

# 腾讯云联邦学习产品使用手册

编制单位	深圳市腾讯计算机系统有限公司
版 本 号	V0.1
级密分类	对客户公开
更新时间	2021 年 7 月

## 目 录

第 1 章	联合建模	3
1.1	账号登录	3
1.2	数据准备	3
1.3	样本对齐	4
1.4	数据上传	4
1.5	数据授权	6
1.6	模型训练	6
1.7	任务查询	10
1.7	模型下载	12

## 第1章 联合建模

本章旨在介绍通过联邦学习进行联合建模的操作流程。

### 1.1 账号登录

模型训练系统登录默认用户名：admin，密码：tadmin。登录后，可以在系统设置修改密码。

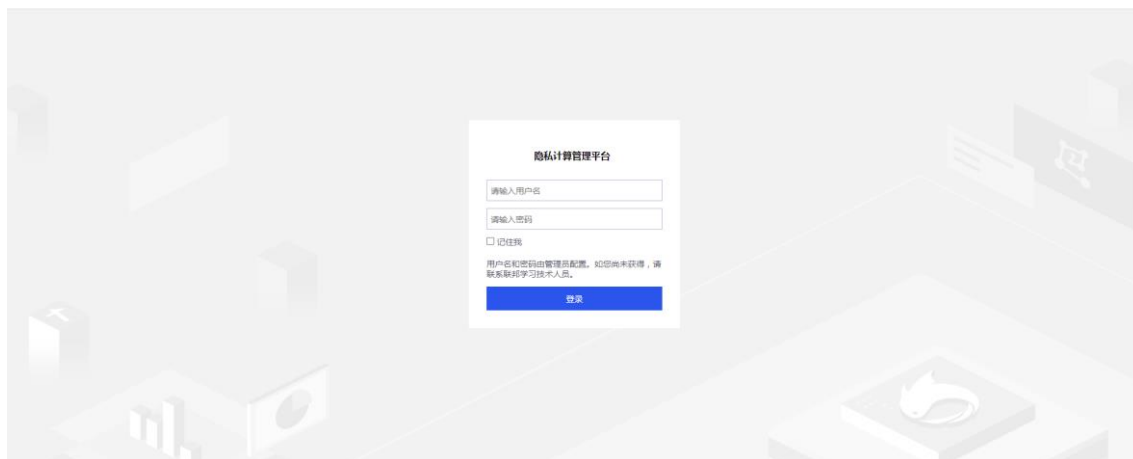


图 1.1 登录页面

### 1.2 数据准备

训练样本格式如图 1.2 所示。csv 文件；id 是样本序号，用于对齐双方样本；label 是标签列（int 型，数值为 0 和 1）；x1 至 x10 是客户方特征列，特征值保留 2 位小数。训练集和验证集格式相同，训练集和测试集需形成两个独立的 csv 文件。如无特征数据，则生成 1 列伪特征（即 x0 列全部填 1）。

id	label	x0	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	
1	1	1	0.25	-1.05	0.21	0.07	-0.44	-0.38	-0.49	0.35	-0.29	-0.73	1
2	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
3	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
4	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
5	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
6	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
7	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
8	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
9	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
10	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
11	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
12	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1
13	1	1	-1.14	-0.78	-1.17	-0.92	0.63	-1.02	-1.11	-0.96	-0.1	-0.12	1

图 1.2 训练样本格式示意图

### 1.3 样本对齐

样本 key 格式为：csv 文件，字段名称：id（从 1 开始编号）、手机号码、身份证号、申请日期（2019 年 7 月份之后）。可明文，如加密建议 md5。客户划分训练集和测试集，形成两个单独的 csv 文件发给腾讯，如图 1.3 所示。

