2019 年（第六届）全国大学生统计建模大赛

运动类 App 使用现状及发展前景研究

——以合肥市高校为例

参赛单位 ： 合肥工业大学

参赛者姓名： 张文钧、殷振豪、张丁洋

目　录

表格清单

插图清单

摘 要

随着全媒体时代媒介融合的不断发展，人们的生活方式也不断发展变化，现如今，手机与移动互联网已经成为当代人的生活必须。新媒体技术的高速发展，驱动着智能手机的应用程序向更加智能化和生活化的方向不断进步，App 应用程序逐步普及。其中，运动健身类 App 凭借其独特的优势，已逐渐成为大学生中流行的运动健身辅助工具，为大学生提供运动健身指导和数据参考服务。但是在 App 应用市场的激烈竞争下，运动健身类 App 在功能完善和精准推送等方面还存在一些问题。

本文的研究将以统计调查分析为主，通过对合肥市高校学生的基本信息、运动类 App 的认识和使用现状进行问卷调查，对运动类 App 现有用户和潜在用户进行分析，进而对运动类 App 未来发展提出具有针对性的建议措施。

在调查方法上，本文采用三阶段不等概率抽样的方法进行抽样调查，提高调查过程的科学性。所有调查数据均通过了信度检验和效度检验，结果真实可靠。在数据分析方法上，本文采用描述性统计，对所调查的人群基本情况等方面进行分析；然后，利用随机森林模型分析影响大学生是否使用运动类 App 的因素，并对因素重要性进行排序；最后，运用 GBDT 模型对使用行为进行预测，并通过用户画像对潜在使用者和目标产品进行刻画。

本报告研究的结论主要有:（1）合肥市在校大学生大多使用过运动类 App，且用户之间存在差异；（2）与记录运动信息相关的功能更被在校大学生所青睐；（3）运动类 App 对在校大学生的运动锻炼效果具有一定的影响；（4）合肥市在校大学生对运动类 App 的满意程度普遍较高。（5）空闲时间、锻炼的频率、偏爱的运动环境对是否使用运动类 App 的影响显著。

根据本文研究结果，运动类 App 市场的未来发展提出改进建议：（1）找准产品定位，明确研发方向；（2）完善产品功能，增强用户粘性；（3）明晰用户群体，精准推送产品。

关键词：运动类 App；三阶段不等概率抽样；随机森林；GBDT；用户画像

第一部分 绪　论

一、研究背景及意义

（一）研究背景

2014 年，随着《关于加快发展体育产业促进体育消费的若干意见》等政策的推出，全民运动，加强体育锻炼上升为国家战略的层次，体育产业成为中国的新兴绿色产业。我国预计到 2020 年， 体育人口有望达到 4.35 亿人。从 20 岁以上参加体育锻炼人数的占比来看，20-29 岁人群占比接近 50%，50 岁以下中青年的整体占比也达到 40%以上，说明有一定消费能力的中青年对运动健身的需求较高，因此，未来运动健身产业的市场规模也有望持续增加。

但是，在 2015 年之后，运动类 App 领域的融资热潮降温，运动类 App 领域所存在的问题也都暴露了出来：App 没有可持续的特色和创新性的功能，很难获得用户黏度，甚至会导致用户放弃使用；App 现有计算方法和统计数据的准确性不够准确，用户同时使用不同 App 可能得到不同结果，导致用户对运动类 App 的信任度下降等。由此看来，现有的运动类 App 用户粘度较差，且现有 App 用户对运动类 App 持有热情不高，运动类 App 市场发展不完善，存在较大问题。

（二）研究意义

国务院在 2016 年发布关于加快发展健身休闲产业的指导意见，欲完善健身休闲服务体系，推动互联网+健身休闲模式的发展，实现健身在线化、产品智能化和数据可视化，以此来推进《全民健身计划 2016-2020 年》的实现，加快推进体育强国建设，实现中华民族伟大复兴的中国梦。

运动类 App 用户量目前已达到 3.64 亿， 未来十几年将会持续增加， 其中，社会白领与高校学生是最主要的用户群体。本文通过对合肥市高校学生基本信息和使用运动类 App 的情况进行问卷调查，运用随机森林模型和 GBDT 模型研究现有消费者对运动类 App 的使用感受和功能期望，得出符合市场和消费者期望的运动类 App 应具有的功能和现有功能的缺陷和问题，生成用户画像，预测消费者未来的使用行为。

本项调查研究可为运动类 App 开发公司提供一定的消费者原始数据和建议，有利于公司据此对 App 现有问题和缺陷进行改善，对未来 App 研发方向进行确定和修改，为今后能开发专业性强、个性化强、可用于指导全民健身的 App 提

供理论基础，使公司在急速扩大的市场中占据有利地位。

二、文献综述

对于我国运动类 App 发展现状而言，林瑶瑶（2019）[1]认为近几年以智能手机为主要载体的移动通信技术以飞快的速度向前迈进，手机 App 已经成为人们生活中不可或缺的一部分，运动健身类 App 凭借其便捷性、互动性强的特点吸引了众多用户，得到了快速的推广和应用；刘璐（2016）[2]认为从现阶段市场中的运动类 App 发展情况来看， 部分运动类 App 的研发为提升开发速度或是抢占市场，没有能够针对 App 的长久发展进行科学论证，导致其自身还存在着一定的缺陷，产品同质化现象严重；李霞（2017）[3]认为在我国根据国民体质监测指标作为健身指导的 App 比较缺乏且亟待开发。

对于影响运动类 App 消费者使用行为的因素，李柔（2017）[4]认为运动类 App 的女性用户数量略高于男性，且 85%的用户为 40 岁以下；陆佳莉（2017）[5]认为运动类 App 用户多集中在超一线城市和一线城市，是一方面大城市拥有丰富的健身资源，另一方面大城市居民的健身习惯和健身意识相对较好；李霞（2017）[3]认为用户对于拥有社交平台的 App 满意度最高，对于运动监测型的 App 满意度最低。

通过对已有文献的分析，我们发现大部分学者主要对于运动类 App 使用情况进行调查研究，而对运动类 App 具体功能满意度和问题缺陷进行调查研究的情况较少，待以补充。本研究报告对运动类 App 现有用户和潜在用户进行分析，对运动类 App 未来发展提出具有针对性的建议措施。具有一定的研究现实意义。

第二部分 调查方案与实施

一、调查方案设计

（一）调查目的

1. 对合肥高校学生进行随机抽样，了解运动类 App 在学生中的使用现状

我们采用随机抽样和三阶段不等概率抽样相结合的方法，在合肥市高校大学生中进行调研。通过对高校学生的性别、年龄、每周运动习惯以及是否使用过运动类 App 等基本情况进行调查，了解运动类 App 在高校学生中的现有市场规模，预测潜在市场规模。

2. 调研使用者对运动类 App 现有功能的满意度，改善功能存在的问题

近年来，运动类 App 研究发展进入寒冬期，根本原因在于使用者粘度不高，产品功能存在较大问题。我们将通过调查，挖掘使用者对具体产品功能的满意程度，从而得到如何改善运动类 App 现有功能问题的解决方案。

