深度之眼
deepshare.net

Kaggle-Riiid! Answer Correctness Prediction竞赛培训

Cookly

关注公众号
获取第一手干货咨询

添加小享
获得Baseline&课件

深度之眼
deepshare.net

个人介绍

All about me

Cookly，推荐算法方向

某大厂互联网公司算法工程师

阿里云天池、DataFountain、京东零售科普讲师

达观推荐1st | 携程出行销售量 1st

携程预订房型 1st | 美年年健康 2nd

阿里里聚安全 3rd | 中国网网络对抗 8th

N次竞赛TOP3

阿里云天池大赛赛题解析　主要作者

一个问题

你为什么要打比赛？

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

竞赛的好处

挣钱

挑战

找工作

交友

我在比赛中获得了什么

深度之眼
deepshare.net

我小白时候的艰辛+
我自己的成长收获+
从工作中发现比赛收益的地方

我们如何带大家学习这门课

课程特色

1、实战Kaggle-riiid比赛（kaggle大数据平台），提供**Lgb和Nffm的baseline方案**

2、**详尽介绍赛题解析，建模过程**，分别了解**传统建模阶段和深度建模阶段**

3、针对Baseline方法如何进一步优化，帮助理解每一个trick后的意义

4、**Ctr类点点击预解决方案**异同点详解

5、如何在下一场类似的比赛中快速的取得好成绩

6、如何将实战中知识用于工作（面试）中

我们如何带大家学习这门课

学习流程

01 赛题理解、Baseline解析

02 数据处理、特征加工
特征筛选、模型训练与验证

03 ctr类模型、类别特征处理、
连续特征处理

04 相似领域Paper讲解

05 拆解赛题、特征进阶
模型优化、模型融合

06 深度模型结构设计技巧
深度模型tick

07 整理知识点和上分点
比赛思路全复盘

深度之眼
deepshare.net

课程计划

直播时间表

	课程专题	知识点	时间	讲师
1	**开营仪式**	赛题内容介绍 Baseline代码讲解	10/17 20:00	Cookly
2	**结构化数据传统建模**	数据处理、特征加工 特征筛选、模型训练与验证	10/18 20:00	Cookly
3	**结构化数据深度建模**	ctr类模型、 类别特征处理、 连续特征处理	10/24 20:00	Cookly
4	**中期直播答疑**	解答比赛过程中的问题	10/25 20:00	Cookly
5	**比赛相关Paper讲解**	相似领域Paper讲解	10/31 20:00	Cookly
6	**传统Baseline模型进阶**	拆解赛题、特征进阶 模型优化、模型融合	11/1 20:00	Cookly
7	**深度Baseline模型进阶**	深度模型结构设计技巧 深度模型tick	11/7 20:00	Cookly
8	**比赛复盘**	比赛思路全复盘 优胜选手方案分享 整理知识点和上分点 比赛经验的面试展现技巧 干货分享	根据赛程和开源情况定期	Cookly

扫码享限时优惠

←←←←←←←

原价298，限时198!

获取Baseline&课件

←扫码加小享!

注意事项

深度之眼
deepshare.net

教练指导带队≠全自动代打

一、针对初学者，教练指导的是思路和方法，打排名的是自己!

二、前期知识的储备很重要，知识不足可参加 **深度之眼相关训练营**

三、积极参加日常群内讨论答疑、比赛组队（报名本次训练营的同学）

四、及时打卡提交作业，作业不仅基于赛事赛程，也会基于面试应用

深度之眼
deepshare.net

目录

1 Kaggle平台简介

2 数据挖掘简介

3 赛题背景分析

4 Baseline思路介绍

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

深度之眼
deepshare.net

1、Kaggle简介

Introduction of Kaggle

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

数据科学比赛简介

深度之眼
deepshare.net

集智平台

Github

TIANCHI天池

Kaggle

DC竞赛
www.dcjingsai.com

DataFountain

Kesci

切忌闭门造车!!!