id	apply_date	card_no_md5	phonenum_md5
1	20190701	0506c	27092456167
2	20190701	06865	381056fa537
3	20190701	07aa3	1aa46cf1a17
4	20190701	07d5d	110de07a49e
5	20190701	0a6ea	32b4bdb84fe
6	20190701	0becc	a9db64d297a
7	20190701	0c634	5d64189ff9d
8	20190701	106ca	012094a5e
9	20190701	10ea7	1dc9c1584d
10	20190701	1188d	0cf04f35325e
11	20190701	11e1c	acdc87da503
12	20190701	13d95	0e53b4b89a50
13	20190701	151f2	0a36589027d8
14	20190701	1686c	391d689ecd95
15	20190701	1b322	6f10d8b6313
16	20190701	1c7d1	15e7153a7f
17	20190701	1d8ae	0d4c75f40d2
18	20190701	1de93	78b527134af
19	20190701	1e3ce	3be279143cc7
20	20190701	1f001	2693a68d6e
21	20190701	1f32d	073b6ff56179
22	20190701	21d4c	adbd366a8
23	20190701	29711	37278892284

图 1.3 样本 key 格式示意图

### 1.4 数据上传

登录模型训练系统，点击数据管理——上传数据，进入上传数据界面，如图 1.4 所示。



图 1.4 客户上传数据界面

训练集和测试集分开上传，步骤如下：

1. 点击数据库名输入框——新建一个数据库——输入客户自定义数据库名(例如 xxx\_space)
2. 点击数据表名输入框——新建一个数据表——输入客户自定义数据表名(例如训练集可名为为 xxx\_train\_data，测试集 xxx\_test\_data)
3. 点击上传数据——上传本地 csv 文件——提示上传成功。数据上传成功之后，数据列表显示数据上传情况，如图 1.5 所示。

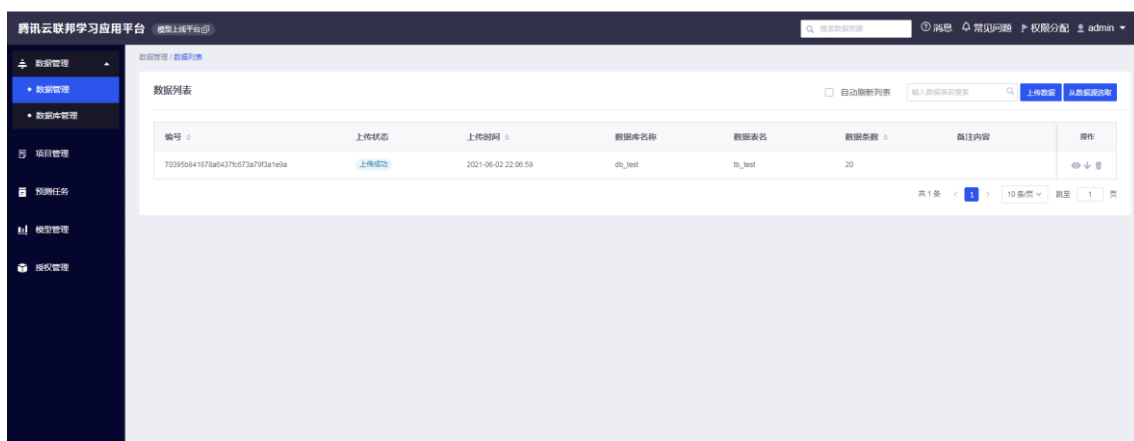
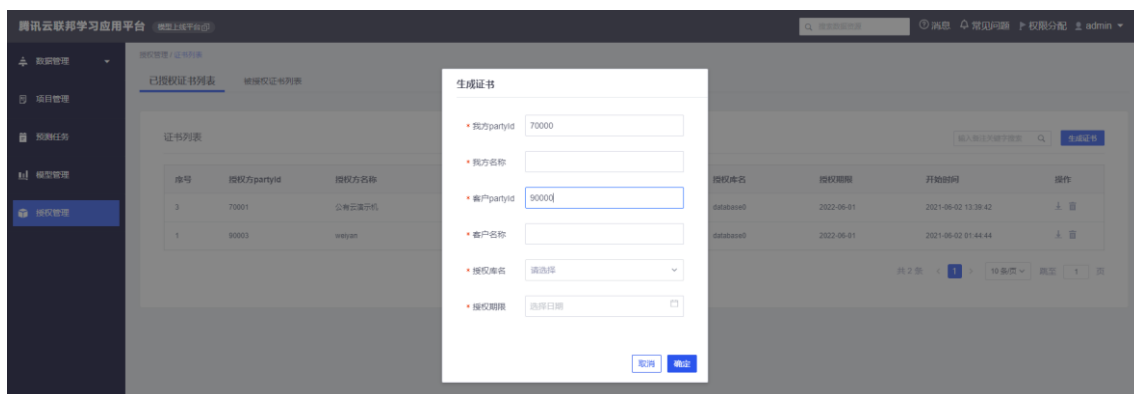