3. 构建运动类 App 的产品画像

通过分析运动类 App 现有功能的重要度和使用者满意度，以及使用者对运动类 App 的功能预期，利用 GBDT 模型来预测目前使用者心目中的理想型 App，以此促进 App 开发公司对产品研究进行调整，使其更满足使用者的心理预期，提高使用者的满意度，来达到扩大市场规模的目的。

（二）调查内容

本文的研究内容主要包含三个方面：首先，对有关运动类 App 的现有文献进行分析，据此设计《大学生运动类 App 使用现状及影响调查》研究问卷，并采用随机抽样与三阶段不等概率抽样相结合的方法确定调查对象及数量；其次，根据回收数据，分析运动类 App 大学生市场使用者对 App 的使用状况及满意度；最后，根据数据分析结果，提出关于如何改进运动类 App 的发展建议。研究的基本思路如图 1 所示：

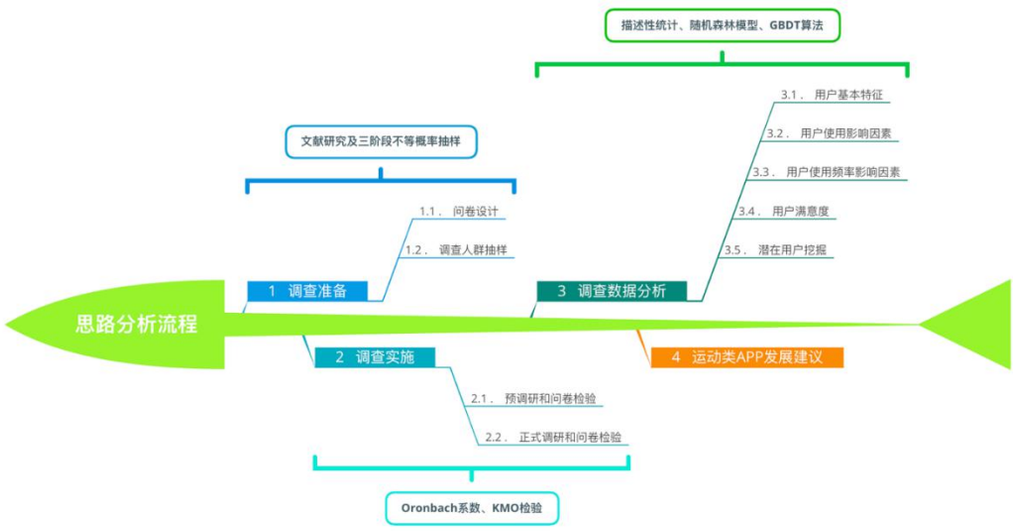

图 1　研究思路流程图

（三）调查和抽查方法

本文选取合肥市 18 所高校全部大学生为调查对象，采用分层抽样和三阶段不等概率抽样相结合的概率抽样调查方式，保证了调查的科学性，减少了抽样误差。考虑到距离问题和人力问题，调查小组主要采用线上与线下相结合发放调查问卷的方法进行问卷调查。

（四）调查方式与方法

考虑到调查总体及各层的财力物力以及人力等多重现实因素后，本文通过线上与线下相结合的调查方式对合肥市高校大学生对运动类 App 的基本了解情况和使用现状进行信息收集，了解用户的满意度，发现潜在用户，开发潜在市场，进而探究其发展潜力与前景。

为了使估计量更加完善、保证调查的科学性，本文采取了三阶段不等概率 PPS 抽样。在第一阶段中从合肥市所有高校中抽取学校；在第二阶段中从学校抽取院系；最后在第三阶段从各个院系中抽取学生进行问卷调查。

在整理数据时将纸质问卷信息录入到线上数据库中，为近一步展开数据分析提供了便利，缩短了项目的执行时间，降低了数据处理的时间成本与保存成本。

1. 调查方法

① 文献研究

在进行问卷设计与问卷分发前，小组主要采用文案调查法，使用计算机检索国内运动类程序与运动类社交媒体的发展历程与发展现状，了解运动类 App 在应用程序市场的成长态势，深度解析运动类 App 开拓高校大学生市场中所遇到的问题，获取对本次调查有价值的二手资料。

② 抽样调查

介于调查的可行性，本文在调研对象中按一定比例抽取一定数量的对象来开展调查研究，再将调研结果按抽取的相应比例进行放大，以便在较短时间内获得较为准确的调研信息。

③ 问卷调查

本文采用问卷调查法作为本次调查的主要调查方法获取一手数据。小组成员按照事先抽取的院系名单，对每个院系按照分配好的最佳样本量在各学院内发放纸质或电子版问卷，收集样本的信息。

2. 调查方式

本文采用三阶段不等概率抽样的方法抽取合肥市高校的调查对象：

① 第一阶段进行 PPS 抽样

利用 PPS 法进行初级抽样单元的抽取时运用代码法进行实施。具体结果如表 1 所示，合肥市高校人数总计 256850 人，最终共抽出 4 个高校。

表 1　第一阶段代码法抽样表

序号	高校	学生人数	所占比例	代码范围	随机产生数	抽中再编码
1	安徽大学	28000	0.109	0—28000	24995	1
2	中国科学技术大学	15500	0.060	28001—43500		
3	合肥工业大学	30450	0.118	43501—64568	36870、59367	2
4	安徽农业大学	21068	0.082	64569—85636		

5	安徽医科大学	19000	0.073	85637—104636		
6	安徽中医药大学	13205	0.051	104637—117841	115369	3
7	巢湖学院	16254	0.063	117842—134095		
8	安徽建筑大学	17700	0.068	134096—151795		
9	安徽三联学院	13800	0.053	151796—165595		
10	合肥学院	15207	0.059	165596—180802	173449	4
11	安徽新华学院	20000	0.077	180803—200802		
12	安徽文达信息工程学院	10500	0.040	200803—211302		
13	安徽外国语学院	12000	0.046	211303—223302		
14	安徽大学江淮学院	1295	0.005	223303—224597		
15	安徽建筑大学城市建设学院	5000	0.019	224598—229597		
16	安徽农业大学经济技术学院	7800	0.030	229598—237397		
17	合肥师范学院	9038	0.035	237398—246435		
18	安徽医科大学临床医学院	1033	0.004	246436—247468		

② 第二阶段进行分层抽样

从第一阶段抽取的合肥市高校中抽取院系。抽样结果如表 2 所示：

表 2　随机数表抽样

序号	高校	所含院系数	入样号码	对应院系名
1	安徽大学	26	24、18、14、	经济学院、外语学院、管理学院、

			03、16、11	艺术学院、哲学系、计算机科学与技术学院
2	合肥工业大学大学	23	31、36、48	化学与化工学院、数学学院、微电子学院
3	安徽中医药大学	16	57、62	药学院、中医学院
4	合肥学院	14	70	机械工程系

