

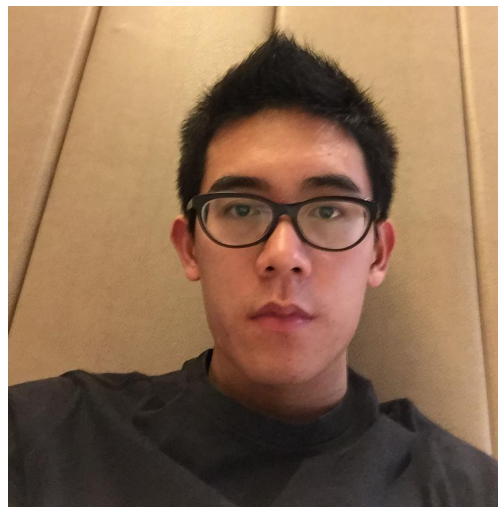
利用机器学习技术**预测**
信贷用户的预期行为

第 1 部分

讲师



讲师介绍



讲师：洪鹏飞

TIANCHIID: Cookly

阿里云天池平台开发者

天池大赛全国季军

微信: skooyboy

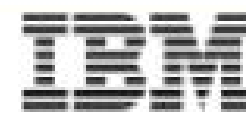
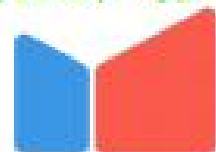
第 2 部分

智能风控



智能风控

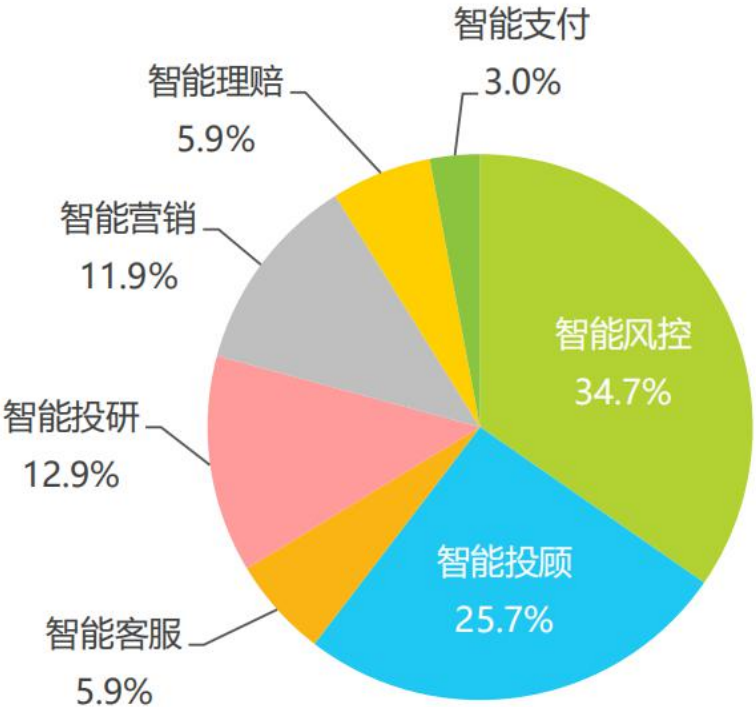
智能风控





智能风控

中国人工智能+金融行业投融资企业类型分布



2018年前三季度中国人工智能+金融行业
亿元以上融资事件

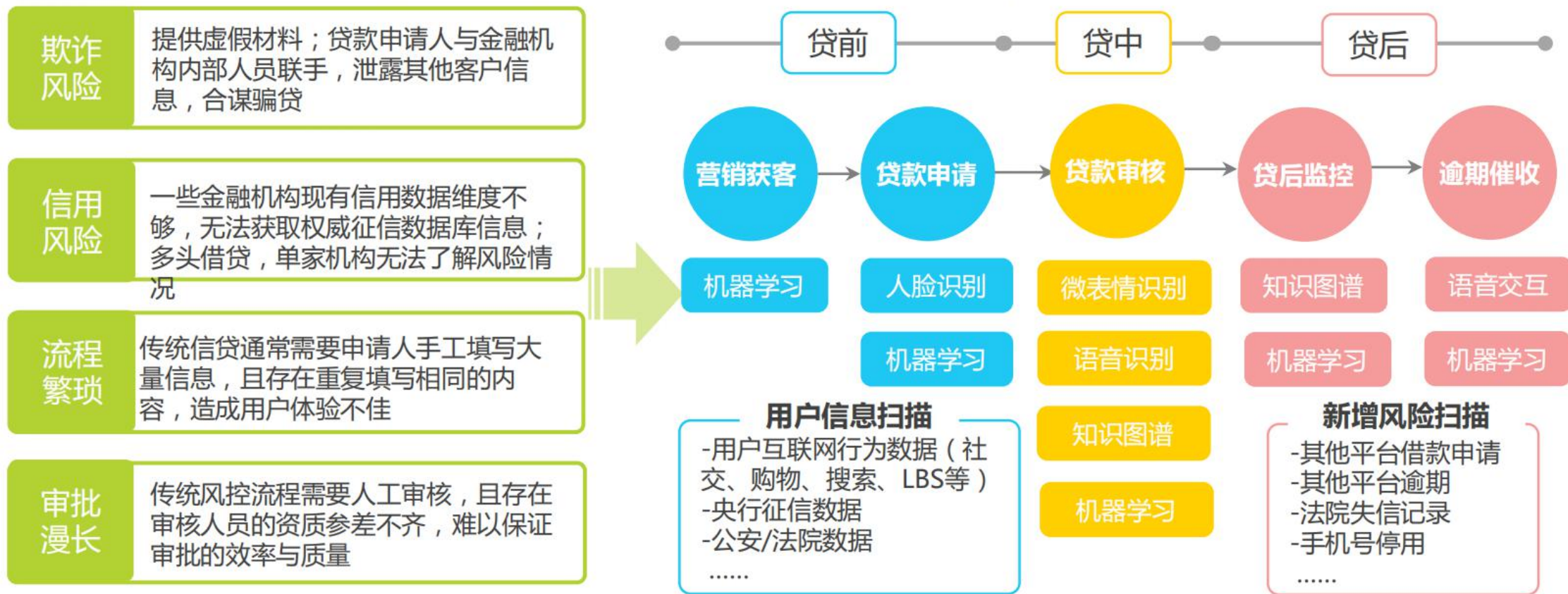
时间	公司	轮次	融资额
2018.8	慧安金科	A轮	1亿人民币
2018.7	京东金融	B轮	130亿人民币
2018.7	Advance.ai	B轮	5000万美元
2018.6	蚂蚁金服	战略投资	16亿人民币
2018.6	PINTEC	A轮	1.03亿美元
2018.6	虎博科技	Pre-A轮	数亿人民币
2018.4	度小满金融	战略投资	19亿美元
2018.4	百融金服	C轮	10亿人民币
2018.4	冰鉴科技	Pre-B轮	1.55亿人民币
2018.3	知因智慧	A轮	1亿人民币
2018.2	蚂蚁金服	战略投资	数亿人民币
2018.2	优品财富	A轮	2亿人民币
2018.1	金融壹账通	A轮	6.5亿美元