图 1.5 数据列表

## 1.5 数据授权

数据授权分为生成我方授权证书，上传对方授权证书两步。

首先，点击授权管理->生成证书提示如下界面。例如我方 partyid 为 70000，客户 partyid 为 90000，则注意表中“**我方 partyid**”应填写 70000，“**客户 partyid**”应填写 90000，同时设置名称和授权库名、期限等。然后点击下载，将证书发给客户上传，同时将客户发来的授权证书上传到“被授权证书列表中”，完成授权。注：数据库表和期限功能在规划中。



生成证书界面



证书生成成功界面



客户证书上传成功界面

## 1.6 模型训练

首先新建项目，然后进入才能模型训练。包括选择算法和模板、设置参数、提交训练三步。点击模型训练—新增训练，进入模型训练页面。如图 3.7 所示。

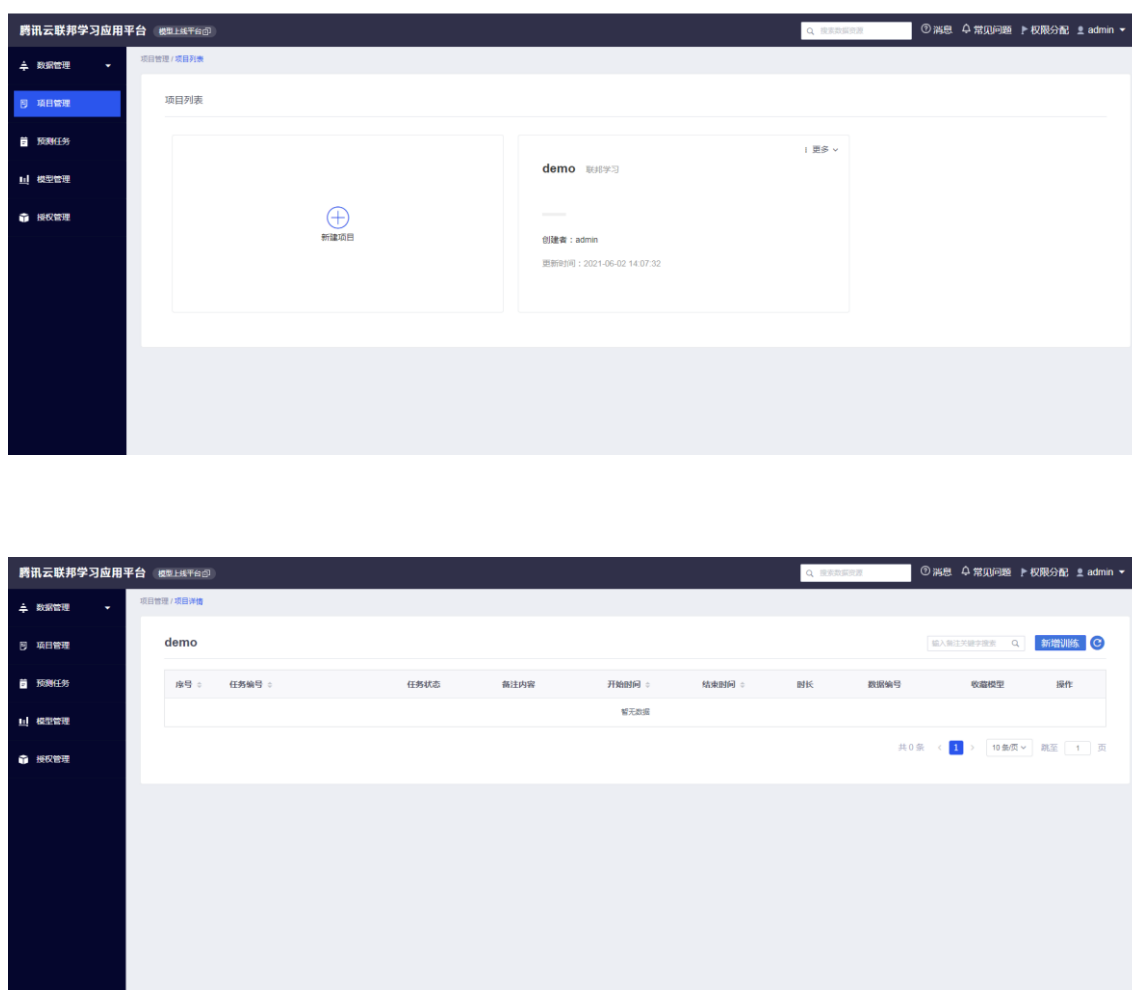


图 1.6 模型训练界面

