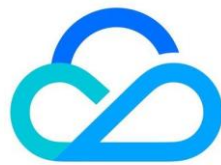


TI-ONE 技术白皮书



腾讯云

目录

1. 产品概述	4
2. 产品定位及用户价值	5
2.1. 产品定位	5
2.1.1. 目标客户	5
2.1.2. 需求与痛点	5
2.2. 用户价值	6
2.2.1. 技术价值	6
2.2.2. 业务价值	7
3. 产品架构及功能	8
3.1. 产品架构	8
3.2. 产品功能	9
3.2.1. 数据服务功能	9
3.2.2. 算法服务功能	9
3.2.3. 模型服务功能	10
3.2.4. 平台管理功能	10
3.2.5. 资源调度功能	11
3.2.6. 系统容灾与恢复	12
3.2.7. 系统安全	12
4. 产品优势	13
4.1. 功能优势	13
4.1.1. 拖拽式任务流设计	13
4.1.2. 内置丰富机器学习算法	13
4.1.3. 多实例调度	13
4.1.4. 便捷的效果可视化	14

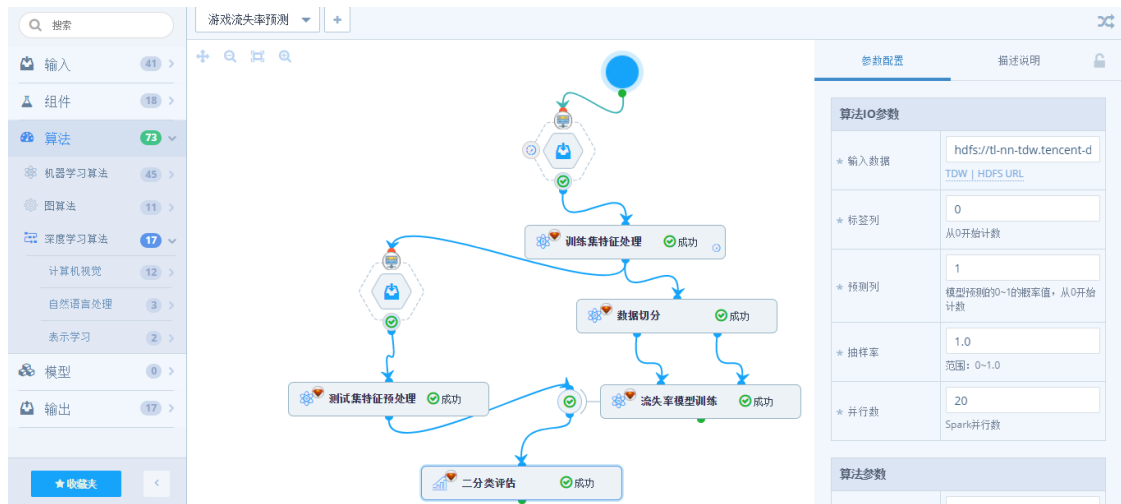
4.1.5.	支持多种机器学习框架	14
4.1.6.	强大的团队写作和分享	14
4.1.7.	模型的完整闭环	14
4.2.	性能优势	15
4.2.1.	批量参数调优	15
4.2.2.	资源利用	15
4.2.3.	多机多卡	15
4.2.4.	通讯加速	15
4.2.5.	高性能参数服务器	16
5.	应用场景及案例	16
5.1.	应用场景	16
5.1.1.	风险控制场景	16
5.1.2.	精准营销场景	17
5.1.3.	各类预测场景	17
5.1.4.	非结构化数据处理	18
5.1.5.	文本分析场景	18
5.1.6.	关系挖掘场景	19
5.2.	案例分析	19
5.2.1.	场景介绍	19
5.2.2.	数据准备	19
5.2.3.	搭建实验流程	20
5.2.4.	模型评估	22

1. 产品概述

智能钛机器学习（TI Machine Learning, TI-ML）是基于腾讯云强大计算能力的一站式机器学习生态服务平台。

它是专为 AI 初学者设计的一站式机器学习平台。其可视化的操作界面（如下图）、具象化的算法结果、拖拽式的任务流、可灵活自定义的特性，以及内置的丰富模型算法与案例。它能够对各种数据源、组件、算法、模型和评估模块进行组合。让 AI 初学者都能在其之上能够方便地进行模型训练、评估和预测，并运用机器学习技术解决各场景的问题。