③ 第三阶段进行随机抽样

在第三阶段中要从入样院系抽取被调查者，考虑到调查难度和人力资源的限制，小组采取随机抽样的方式，在每个院系内随机选取学生进行调查。

（五）抽样调查基本步骤

1. 编制抽样框

本文使用三阶段不等概率抽样的方法进行抽样。在第一阶段中，将合肥市的所有高校作为一级单元的抽样框，抽出所选中高校作为一级抽样单位；在第二阶段中，二级单元的抽样框是第一阶段入样的高校的所有院系，抽出所选院系作为二级抽样单位；在第三阶段中，三级单元的抽样框为每个入样院系的所有学生。具体抽样框如表 3 所示。

2. 设计调查问卷

调查问卷由三个部分组成，分别是调查对象的基本情况调查、对运动类 App 认知情况的调查和对运动类 App 使用现状的调查。问卷排布逻辑清晰，并且问卷内容设计是基于对大学生使用运动类 App 的现状即问题的相关文献资料的充分学习，问卷具有很强的科学性和代表性。

3. 预抽样调查

在进行正式调查之前，先根据抽样方案从抽样框中抽取较小样本进行预调查，以检验所设计的抽样方案、抽样框和调查问卷的科学性和可操作性，对检测效果较差的项目进行调整修改。

表 3　高校院系抽样框

一级单元	入样	二级单元的抽样框	入样	三级元
合肥市的所有高校	安徽大学	安徽大学的所有院系	经济学院	经济学院所有学生
			外语学院	外语学院所有学生
			管理学院	管理学院所有学生
			艺术学院	艺术学院所有学生
			哲学系	哲学系所有学生
	合肥工业大学	合肥工业大学的所有院系	计算机科学与技术学院	计算机科学与技术学院所有学生
			化学与化工学院	化学与化工学院所有学生
			数学学院	数学学院所有学生
			微电子学院	微电子学院所有学生
	安徽中医药大学	安徽中医药大学的所有院系	药学院	药学院所有学生
			中医学院	中医学院所有学生
	合肥学院	合肥学院的所有院系	机械工程系	机械工程系所有学生

4. 样本容量的确定

将预调查问卷填写的运动类 App 使用者占比作为估计对象，因此关注总体认知比例的样本方差。修正前最佳样本量n_0的计算公式如下：

$$n_0=\frac{t^2PQ/d^2}{1+\frac{1}{N}\left[\frac{t^2PQ}{d^2}-1\right]}$$

N 为总体数量，取置信度为 95%时的 t 值，t=1.96，P 为样本比例，d 为绝对允许误差，d=0.05，P 取 0.5。则可以近似得出最佳样本量为 384：

$$n_0 = \frac{t^2 p(1-p)}{d^2} = \frac{1.96^2 \times 0.5 \times 0.5}{0.05^2} \approx 384$$

根据预调查和文献资料的结合，假设我们采取的多阶段抽样的设计效应为 1.8，则应回收的有效样本量为 691 份。

$$n = n_0 \times \text{deff} = 384 \times 1.8 \approx 691$$

考虑到样本无效问题，在征集指导老师的意见和经验后，我们假设无效比例为 10%，则实际应调查的样本量为 767 份。

在确定了样本量之后，我们根据入样高校的在校生比例进行了样本量的分配，具体分配情况如表 4 所示：

表 4　样本量分配表

入样高校	在校生数量	人数所占比例	样本量
安徽大学	28000	32.24%	247
合肥工业大学	30450	35.05%	269
安徽中医药大学	13205	15.20%	117
合肥学院	15207	17.51%	134

5. 正式抽样调查

本次共发放 821 份问卷，有效问卷为 779 份，有效问卷回收率 94.88%。

二、调查实施

（一）调查实施进度

本次调查的工作期限为 2019 年 6 月 1 日-2019 年 6 月 12 日，具体实施进度见图 2 所示。

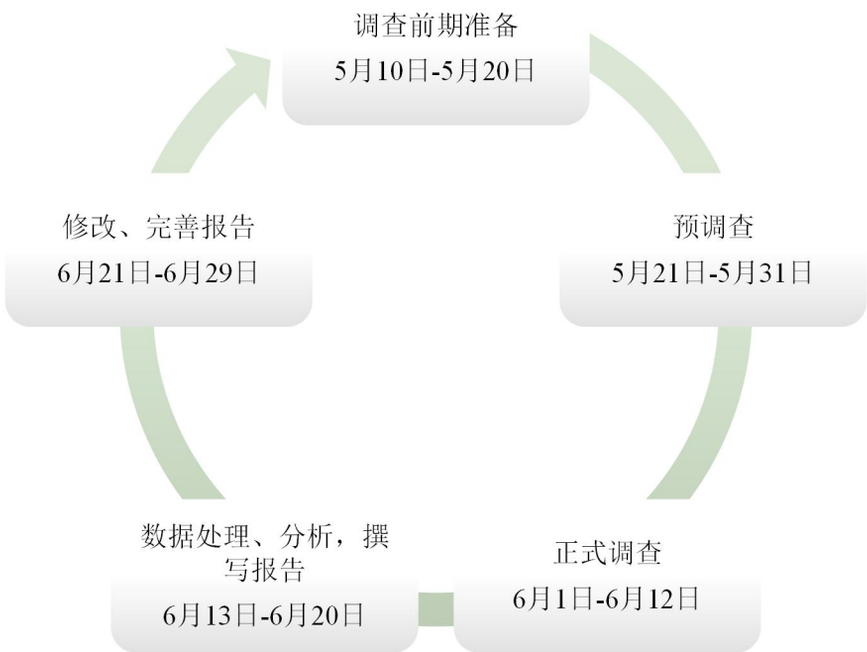

图 2　实施进度图

（二）实际操作过程

在实际问卷发放的过程中，我们小组三人分工合作，对所抽取的安徽大学、合肥工业大学、安徽中医药大学、合肥学院这四所高校入样院系的学生进行问卷调查。我们的问卷形式有纸质版问卷与电子版问卷两种，其中电子版问卷是借助问卷星平台构建而成。考虑到物力财力的约束与发放效率，本小组在合肥工业大学（本学校）各入样院系学生里进行大量的随机实体问卷调查，并结合电子问卷进行发放，在安徽大学也同样进行了部分随机实体问卷发放与电子问卷发放。在发放纸质版问卷时，小组成员在被调查者完成问卷填写之后，与被调查者进行交流与询问，了解其对运动类 App 认识，并及时回收问卷信息。之后，小组联系了安徽中医药大学与合肥学院入样学院各班级的班干部，委托班干部于各班级采用电子问卷的形式进行随机调查，并及时回收信息。

（三）质量控制

1. 本次调研过程中，以纸质版问卷和网络问卷相结合的方式进行调查，网络问卷具备的高质量的数据回收的特点，能够避免数据缺失情况的发生；
2. 网络问卷设置省去了部分数据录入环节，缩短了项目周期和时间，方便快捷；
3. 为使样本更具有代表性，团队成员在微博、朋友圈以及 QQ 空间等网络平台进行问卷的发放与收集；

