

2021 中国高校计算机大赛——微信大数据挑战赛

竞赛规程

2016 年，教育部高等学校计算机类专业教学指导委员会、教育部高等学校软件工程专业教学指导委员会、教育部高等学校大学计算机课程教学指导委员会、全国高等学校计算机教育研究会联合创办了“中国高校计算机大赛”（*China Collegiate Computing Contest*, 简称 C4），目前“中国高校计算机大赛”继续由全国高等学校计算机教育研究会主办。大数据挑战赛是其中的一项重要赛事，在 2018-2020 年期间均入选全国普通高校学科竞赛排行榜，获得社会各界的高度关注和广泛好评。

2021 中国高校计算机大赛——微信大数据挑战赛（以下简称“大赛”）由清华大学和腾讯微信事业群联合举办，由腾讯云提供大赛资源支持。本次大赛是以企业真实场景和实际脱敏数据为基础、面向全球开放的高端算法竞赛。大赛旨在通过竞技的方式，提升人们对数据分析与处理的算法研究与技术应用能力，探索大数据的核心科学与技术问题，尝试创新大数据技术，推动大数据的产学研用。

一、参赛对象

本次大赛面向全球开放，不限年龄国籍，高等院校在校学生（包括高职高专、本科、研究生）以及科研机构和企业从业人员均可参赛。具体要求如下：

- 可以自由组队参赛，具体组队要求见报名&组队的相关说明；
- 参赛选手应保证报名信息准确有效，如队伍中的选手信息不符合要求，组委会有权取消整个队伍的参赛资格及奖励。

为了保证大赛的公平性，大赛主办和技术支持单位如有机会接触赛题和相关数据的人员不允许参赛。

腾讯公司内部非接触赛题和相关数据的人员可报名参赛，具体规定如下：

- 实习生不受限制，以所在学校报名参赛；
- 参赛队伍中如有一位腾讯公司正式员工，即视为腾讯公司参赛队伍；
- 是否正式员工以大赛启动报名时间的状态为准；
- 腾讯公司参赛队伍的成绩可参与排名，入围复赛队伍数量不超过 10 支，不允许入围决赛以及获得复赛与决赛奖项，可发放纪念奖。

二、赛制说明

本次大赛分为报名&组队、初赛、复赛和决赛等四个阶段，其中初赛阶段由参赛队伍下载数据在本地进行算法设计和调试，并通过大赛报名官网提交结果文件；复赛阶段要求参赛队伍在大赛官网平台上进行数据处理、算法调试和生成结果，数据不可下载，可使用平台提供的计算资源和工具包；决赛要求参赛者进行现场演示和答辩。本次大赛所提供数据均为脱敏数据。

1. 报名&组队 (4月30日 – 6月28日)

参赛选手须在大赛官网报名并且组队参赛（即使单人参赛也要组建单人队伍），大赛不收取任何报名费用。大赛报名系统开放时间为北京时间 2021 年 4 月 30 日 10:00，截止时间为北京时间 2021 年 6 月 28 日中午 12:00。

- 报名方式：登录比赛官网，完成个人信息注册，即可报名参赛；
- 每个选手可单人成队或 2-3 人组队参赛，且每人只能参加一支队伍。

大赛官方渠道主要包括：

- 大赛官网：<https://algo.weixin.qq.com/>
- 大赛邮箱：data@tsinghua.edu.cn
- 大赛 QQ 群：566353409 / 758344321 / 762146461

报名截止之后，不再允许添加或更改任何队伍成员。如有中途退出情况，只允许在参赛队伍内部更换队长或删除队员。参赛队伍须应在决赛开始前向大赛组委会提交成员更换申请，由参赛队伍全部成员亲笔签名，经由大赛组委会审核后变更生效。

2. 初赛 (5月20日 – 6月30日)

参赛队伍可从大赛官方网站下载数据，在本地进行算法调试，并在线提交结果。

5月20日 10:00 开始，选手可以从竞赛平台下载初赛训练数据集，用于参赛队伍训练模型以及制定预估策略；同时，平台提供测试数据集，用于参赛队伍在比赛中的模型评估和排名。

初赛采用 AB 榜形式：

- 初赛 A 阶段：5月20日 10:00 – 6月29日 20:00，每个参赛队伍每天可以有 3 次提交结果机会，系统实时评测并返回成绩。排行榜每小时更新，将选择参赛队伍在本阶段的历史最优成绩，按照评测指标从高到低排序。
- 初赛 B 阶段：6月30日 10:00-20:00。系统将在 6 月 29 日 12:00 更换测试数据，参赛队伍需再次下载数据文件。本阶段提供 2 次提交结果的机会，系统进行实时评测并返回成绩。排行榜每小时进行更新，并选择参赛队伍在本阶段的历史最优成绩进行排名展示。