赛题分类1：

01 **Featured**
商业或科研难题，奖金一般较为丰厚

02 **Recruitment**
比赛的奖励为面试机会

03 **Research**
科研和学术性较强的比赛，也有一般需要较强的领域和专业知识

04 **Playground**
提供一些简单的任务用于熟悉平台和比赛

05 **Getting Started**
提供一些简单的任务用于熟悉平台和比赛

06 **In Class**
用于课堂项目作业或者考试

在线提交比赛

赛题分类2：

离线提交比赛

赛题分类3：

数据挖掘　　图像

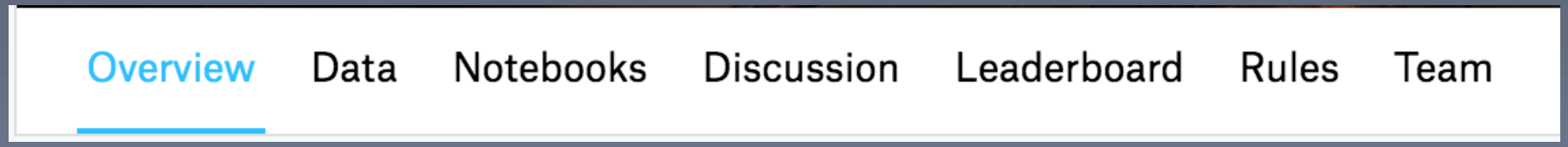
语音　　自然语言

称号和奖牌

金、银、铜

GM，M，EX，Con，Nov，User

初学者怎么用Kaggle



多看，多学，多沟通！！！

加小享回复【Kaggle入门】
获取Kaggle比赛入门资料

关于 A/B 榜

A 榜成绩好≠B 榜单成绩好

目的：防止过拟合

深度之眼
deepshare.net

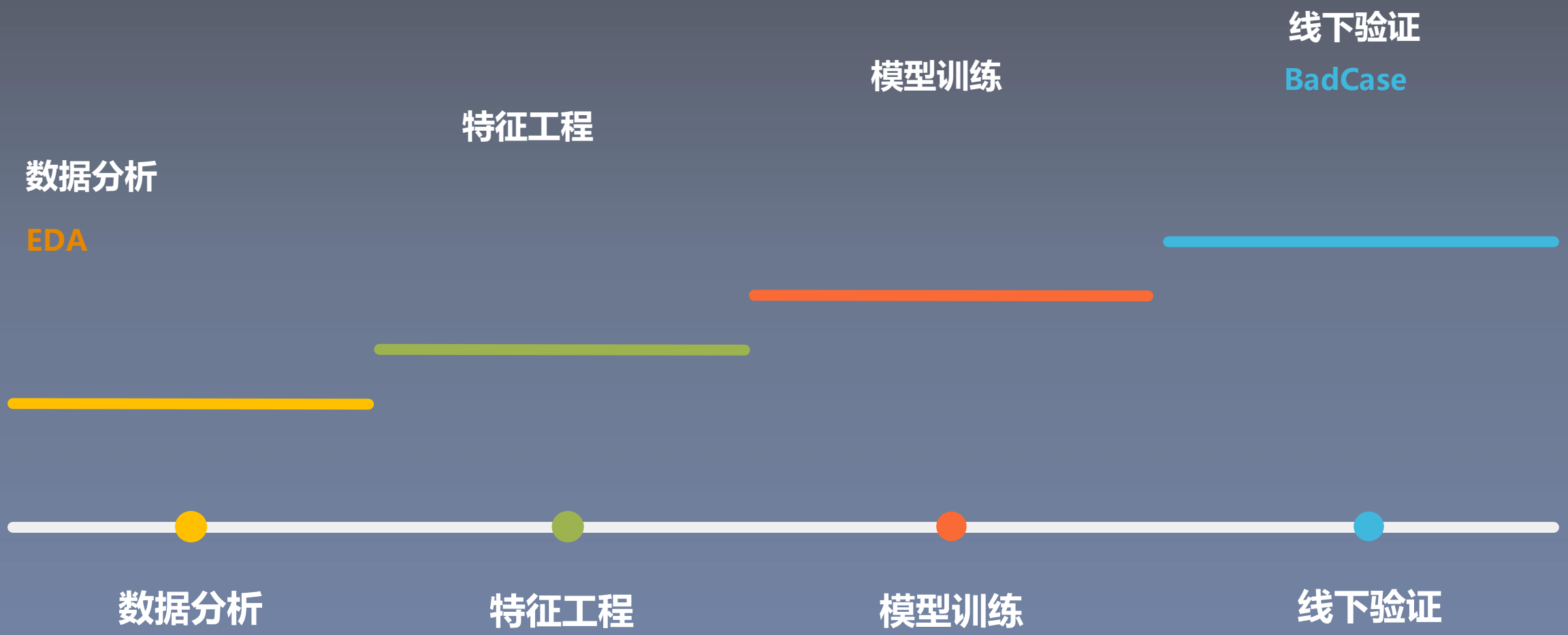
2、比赛通用流程

General process of competition

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

比赛通用流程



多找工具去复用　Pipeline很重要

深度之眼
deepshare.net

3、赛题介绍

Introduction to competition questions

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

赛题背景

深度之眼
deepshare.net

Riiid! Answer Correctness Prediction

在2018年，有2.6亿儿童没有上学。同时，这些年幼的学生中有一半以上没有达到最低的阅读和数学标准。COVID-19迫使大多数国家暂时关闭学校，此时教育已经陷入困境。这不仅进一步耽误了学生们的学习机会和智力发展。而且每个国家的教育公平差距可能会越来越大。我们需要从出勤率、参与度和个性化学习的关注度等各方面重新思考当前的教育体系。

在这次比赛中，你所面临的挑战是创建"知识追踪"的算法，即随着时间的推移对学生知识进行建模。目标是准确预测学生在未来互动中的表现。你将使用Riiid的EdNet数据来搭配你的机器学习技能。

赛题思路

传统建模

数据清洗

特征工程

数据切分

模型训练

模型验证

模型预测

深度建模

数据清洗

网络设计

数据切分

模型训练

模型验证

模型预测

赛题难点

1. kernel 16G较小，容易死

2. 传统特征工程比较难操作

3. 深度模型需要较多的网络设计技巧

怎么办？？？

赛程注意事项

时间线

开始日期 ：2020/10/5
合并截止日期 ：2020/12/31
报名截止日期：2020/12/31
结束日期（最终提交截止日期）：2021/1/7

注：除非另有说明，所有截止日期均为相应日期的11:59 UTC。大赛主办方保留更改比赛赛程的权利。UTC为世界标准时。 中国大陆、中国香港、中国澳门、中国台du湾、蒙古国zhi、新加坡、马来西亚、菲律宾、西澳大利亚州的时间与UTC的时差均为+8，也就是UTC+8。