4. 实地调研前调研小组各成员集中学习了解运动类 App 的相关内容，严格按照统一计划和填表说明的要求执行；

5. 通过合理编制抽样框、现场调查环节检查的全面性和对于回收问卷审查的严谨性，有效保证了调查的质量；

6. 在数据录入时，采取两次录入的模式，保证录入数据的准确性。

（四）调查误差的控制

调查误差（Error in Survey）是指在取得样本数据资料过程中产生的误差。这部分误差通常与调查者、回答者、资料搜集方式和问卷等因素有关，它们会形成在调查过程中出现无回答和回答出现偏误等情况，进而形成系统性误差。

因此，我们采用以下方式来减小并控制调查误差：

1. 问卷和表格设计

在预调查中，通过信度检验，问卷各部分的 Cronbach 系数均大于 0.8，同时问卷内部一致性系数为 0.965，而在问卷结构效度检验中，KMO 值为 0.617，大于 0.5。我们认为问卷设计及调查结果符合我们的预期结果和目标。

2. 抽样

我们采用了分层抽样和三阶段不等概率抽样的方法，降低主观判断对样本的抽选，有利于提高估计的精度,减少抽样误差,使估计量更加完善、保证调查的科学性。

3. 数据的收集

实地调研前调研小组各成员集中学习了解运动类 App 的相关内容，严格按照统一计划和填表说明的要求执行，尽量减低数据收集过程中可能产生的调查误差。

4. 数据录入

在发放问卷的过程中，将被调查者配合度较低、填写较随意的问卷及中途放弃填写的问卷标注为无效问卷，将最终有效的纸质问卷和电子问卷汇总，并录入数据。而在数据录入时，我们采取两次录入的模式，保证录入数据的准确性。

5. 调查计划

调查问卷由三个部分组成，逻辑清晰、合理，问卷在问题设计、排版布局等方面具有很强的科学性和代表性。

三、预调查数据处理及检验

（一）信度检验

问卷中相关量表的 Cronbach 系数如表 5 所示：

表 5　预调查问卷相关量表 Cronbach 系数表

层面	Cronbach 系数	项数
App 具体功能满意度	1.0000	9
问卷整体	0.9650	92

各部分的 Cronbach 系数均大于 0.8，说明每个部分的项目描述都很精确；问卷内部一致性系数为 0.9650，说明问卷整体信度很好。

（二）结构效度检验

为反映问卷整体的结构效度，本文进行了 KMO 和 Bartlett 球形检验，结果如表 6 所示。

表 6　预调查 KMO 和 Bartlett 球形检验结果

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.6170
Bartlett's Test of Sphericity	Approx. Chi-Square	1765.1190
	df	528
	Sig.	0

KMO 值为 0.6170，大于 0.5，且 Bartlett 球形检验的 *P* 值为 0，表明高度显著，问卷整体效度较好。

（三）卡方检验

为判断使用线上调研和线下两种调研方式之间是否存在差异，我们使用卡方分析比较线上调研结果和线下调研结果两组之间是否存在显著差异。

表 7　卡方检验结果

	值	df	渐进 Sig.（双侧）	精确 Sig.（双侧）	精确 Sig.（单侧）
Person 卡方	0.3890	1	0.5330		

连续校正	0.1450	1	0.7030		
似然比	0.3880	1	0.5330		
Fisher 的精确检验				0.6300	0.3510
线性和线性组合	0.3830	1	0.5360		
有效案例中的 N	70				

a. 0 单位格（.0%）的期望计数少于 5。最小期望计数为 13.71。
b. 仅对 2×2 表计算

检验结果表明，采用线上和线下相结合的调研方式对用户是否会使用运动类 App 无显著影响，同时对调研样本的一致性和相关性无影响。

四、正式调查数据处理

（一）问卷处理与数据录入

在发放问卷的过程中，将被调查者配合度较低、填写较随意的问卷及中途放弃填写的问卷标注为无效问卷，将最终有效的纸质问卷和电子问卷汇总，并录入数据。

（二）信度与效度检验

表 8　正式调查问卷相关量表 Cronbach 系数表

层面	Cronbach 系数	项数
App 具体功能满意度	0.9990	9
问卷整体	0.9580	92

Cronbach 系数在 0.9～1 之间时，说明量表的可信度非常好。根据 Cronbach 系数表的结果说明问卷分类合理，量表内在一致性高。

对正式数据进行 KMO 和 Bartlett 球形度检验，结果如表 9 所示：

表 9　正式调查 KMO 和 Bartlett 球形检验结果

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.6170
Bartlett's Test of Sphericity	Approx. Chi-Square	1765.1190
	df	528
	Sig.	0

检验结果 KMO=0.933>0.5，Bartlett 球形度检验 *P* 值小于 0.05，非常显著，说明数据可以进行因子分析。

五、研究创新点

本文在研究内容方面进行了创新：

（一） 自 2014 年运动类 App 出现后，绝大多数文献的研究内容均是对运动类 App 的整体而言，对具体功能满意度的调查研究内容较少，而本文在此方面进行了开拓创新。

（二） 构建产品画像，得出最符合使用者心理预期的运动类 App 产品画像，有助于对运动类 App 进行精准的研发设计。

第三部分 运动类 App 用户调查结果统计及分析

一、大学生对运动类 App 认知统计学特征分析

（一）高校学生使用 App 原因（第 9 题）

选项	平均综合得分	比例
A.锻炼兴趣	5.58	
B.空闲时间	5.29	
C.场地器材	3.12	
F.集体效应	2.27	
D.专业辅导	1.99	
E.媒体宣传	1.26	
G.其他	0.65	

图 3　高校学生使用 App 原因分布图

通过计算各影响因素的平均综合得分可得，合肥市高校学生使用运动类 App 主要影响因素是“锻炼兴趣”和“空闲时间”。其中“锻炼兴趣”这一因素的综合得分为 5.58，“空闲时间”的得分为 5.29。这说明影响合肥市高校学生使用 App 进行体育锻炼的首要影响因素是锻炼兴趣，其次是空闲时间，有无空闲时间是大学生是否进行体育锻炼的重要前提。

（二）高校学生对运动类 App 功能重要程度的认知（第 10、11 题）

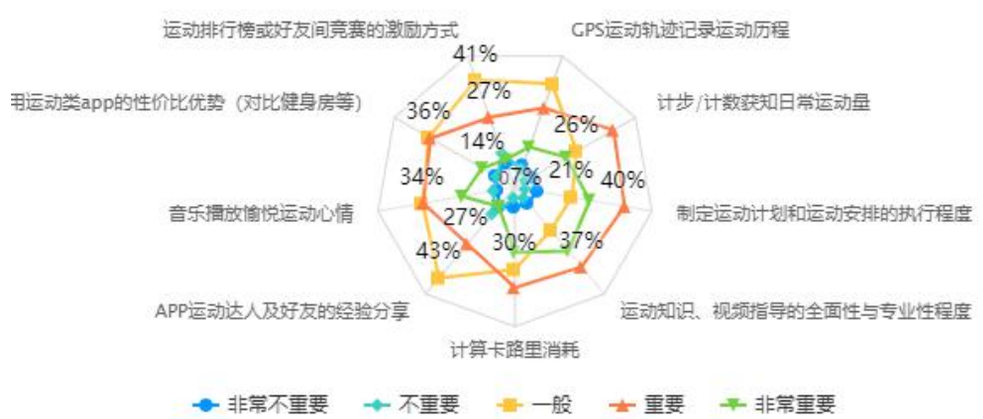

图 4　运动类 App 功能重要度分布雷达图

选项	平均综合得分	比例
卡路里消耗	5.86	
运动计划	5.84	
计步/计数	5.75	
GPS记录运动轨迹	5.57	
视频指导/运动知识	5.1	
音乐播放	3.64	
运动圈互动	2.7	
运动榜单/好友竞赛	2.51	
提供课程性价比高	1.94	
其他	0.66	

