# 机器学习实践全流程演示

## 赛题介绍

本次比赛为数据挖掘类型的比赛，聚焦于工业场景。本赛题实质上为回归任务，其中会涉及到时序预测相关的知识。

通过电炉空间温度推测产品内部温度，设计烧结过程的温度场和浓度场的最优控制律：

* **任务输入**：电炉对应17个温区的实际生产数据，分别是电炉上部17组加热棒设定温度T1-1~T1-17，电炉下部17组加热棒设定温度T2-1~T2-17，底部17组进气口的设定进气流量V1-V17；
* **任务输出**：电炉对应17个温区上部空间和下部空间17个测温点的测量温度值。

**值得注意的是预测目标为34个，所以需要我们进行34次模型训练和预测。**

**任务解析：**

本次问题是回归预测，常规思路为使用机器学习，如LightGBM、XGBoost，或者使用深度学习进行实践，在模型的搭建上就比较复杂，需要自己构建模型结构，对于数值数据需要进行标准化处理；

最简单的方式可使用机器学习模型解决本次问题，该方法模型使用简单，数据不需要过多预处理；

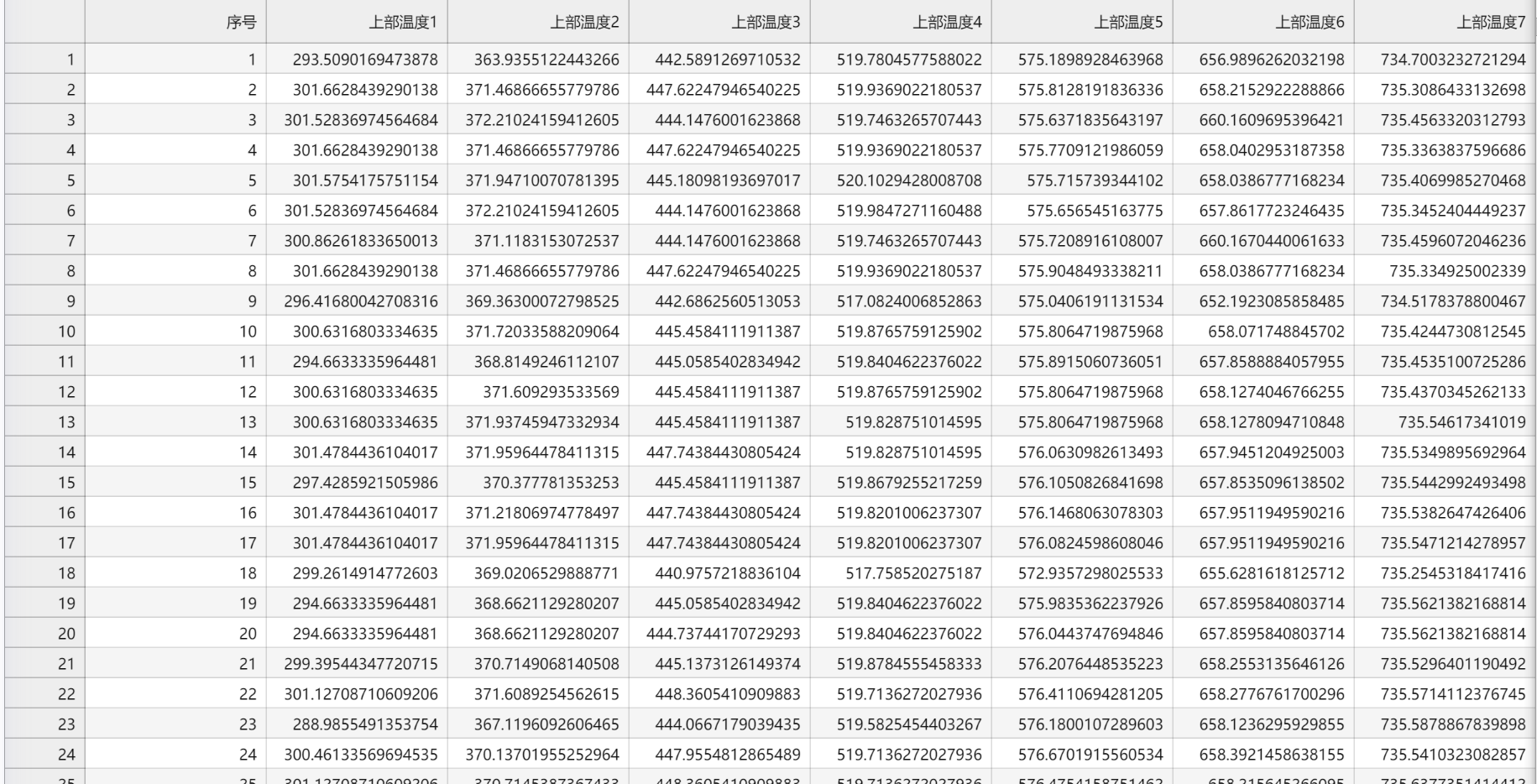
主要步骤为数据预处理、切分训练集与验证集、训练模型、预测结果。

**数据解析：**

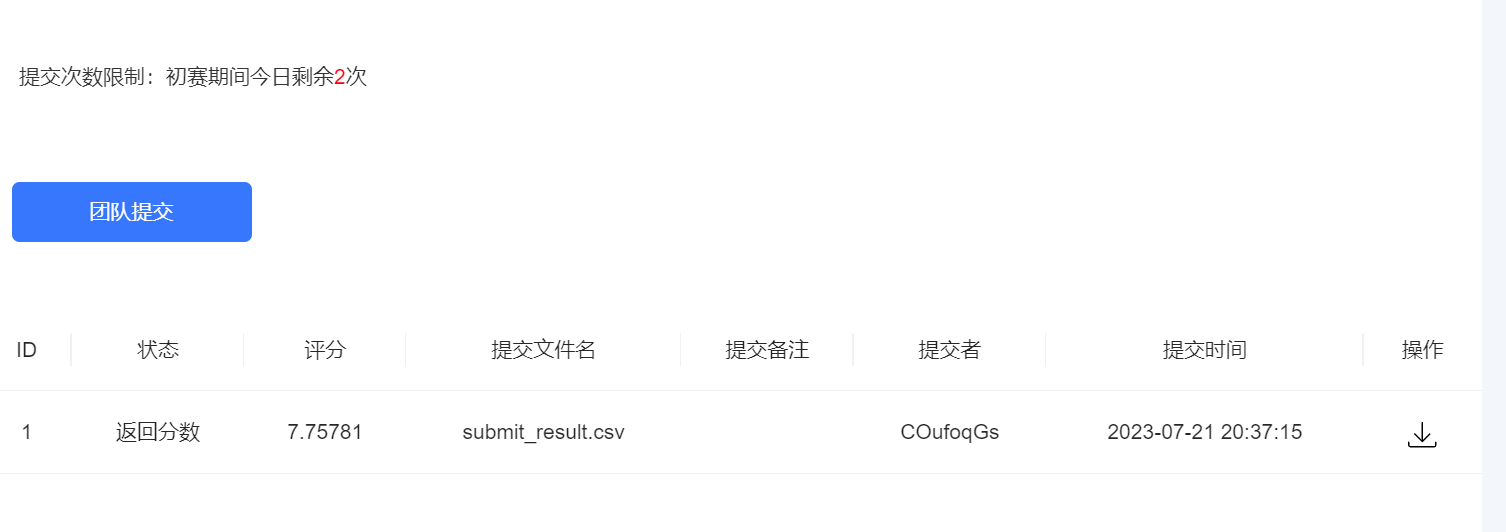
数据规模比较小，可以快速处理数据和搭建模型，对于机器要求8g内存即可。本次为结构化赛题，包含电炉烧结每个时间段的流量、上下部设定温度，以及预测目标上下部测量温度值。

## ML实践全流程演示

1. Shell指令需要前面加!才能运行
2. Run all cell 跑出来结果，右键下载保存到本地后



1. 获得比赛评分:



## 附录：Python语言编程及pytorch入门(以cv为例) （有事暂未更新完全）

### Python语法速览

#### 3.1.1 print和注释

print(“x”)打印内容x，每一个print会默认换行

print(“x”,end=“”)这样取消换行