第 3 部分

案例场景

智能风控

传统业务面临的问题与智能风控流程图（以信贷为例）





智能风控

核心AI 技术

图像识别

文本识别

语音交互

机器学习

知识图谱

自然语言处理



核心产品与解决方案

反欺诈服务

业务风控服务

欺诈洞察

风控数据产品

人机识别

信贷风控服务

贷前反欺诈

贷前审核

智能信审

智能催收

贷后监控

信贷保镖

同盾智察分

同盾智信分

同盾定制分

用户增长服务

营销赋能服务

运营赋能服务

银行风控解决方案

保险风控解决方案

基金理财风控解决方案

汽车金融风控解决方案

非银行信贷风控解决方案

小微企业风控解决方案

三方支付风控解决方案

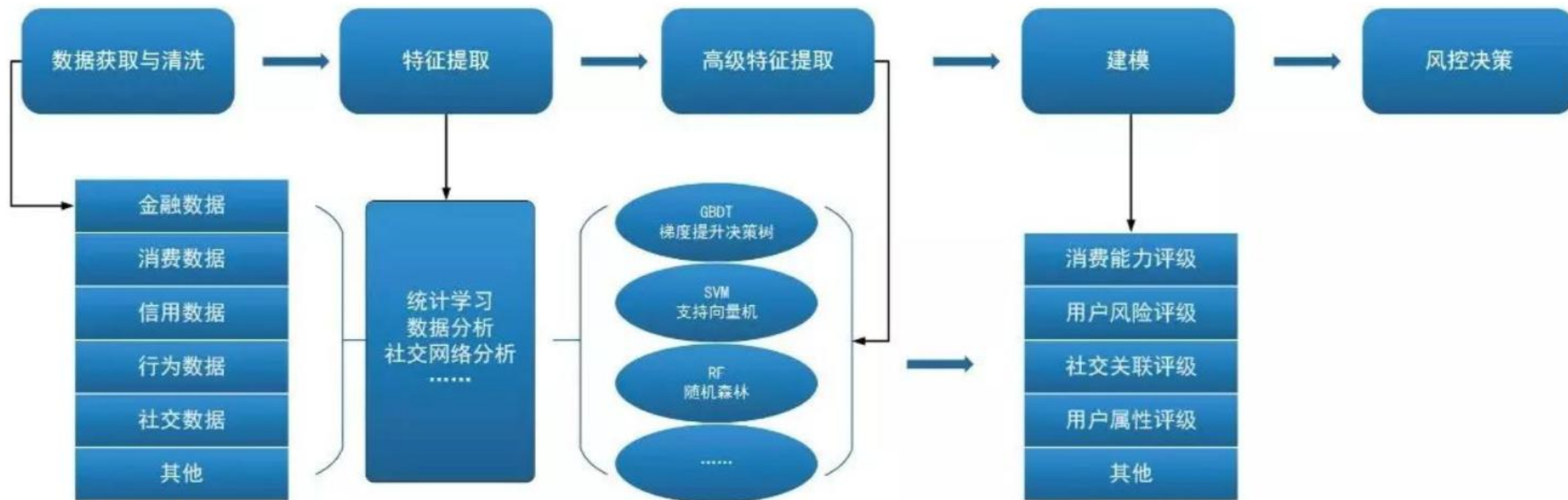
第 4 部分

风控建模