图 5　运动类 App 功能重要度选项分布图

可以看出，合肥市高校在校大学生对“运动知识、视频指导的全面性与专业性程度”、“制定运动计划和运动安排的执行程度”与“计步/计数获知日常运动量”这三项功能的重要性评分较高，在满意及以上程度占比分布为 67.21%、67.21%、62.34%。“App 运动达人及好友的经验分享”和“运动排行榜或好友间竞赛的激励方式”对合肥市在校大学生使用运动类 App 的影响不高。可以分析出，大部分在校大学生更看重运动类 App 作为运动锻炼辅助工具所具有的功能和作用，对分享和竞赛的形式并不看重。

二、使用者的人口统计学特征分析

（一）性别分布情况（第 1 题）

性别分布情况见图 6 所示

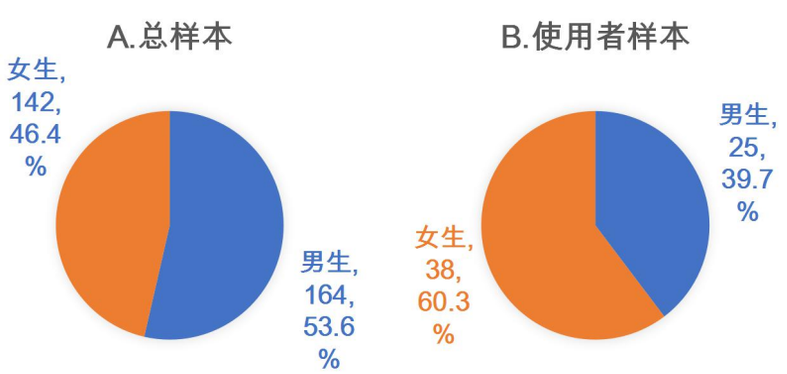

图 6　性别分布情况图

由图 6 可知，在被调查的正在使用运动类 App 的群体中，男性所占比例为

39.68%，女性为 60.32%；而在整体样本中，53.6%为男性，46.4%为女性。由此可以看出，在使用运动类 App 的人群中，女性所占比例远高于男性，大约为其两倍。性别上存在了明显的使用需求差异，这在一定程度上表明，在实际生活中，女性更愿意使用运动类 App 进行体育锻炼。

（二）各年级使用者占比情况（第 2 题）

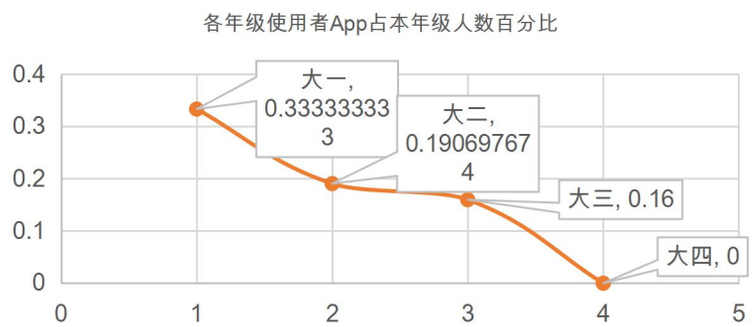

图 7　年级分布图

从图 7 反映出的数据情况来看，在被调查的各年级在校大学生中，大一和大二的大学生大多都会使用运动类 App。大三及大四大学生中使用运动类 App 的占年级被调查者总数较少，这主要是因为一二年级在校大学生相比较高年级大学生，有更多的课余空闲时间和更强的锻炼意识。高年级的在校大学生面临着毕业工作或者升学深造的压力，其可支配剩余时间较低年级在校大学生而言较为不足。

（三）使用比例及意愿分布情况（第 12 题）

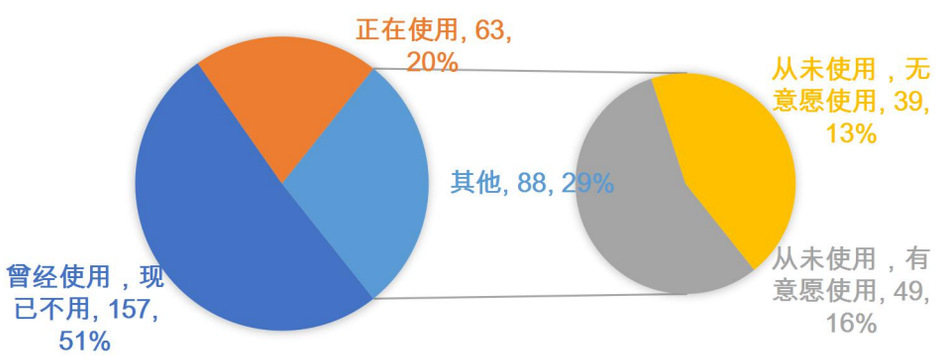

图 8　使用情况及意愿分布图

从图 8 中可以看出有 15.91%的人没有使用过运动类 App 但有意愿在未来使用运动类 App。综合图 8 所得出的数据，合肥市在校大学生中，有 71.43%的人曾经使用或正在使用运动类 App，这一数据也能够反映出合肥市在校大学生对运

动类 App 较为熟悉。对于未使用过运动类 App 的在校大学生学生而言，超过 50%的人选择愿意对运动类 App 进行尝试，这也说明了运动类 App 对在校大学生有相当的吸引力。

（四）每周使用频率情况（第 13 题）

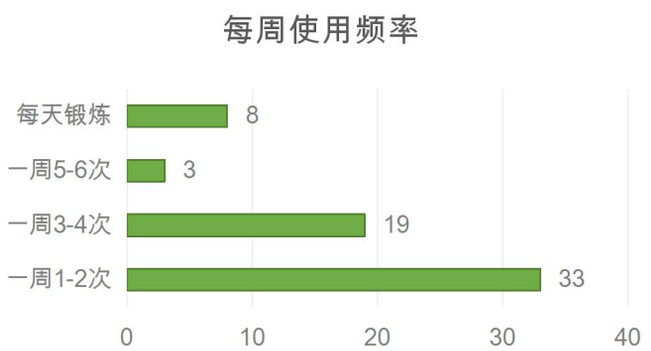

图 9　每周使用频率分布图

在使用运动类 App 的群体中，每周锻炼 1-2 次的群体占比 52.38%，每周锻炼 3-4 次的群体占比 30.16%，而每周运动 5-6 次的使用者为 4.76%，每天坚持锻炼的群体比例只占到了 12.70%。由此可见，使用运动类 App 来进行体育锻炼的人大部分都将每周锻炼的频率保持在每周 1-2 次，绝大多数的人会在一周内进行不超过 5 次的体育锻炼，进一步可以说明，运动类 App 的使用者有一定的体育锻炼意识，且能够将锻炼保持适度的频率。