print(“a”,“b”)多个内容以空格分隔 a b

print(“a”，“b”,sep=“\*”) 输出a\*b

单行注释 #后面不会被当做代码

#### 3.1.2 列表与字典

容器的唯一作用就是打包、解包、内容传递

列表是一种可变序列（容器）

列表索引从左到右0开始，从右到左-1开始

切片默认左闭右开区间切片

集合是不包含重复元素的容器（去重）

字典是键值对的集合

Python3支持unicode 可以输出emoji表情（狗头）

#### 3.1.3 匿名函数

Relu = lambda x:max(x,0)

print(Relu(-9)) 输出0

#### 3.1.4 回调函数

回调函数就是一个通过函数指针调用的函数。如果你把函数的指针（地址）作为参数传递给另一个函数，当这个指针被用来调用其所指向的函数时，我们就说这是回调函数。回调函数不是由该函数的实现方直接调用，而是在特定的事件或条件发生时由另外的一方调用的，用于对该事件或条件进行响应。

#### 3.1.5 浅拷贝与深拷贝

id（a）表示a对象真实物理地址

c = a.copy() 浅拷贝，对于列表内部的列表其实是共用一个物理地址的

import copy

d = copy.deepcopy(a)深度拷贝，真正的复制了另一个对象在其他物理地址里

### 3.2 Pytorch入门

#### 3.2.1 dir和help

dir用来探索工具包的组成

help用来理解某个工具包的函数/类的使用

Dateset ??也可以用来看说明

#### 3.2.2 pytorch加载数据 对应read\_data.py

Dataset类:提供一种方式去获取数据及其label，告诉使用者如何获取每一个数据及其label，告诉使用者总共有多少数据

数据集几种组织形式：文件夹名称为label，内部为数据集



一个文件夹作为数据集，另一个文件夹作为label



直接将label写在图片的名称上一般是目标检测