



赛题 1：企业投资价值评估

赛题说明：

1. **赛题目标：**参赛团队需要预测出**企业评分.xlsx** 中第 2 列：**企业总评分**的实际数值。
2. **企业评分.xlsx** 为本赛题的核心数据表格。该表格分为 2 列，第 1 列为：3000 家上市公司的**企业编号**。第 2 列为：相应企业的**投资价值综合评分**（评分为整数，范围：1~100）。
3. 该赛题主办方共提供 38 张 **xlsx** 格式的数据表格，其中包含上市企业的：基本工商信息、购地信息、年报信息、财务信息、纳税信息、知识产权信息等各维度全量企业数据。
4. 每张表格中只包含 1 个 **Sheet** (统一命名为：**Sheet**)，每个 **Sheet** 中包含经过处理的 2~n 列、2~m 行参考数据。参赛团队需要根据实际需求，对数据进行清洗、整合、筛选、拆分、抽取特征、调整参数、训练模型等数据分析工作。
5. 出于对企业机密、知识产权、债务信息等数据信息的保护，主办方对部分信息进行了匿名化、离散化、归一化处理。具体事例可参照说明 6。
6. 主办方提供了 3000 家企业作为训练数据集，企业名称进行了匿名化处理，**企业编号：1001~4000**。同时对企业部分信息进行了归一化处理，如：注册资本等。
7. 主办方提供了约 500 家企业作为测试数据集，**企业编号：4001~4500+**。相关材料会于 2019 年 4 月 16 日下发。
8. 部分 **excel** 表格中的数据不全，参赛团队需要灵活处理各维度数据。
9. 主办方统一于 2019 年 3 月 1 日将训练数据集等信息发送到参赛团队指定邮箱。2019 年 3 月 1 日 - 2019 年 3 月 15 日报名的团队，主办方将于团队报名后的 3 个工作日内将相关数据集和参考信息下发到参赛团队指定邮箱。（**指定邮箱为：报名表中报名邮箱，今后通讯我们只回复报名邮箱发来的邮件**）
10. 测试集数据主办方将统一于 **2019 年 4 月 16 日** 发送到参赛团队的指定邮箱中。参赛团队需要在 **2019 年 4 月 23 日 23:59:59** 前将评审材料发送到大会指定邮箱：ibasic_cas@163.com
11. 评审材料分为以下三部分：（部分信息与大赛官网有细微出入，具体以本赛题



说明为主)

● **xlsx 格式结果数据集**

该文件由 2 列组成，第 1 列为评测企业的企业编号，由赛题方提供；第 2 列为预测的企业投资价值，以整数表示（范围：1~100）。

命名格式：赛题 1 结果_团队名.xlsx。

（注：2 列数据顶格，没有标题。赛会主办方会于 2019 年 4 月 16 日下发 xlsx 格式结果数据集，选手可将运行结果按照结果数据集格式提交。）

● **模型源代码压缩文件**

参赛团队需要提交本次比赛对应赛题的完整模型源代码（建议使用 Python、Java 语言）。需保留依赖的类库、Jar 包等；需补全代码注释，标示出文件读入、数据处理、核心算法、结果输出等部分；需保证代码可运行，以便验证 csv 结果。最后打包成压缩文件。

命名格式：赛题 1 代码_团队名.zip。

● **PPT 报告**

参赛团队需要提交一份 15~20 页的 PPT 报告。

命名格式：赛题 1 报告_团队名.ppt。

报告中需涵盖以下内容：

- 模型源代码相关文档说明。
- 模型算法思路描述，包含数据处理、特征抽取、参数调整等部分。
- 此次比赛的探索历程和心得体会。
- 对本次比赛的建议。

12. 对本赛题有任何疑问，请各参赛团队及时和赛会主办方联系。邮箱：
ibasic_cas@163.com

13. 主办方拥有本次大赛的最终解释权。