**课题描述：**

本次实习课题为“预测事件对上市公司经营业绩的影响”，使用提供的业绩预告文本，从语料中抽取出事件用于训练自然语言模型，最终输出的模型需要具有判断事件对上市公司业绩变化方向的能力。

**课题说明：**

1. 字段说明：

|  |  |
| --- | --- |
| 字段代码 | 字段释义 |
| S\_INFO\_WINDCODE | 股票代码 |
| S\_PROFITNOTICE\_DATE | 发布日期 |
| S\_PROFITNOTICE\_PERIOD | 业绩预报描述期 |
| S\_PROFITNOTICE\_REASON | 业绩预报原因 |
| Label | 业绩变动分类 |

选择本课题的候选人，只需要用到最后两个字段。基于“业绩预报原因”的文本内容，从中抽取出事件，形成合理的描述，结合所给的label进行建模。

1. 标签说明：

label为根据上年同期净利润与今年同期净利润计算得出的

|  |  |  |
| --- | --- | --- |
|  | | |
| 首亏 | 上年同期盈利，预告期亏损 |
| 续亏 | 上年同期亏损，预告期亏损 |
| 扭亏 | 上年同期亏损，预告期盈利 |
| 不确定 | 预告期盈亏不定 |
|  |  |  |
| 上年、预告期皆为盈利时，按预告期盈利幅度细分: | | |
| 预增 | 1、增幅上限>50,下限>=0；2、“大幅度增长”；3、增幅50%左右 |
| 略增 | 1、0<=增幅<=50；2、“较大幅度增长”或“一定幅度增长” |
| 续盈 | 无重大变化，预报中描述增幅上限>0,下限<0 |
| 略减 | 1、-50<=降幅<=0；2、“较大幅度下降”或“一定幅度下降” |
| 预减 | 1、降幅上限<=-50，上限<=0；2、“大幅度下降” |

**本课题只需要判断出事件对于公司业绩变化方向的影响：向上、向下、无影响（或者影响方向不确定）**。因此需要对提供的label进行合并，具体合并方式由候选人自行决定。

1. 评价标准：

候选人建好模型后，对实习课题中的test.xlsx中的2500个事件进行预测，**预测结果为”向下”、”向上”、”无影响或不确定”三个类别**。我们基于分类准确率进行评估。

建议候选人在训练模型时事先预留一部分测试集，用于材料撰写过程中的模型准确率展示。

1. 提交结果：

1) 对给定事件的方向预测结果，格式为xlsx，第一列为test.xlsx中的reason,第二列为模型预测结果。

2) 模型代码，要求能够运行并得出相同结果。

3) PPT，用于课题答辩，答辩时长约为10到15分钟。PPT长度无要求，但需要符合公司PPT模板，建议为10页上下。