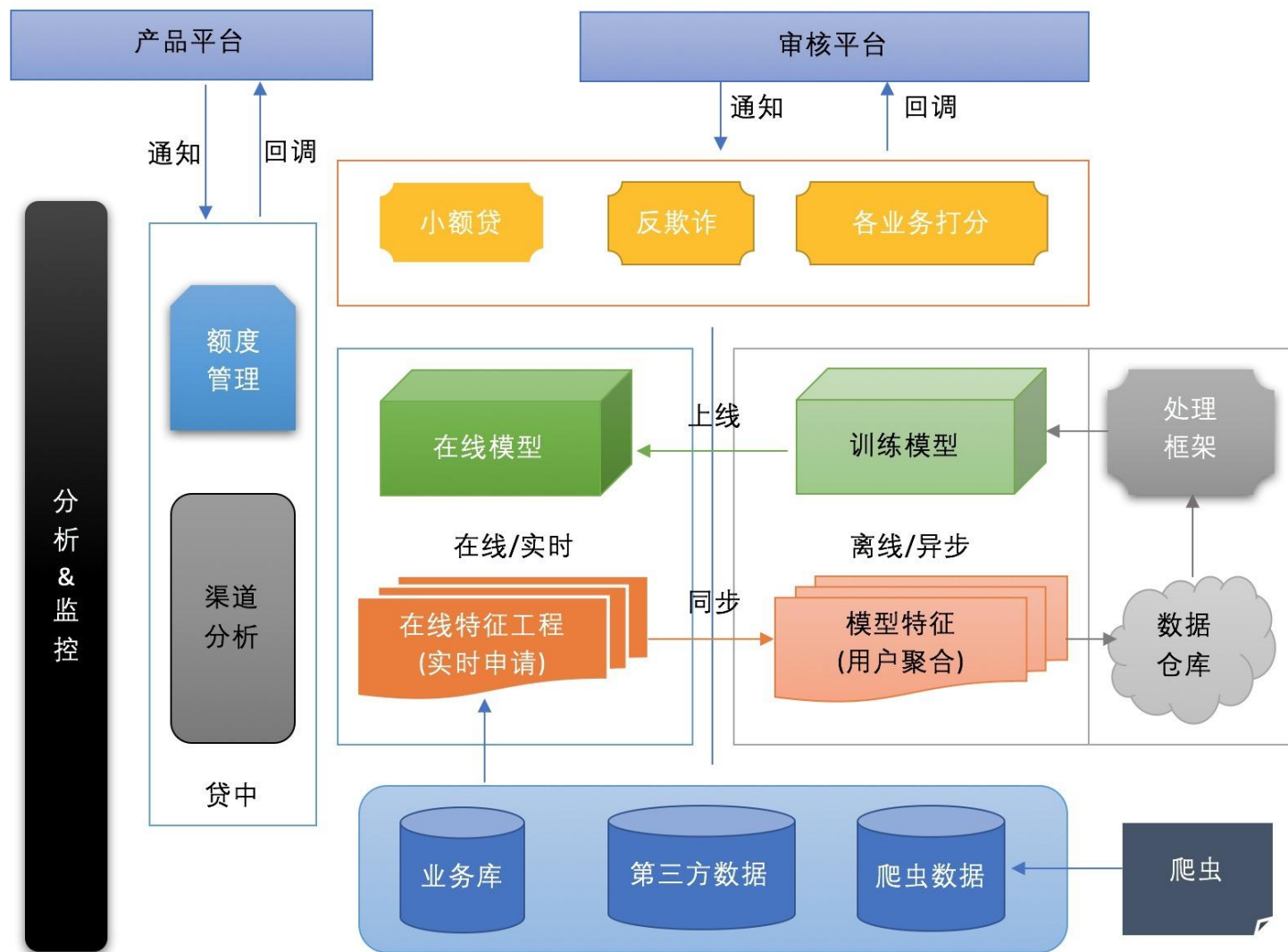
风控建模

风控建模中的数据科学



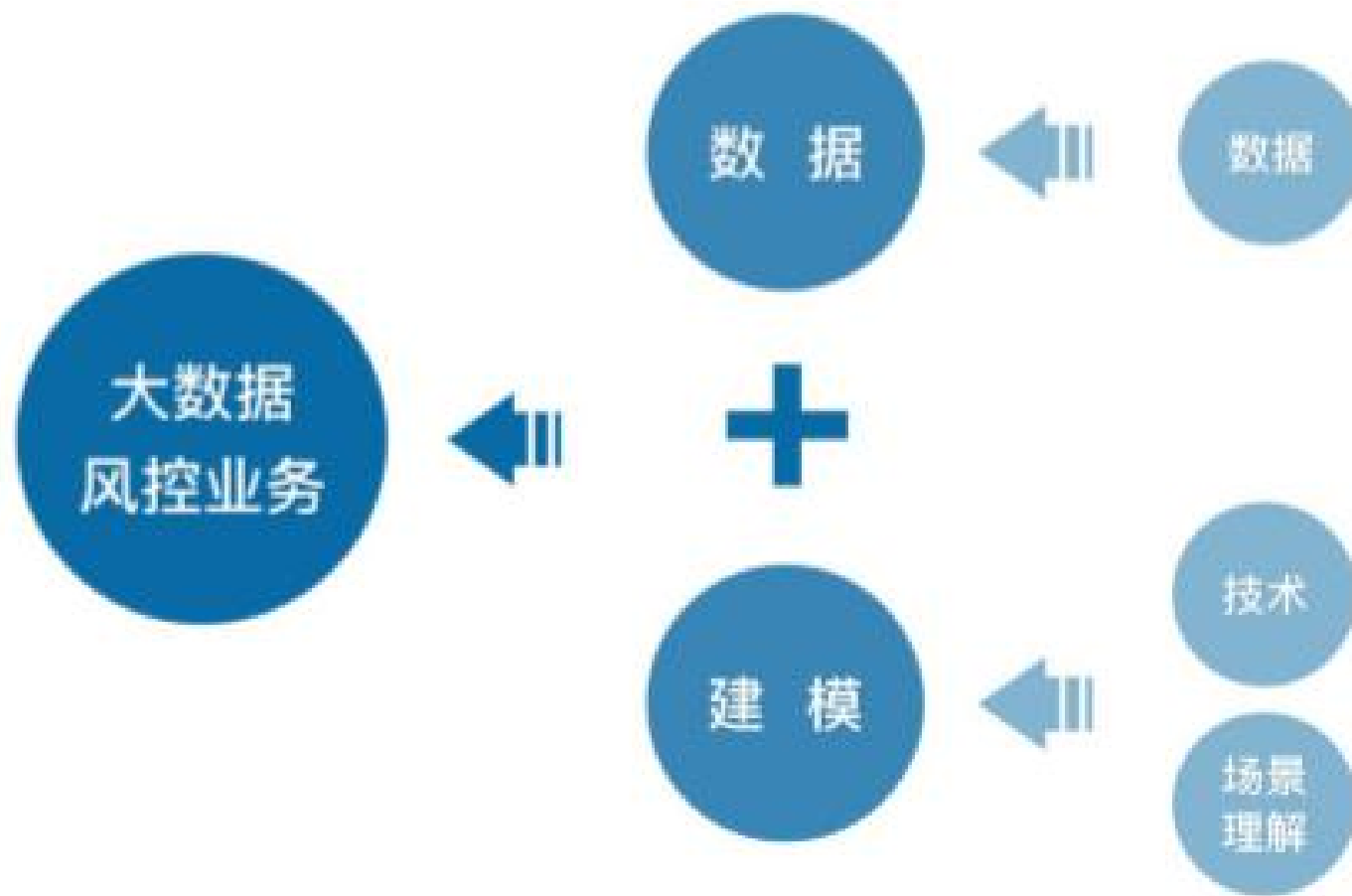


方案架构



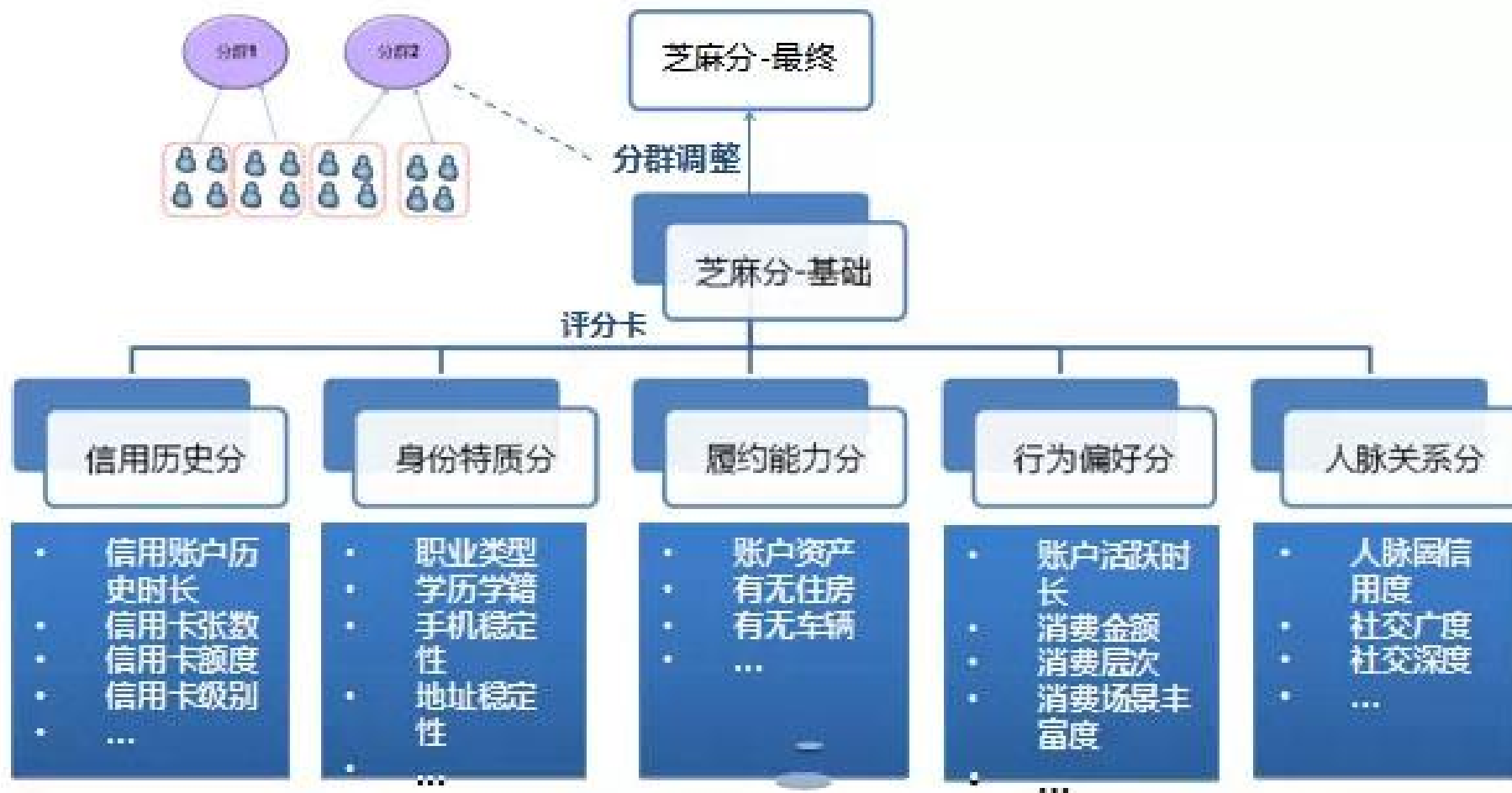


风控建模





风控建模



第 5 部分

案例风控数据



风控数据

什么是用户画像?