智能钛机器学习适用于作为 AI 科研教学平台（如企业、科研机构、高校等），并提供顶级人工智能平台并开放接口与能力。为企业、高校、科研机构，开展科技企业培训、AI 能力提升、公共服务优化等多层次合作提供技术支撑。进一步帮助有效带动区域人工智能产业快速发展。



图：智能钛机器学习平台平台界面

2. 产品定位及用户价值

2.1. 产品定位

2.1.1. 目标客户

TI-ONE 的目标客户主要包括以下类型：

- 经验较浅的机器学习玩家，例如刚开始学习人工智能建模的初学者、爱好者、学生。
- 企业用户 - 想要尝试、体验 Tesla 的客户部门；企业 IT 部门需要为将来各个业务部门提供 AI 平台能力；因监管、公司战略需求采购 ML 平台。
- 开设 AI 课程的院校、算法大赛等。

2.1.2. 需求与痛点

TI-ONE 用户的主要需求包括以下几点：

- 尝试体验，想了解机器学习是怎么回事或入门
- 想从事 AI，但缺乏模型构建能力，需要各类现成的算法模型
- 需要提供各类 AI 能力，但当前需求仍未知，需要储备能力

此类用户的痛点包括：

- 难以快速入门机器学习
- 缺乏模型构建能力
- 在需求未知的情况下要准备 AI 能力

2.2. 用户价值

智能钛机器学习平台的客户价值包含技术价值和业务价值两个方面：

2.2.1. 技术价值

- 上手门槛低，用户界面友好。即使 AI 初学者也能迅速上手。适用于作为 AI 科研教学平台（如企业、科研机构、高校）等。
- 提供顶级人工智能平台并开放接口与能力。为服务企业、高校、科研机构，开展科技企业培训、AI 能力提升、公共服务优化等多层次合作提供技术支撑。

- 云端的高可用 GPU 分布式集群服务器，满足大规模深度学习模型训练对性能的要求且可随时购买随时使用；
- 基于 GPU 的分布式机器学习平台，兼容 tensorflow、caffe 等主流开源机器学习框架，实现底层框架管理对算法实现的透明化；
- 用户可在平台上灵活地定义模型和算法模块。
- 智能钛机器学习平台对 GPU 分布式集群服务器上的深度学习模型训练算法进行优化，能够大幅提升训练速度，从而缩短模型训练的时间。与原生 tensorflow 框架相比，智能钛机器学习平台在 128 块 GPU 卡的集群上的训练速度能提高 1~3 倍；

2.2.2. 业务价值

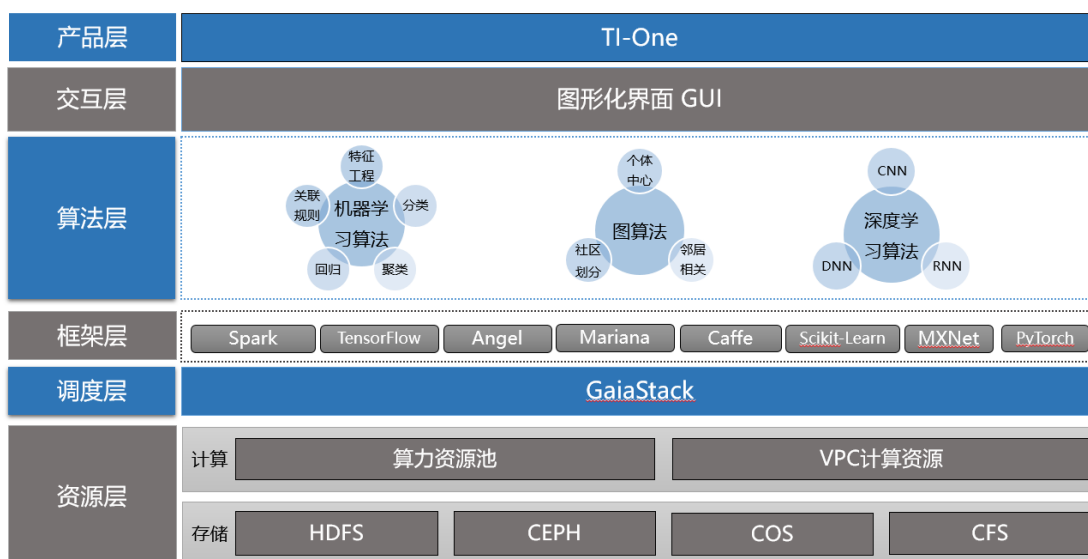
- 适用于作为 AI 科研教学平台（如企业、科研机构、高校等）。帮助有效带动区域人工智能产业快速发展，加快实现新旧动能转换，服务区域软件行业建设；
- 利用已有数据，几乎能在智能钛机器学习平台上训练出用户所需的大部分模型。结合用户实际需求，如医疗、教育、安防等领域大数据资源，培育人工智能企业、帮助企业做大做强。

