

GLEAM 大学生大数据竞赛平台

项目远景说明书

版本：0.1

主要作者

朱泓宇

Hongyu Zhu

2015010494

aglax2357@icloud.com

目录

- 1. 文档修订 - 3 -
- 2. 简述 - 4 -
- 3. 定位 - 4 -
 - 3.1. 市场需求 - 4 -
 - 3.2. 问题描述 - 4 -
 - 3.2.1. 比赛信息 - 4 -
 - 3.2.2. 用户管理 - 5 -
 - 3.2.3. 比赛流程 - 5 -
 - 3.2.4. 用户交流 - 5 -
 - 3.3. 产品定位 - 6 -
- 4. 涉众和目标用户 - 6 -
 - 4.1. 涉众概要 - 6 -
 - 4.2. 用户环境 - 7 -
- 5. 系统特性 - 7 -
 - 5.1. 基础 - 7 -
 - 5.2. 数据下载 - 7 -
 - 5.3. 结果提交 - 7 -
 - 5.4. 评测 - 8 -
 - 5.5. 反作弊 - 8 -
- 6. 优先级 - 8 -
 - 6.1. 第一阶段 - 8 -
 - 6.2. 第二阶段 - 9 -
- 7. 其他产品需求 - 9 -
 - 7.1. 性能要求 - 9 -
 - 7.2. 多语言支持 - 9 -
 - 7.3. 推广 - 9 -
 - 7.4. 奖金和证书 - 10 -

1. 文档修订

| 版本 | 主要作者 | 简述 | 完成时间 |
|-----|------|-------------------------------|------------|
| 0,1 | | 对项目整体进行分析，确定了项目的需求愿景，本文档撰写完成。 | 2017/10/22 |

2. 简述

本文档旨在对大学生大数据平台 Glean 的定位、目标用户、系统特性、开发优先级等信息进行阐述，以方便用户和项目未来的维护者了解平台的基本信息。

Glean 是一个以大学生为主要参赛方，以学校和企业为主要比赛发布方，以大数据比赛为主要内容，覆盖从比赛的创建报名到评审颁奖整个流程的竞赛平台。

Glean 大学生大数据竞赛平台由 Candle 小组开发。

3. 定位

3.1. 市场需求

机器学习方向的系统性研究早在半个世纪前就已经开始，但直到最近，其技术前提才被满足。计算机计算能力的飞升和数据体量的膨胀将机器学习带入大数据时代，这项古老的技术也开始真正的施展拳脚。

为鼓励大学生学习大数据知识，推广相关技术，越来越多的学校和企业开始组织面向大学生群体的大数据比赛。但现有的大数据比赛平台存在比赛不专一、管理混乱、用户体验差等问题。针对这些问题，我们计划开发只专注大数据比赛的 Glean 平台，方便使用者，弥补这一市场空缺。

3.2. 问题描述

3.2.1. 比赛信息

| | |
|----|---------------------------|
| 问题 | 比赛信息冗杂，除了大数据比赛，还有其他类型的比赛。 |
| 影响 | 多而不精，没有能够针对大数据比赛的特性设计平台。 |

| | |
|------|------------------------------------|
| 解决方案 | 明确平台的比赛类型，只做大数据比赛。 |
| 优势 | 提供针对大数据比赛的专业平台服务，让参赛方和赛事方享受到专业的服务。 |

3.2.2. 用户管理

| | |
|------|-------------------------------|
| 问题 | 用户管理混乱，对于用户的增删改查操作不方便。 |
| 影响 | 错误信息难以修改，用户违规行为得不到处理，平台公正性受损。 |
| 解决方案 | 完善用户管理系统，明确注册、修改、封禁、申诉等操作的流程。 |
| 优势 | 方便平台管理，维护公平。 |

3.2.3. 比赛流程

| | |
|------|---|
| 问题 | 比赛流程不明确，仅能设定开始时间和结束时间，不能增加重要节点。 |
| 影响 | 参赛者不清楚比赛流程，错过重要节点；组织者需要手动操作，工作量大，不准确。 |
| 解决方案 | 允许自定义比赛流程，将比赛流程明确显示在比赛信息页面上，在重要节点前夕通知参赛者，防止其错过截止日期。 |
| 优势 | 减少人为干扰，减轻比赛组织者负担。 |

3.2.4. 用户交流

| | |
|------|---|
| 问题 | 参赛方之间，参赛方和赛事方没有交流渠道，遇到问题难以解决，团体比赛组队不方便。 |
| 影响 | 赛事方的重要通知难以传达，参赛方的问题无法被解答，组队之后队员之间难以交流。 |
| 解决方案 | 搭建论坛。 |
| 优势 | 方便用户之间的交流，同一问题不必重复回答。 |