简而言之，用户画像是根据用户社会属性、生活习惯和消费行为等信息而抽象出的一个标签化的用户模型。构建用户画像的核心工作即是给用户贴“标签”，而标签是通过对用户信息分析而来的高度精炼的特征标识。

```
{  
  "ID": 123456,  
  "姓名": "张建国",  
  "性别": "男",  
  "出生年月": 631123200,  
  "籍贯": "北京",  
  "居住地": "北京",  
  "教育背景": {  
    "学校": "北京大学",  
    "专业": "CS",  
    "入学年月": 1220198400  
  }  
}
```




风控数据

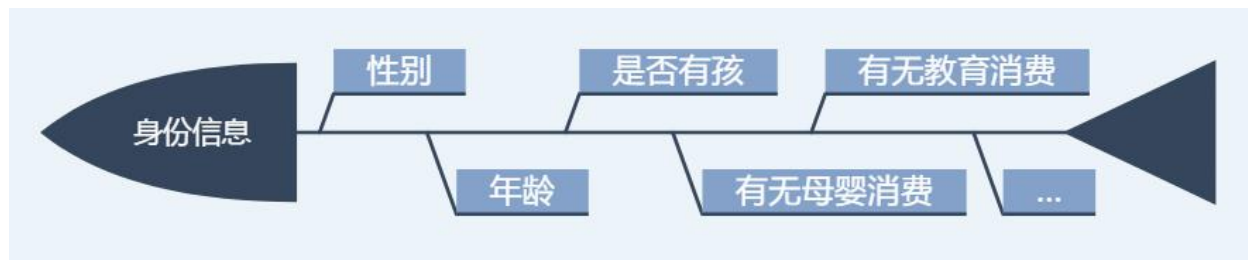
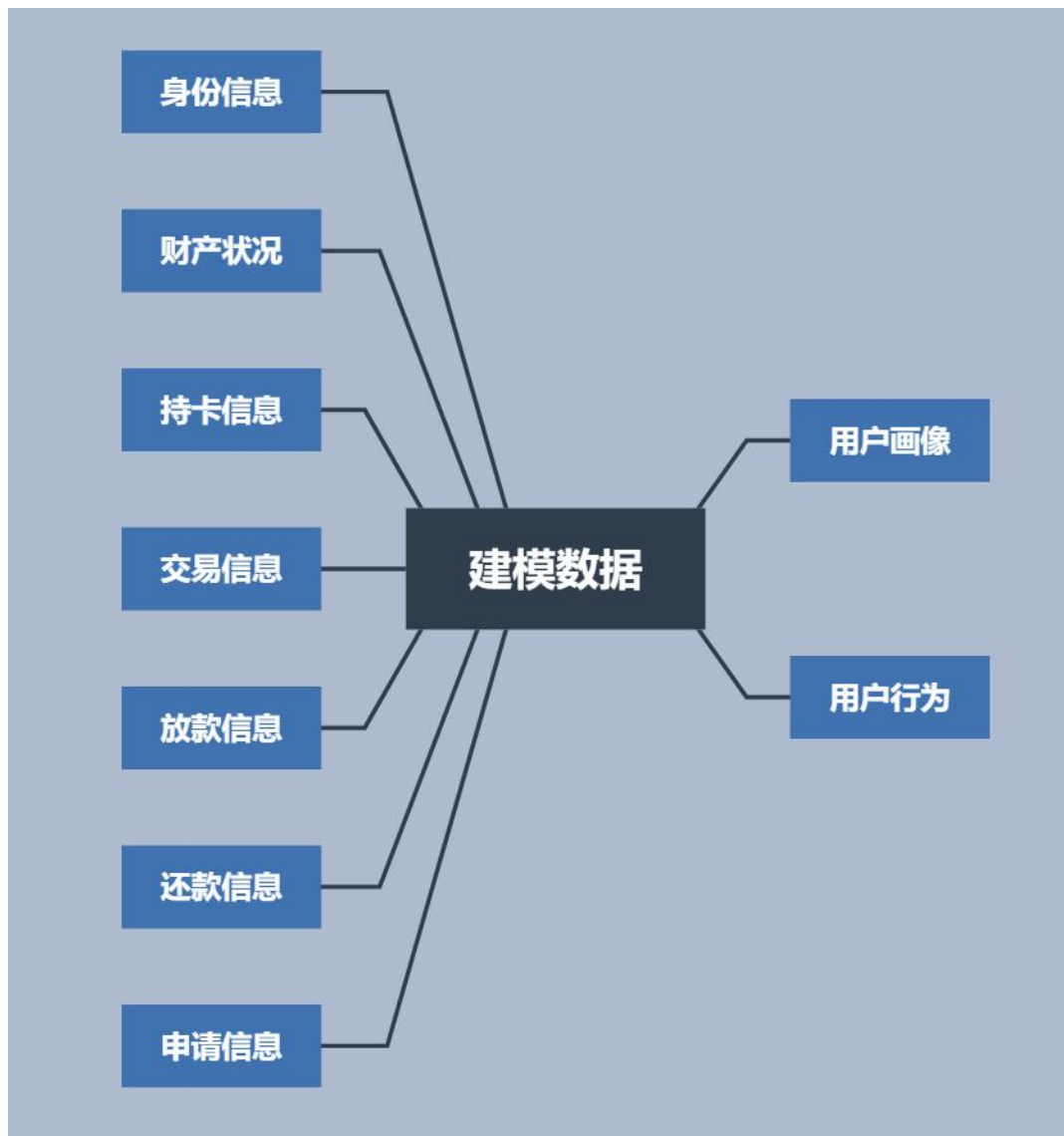
一、什么是用户行为？