奖金

- 第一名-$ 50,000
- 第二名-$ 30,000
- 第三名-$ 10,000
- 第四名-$ 5,000
- 第五名-$ 5,000

获奖团队将被邀请2021年二月在“AAAI-2021 Workshop on AI Education - Imagining Post-COVID Education with AI” 展示他们的模型。

深度之眼
deepshare.net

4、数据介绍

Data introduction

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

数据介绍

训练数据集：train.csv

row_id:	int64	行ID
timestamp:	int64	从该用户交互到该用户第一次事件完成之间的时间（以毫秒为单位）
user_id:	Int32	用户的ID
content_type_id:	Int16	问题或讲座的ID
user_answer:	Int8	用户对问题的答案(如果有)
answered_correctly:	Int8	用户是否正确(Label)
previous_question_elapsed_time:	Float32	用户回答上一个问题包中的每个问题所花费的平均时间
before_question_had_explanation:	Bool	是否看了演讲，是否看到了说明和正确的回答

数据介绍

问题数据：questions.csv

question_id:	train / test content_id列的外键
bundle_id:	解决问题ID
correct_answer:	问题的答案
part:	TOEIC考试的相关部分
tags:	问题的一个或多个详细标签ID

数据介绍

用户在学习讲座：lectures.csv

question_id:	讲座ID
part:	讲座的类型
tag:	讲座的标签ID
type_of:	类型

数据样本分析

example_test_rows.csv： 时间序列API会传递测试集数据的三个样本组。 格式与train.csv大致相同。 有两个不同的列反映了AI辅导员在任何给定时间实际可获得的信息，但是为了API性能的考虑，用户交互被分组在一起，而不是一次严格地显示单个用户的信息。 一些问题将出现在隐藏的测试集中，而训练集中未出现这些问题，从而模拟了快速适应新引入的问题的建模挑战。 像往常一样，它们的元数据仍在question.csv中。

previous_group_responses（字符串）以该组第一行中列表的字符串表示形式提供前一个组的所有user_answer条目。 每个组中的所有其他行均为空。 如果您使用的是Python，则可能需要在非空行上调用eval。 有些行可能为空，也可能为空列表。

previous_group_answers_correct（字符串）为上一组提供所有Answer_correctly字段，格式和警告与先前的group_responses相同。 有些行可能为空，也可能为空列表。

数据API

时序API详细信息

请参阅入门笔记本，以获取有关如何完成提交的示例。时间序列API与以前的竞赛相比有所变化!