（五）每次使用时间分布情况（第 14 题）

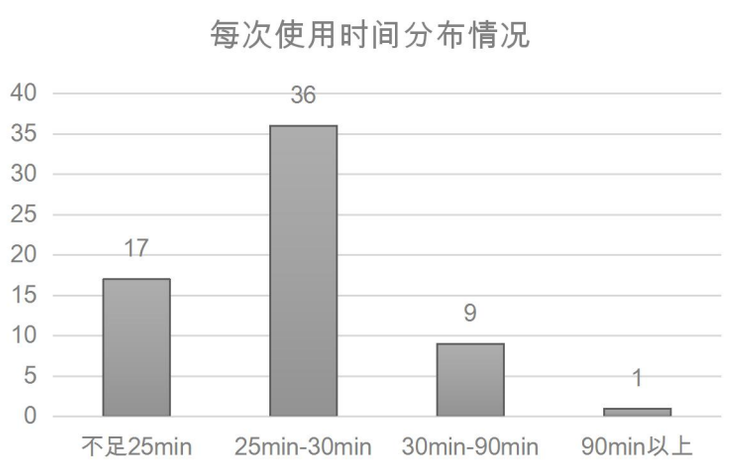

图 10　每次使用情况分布图

使用运动类 App 进行体育锻炼的群体中，每次使用的时长在 25min-30min 的人群占比最多，高达 57.14%；不足 25min 的使用者占比 26.98%；锻炼时间超过 30 分钟的人很少。由此可见，使用运动类 App 进行锻炼的人基本能够将锻炼的时长保持在合理的范围内。

（六）使用产品分布情况（第 16 题）

图 11　使用产品分布图

从图 11 可以了解合肥市高校大学生使用的运动类 App 产品的使用情况。Keep 为最常用的运动类 App，有 68.25%的大学生都在使用，其次为悦跑圈、小米运动等产品。另外有 15.87%的学生使用其他产品进行体育锻炼，经了解，主要是一些可穿戴设备，如小米手环等。随着科技的进步和生产力的发展，越来越多的运动类产品通过各种各样的方式和渠道为人所知并且开始逐渐被使用。首先运动类 App 作为以智能手机为载体的智能应用程序，其便携性、稳定性、快捷性容易被在校大学生所接受，而智能手环、智能手表等智能产品在大学生中的使用相对很少，这主要是由于大学生的经济水平不高，且对手机的依赖性强。

（七）使用功能的分布情况（第 17 题）

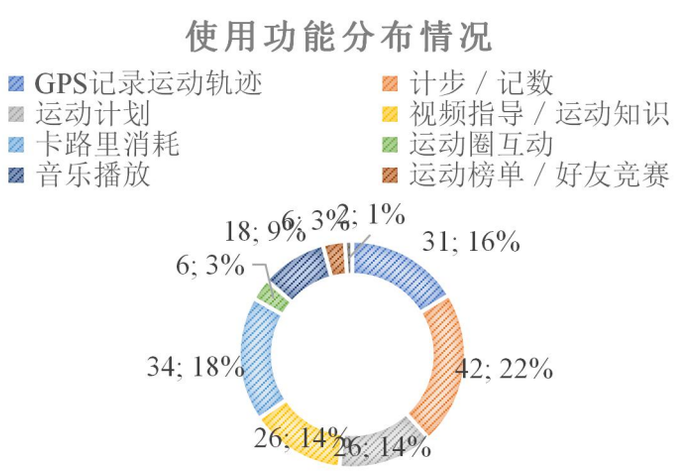

图 12　使用情况分布图

由图 12 可知，合肥市大学生使用运动类 App 主要功能还是用于记录步数、记录卡路里消耗和运动轨迹这些相关的运动数据，运用其进行健身指导和学习运动技能的人也较多，而朋友圈互动和好友竞赛这些功能不经常被使用。这符合合肥市在校大学生对运动类 App 功能重要程度的调查分析结论。

（八）运动类 App 功能满意度调查（第 18 题）

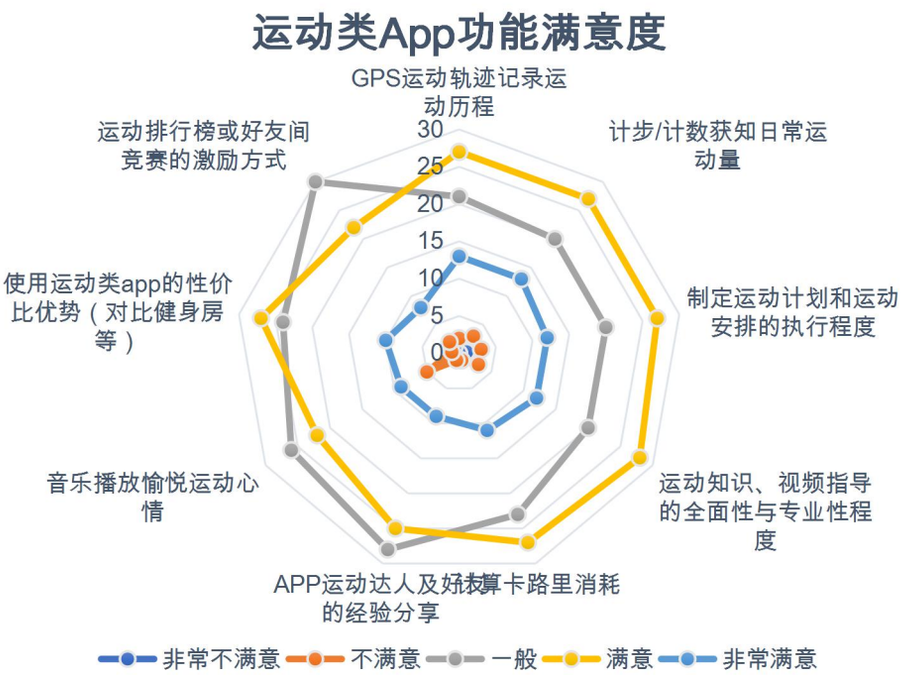

图 13　运动类 App 功能满意度雷达图

可以看出，使用运动类 App 的人对 9 种功能均达到了认可的态度。其中对“GPS 运动轨迹记录运动历程”这一功能最为认可，对“播放音乐”这一功能满意度相对较低，希望其进行改善，其他功能的满意程度差别不大。

