##代码审核规则 1、选手需要提交整个建模过程代码,其中包括但不限于:赛 题数据分析与预处理、特征工程、模型构建、结果预测等代码,以及训练好的模型。 所有文件打包在一个 .zip 压缩包中,具体的压缩包文件结构详见代码打包规范内 容。

2、赛题主办方将对选手提交的内容进行审核、复现、查重,如果发现选手提交的内容存在无法复现结果(允许复现成绩和排行榜成绩出现偏差,但是不得超过主办方规定的偏差范围)、抄袭他人代码、作弊等情况,将会取消选手的排行榜成绩,并将该账号拉入黑名单,不得参加平台任何其他比赛。所有规则解释权由主办方所有。

代码打包规范

1、原始数据文件夹 init_data/

选手无需打包原始的数据文件,初始时我们会清空此文件夹,然后把与官网上提供的同样的数据文件放到此文件夹下。

####2、中间数据文件夹 temp_ data/ 选手可以在此文件夹下存放一些在数据处理、训练模型中需要产生的中间数据。但是请注意,如果选手在此文件夹下存放了中间数据,需要在代码说明文件夹 doc/Readme.md 文件中说明此中间数据的生成过程,并且在代码文件夹 code/ 中存放可以生成这些中间文件的代码。

3、代码文件夹 code/

选手需要在此文件夹下存放所有的算法源代码和最优模型文件,包括(但不限于):

- 数据预处理
- 特征构建和选择
- 模型建立和训练
- 预测结果
- 训练好的模型文件(需要保证在预测结果代码中调用的就是该模型)

源代码中必须包含一些必要的注释,可方便主办方更好的进行复现。

请注意:为了防止模型训练出现随机性,代码中所有影响模型的参数必须固定(比如训练轮次、学习率、树的深度、随机种子等)。

```
    数据结构示例如下(.py结构和.ipynb结构):
    ├─ code
    │ ├─ train.py(.ipynb) # 用于训练的文件
    │ ├─ test.py(.ipynb) # 用于预测的文件
    │ ├─ xxx.py(.ipynb) # 其他文件,此文件可以有多个,命名选手自己指定即可
    │ └─ model # 此文件夹为模型文件夹,可以存放模型或者模型相关的权重文件
    │ └─ model.h5(.pt/.pb/.pth等) # 最终生成最优预测结果的模型
```

根据选手编写代码的习惯不同,选手可以选择提交.py文件结构或者.ipynb文件结构,不同的提交结构在复现的时候会有所区别,所以下面详细列出了对于不同结构提交的要求。

• 选手提交.py 文件的结构 主办方在复现时会在 code 路径下分别调用 train.py 和 test.py, 用于模型训练和结果预测。.py 文件调用函数命令如下:(请选手提交代码之前确保调用命令能够正确执行)

```
python train.py trainSetDir modelPath
python test.py testSetDir resultPath
```

• 选手提交.ipynb 文件的结构 主办方在复现时会在 code 路径下直接打开 jupyter notebook 分别运行 train.ipynb 和 test.ipynb,用于模型 训练和结果预测。同样的选手需要固定所使用的路径。

如果以当前所在路径为 code 文件夹为例,以上两种提交结构的四个相对路径分别为:

```
1 trainSetDir = ../init_data/toUser/train # 训练数据加载根路径
2 modelPath = ./model/model.h5 # 模型存放路径
3 testSetDir = ../init_data/toUser/test # 测试数据加载根路径
4 resultPath = ../result/result.csv # 结果生成路径
```

4、代码说明文件夹 doc/

选手需要在此文件夹下存放以下两个文件,每个文件所需要包含的内容如下:

```
1、**Readme.md 文件**
```

- 整个赛题的解决思路以及用到的数据处理方法;
- 使用的算法模型简要介绍;
- 整个模型的实现过程的详细步骤说明(主办方会根据这个说明进行每一步的 复现);
- 算法运行(环境)说明;

2、** requirements.txt **文件 此文件中需要列出所有选手在建模过程中使用的依赖以及其明确的版本,例如TensorFlow、PyTorch、LightGBM、XGBoost等。

5、预测结果文件夹 result/

此文件夹下存放最终预测的结果 result.csv, 注意此结果必须由code文件夹中提交的模型和预测代码产出, 且此结果要与排行榜上的最高成绩文件保持一致。初始时我们会清空此文件夹。

####6、整体提交文件夹结构

```
- my_project # 整个需要提交的文件夹
1
      ├─ init_data # 原始数据文件夹
2
         L— toUser # 命名与主办方提供的数据集一致
3
            ├─ train # 训练集数据
            └─ test # 测试集数据
5
      ─ code # 代码文件夹
6
         ├─ train.py(.ipynb) # 用于训练的文件
7
         ├─ test.py(.ipynb) # 用于预测的文件
8
         ├─ xxx.py(.ipynb) # 其他文件,此文件可以有多个,命名选手自己指
9
   定即可
         — model # 此文件夹为模型文件夹,可以存放模型或者模型相关的权
10
   重文件
            — model.h5(.pt/.pb/.pth等) # 最终生成最优预测结果的模型
11
      ├─ doc # 文档文件夹
12
         - Readme.md
13
         └─ requirements.txt
14
      └─ result # 结果文件夹
15
         └─ result.csv
16
```