初赛提交的截止时间是 6 月 30 日 22:00，初赛以 B 榜成绩作为初赛成绩依照，要求 TOP150 团队提交代码审核，具体要求见 <https://git.code.tencent.com/perryprli/Code-Requirements>，代码提交截止时间是 7 月 4 日上午 10:00。

组委会将审核并取消存在人工标注、相互抄袭等行为队伍的比赛资格，晋级空缺名额后补。对于初赛成绩符合要求且通过实名认证的参赛队伍，排名前 100 名的参赛队伍将进入复赛，所有通过审核的队伍将获得初赛名次证书。

3. 复赛 (7月12日 – 8月9日)

复赛参赛队伍需要在大赛官网平台上完成数据处理、建模、算法调试、生成结果等，所有比赛数据不可下载，可使用平台提供的计算资源和工具包。

预装环境阶段（7月12日 12:00 – 7月15日 12:00）：选手可以在复赛平台上安装和调试自己需要的环境，平台提供初赛数据集给选手测试，不提供系统实时评测。截止时间之后，禁止选手在平台上进行上传和下载操作。

正式阶段（7月16日 12:00 – 8月9日 22:00）：复赛采用AB榜形式。

- 复赛A阶段：7月16日 12:00 – 8月8日 12:00，每个参赛队伍每天可以有3次提交结果机会，系统实时评测并返回成绩。排行榜每小时更新，将选择参赛队伍在本阶段的历史最优成绩，按照评测指标从高到低排序。
- 在复赛A阶段，所有参赛队伍需要按照代码规范整理好代码，具体要求见<https://git.code.tencent.com/perryprli/Code-Requirements-Semi>。在7月23日12:00 – 8月8日12:00期间，参赛队伍将所有代码和模型以及必要的特征打包上传，系统自动记录上传文件的md5。在8月8日12:00之后，所有上传文件均不允许再更改。
- 复赛B阶段：8月8日 15:00 – 8月9日 22:00，系统将在8月8日15:00提供B榜测试数据集。每个参赛队伍限使用A榜提交的模型和代码进行预测和提交，共有2次提交结果机会，系统实时评测并返回成绩。排行榜每小时更新，将选择参赛队伍在本阶段的历史最优成绩，按照评测指标从高到低排序。

组委会将审核并剔除存在人工标注、相互抄袭、未提交代码、结果无法复现等问题队伍以及腾讯公司参赛队伍，另外将评估队伍算法模型的性能，具体如下：

- 性能评估方法：在组委会指定的机器上（2卡P40 48G显存 14核CPU 112G内存），预测给定测试数据集7个目标行为的概率，记录预测时长，并计算单个目标行为2000条样本的平均预测时长（单位：毫秒）。
- 单个目标行为2000条样本的平均预测时长（单位：毫秒）=预测总时长（单位：秒） $\times 1000 \times 2000 / (7.0 \times \text{测试集样本数})$ 。
- 最低性能要求：单个目标行为2000条样本的预测时长不超过300毫秒。通过代码审核且满足最低性能要求的前6名队伍将晋级全国总决赛，第7-30名队伍将获得复赛奖项。

4. 决赛 (8月下旬)

决赛将以现场答辩会的形式进行，具体要求和安排另行通知。受邀参加决赛的选手在决赛期间的食宿由大赛组委会安排，往返交通费及其他费用自理。

晋级决赛团队需提前准备答辩材料，包括路演 PPT、参赛总结、算法核心代码。在决赛答辩会上，每支队伍面对评委有 20 分钟的路演时间和 10 分钟的答辩时间。评委将根据选手的技术思路、理论深度、算法性能和现场表现进行综合评分。

决赛分数将根据参赛队伍的算法成绩和答辩成绩加权得出，评分权重为复赛阶段 70%，决赛答辩 30%。

三、奖项设置

1. 初赛奖项

初赛 TOP150 且通过代码审核的团队将颁发初赛名次证书，此项奖励以大赛官网初赛最终排行榜为准。

2. 复赛与决赛奖项

大赛奖金池总额为 66 万元人民币，所有奖金均为税前金额。同时，所有队伍将颁发复赛或决赛名次证书。此项奖励以大赛官网复赛最终排行榜和决赛结果为准。

奖励对象	数量	奖金
决赛第 1 名队伍	1	30 万元
决赛第 2 名队伍	1	10 万元
决赛第 3 名队伍	1	7 万元
决赛第 4-6 名队伍	3	3 万元
复赛第 7-10 名队伍	4	1 万元
复赛第 11-30 名队伍	20	3 千元

3. 在校学生队伍奖项

在校学生队伍要求所有参赛队员必须全部为在校学生，如果队伍中有一名在职人员，则整个队伍视为在职人员队伍。其中中国大陆在校学生提供学信网的教育部学籍在线验证报告编号进行身份验证，其余学生提供相关在读证明进行身份验证，在校学籍以大赛报名时间 2021 年 4 月 30 日为准。