### 1.6.1 选择算法和模板

目前可供选择的算法有纵向逻辑回归、纵向提升树。选择模板时，可沿用已保存的模板（LR 模板 2 暂不可用），也可从右侧可选组建中向左拖入模型训练流程，形成新的训练模板，如图 1.7 所示。

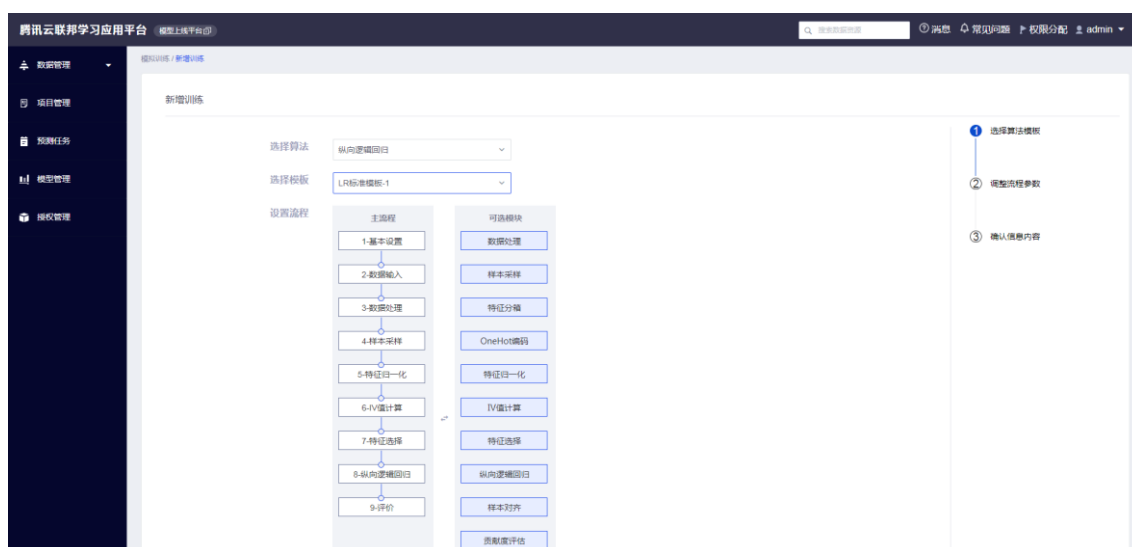


图 1.7 选择算法和模板界面

### 1.6.2 设置参数

模型训练流程中支持自定义参数，系统对参数设置方法做了提示说明。为了用户能快速上手，本系统除了基本设置、数据输入需要用户填写之外，其余步骤均设置了默认参数，以下对参数设置做举例说明。

参数设置举例：partyid。该参数是参与方的唯一识别号，如 20001；直接向腾讯技术人员获取。



图 1.8 partyid 设置界面



参数设置举例：数据输入。我方数据直接从本地已上传的数据集中选择，对方数据表名称由腾讯技术人员提供，注意红框处对方数据需要填写对方提供的绝对路径（一般为/data/examples/data/\*\*\*.csv）。如图 1.9 所示。