- 在智能钛机器学习平台上建模，用户可以节省搭建机器学习平台和管理物理资源的时间，把精力聚焦在更有业务价值的建模工作上。

3. 产品架构及功能

3.1. 产品架构

腾讯云机器学习平台的基础架构如下图所示。



说明如下：

- 储存层：腾讯云 COS 储存等
- 调度层：GaiaStack
- 框架层：包括 Tensorflow, PyTorch, Caffe 等多种框架。
- 算法层：包括 CNN、RNN、DBN 等深度学习算法，以及传统机器学习算法等。
- 交互层：图形化界面 GUI

3.2. 产品功能

3.2.1. 数据服务功能

TI-ONE 机器学习平台提供以下数据服务功能：

- 平台数据管理：支持 hdfs、视频、音频等多种数据格式的导入和处理，以及维护管理等。IO 黑盒化及组件 IO 封装，且中间结果可另存。此外，还支持可视化数据透视。
- 数据集预处理：支持 sql 语法处理数据集。
- 特征工程：支持特征提取、特征转换和特征选择等处理。

3.2.2. 算法服务功能

TI-ONE 为用户提供多种算法功能：

- 平台免编程算法：TI-ONE 不仅提供分类、回归、无监督学习、特征提取、数据预处理、模型评估、深度学习等传统算法，还提供部分拥有独立知识产权的专利算法，且算法框架向下兼容。
- 自主编程功能：平台支持多种语言交互式编程环境，如 Scala、R、SQL 等，还提供深度学习、机器学习的调试环境。
- 内置镜像：平台内置多种机器学习和深度学习框架的标准

镜像，可引入外部主流算法包，用户可以直接使用，如 TensorFlow、torch、pytorch、caffe2 等。

- 乐高化自定义算法：支持用户进行算法和模型的发布，算法和模型可按版本发布，算法和模型的发布可以控制分享粒度和级别。

3.2.3. 模型服务功能

TI-ONE 在模型的训练、评估、应用等过程中分别提供以下模型服务功能：

- 模型训练：支持对算法的训练参数与运行参数进行调整。支持批量参数调优，支持参数集功能，一键设置多种参数组合进行调优，提高训练效率。
- 模型评估：支持模型训练评估的主要衡量指标体系，包括 ROC、混淆矩阵等。TI-ONE 还可以根据不同类型算法提供评估指标。此外，深度学习训练过程支持动态指标的可视化展现。
- 模型应用：模型服务 saas 化，支持模型一键部署和自定义部署。模型可发布为 API 服务，还可管理 API 的调用状况、性能指标等。

3.2.4. 平台管理功能

TI-ONE 机器学习平台提供以下平台管理功能：

- 工程管理：具备完整的模型维护管理流程，比如模型保存、下载、导入等。具备完整的建模工程管理流程，比如工程保存、分享等。
- 系统运维管理：支持可视化界面对平台数据集的资源管理、权限管理、信息维护管理；以及对平台各深度学习、机器学习任务进程进行监控和管理。支持实时可视化展示系统任务状态和集群资源状态、还可对集群资源状态定时采样，并记录建模过程各环节详细日志，并提供图形化查看。此外，还可满足用户维度的运营需求。
- 用户管理：提供对平台操作的审计管理功能。对于企业用户，TI-ONE 支持用户主-子账号模式，如主账号记账、子账号使用。
- 团队协作：提供节点收藏、模型收藏功能，收藏后可直接复用。支持工作流、工程、模型的分享并可进行不同级别的授权，并支持多人关注同一工作流，进行群组监控。此外，可通过乐高化自定义算法功能，高级用户间共享算法研究成果。

3.2.5. 资源调度功能

TI-ONE 支持 CPU 和 GPU 计算资源，通过 GaiaStack 进行资源调度。GaiaStack 赋予 TI One 很多商业特性。当用户数据量较大时，GaiaStack 提供多个完整集群，当用户的