您不应尝试为包含讲座的行提交任何内容。

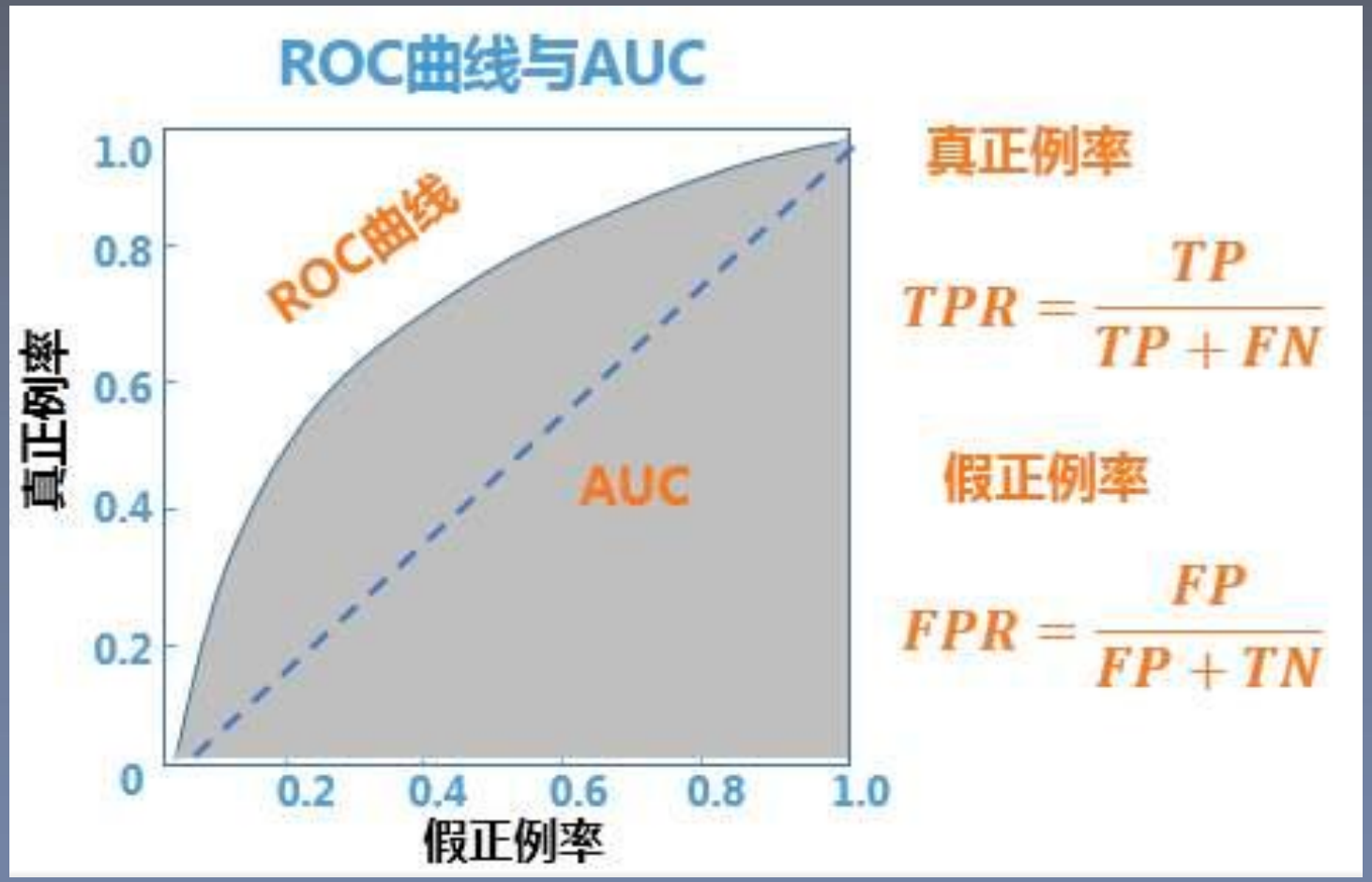
该API按发生的顺序提供用户交互组。每个组将包含来自许多不同用户的交互，但是来自任何单个用户的问题的总数不得超过task_container_id。每个组有1到1000个用户。