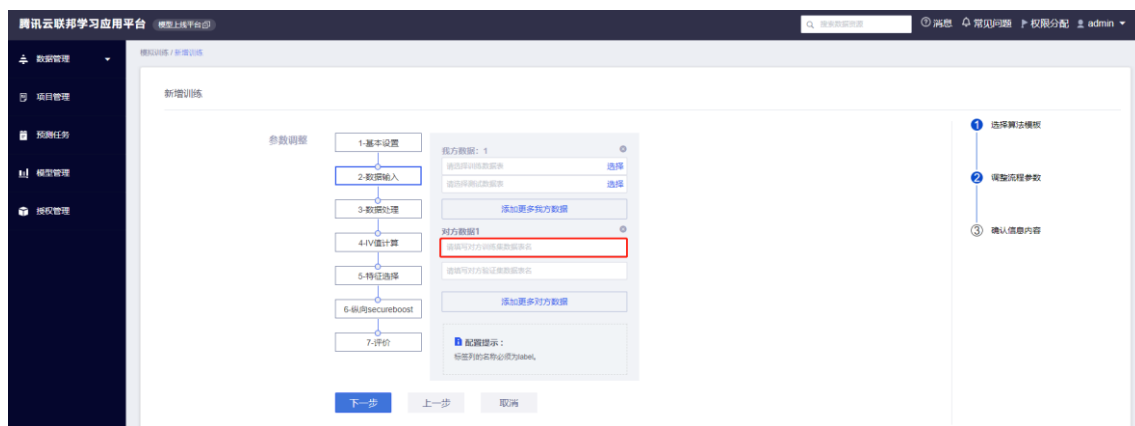


图 1.9 数据输入界面

参数设置举例：特征选择。特征选择列名默认值：全部，IV 阈值默认值：0.01。如用户无需修改系统设置的参数默认值，直接点击下一步，开始训练模型，如图 3.11 所示。