（九）使用者对运动类 App 感知调查（第 19 题、20 题）

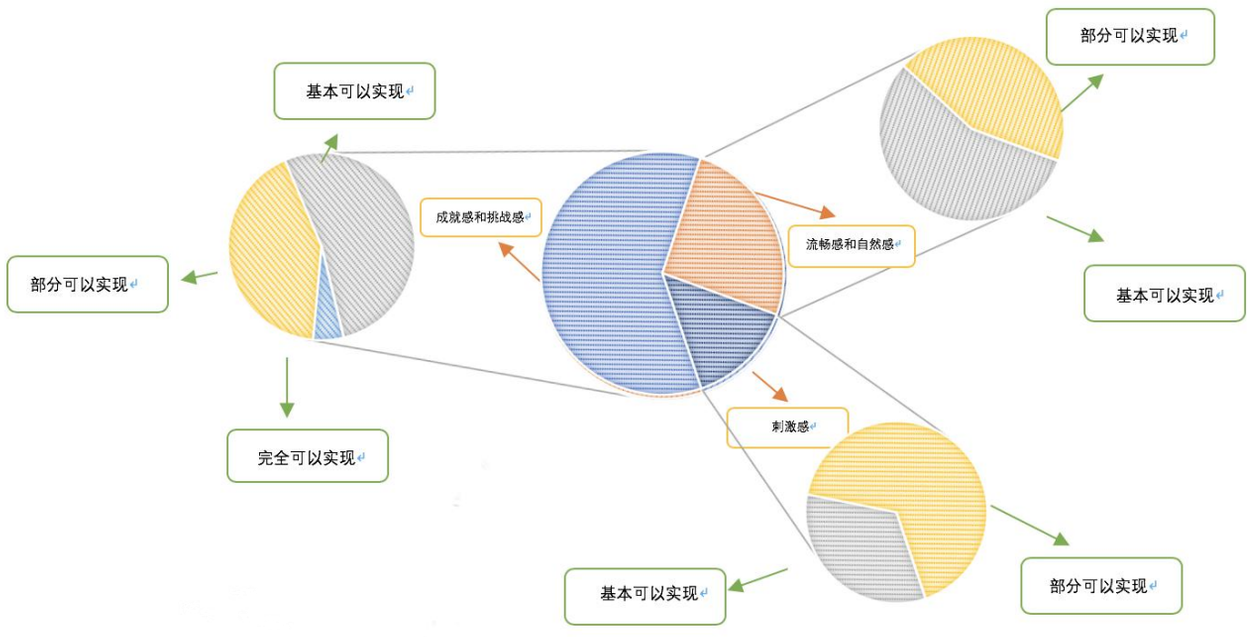

图 14　使用者对运动类 App 感知调查图

通过对使用者对运动类 App 的感知与满足程度的调查，可以看出有 60.32%的使用者认为通过使用运动类 App 来完成运动计划、在运动圈内分享胜利成果等可以给其带来的自我成就感和挑战感。较少的使用者认为运动类 App 中的竞争排名、每日目标等环节可以为其带来刺激感这一体验。进一步分析可知，在满足使用者成就感与挑战感方面，大部分人认为现有运动类 App 基本可以实现，少部分人认为完全可以实现；在满足使用流畅感与自然感方面，大部分使用者认为基本可以实现，其余认为部分可以实现；另外，大部分人认为现有运动类 App 功能可以部分实现其使用过程中所寻求的刺激感。

（十）使用者的学习需求方向（第 22 题）

图 15　使用者的学习需求方向分布图

由图 15 可知，在运动类 App 运动指导的专业区域方面，使用者大部分会选择运动训练和健身塑型这两个专业区域进行学习，对运动装备的学习需求较少，且冬季运动、运动损伤防护的专业知识需求最低。随着人们生活水平的提高，健身已经成为一种时尚的生活方式，健身塑形也成为一种潮流，但人们大多不会关注运动中的损伤防护，相关知识较为欠缺。

（十一）运动类 App 贡献及现有缺陷（第 21、23 题）

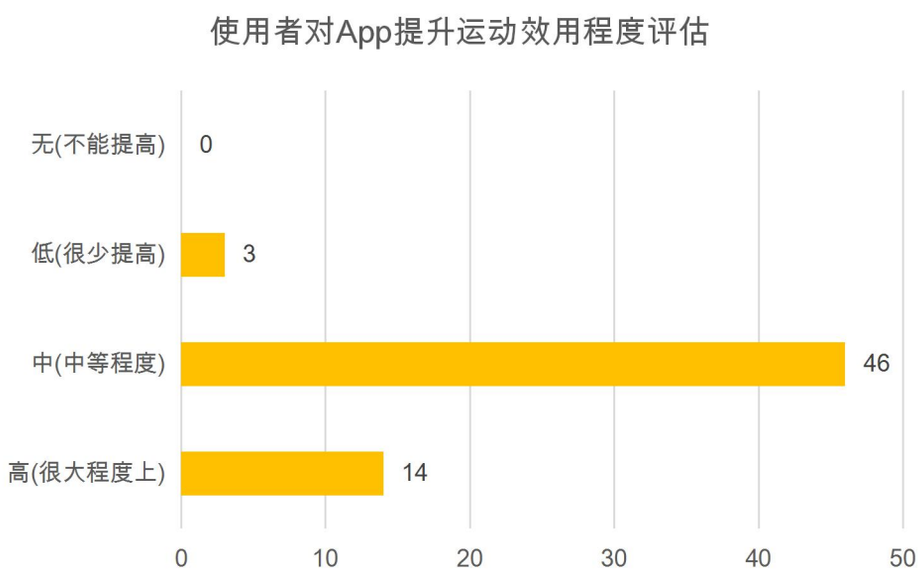

图 16　使用者对 App 提升运动效用程度评估分布图

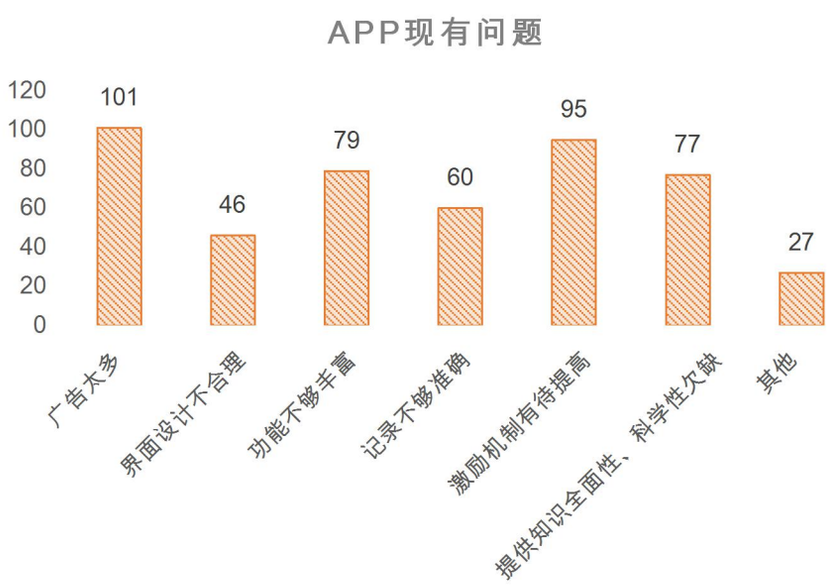

图 17　使用者对 App 现有问题分布图

由图 16 可知，73.02%的使用者认为运动类 App 能够对其运动效果的提升有一定的贡献，但是其在广告数量的控制、运动激励形式的改善、功能的扩展和丰富等方面需要提高和加强。可以看出，正在使用运动类 App 的用户数据不理想，这表明在校大学生认为运动类 App 的使用尚存在提高空间，曾经使用过运动类 App 但现在放弃使用的群体大都认为，这一运动辅助工具在运动激励机制方面还有待改善，且现有运动类 App 广告太多，影响大学生使用体验，功能也不够丰