预计在隐藏的测试集中将看到大约250万个问题。

初始化后，API将在内存中最多加载1 GB的测试集数据。初始化步骤（env.iter_test（））实际需要更多的内存;我们建议您在调用之后才加载模型。该API还将花费大约15分钟的运行时间来加载和提供数据。

API使用上面指定的类型加载数据（对于user_id，int32；对于content_type_id，int8，等等）。

评价指标



提交：

ROC下曲线面积

AUC

注意事项

参赛者必须通过Notebooks提交　注：本次比赛不要求在Notebooks训练

为了使提交后的“提交竞赛”按钮生效，必须满足以下条件:

- CPU Notebook <= 9 hours run-time
- GPU Notebook <= 9 hours run-time
- TPU Notebook <= 3 hours run-time
- 允许免费和公开的外部数据，包括预先训练的模型
- Submission file must be named submission.csv
- 请查看Code Competition FAQ了解关于提交的更多信息。

机器配置

Competition Introduction

机器配置：

✓ 传统机器学习（CPU机器）：**16G 3Cpu**

✓ 深度学习（GPU机器）：**1080Ti**

✓ **所有深度之眼比赛AI年度会员，将提供一键运行的GPU环境；**

✓ **赠送：5000BDC**

深度之眼
deepshare.net

5、Baseline思路介绍

Baseline

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

Baseline思路介绍

深度之眼
deepshare.net

Baseline选型

①传统模型

Lgb模型

②深度模型

Nffm模型

免费获取：

Lgb开源提升比赛方案

报名课程获取：

Lgb传统特征模型比赛

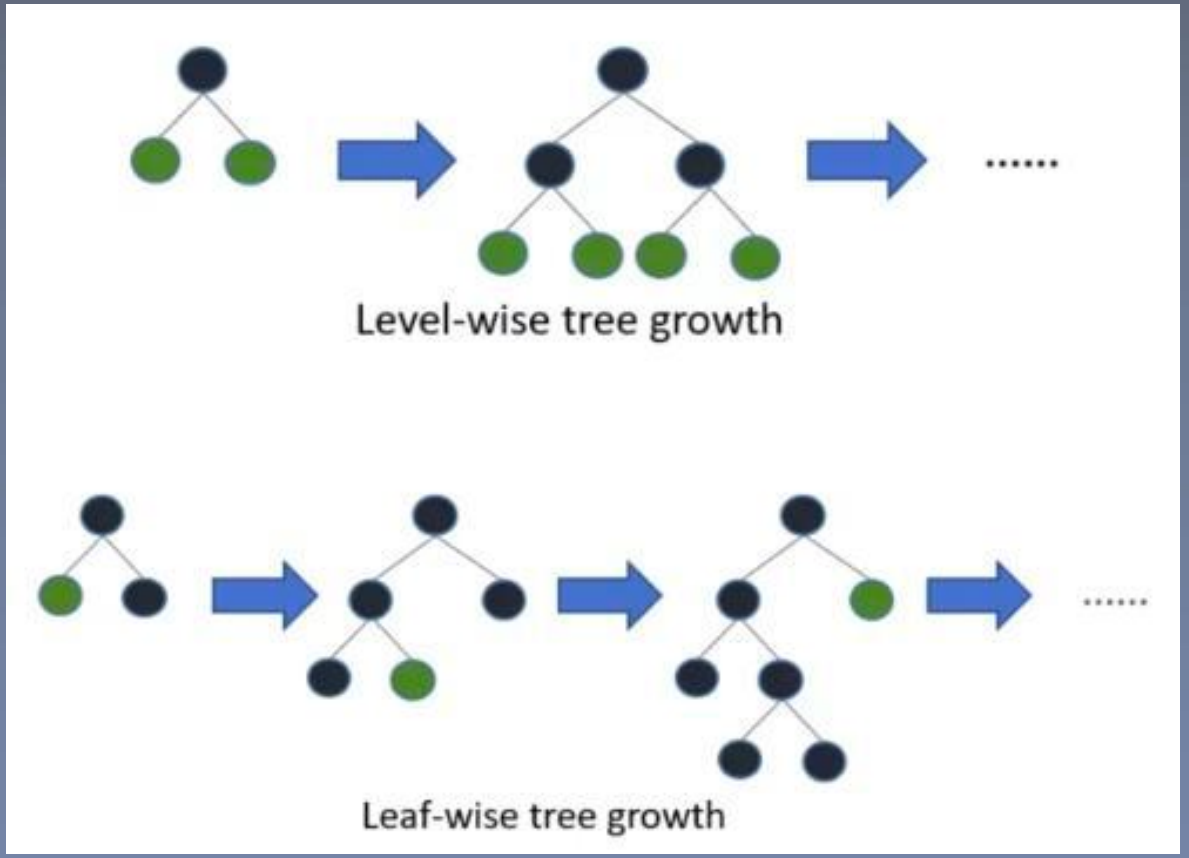
Nffm深度模型比赛

Baseline思路介绍