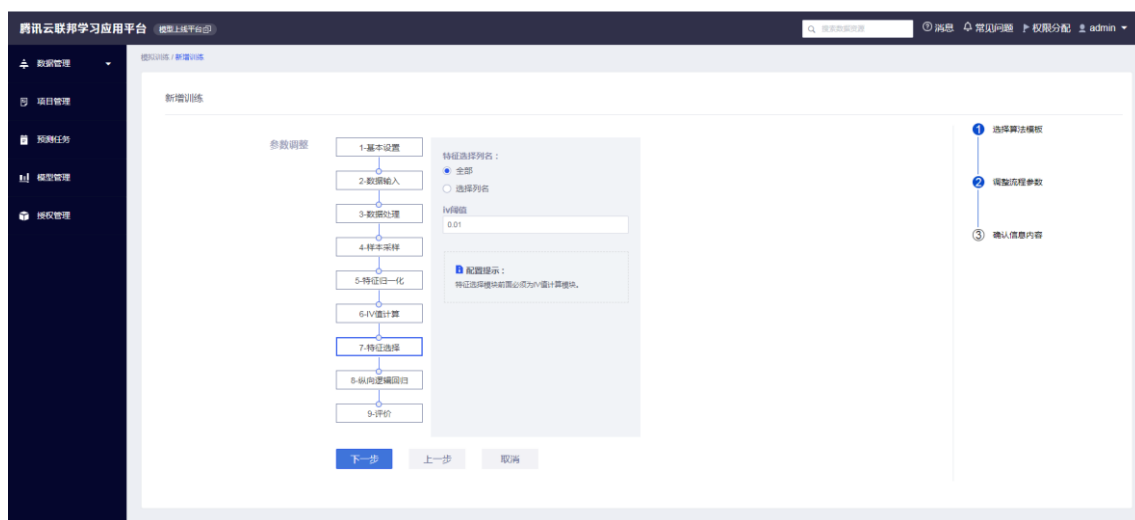


图 1.10 特征选择界面界面

### 1.6.3 提交训练

参数设置完毕之后，点击开始训练，系统会提示是否保存模板，用户如选择保存，下次训练模型选择模板时可沿用该模板，如图 3.12 所示。

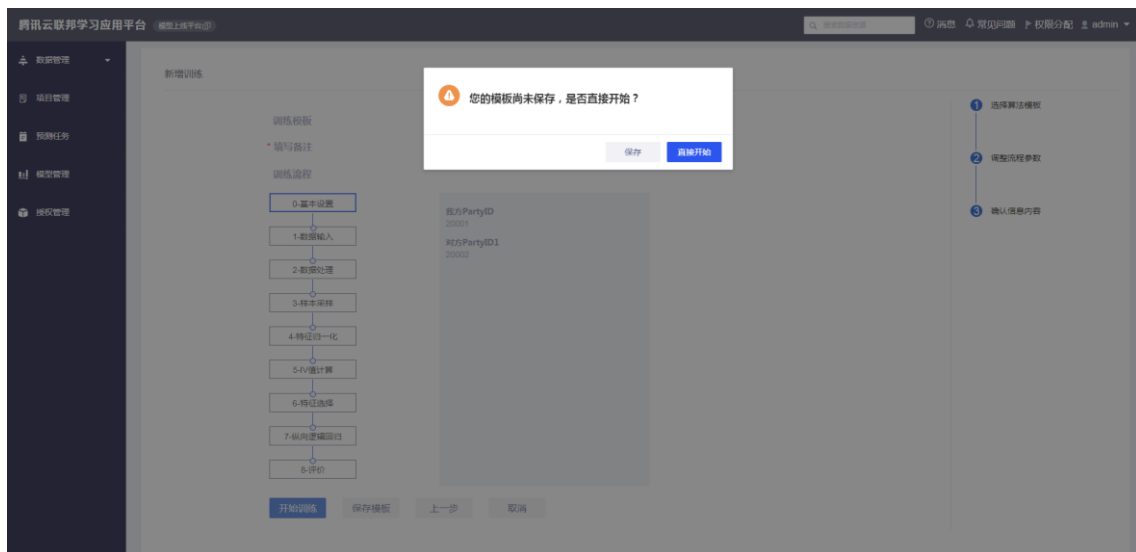


图 1.11 提交训练界面

训练任务提交后，用户在点击左侧的模型训练，在训练列表中能看见训练任务的基本情况。如图 1.12。

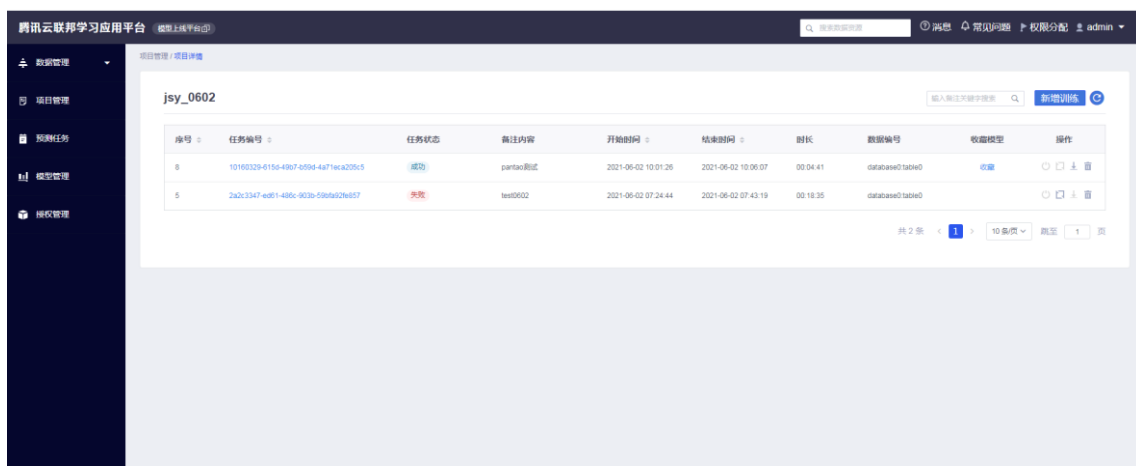