3.3. 产品定位

Gleam 产品描述

| | |
|------|--|
| 目标用户 | 对大数据感兴趣的大学生、想举办大数据比赛的学校或企业 |
| 简介 | 为大数据比赛提供一站式平台 |
| 产品理念 | 简洁、专业 |
| 主要功能 | 赛事方：比赛的创建\管理\维护、管理\处罚参赛者、颁奖、论坛 参赛者：比赛的参加\退出、修改个人信息、论坛 |
| 卖点 | 专注大数据比赛、提供专业服务、用户体验优秀的一站式比赛平台。 |

4. 涉众和目标用户

4.1. 涉众概要

| 涉众 | 描述 | 职责 |
|-----------|-------------------|--------------------------------------|
| Candle 小组 | 开发、管理 Gleam 平台的团队 | 对 Gleam 平台进行设计、开发、测试，与用户进行沟通，对申诉进行仲裁 |
| 参赛方 | 希望参加大数据比赛的在校大学生 | 参加比赛 |
| 赛事方 | 希望举办大数据比赛的学校和企业 | 管理比赛，回答选手问题 |

4.2. 用户环境

赛事方可以通过后台对比赛进行管理，处罚违规选手，用官方认证的账号在论坛中回答问题，查看用户分数排行。

参赛方参加和退出比赛，提交结果并查看名次，修改个人信息，参与论坛讨论，对处罚提出申诉。

5. 系统特性

5.1. 基础

- Gleam 的一些文本将会支持 Latex 语法，方便赛事方对比赛进行描述；

5.2. 数据下载

- 提供可高速下载的比赛数据；
- 整个平台上的所有比赛的数据在储存格式上保持统一，参加过之前比赛的选手不用再花时间熟悉数据储存格式；
- 提供针对各个架构 (Tensorflow、PyTorch 等) 的数据预处理代码，选手可以调用这些代码来获取数据；

5.3. 结果提交

- 可以高速提交结果，对于结果文件较大的，Gleam 还支持将大文件压缩成多个小文件上传，全部上传成功后再由平台解压，减少因网络不稳定造成的上传不成功。
- 整个平台上的所有比赛的结果在储存格式上保持统一，参加过之前比赛的选手不用再花时间熟悉数据储存格式；

- 提供针对各个架构 (Tensorflow、PyTorch 等) 的结果保存代码, 选手可以直接调用这些代码来保存结果;

5.4. 评测

- Gleam 提供多种常用模型评测函数供赛事方选择, 同时也支持自定义评测函数;
- 评测分为公开评测和非公开评测, 公开评测在提交后立刻公布评测结果, 而非公开评测仅在提交阶段结束之后才会统一公布结果, 只有后者才会被作为最终名次的评判依据;

5.5. 反作弊

- 平台注册时需要申请人提供身份证件 (如身份证、护照、台胞证等), 防止一人注册多个账户;
- 每一名选手在一场比赛中只能提交有限次, 且两次提交之间必须有一定的时间间隔;
- 参赛者需要同时提交结果和代码, 代码备份, 以备查验;
- 对于相邻两次提交结果差距较大的参赛者, 将会被放入 “特别关注” 列表, 供赛事方核对。

6. 优先级

首先完成核心系统特性, 然后在完善核心特性的同时添加新的特性。

6.1. 第一阶段

- 参赛方核心功能

注册、登录、查看比赛列表、查看比赛详情、参加比赛、下载数据集、提交结果、更新 排名、登出。

- 赛事方核心功能

赛事方能够注册、登录、登出、修改张数；创建比赛、修改比赛、上传数据集、提供评测方法、发布比赛、登出。

- 论坛核心功能

赛事方能够注册、登录、登出、修改张数；创建比赛、修改比赛、上传数据集、提供评测方法、发布比赛、登出

6.2. 第二阶段

- 参赛方进阶功能

搜索比赛、分享比赛、接收邮件提醒，查看已有荣誉

- 赛事方进阶功能

- 设置比赛节点、管理参赛者、发布通知、生成证书、发放奖金

7. 其他产品需求

7.1. 性能要求

能够应付大量的用户同时提交。

7.2. 多语言支持

支持多种主要大数据编程语言和架构。

7.3. 推广

能够通过多种社交平台（如：微博、微信朋友圈、QQ 空间）对比赛进行分享，推广。

7.4. 奖金和证书

能够根据比赛结果分发奖金；

能够自动生成电子证书。