说明（**必需**）等内容。报告可用中文或英文书写，要求不超过15页A4的篇幅（小四号字、宋体、单倍行距；若不按格式提交将扣分）。若算法或实现代码是借鉴或套用了部分他人成果的，请在论文中详细说明并附上参考文献。现场演示时展现的效果/结果与项目报告不符的，须做解释说明，否则会被扣分。（注意请尽量避免报告中出现一个占一页的大表格/大图，放入的最好是清晰一点的图，图和表格缩到一样的尺度）。

4）成果提交要求：

2022年6月23日晚11点前将以下文件发送至助教邮箱，邮件命名为“第xx小组作业”：

a） 项目报告，以“第xx小组\_项目报告”命名；

b） 项目PPT，以“第xx小组\_项目PPT”命名

c） 项目的全部源代码和其它相关资料请压缩成zip或tar.gz格式，并以“第xx小组\_源代码”命名。文件过大的可以通过网盘发送；

d） 所使用的数据的链接 (百度盘/钉盘/google driver/dropbox等)，注意是可以使代码运行的数据，**不是源数据的网址链接**

**以上成果的提交每超时1天，扣除该项目成绩最终得分的10%。**

**1.3项目评分标准**

该项目评分共占总成绩的40%，以满分100分计算，其中：

* 报告(60分)
  + 选题、背景介绍（15分）：选题具有创新性、可行性;背景介绍逻辑清晰、调研深入；数据的独特性、是否自己收集
  + 数据说明（5分）：样本量充足、变量维度丰富;表格简洁、变量描述清晰
  + 描述性分析(15分)：图形简洁、规范、美观;文字说明解读得当，逻辑清晰
  + 建模分析(20分): 建模深入，需包含多模型的比较与解读
  + 结论与建议(5分): 结论深入、建议可行；具备应用前景
* 展示（25分)：
  + PPT美观、结构简洁、逻辑清晰、图表符合规范、建模结论安排合理
  + 讲解生动、逻辑顺畅、回答问题简洁清晰（每组派1人讲解，可多人回答问题）
  + 严禁超时（超时扣10分）
* 其他 (15分)：代码简洁、注释充分、可复现；独立成组会适当加分；报告未按格式要求要减10分

**特别注意：建模部分的模型必须包含本课程涉及的模型，且主要建模分析部分以本课程涉及建模部分为主，项目内容可以包含部分深度学习模型，但项目整体分析不可以以深度学习模型为主！**