数据比较小的时候，可以多个用户共享集群。TI-ONE 对用户训练和推理使用的资源分离管理，资源配置可按优先级控制排序。此外，在线离线混部，支持热升级，业务不中断，用户无感知。支持主备自动切换的高可用，当服务量增大时，会自动加载新的实例，并自动负载均衡。

3.2.6. 系统容灾与恢复

TI-ONE 具有完备的系统容灾与恢复机制。用户任务可运行在多集群多 IDC，同时，系统具备断点恢复功能，可以保存任务中断及停止后的状态，并在恢复后可按现场继续执行。此外，TI-ONE 支持自动化安装部署，能够在 2 台物理机上实现最小化平台安装，满足基本的建模、推理应用。

3.2.7. 系统安全

TI-ONE 机器学习平台具有安全性和稳定性。一方面，我们在架构设计中考虑平台中模型、算法、数据的安全控制方案，有效的防止模型、算法、数据的非法访问和下载，防止数据泄露。另一方面，TI-ONE 系统可对后台异常按严重程度分级，不严重的异常后台自动重跑解决，严重异常及时抛出到前台告知用户，常见的抛出异常（如空间不足、路径不存在等）在抛出时会转换成用户容易理解的错误信息。此外，用户还可通过 TI-ONE 可视化控制台，看到自己的任务日志。

4. 产品优势

4.1. 功能优势

4.1.1. 拖拽式任务流设计

智能钛机器学习平台采用拖拽式任务流设计，各元素可以自由地通过可视化拖拽，“搭积木”般地绘制任务流，大数据、小模型在其中融合、灵动，让 AI 初学者也能轻松快速地上手。

4.1.2. 内置丰富机器学习算法

智能钛机器学习平台内置机器学习算法，通过拖拽就能灵活使用，各类丰富的算法包含特征工程、机器学习、深度学习、图算法等，能充分支持用户在不同场景和数据情况下的需求。

- 特征工程：离散、归一、降维...
- 非深度学习：回归、聚类、分类...
- 深度学习：CNN、DNN、RNN...

4.1.3. 多实例调度

智能钛机器学习平台支持四种实例调度方式：手工、定时、批量参数、重跑，能方便用户在各场景需求下灵活调度的需求，降低手工调度的次数与时间成本。

4.1.4. 便捷的效果可视化

智能钛机器学习平台借助强大的 Python\Spark，集结流行的可视化效果，为用户提供科学美观的数据可视化服务，且模型训练效果直接悬浮呈现，用户无需点击即可直观方便地辨别模型的质量，判断优化的方案。

4.1.5. 支持多种机器学习框架

小到 Python、R，大到当红的 Spark、TensorFlow、xgBoost (DMLC)，从准流式计算的 streaming 系列到图算法，以及腾讯自研的 Angel 都灵活可供使用，以满足不同用户对不同框架的偏好习惯。

4.1.6. 强大的团队写作和分享

面向算法工程师，支持工程分享、协同调试、任务收藏复用，亦能自助配置/发布自研精妙算法，贡献至集团算法库，为子公司或部门提供赋能。

4.1.7. 模型的完整闭环

智能钛机器学习平台为用户提供“一站式”的机器学习平台体验，从模型训练、评估、服务部署到在线推理，覆盖全工作流程，形成机器学习训练的完整闭环。

4.2. 性能优势

4.2.1. 批量参数调优

智能钛机器学习平台支持参数集功能，一键设置多种参数组合进行调优，大大提高您的训练效率。

4.2.2. 资源利用

智能钛机器学习平台通过队列管理资源和数据，队列之间实现数据的隔离，可灵活支持共享资源和专用资源方案。而且，通过削峰填谷、任务排队和失败重试机制，使计算资源达到最大化利用，理想情况下能够进行 7*24 任务训练，提高集群资源的整体利用率。

4.2.3. 多机多卡

智能钛机器学习平台深度学习支持多机多卡调度 (tensorflow)，突破单机限制。

4.2.4. 通讯加速

智能钛机器学习平台采用 rmda 网络通信替换 tensorflow 原生的 gRPC 通信，使 Tensorflow 的集群间通讯速度更快 (50% 提升)。

4.2.5. 高性能参数服务器

智能钛机器学习平台搭载腾讯自研的高性能分布式机器学习组件 Angel，基于参数服务器架构，将高维度的大模型切分到多个参数服务器节点，并通过高效的模型更新接口和运算函数，以及灵活的同步协议，轻松实现各种高效的机器学习算法，具有处理 T 级大数据和训练十亿级维度大模型的能力。