深度之眼
deepshare.net

① 传统模型

特征工程: 统计特征

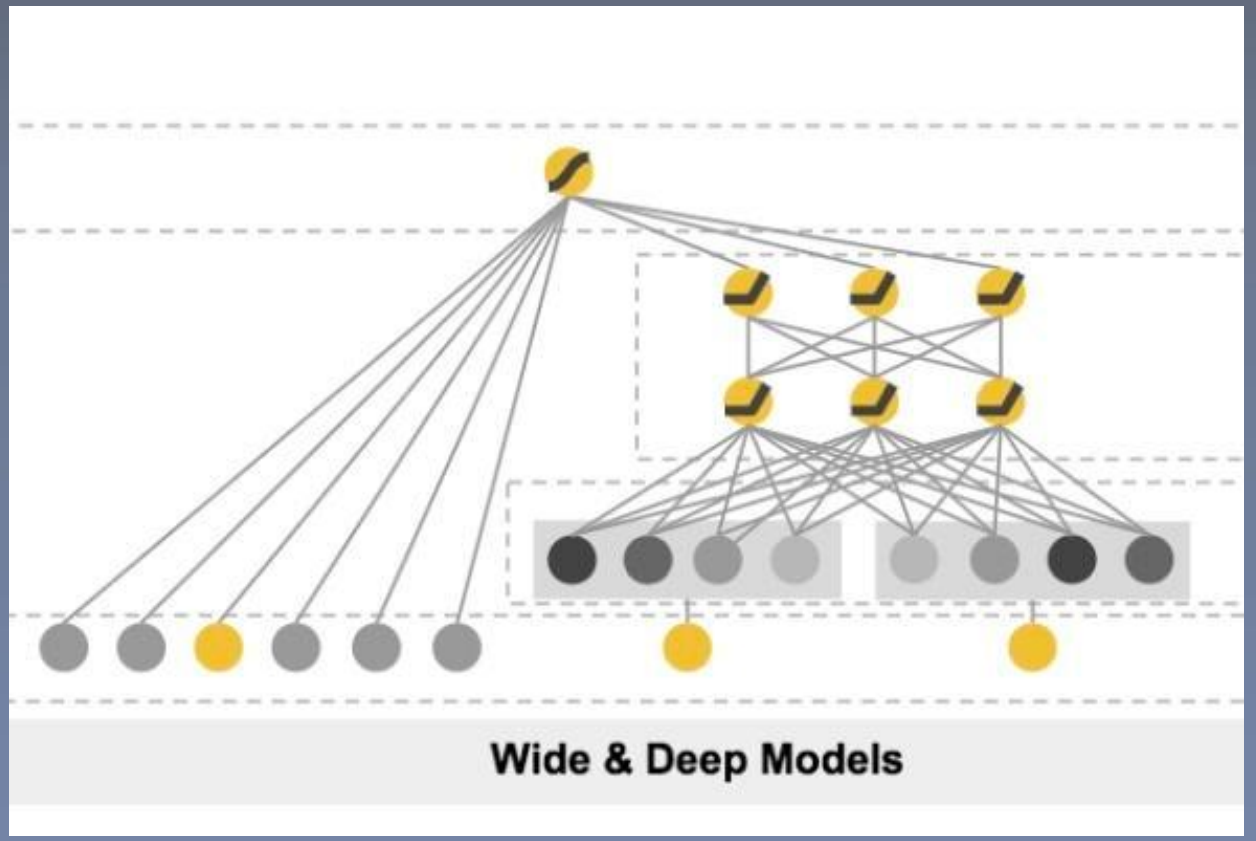


Baseline思路介绍

深度之眼
deepshare.net

② 深度模型

模型结构: ctr类推荐深度模型



Baseline思路介绍

个人建议：

深度模型

深度模型抽取特征

模型融合

深度之眼
deepshare.net

5、互动时间

限时优惠最后一天！
扫码即可报名本次比赛
今晚仅限50个名额！

扫码，回复“Kaggle”
加入直播讨论群
获取Baseline&课件

Q&A

Ask me anything

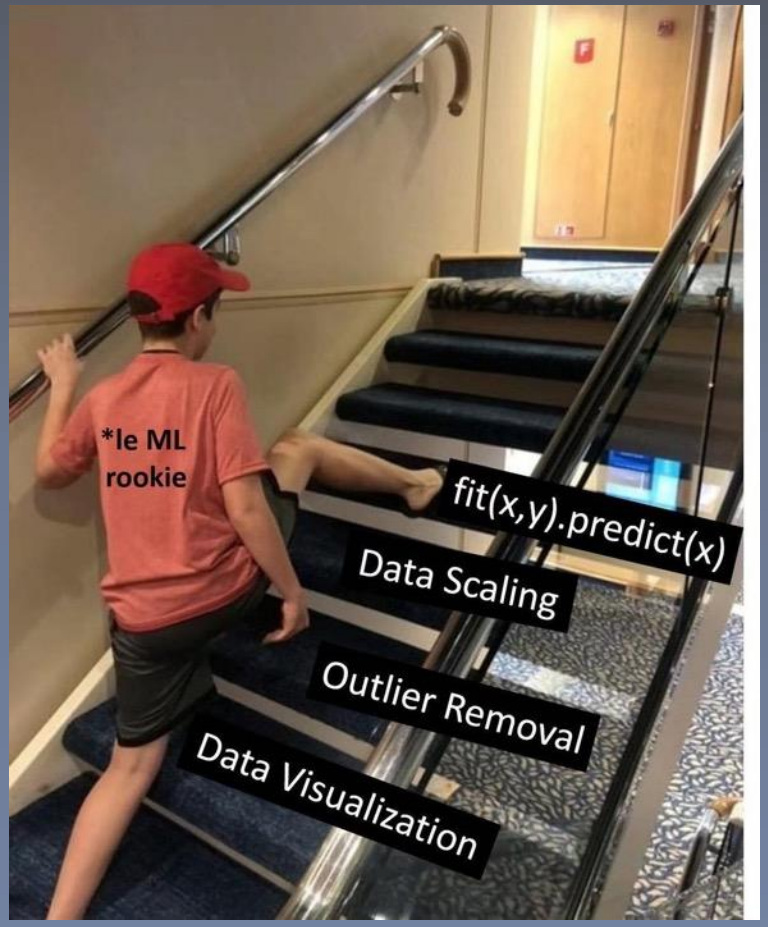
深度之眼
deepshare.net

■ 小白如何入门算法竞赛?

如果你想从零学习竞赛，可能会遇到以下问题:

- ✓ 对Python和操作系统不太熟悉;
- ✓ 对机器学习理论和库使用不太熟悉;

强烈建议参加我们的课程，低难度无门槛进入赛圈!



