

大数据比赛汇总

杨福康*

2020 年 4 月 12 日

1 unstructured

1. 零基础入门数据挖掘 - 二手车交易价格预测

- adress : <https://tianchi.aliyun.com/competition/entrance/231784/introduction>
- time: 2020
- top solution:

2. 离散制造过程中典型工件的质量符合率预测

- <https://www.datafountain.cn/competitions/351>
- 赛事方向：分类预测、数据挖掘
- 赛事简介：由于在实际生产中，同一组工艺参数设定下生产的工件会出现多种质检结果，所以我们针对各组工艺参数定义其质检标准符合率，即为该组工艺参数生产的工件的质检结果分别符合优、良、合格与不合格四类指标的比率。相比预测各个工件的质检结果，预测该质检标准符合率会更具有实际意义。本赛题要求参赛者对给定的工艺参数组合所生产工件的质检标准符合率进行预测。
- 方案分享:https://mp.weixin.qq.com/s/MYbQ_J0XQwA2zs9VxO1IEA
- 代码: <https://github.com/CcIsHandsome/-TOP1->

3. 乘用车细分市场销量预测

*E-mail:1766084780@qq.com

- <https://www.datafountain.cn/competitions/352>
- 赛事方向：预测回归、数据挖掘
- 赛事简介：本赛题需要参赛队伍根据给出的 60 款车型在 22 个细分市场（省份）的销量连续 24 个月（从 2016 年 1 月至 2018 年 12 月）的销量数据，建立销量预测模型；基于该模型预测同一款车型和相同细分市场在接下来一个季度连续 4 个月份的销量；除销量数据外，还提供同时期的用户互联网行为统计数据，包括：各细分市场每个车型名称的互联网搜索量数据；主流汽车垂直媒体用户活跃数据等。参赛队伍可同时使用这些非销量数据用于建模。除了模型的准确性外，参赛队伍需对本赛题任务有系统性的思考和设计，在决赛阶段，参赛队伍对于所提交的模型的适应性、可扩展性、代码的工程性等方面也会影响参赛队伍的最终名次。
- 方案分享：<https://mp.weixin.qq.com/s/-tT9BKrANTwJK9-N1K4j9g>
- 代码：https://github.com/cxq80803716/2019-CCF-BDCI-Car_sales

4. 混凝土泵车砼活塞故障预警

2 graph

1. 基于 OCR 的身份证要素提取

- 赛事方向：分类预测、数据挖掘
- adress: <https://www.datafountain.cn/competitions/346>
- 赛事简介：设计针对商业银行身份证识别的 OCR 系统，识别身份证中姓名、地址、身份证号码和身份证有效日期等信息。
- 方案分享：https://mp.weixin.qq.com/s/_Vxg5zRAzeUNoF9N3ungXg

2. 多人种人脸识别

- 赛事方向：
- adress:
- 赛事简介：
- 方案分享：

3.
 - 赛事方向:
 - adress:
 - 赛事简介:
 - 方案分享:

3 NLP

1. 互联网新闻情感分析

- 赛事方向: 自然语言处理、机器学习
- adress:
- 赛事简介: 参赛者需要对我们提供的新闻数据进行情感极性分类, 其中正面情绪对应 0, 中性情绪对应 1 以及负面情绪对应 2。根据我们提供的训练数据, 通过您的算法或模型判断出测试集中新闻的情感极性。
- 方案分享:https://mp.weixin.qq.com/s/SgkQB7t0j2_kqHeotspHBQ

2. 金融信息负面及主体判定

- 赛事方向: 情感识别、自然语言处理
- adress:
- 赛事简介: 该赛题分为两个子任务: 给定一条金融文本和文本中出现的金融实体列表, 负面信息判定: 判定该文本是否包含金融实体的负面信息。如果该文本不包含负面信息, 或者包含负面信息但负面信息未涉及到金融实体, 则负面信息判定结果为 0。负面主体判定: 如果任务 1 中包含金融实体的负面信息, 继续判断负面信息的主体对象是实体列表中的哪些实体。
- 方案分享:https://mp.weixin.qq.com/s/df8uxXenu_gRLjbKIwm12g