

赛题 A 僵尸企业智能识别

一、赛事背景

“僵尸企业”指丧失自我发展能力，必须依赖非市场因素即政府补贴或银行续贷来维持生存的企业。尽管这些企业不产生效益，却依然占有土地、资本、劳动力等要素资源，严重妨碍了新技术、新产业等新动能的成长。僵尸企业会严重浪费社会资源，扰乱市场秩序，脱离资产监管，损害职工权益，形成金融风险。随着人工智能技术的不断发展，可以通过构建识别预警模型来有效识别出僵尸企业，科学反映公司经营状态，让政府和银行提供更具针对性的支撑服务，让僵尸企业从困境走向复活。

二、赛事安排

比赛分初赛、决赛两个阶段。初赛成绩前 15 组进入决赛。决赛邀请业界专家对决赛的队伍根据初赛成绩和决赛答辩成绩综合评分。

奖项设置：

- 一等奖 1 组
- 二等奖 2 组
- 三等奖 3 组
- 纪念奖 9 组（共 15 组）

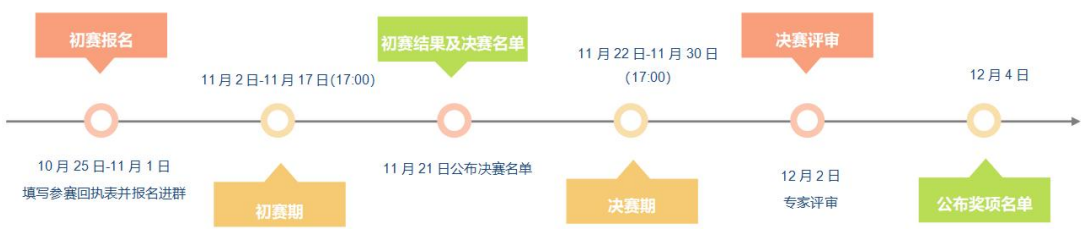
1. 报名 10 月 25 日-11 月 1 日 17：00

2. 初赛

11 月 2 日 10：00-11 月 17 日 17：00，参赛选手按照比赛微信群内通知下载初赛原始数据集和对应题目要求，按照提交需求在规定时间内发送给评选老师。

3. 决赛

材料提交 11 月 22 日 10：00-11 月 30 日 17：00，
答辩评审 12 月 2 日
进入决赛的团队参加答辩，阐述赛题思路和亮点。评审专家根据初赛成绩和决赛答辩成绩进行综合评分（初赛成绩占比 70%，答辩分数占比 30%）。



三、参赛对象

本次竞赛面向北京信息科技大学全体学生，鼓励不同层次、不同专业学生以小组（3-5人）为单位报名参加比赛，每位学生只允许参加一支团队。参赛选手需具备基本 Python 基础。

四、赛事任务

本赛题数据来自国内某地区的真实企业，每一条数据表示一家企业。数据中包含企业的基本工商信息、企业经营信息和知识产权等维度数据以及是否为异常企业的标签等 21 个字段，具体字段说明见表 1。

表 1 数据字段说明

字段名称	说明
registered_fund	注册资金
ent_type	企业类型
employee_cnt	员工数量
anche_year	年报年度
assgro	资产总额
liagro	负债总额
vendinc	营业总收入
maibusinc	主营业务收入
progro	利润总额
netinc	净利润
ratgro	纳税总额
totequ	所有者权益合计
pat_cnt	专利数量
mark_cnt	商标数量
soft_cnt	软著数量
works_cnt	作著数量
ass_lia_rat	资产负债率
ROA	资产收益率
IP	知识产权
loss_cnt	连续亏损的年数
Label	是否是异常企业（1 代表是，0 代表不是）

请以是否是异常企业为建模目标，构建僵尸企业识别模型。

五、评审规则

1. 评估指标

评审阶段，根据参赛者上传的模型文件以及推理脚本，由主办方提供的验证数据集得到预测结果。

评估指标为 **F1-score**，取值范围为[0, 1]，公式如下：

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

其中 *Precision* 表示精确率，*Recall* 表示召回率

2. 评测及排行

本赛题均提供下载数据，选手在本地进行算法调试，排名按得分从高到低排序。

六、作品提交要求

1、文件格式：zip 格式，压缩包命名：组长名-赛题名（初赛）

2、文件大小：无要求

3、提交形式：百度云上传，初赛截止时间（11 月 17 日 17：00）前可修改

4、文件详细说明：

（1）验证模型效果的测试文件，命名为 **test.ipynb**（Jupyter Notebook 文件格式）

（2）**requirement.txt**，其中包含可以保证 **test.ipynb** 正常运行的所有包的安装命令（pip 方式，包含版本号）

（3）保证 **test.ipynb** 正常运行的其它所有文件

（4）赛期组内互动照片 3 张+一段视频（3 分钟内即可）

备注：（1）**requirement.txt** 示例如下：

```
# torch and torchvision 包名称
pip install torch==1.10.1+cu102 torchvision==0.11.2+cu102 torchaudio==0.10.1 -f https://download.pytorch.org/whl/torch_stable.html

# PIL
pip install pillow==9.0.1

# NumPy
pip install numpy==0.21.0

# Pandas
pip install pandas==1.0.5
```

安装命令

（2）关于测试文件 **test.ipynb**

用于对比提交的验证集结果，验证模型的真实效果；要求从上到下依次运行，即可成功得到验证集预测标签，若中途报错，则无法验证结果，将按照没有成绩处理！！！！

需要包含下列内容：

验证集数据读取与预处理

加载模型结构和权重

前向推断出预测结果