你能从本次课程中获得什么

深度之眼
deepshare.net

Supplementary teaching

比赛前

赛题指导
+
Baseline思路

比赛中

针对trick的提分思路

不同方案如何进一步优化

针对赛题前沿Paper+资料解析

比赛后

复盘优秀方案+思路

工程&面试应用技巧

你能获得什么

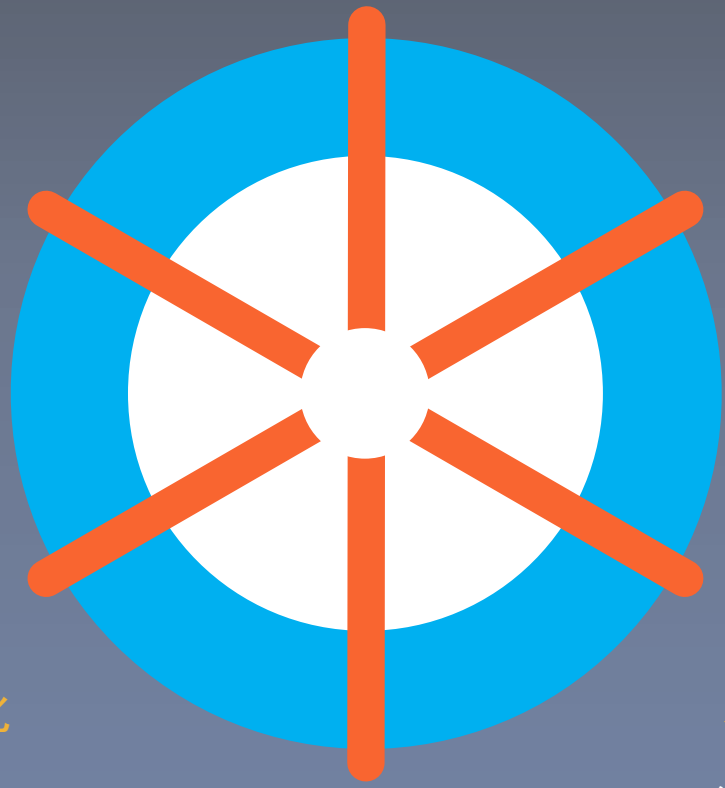
Supplementary teaching

【Kaggle大赛】回答准确性预测竞赛
限时限量课程优惠券
只有50张！速扫码！

深度之眼
deepshare.net

赛题指导+Baseline思路

深入解析比赛对应场景，数据，让大家站在出题立场看问题，深入理解老师给出的baseline思路，为什么要这样写，这样写好在哪里，在这个基础上还有什么提升的点。在参加比赛时站在更好的起跑线上。



针对赛题前沿Paper+资料解析

拒绝调包侠，通过前沿paper及重要理论的算法原理，不仅让你知道这个方法可以行，更要让你知道背后行的道理。

针对trick的提分思路

根据赛题方向提供通用的提升思路，举一反三，帮助大家不仅在此次比赛中提分，还可以把这些方法迁移到其他比赛中。

工程&面试应用技巧

除了提分思路分享，老师会分享工业和比赛的异同，比赛思路如何用于工业项目中，比赛经验在面试中的展现。帮助大家在工作或求职中更好地展现自己。

不同方案如何进一步优化

先锋队对比赛提前探索，提取Trick。由老师分析每个Trick背后的意义，针对不同Trick，提供下一步优化的方案。

复盘优秀方案+思路

复盘比赛全流程，梳理提分思路的完整逻辑，并对好的开源方案做案例分析。

如何学习AI竞赛？

How to learn AI competition?

深度之眼
deepshare.net

会员打包价仅需1498
领券还能立减200

注：已购买Kaggle比赛的同学还可返学费！

Step0：选修知识

- 数学基础
- Python基础
- 图像基础
- NLP基础

深度方向

解决**基础不牢固**
替你**查漏补缺**

Step1：参加经典赛练习

四大方向+十二场经典赛

- 数据科学
- NLP方向
- CV方向
- 综合方向

按照个人学习能力和技术深度，设计了不同阶段课程，带你**层层提升**。

https://ai.deepshare.net/all/3279059

Step2：参加进行的新比赛

Kaggle
DC竞赛 www.dcjingsai.com
TIANCHI天池
DataFountain
Kesci

轻松入门CV /NLP
扎实**细分领域**

Step3：上TOP

拿奖金
奖励/内推/实习

PS欢迎来当讲师（长期跪舔TOP大神）

结 语

爱折腾的你，追逐你的梦想！

深度之眼
deepshare.net

联系我们：

电话：18001992849

邮箱：

Q Q： 2677693114

公众号

客服微信