图 1.12 模型训练任务列表

## 1.7 任务查询

在浏览器 PC 设置 hosts 文件（例如 windows 系统位置：

C:\WINDOWS\system32\drivers\etc）：

外网 IP powerfl-xxx.com（例如，159.75.xx.xx powerfl-70000.com）。

1. 在模型训练页面点击任务编号，进入任务流程图界面。

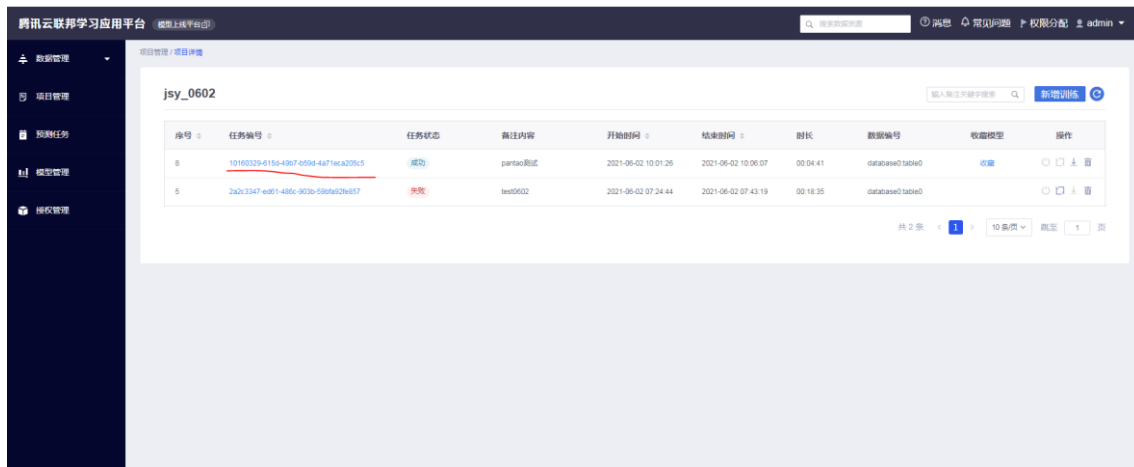


图 1.13 任务流程图入口

2. 点击任务链接 ID，查看任务流程图，全部为绿色为成功模块，红色为失败模块。