此奖项仅颁发给进入复赛的在校学生队伍，要求队伍提交复赛代码并获得有效的复赛成绩，根据所有在校学生队伍复赛成绩的单独排名结果进行颁发。

奖项名称	数量	对象
全国一等奖	10	单独排名第 1-10 名

全国二等奖	20	单独排名第 11-30 名
全国三等奖	N	单独排名第 31 名及之后的队伍

另外，如果入围决赛的队伍中没有在校学生队伍，大赛将增补一个决赛名额，提供给单独排名第一的在校学生队伍入围决赛。

4. 周周星

在初赛阶段，设立周周星奖励。从初赛第三周开始，以每周一中午 12 点的排行榜为准，取前两名参赛队伍发放周周星纪念礼物；另外视情况会额外增加一个名额，以保证当期周周星队伍中至少有一个在校生队伍或一个在职人员队伍。对于前面已经获得周周星的队伍，不重复发放，名额按名次顺延。

5. 其他激励

招聘绿色通道：复赛排名前 30 队伍的在校学生以及在校学生队伍单独排名前 10 名队伍的在校学生，将获得 2021 年腾讯集团微信事业群校园招聘和实习招聘绿色通道资格，具体细则另行通知。

四、违规处理

参赛者应本着诚实、公平的态度参加比赛，如在以下情况出现违规，大赛组织委员会（简称“大赛组委会”）有权取消参赛者所在队伍的参赛资格，情节严重者将通报参赛者所在高校并追究其违法责任。

1. 账号使用：参赛者有义务保证账号信息的真实性和有效性，且账号仅限于参赛者本人使用；参赛者禁止使用多账号参赛，同一参赛者不可使用多个账号进行提交、刷分操作；如根据判断认为参赛账号存在异常或违背正常使用条例，组委会可以单方面暂停或终止该账号登录大赛平台。
2. 比赛成果：
 - 严禁参赛队伍之间相互抄袭。如不同参赛队伍提交结果高度相似，经判定存在抄袭行为的，组委会将取消相关参赛队伍的参赛资格，相关参赛成绩无效。
 - 参赛者应保证其在比赛过程中所产出的所有成果未侵犯任何第三方的知识产权、商业秘密及其他合法权益。如第三方因为参赛者侵权行为提出索赔、诉讼等，参赛者应承担由此产生的全部责任及损失。
 - 如大赛主办方及其关联公司有意取得参赛者在本次大赛中独立开发的依约定享有完整知识产权的研究成果，参赛者同意大赛举办方及其关联公司在同等条件下享有优先受让权，相关转让事宜由双方另行协商确定。
3. 数据使用：对于大赛提供的数据（数据集），参赛者须仅在比赛场景下使用，

并应妥善保存已下载的数据（数据集），避免泄露；在完成比赛使用后应及时销毁已下载数据（数据集）；如使用比赛之外的任何数据应获得组委会许可。对于不提供下载的比赛数据，参赛者不得以任何形式擅自复制、下载或获取。参赛者如发现任何出现数据未授权访问的可能，应立即通知组委会并积极提供相关信息。

4. 代码分享：在大赛举办期间，未经组委会同意，参赛者禁止公开分享与赛事相关的数据、模型和代码；大赛结束之后，参赛者可以在拥有模型和代码的知识产权的情况下自行选择公开分享，但需要确保此类公开共享不会侵犯任何第三方的知识产权、商业秘密及其他合法权益。
5. 参赛者若在参赛过程中发现相关规则漏洞或技术漏洞，有义务及时告知资委会相关漏洞的信息，组委会将对提供相关信息的参赛者表示相关感谢；若参赛者利用相关漏洞进行参赛，经判断查证后，成绩将会被判断为无效成绩。

五、申诉与仲裁

1. 参赛团队或选手对不符合大赛规定的设备、工具和软件，有失公正的评判和奖励以及工作人员的违规行为等，均可向大赛组委会提出申诉。组委会负责受理比赛中提出的申诉并进行调解仲裁，以保证大赛的顺利进行和大赛结果的公平公正。组织委员会作出的仲裁结果为终局决定。
2. 申诉报告应明确申诉内容，指定一名成员作为联系人，通过大赛邮箱以邮件发送，否则申诉将不予以受理。
3. 组织委员会将在收到申诉之日起 5 个工作日内受理，并认真核查和处理。

六、其他说明

1. 在大赛举办过程中，竞赛规程可能会有少量的变更和调整，所有内容均以大赛官网为准。
2. 本大赛规程的最终解释权归“中国高校计算机大赛——微信大数据挑战赛”组织委员会所有。