用户行为由最简单的五个元素构成：时间、地点、人物、交互、交互的内容。

607045#成都市武侯区曾女士借款（第二标）	1	2018/7/15	还息	展期（预计还本时间2018.10.10）
607044#成都市武侯区曾女士借款（第一标）	1	2018/7/15	还息	展期（预计还本时间2018.10.10）
607042#成都市青羊区程女士借款（第一标）	1	2018/7/15	还息	展期（预计还本时间2018.10.10）
607046#成都市青羊区程先生借款（第一标）	1	2018/7/15	还息	展期（预计还本时间2018.10.10）
607047#成都市青羊区程先生借款（第二标）	1	2018/7/15	还息	展期（预计还本时间2018.10.10）
606430#成都市锦江区王先生借款	2	2018/7/15	还息	展期（预计还本时间2018.10.10）
607039#重庆市万州区李先生借款（第一标）	1	2018/7/15	还本付息	完成
606932#重庆市永川区饶先生借款	2	2018/7/15	还息	完成
606468#重庆市綦江区罗女士借款	3	2018/7/15	还息	完成
606944#重庆市铜梁区陈先生借款	3	2018/7/15	还息	完成
606943#重庆市铜梁区范女士借款	3	2018/7/15	还息	完成



风控数据





风控数据

用户画像

用户行为



第 6 部分

案例逾期建模

逾期预测建模





处理数据



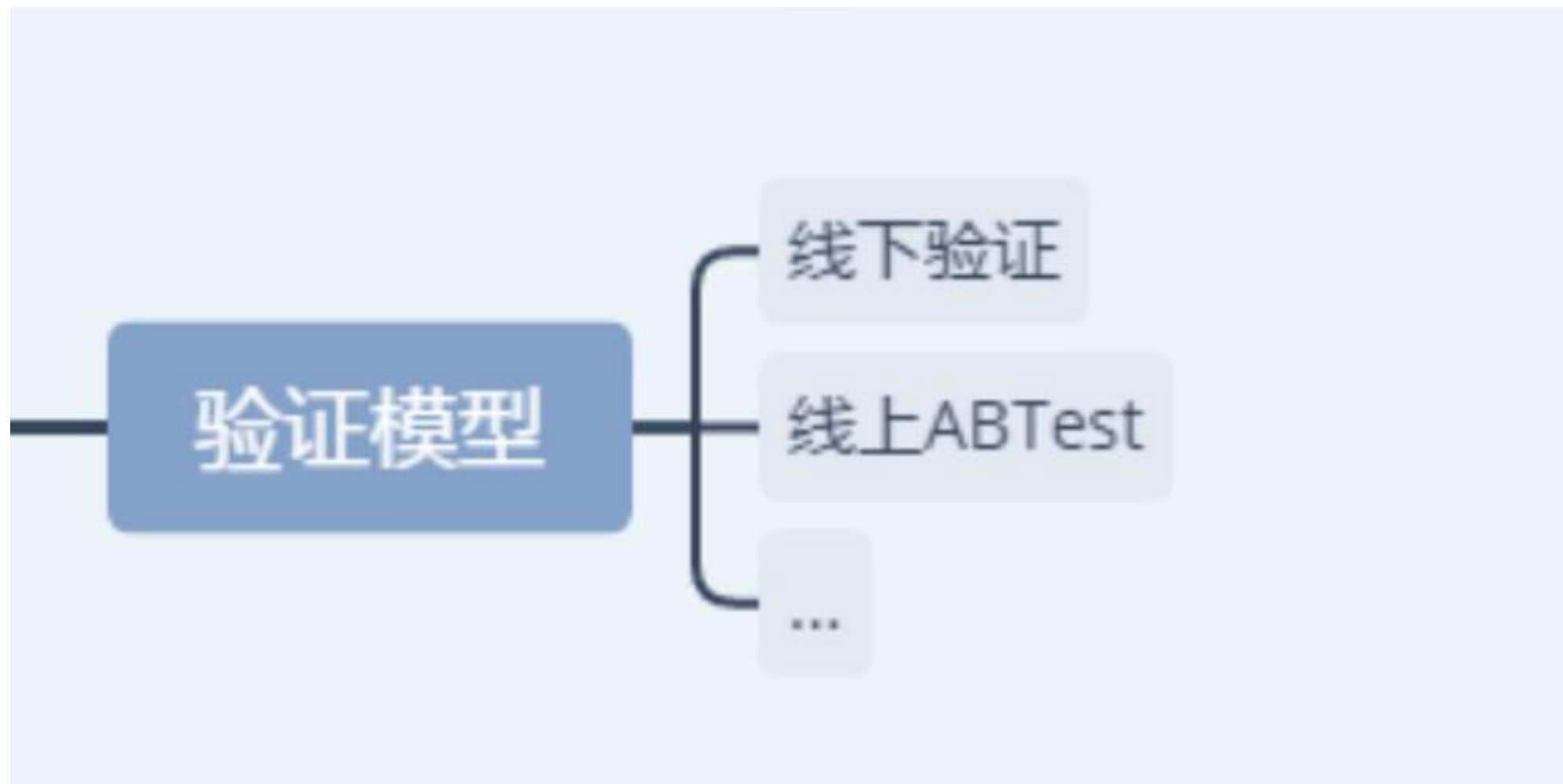


训练模型





验证模型



第 6 部分

天池平台



阿里云天池

阿里云 TIANCHI 天池

55 Cookly | 退出

首页

天池大赛

AI 学习

天池实验室

正式开放

数据集

技术圈

中
En

Guide

新手指南

一张小地图带你玩转天池
从入门到进阶，找到自己的
数据圈子

Hold a competition

我要参赛

百万奖金池，真实业务场
景演练，与全球ai人才比
拼，挑战业界排名

FLINK FORWARD

The Apache Flink® Conference

不仅仅是流计算 | More Than Streaming

2018年12月20日-21日 · 北京

现在注册

获得 **20+**
免费阿里内部培训课程
及往年大赛优质数据集

立即注册



天池实验室

3秒创建属于你的实验室
在线编辑分享，享受免费计算资源



AI课程

每天10分钟，成为下一个data hero
阿里内外部顶尖师资，帮你找到专属学习路径



技术圈

解答你的每一个AI问题
各领域技术专家在线响应



数据集

探索不同行业真实场景数据
阿里系唯一对外开放数据分享平台



AI学习

阿里云 TIANCHI天池

44 Cookly | 退出

首页

天池大赛

AI学习

天池实验室

正式开放

数据集

技术圈

其他

中 En

The diagram illustrates a learning path for AI, starting from a rocket launch at the bottom and branching out into various topics. The path is divided into two main sections: '机器学习快速入门' (Machine Learning Quick Start) and '机器学习进阶' (Machine Learning Advanced). The '机器学习快速入门' section includes '机器学习入门' (Machine Learning Introduction), '数据可视化' (Data Visualization), '数据科学入门' (Data Science Introduction), '数据科学进阶' (Data Science Advanced), '数据科学实战' (Data Science Practical), '数据科学理论' (Data Science Theory), '数据科学应用' (Data Science Application), '数据科学案例' (Data Science Case Studies), '数据科学项目' (Data Science Projects), '数据科学竞赛' (Data Science Competitions), '数据科学社区' (Data Science Community), '数据科学资源' (Data Science Resources), '数据科学工具' (Data Science Tools), '数据科学框架' (Data Science Frameworks), '数据科学库' (Data Science Libraries), '数据科学接口' (Data Science APIs), '数据科学服务' (Data Science Services), '数据科学平台' (Data Science Platforms), '数据科学生态' (Data Science Ecosystem), '数据科学未来' (Data Science Future). The '机器学习进阶' section includes '机器学习进阶' (Machine Learning Advanced), '机器学习项目' (Machine Learning Projects), '机器学习竞赛' (Machine Learning Competitions), '机器学习社区' (Machine Learning Community), '机器学习资源' (Machine Learning Resources), '机器学习工具' (Machine Learning Tools), '机器学习框架' (Machine Learning Frameworks), '机器学习库' (Machine Learning Libraries), '机器学习接口' (Machine Learning APIs), '机器学习服务' (Machine Learning Services), '机器学习平台' (Machine Learning Platforms), '机器学习生态' (Machine Learning Ecosystem), '机器学习未来' (Machine Learning Future). The diagram also features a legend for '待上线课程' (Upcoming Courses) and '已上线课程' (Released Courses), a rocket launch illustration, and a small figure of a person standing on a globe.

机器学习快速入门

机器学习入门

数据可视化

数据科学入门

数据科学进阶

数据科学实战

数据科学理论

数据科学应用

数据科学案例

数据科学项目

数据科学竞赛

数据科学社区

数据科学资源

数据科学工具

数据科学框架

数据科学库

数据科学接口

数据科学服务

数据科学平台

数据科学生态

数据科学未来

机器学习进阶

机器学习项目

机器学习竞赛

机器学习社区

机器学习资源

机器学习工具

机器学习框架

机器学习库

机器学习接口

机器学习服务

机器学习平台

机器学习生态

机器学习未来

待上线课程

已上线课程

推荐讲师

我要开讲



天池实验室

阿里云 TIANCHI天池

☰

我的notebook >

我的数据 >

使用说明 >

全部notebook >

Draft DefaultWed Dec 19 15:53:17 CST 2018 私有

编辑 删除

版本 1 ^

数据列表

Content forks (0) comments (0)

```
[ ]: 1 # 查看数据文件目录 list datalab files
      2 !ls datalab/

[ ]: 1 # 查看个人永久空间文件 list files in your permanent storage
      2 !ls /home/myspace/

[ ]: 1 # 查看当前kernel下已安装的包 list packages
      2 !pip list --format=columns

[ ]: 1 # 绘图案例 an example of matplotlib
      2 %matplotlib inline
      3 import numpy as np
      4 import matplotlib.pyplot as plt
      5 from scipy.special import jn
      6 from IPython.display import display, clear_output
      7 import time
      8 x = np.linspace(0,5)
      9 f, ax = plt.subplots()
     10 ax.set_title("Bessel functions")
     11
     12 for n in range(1,10):
     13     time.sleep(1)
     14     ax.plot(x, jn(x,n))
     15     clear_output(wait=True)
     16     display(f)
     17
     18 # close the figure at the end, so we don't get a duplicate
     19 # of the last plot
     20 plt.close()

[ ]: 1
```



天池实验室用户手册

阿里云 TIANCHI天池

我的notebook

我的数据

使用说明

全部notebook

GPU使用方法

2018-09-18

天池实验室用户手册

2018-03-30

天池实验室用户手册 公开

天池小T 2018-03-30 PYTHON3 33234

内容

fork记录 (37)

评论 (149)

天池实验室用户手册

天池实验室简介

亲爱的天池用户大家好!

天池notebook正式升级为天池实验室啦。后续天池实验室将会为大家提供包括notebook在内的多种简单易用的数据分析工具。在这里,你将能够和大牛们一起学习交流,分享见解,不断进步。不要错过哦~

快捷键 & 代码补全 (Hot!)