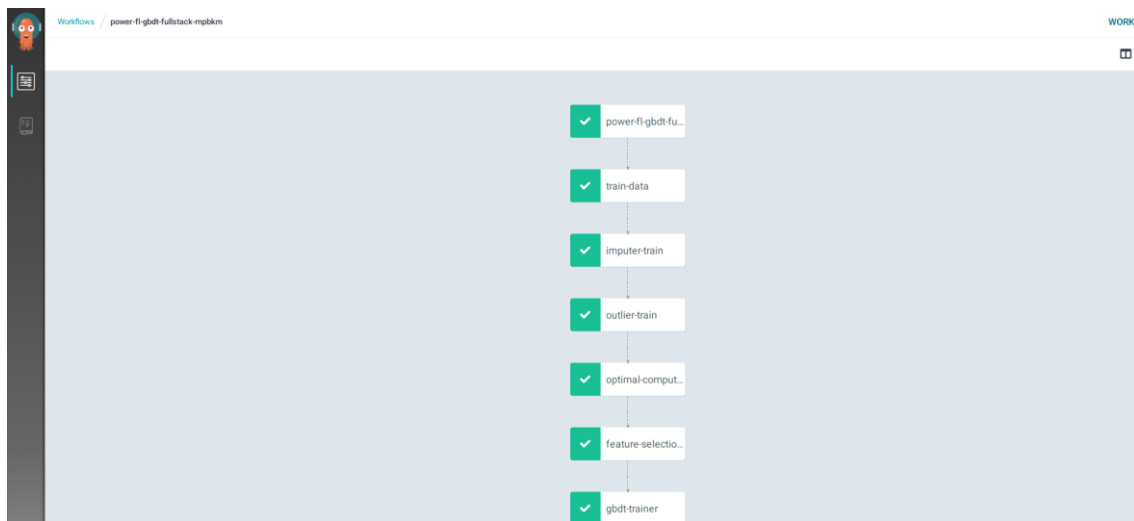


图 1.14 模型任务流程界面

在浏览器输入：<http://powerfl-70001.com> 可以打开模型结果页面，如下：

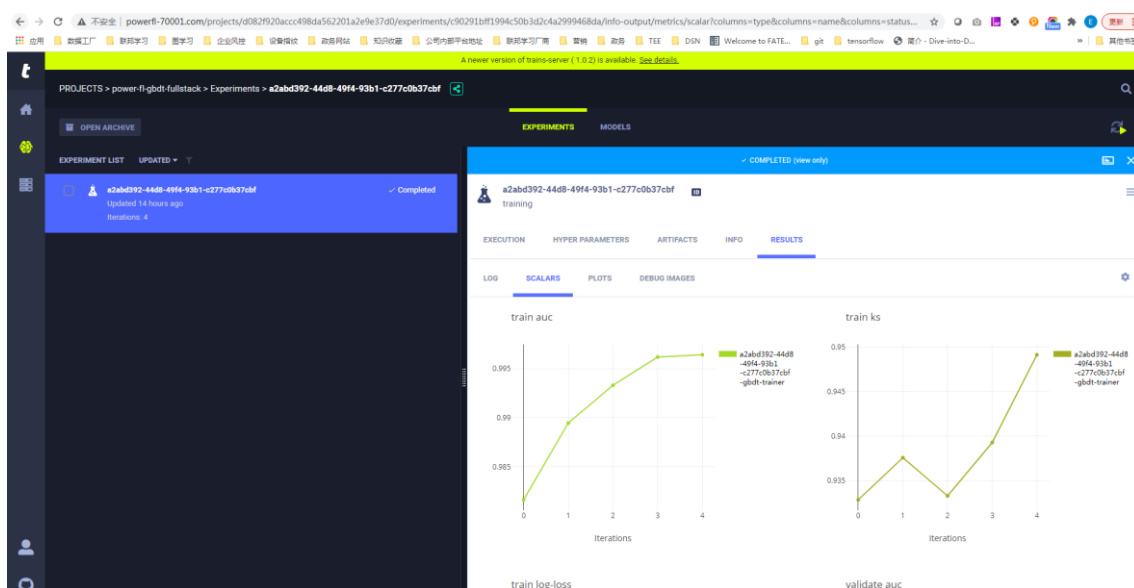


图 1.15 模型训练任务结果展示

## 1.7 模型下载

点击模型训练，进入训练列表，在操作栏中点击下载，即可下载模型文件，如图 1.16 所示，然后就可以上线生产。

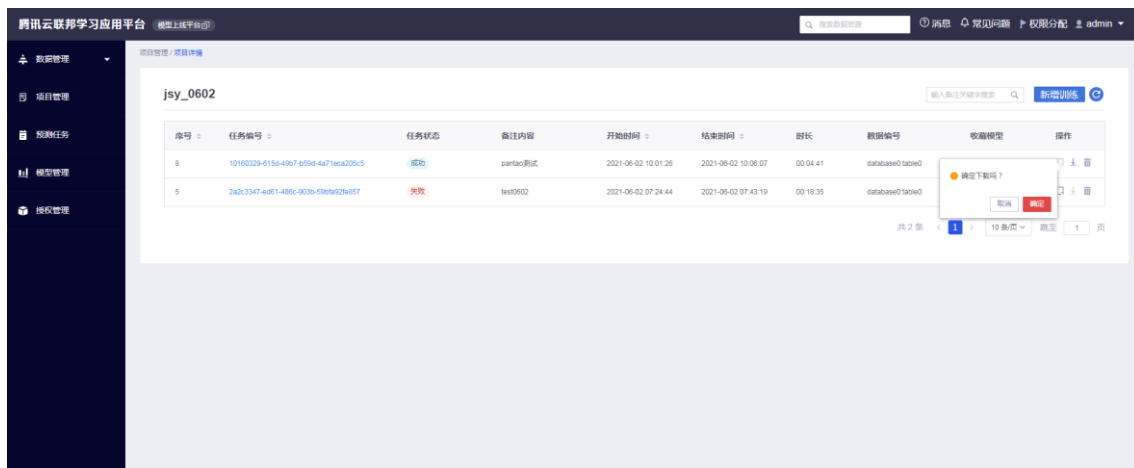


图 1.16 模型下载页面