5. 应用场景及案例

5.1. 应用场景

智能钛机器学习平台适合所有需要使用机器学习或深度学习平台进行定制建模的场景。TI-ONE 的用户涵盖了金融、教育、政府、医疗、商业等多个领域，TI-ONE 在以下多个场景中为各领域用户提供帮助及创造价值。

5.1.1. 风险控制场景

金融机构存有大量与风险有关的高质量数据，且有从这些数据中挖掘规律的需求。从贷前的额度审批、贷中的交易反欺诈到贷后的催收都可以使用机器学习提高风控的时效性、准确率和覆盖率，从而减小金融机构的风险损失和管理成本，甚至带来直接收益。

典型的风控场景包括信用卡交易欺诈预测、用户信用评级等。

5.1.2. 精准营销场景

营销推荐本质上是根据历史成交数据预测特定场景下客户和商品最优匹配，从而实现客户价值最大化，通过对历史数据的学习，机器能够预测客户和商品的匹配概率，并预测每种匹配带来的预期客户价值，从而实现提升营销效果、降低营销成本、挖掘潜在客户、实现交叉销售、最大化客户价值等目的。

典型的精准营销场景包括广告投放、商品推荐等。

5.1.3. 各类预测场景

预测问题是指使用历史数据通过建模对未来做出预测，在政府、医疗、投资机构等均有所应用。预测的结果可能是类别或数值，这类问题非常适合使用机器学习训练分类或回归模型来解决。

典型的预测场景包括天气预测、交通路况预测、健康状况预测、投资收益预测等。

5.1.4. 非结构化数据处理

非结构化数据是针对结构化数据而言，指没有以行的形式保存在关系型数据库中的数据，典型的非结构化数据包括图像、音频等，其中图像包括物体、人、人脸、指纹、虹膜等类别，音频包括语音、音乐、非语音音频等类别。此类数据的处理方式包括非结构化数据的超分辨率处理、从非结构化数据中提取结构化信息等。

典型的非结构化数据处理场景包括图像识别、语音识别等。

5.1.5. 文本分析场景

文本分析主要指需要从文本、后台日志、接口日志等半结构化数据中提取结构化信息的问题，例如语义分析、情绪情感识别等。传统的模板、关键词等解决方案已无法满足用户需求，深度学习方法越来越多的应用到文本分析中。

典型的文本分析场景包括文本分类、舆情分析、人机交互的自然语言入口、搜索引擎等。

5.1.6. 关系挖掘场景

关系挖掘主要是对图关系数据进行处理，包括对节点和关系两类数据进行挖掘和分析，从而发现隐藏在图关系中的规律，由于数据量较大且数据间规律不明显，故使用机器学习方法来解决此类问题。

典型的关系挖掘场景包括用户画像、社交关系分析、社交风险控制等。

5.2. 案例分析

5.2.1. 场景介绍

科比投篮预测：科比在 2016 年 4 月 12 日星期三的最后一场比赛中为洛杉矶湖人队得到 60 分，标志着他从 NBA 退役。从 17 岁入选 NBA，科比在他的职业生涯中获得了无数的赞誉。使用 20 年有关科比的投篮命中和投失数据，你能预测哪些投篮能命中篮筐吗？本案例非常适合实践分类基础知识、特征工程和时间序列分析。训练数据包含了科比在他 20 年职业生涯中所尝试的每个投篮命中的位置和情况。我们的任务是训练模型来预测每次投篮是否会投进。

5.2.2. 数据准备

字段名	取值类型	取值举例
action_type	枚举	Jump Shot Running Jump Shot Layup Shot Reverse Dunk Shot Slam Dunk Shot