天池实验室中的notebook已完全采用jupyter notebook的原生快捷键,并且将支持tab代码补全哦~

入口

你可以从天池网站首页一级导航找到天池实验室

阿里云 TIANCHI天池

user08984 | 退出

首页 天池大赛 AI 学习 天池实验室 数据集 技术圈

新手指南

一张小地图带你玩转天池从入门到进阶,找到适合自己的数据圈子

我要参赛

百万奖金池,真实业务场景挑战,联合全球人才比拼

“飞凤数创”2018全球物联网数据创新大赛

万物互联,云启无限!可视化创新挑战赛等你来战!

查看详情

现在注册

获得20+免费阿里内部培训课程及往年大赛优质数据集

立即注册

版本 9

数据列表

第 7 部分

天池实验室实战



天池试验

阿里云 TIANCHI天池



Cookly | 退出

首页

天池大赛

AI 学习

天池实验室

正式开放

数据集

技术圈



Guide

新手指南

一张小地图带你玩转天池
从入门到进阶，找到自己的数据圈子

Hold a competition

我要参赛

百万奖金池，真实业务场景演练，与全球AI人才比拼，挑战业界排名

津南数字制造算法挑战赛

聚焦收率预测以及智能预警 助力津南制造数字化转型

Dec.8, 2018- Mar.15, 2019

立即报名

现在注册

获得 **20+**
免费阿里内部培训课程
及往年大赛优质数据集

立即注册



天池实验室

3秒创建属于你的实验室
在线编辑分享，享受免费计算资源



AI课程

每天10分钟，成为下一个data hero
阿里内外部顶尖师资，帮你找到专属学习路径



技术圈

解答你的每一个AI问题
各领域技术专家在线响应



数据集

探索不同行业真实场景数据
阿里系唯一对外开放数据分享平台



上传数据

阿里云 TIANCHI 天池

我的notebook

我的数据

使用说明

全部notebook

私有

公有

Draft Thu Dec 20 17:20:37 CS...

2018-12-20

YinLianData

2018-12-19

Draft Tue Dec 18 19:20:45 CS...

2018-12-18

第一步

+新建

Draft DefaultWed Dec 19 16:46:44 CST 2018 私有

Cookly

2018-12-19

</> PYTHON3

内容

fork记录 (0)

评论 (0)

[17]:

1 # 查看数据文件目录 List datalab files

2 !ls datalab/

7572

[18]:

1 # 查看个人永久空间文件 List files in your permanent storage

2 !ls /home/myspace/

[19]:

1 # 查看当前kernel下已安装的包 List packages

2 !pip list --format=columns

Package	Version
abs1-py	0.2.2
alabaster	0.7.11
alembic	0.9.9
altair	1.2.0
arrow	0.12.1
asn1crypto	0.24.0
astor	0.6.2
attrs	18.1.0
Babel	2.5.3
backcall	0.1.0
backports.functools-lru-cache	1.5
beautifulsoup4	4.6.0
blaze	0.10.1
bleach	1.5.0
blinker	1.4
bokeh	0.12.16
boto	2.48.0
boto3	1.7.48
botocore	1.10.48
bunch	1.0.1
bz2file	0.98
certifi	2018.8.24
cffi	1.11.5
chardet	3.0.4
click	6.7
cloudpickle	0.5.3
colorama	0.3.9

版本 1

目录

数据列表

YinLianData



上传数据

阿里云 TIANCHI 天池

YINLIAN

公开 私有

发布

数据名称

内容

数据列表 只支持csv/txt/zip/tar/gz/md/jar/tif格式，上传数据最大不超过1.5G/个文件，超大数据最大不超过10G/个文件 超大文件上传

拖拽上传

上传数据

文档

markdown 图片 预览

打开

组织 新建文件夹

名称 修改日期

model_sample.csv 2018/4/3 14:5

参数表.xlsx 2018/4/3 14:5

字段解释.xlsx 2018/4/3 15:0

文件(N): model_sample.csv 所有文件

打开(O) 取消

选择数据

点击上传



创建notebook

阿里云 TIANCHI天池

中En 4.4 Cookly | 退出

我的notebook >
我的数据 >
使用说明 >
全部notebook >

YINLIAN

Cookly 2018-12-20 0 0 0 0

创建notebook

编辑

删除

内容 notebook 评论

数据列表

数据名称	上传时间	大小	删除	下载
model_sample.csv	2018-12-20 17:40	4.99MB		

文档

暂无

描述:

暂无描述

点击创建notebook



创建notebook

阿里云 TIANCHI天池

中 En 44 Cookly | 退出

Draft Thu Dec 20 17:42:50 CST 2018 PYTHON3 关联比赛 发布版本

connected kernel将在 01:59:32 断开

[2]:
1 # 查看数据文件目录 List datalab files
2 !ls datalab/

7605

[3]:
1 # 查看个人永久空间文件 List files in your permanent storage
2 !ls /home/myspace/

[6]:
1 # 查看当前kernel下已安装的包 List packages
2 !pip list --format=columns

[5]:
1 # 绘图案例 an example of matplotlib
2 %matplotlib inline
3 import numpy as np
4 import matplotlib.pyplot as plt
5 from scipy.special import jn
6 from IPython.display import display, clear_output
7 import time
8 x = np.linspace(0,5)
9 f, ax = plt.subplots()
10 ax.set_title("Bessel functions")
11
12 for n in range(1,10):
13 time.sleep(1)
14 ax.plot(x, jn(x,n))
15 clear_output(wait=True)
16 display(f)
17
18 # close the figure at the end, so we don't get a duplicate
19 # of the last plot
20 plt.close()

添加数据集

已有数据集

代码区

申请GPU

数据列表

+
YINLIAN

Bessel functions

第 8 部分

模型应用



芝麻信分



芝麻分是什么？

芝麻分是一个代表您信用程度的数字
参考国际通用的信用分体系



分数越大，代表信用程度越好





风控运用



谢

谢