combined_shot_type	枚举	Jump Shot Layup Dunk
game_event_id	数值	1 2 3 ……
game_id	数值	20000012 20000047 ……
lat	数值	33.9343 34.0163 ……
loc_x	数值	-157 138 0 ……
loc_y	数值	175 -11 0 ……
lon	数值	-118.1028 -118.2938 ……
minutes_remaining	数值	0 1 2 3 4 5 6 7 8 9 10 11
period	数值	1 2 3 4 5 6 7
playoffs	数值	0 1
season	字符串	2004/5/1 2007/8/1 2015-16 ……
seconds_remaining	数值	0~59
shot_distance	数值	0 45 79 ……
target	label	0 1
shot_type	枚举	2PT Field Goal 3PT Field Goal
shot_zone_area	枚举	Right Side(R) Left Side(L) Left Side Center(LC) Right Side Center(RC) Center© Back Court(BC)
shot_zone_basic	枚举	Mid-Range Restricted Area In The Paint (Non-RA) Above the Break 3 ……
shot_zone_range	枚举	16-24 ft. 8-16 ft. Less Than 8 ft. 24+ ft. ……
team_id	数值	1610612747
team_name	字符串	Los Angeles Lakers
game_date	日期	2000/10/31 2014/12/21 ……
matchup	字符串	LAL @ IND LAL @ SAC LAL vs. UTA LAL vs. SAC ……
opponent	字符串	SAC PHX ……
shot_id	数值	1 2 3 4 5 ……

5.2.3. 搭建实验流程

整体流程如下图所示：

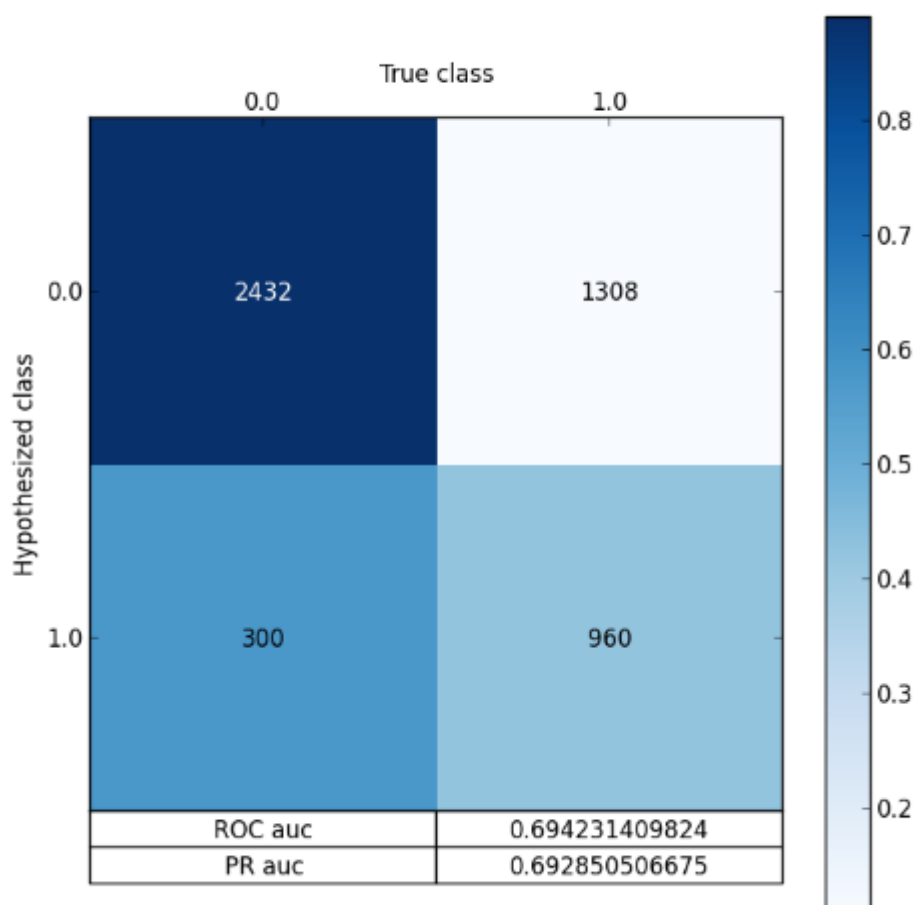


流程共包含 8 个环节，分别是：

1.	将训练数据从本地上传到 cos
2.	对训练数据做特征处理
3.	将特征处理后的训练数据切分成训练集和测试集
4.	训练 kobe 投篮预测模型
5.	将验证数据从本地上传到 cos
6.	对验证数据做特征处理
7.	在验证数据集上验证 kobe 投篮预测模型
8.	生成模型评估结果

5.2.4. 模型评估

TI-ONE 内置多种可视化与模型评估组件，方便用户快速辨别模型的质量，并进行优化，拖拽组件即可使用。单击评估节点即可看到模型评估的可视化结果，效果直接悬浮呈现，模型优劣，一眼立断。本模型采用混淆矩阵来评估效果，结果如下：





腾讯云

TI-ONE 体验地址: <https://cloud.tencent.com/product/TI>

欢迎试用!