富和新颖。

第四部分 基于新型决策树算法的运动类 App 使用行为研究

一、基于随机森林的使用者主要影响因素研究

（一）模型的选择

1. 调查数据特征

“是否坚持使用运动类 App”是二分类问题，并且此问题的结果由使用者多个特征综合决定。但是收集的数据不符合具有严格假设的传统统计计量模型，主要表现在以下两个方面：

① 通过对数据进行共线性诊断，结果显示自变量间存在严重共线性；

② 对数据进行正态性检验，输出结果为 Sig 值均小于 0.05，所以认为数据不属于严格的高斯分布。

查阅资料可知，一般用传统的数据分析方法如多元 Logistic 回归分析，对数据处理具有严格的假设前提。但通过对数据进行验证可以看出，收集的数据不符合传统的具有严格假设的分析方法，所以我们选取随机森林对数据进行分析，使得出的结果更具有稳健性。

2. 随机森林优势

随机森林是由 Leo Breiman[6] 和 Adele Cutler 发展推论出来的一种基于分类树的算法，在变量（列）的使用和数据（行）的使用上进行随机化，生成很多分类树，再汇总分类树的结果，是能较好取代传统机器学习方法的新的模型。

它有许多优点[7] ：

① 对于很多种资料，它可以产生高准确度的分类器；

② 它可以处理大量的多达几千个的自变量；

③ 它可以在决定类别时，评估变数的重要性不需要考虑在一般回归问题中发生的多元共线性的问题；

④ 它包含估计缺失值的算法，例如：如果有很大一部分的数据失去，仍可以维持较高的准确度；

⑤ 对于非平衡的分类数据集来说，它可以较好的平衡误差；

其算法流程如 18 图所示：

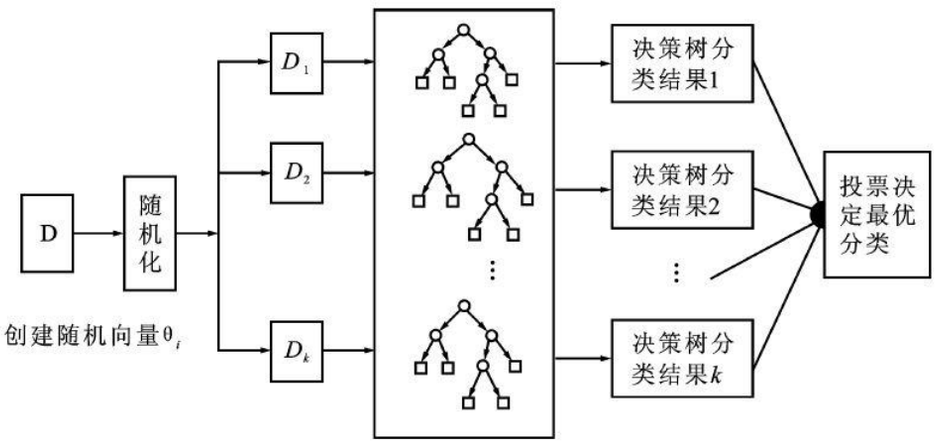

图 18　随机森林算法流程图

在随机森林模型进行工作时，利用 Bootstrap 从原始数据集中抽取与原始样本量相同的 k 个样本作为训练集，未抽取到的 30%-40%的数据为 OOB 数据，形成天然测试集。通过对 k 个样本分别建立决策树模型，得到 k 种分类结果。最后采用简单多数投票法对 k 种分类结果进行投票得到最终分类结果：

多分类模型系统：　$\{h_1(x), h_2(x), h_3(x), \cdots, h_k(x)\}$

最终分类决策：　$H(X) = \arg\max_Y \sum_{i=1}^{k} I\left(h_i(X) = Y\right)$

其中，X 表示给定自变量，$h(x)$ 为单个决策树分类模型，Y 表示输出变量，$I\left(h_i(X) = Y\right)$为示性函数。

（二）随机森林变量预处理

1. 类别标签的确认。将“您是否正在使用运动类 App”这种二元选择行为表示为分类标签（即因变量），当选择“是”时，取值为 1；否则，取值为 0。

2. 为了问卷中的特征值属性（如满意，不满意等），计算机难以识别处理，此处依照问卷中选项的顺序对每一个特征值属性的具体内容进行数值转化，如：在性别属性中，1 代表男性，2 代表女性。

（三）模型参数的确定与程序运行

运动类 App 使用情况可看作是一个二分类的分类因变量（正在使用=1，否

则=0），为了使模型准确率更好，在将其转化为因子型变量时，本文通过使用 Python 超参数自动搜索模块 GridSearchCV 进行部分参数调优，选取的最优参数如下(未提到的说明最优为缺省值)：

```
n_estimators=529,min_samples_split=88,min_samples_leaf=1,max_depth=5,min_samples_leaf=1
```

基于以上参数，运行随机森林分类模型可以得到训练集的分数为 0.778 测试集的分数为 0.824，说明模型训练效果较好。 得出特征值相对重要性排序如表 10：

表 10　基于随机森林模型的特征值重要性排序表

特征值	重要性	特征值	重要性
空闲时间	0.0856	锻炼兴趣	0.0146
体育锻炼的频率	0.0787	使用运动类 app 的性价比优势（对比健身房等）	0.0142
运动环境	0.0632	体育锻炼的强度	0.0129
视频指导/运动知识	0.0576	卡路里消耗	0.0123
性别	0.0554	篮球	0.0121
排球	0.0410	音乐播放愉悦运动心情	0.0119
计算卡路里消耗	0.0324	制定运动计划和运动安排的执行程度	0.0108
计步/计数获知日常运动量	0.0305	运动计划	0.0102
运动排行榜或好友间竞赛的激励方式	0.0294	场地器材	0.0083
提供课程性价比高	0.0286	运动榜单/好友竞赛	0.0078
运动圈互动	0.0286	羽毛球	0.0076
App 运动达人及好友的经验分享	0.0281	媒体宣传	0.0065
瑜伽	0.0255	专业辅导	0.0057
喜欢的颜色系	0.0236	跑步	0.0051
体育锻炼的时间	0.0231	跳绳	0.0045
运动知识、视频指导的全面性与专业性程度	0.0226	网球	0.0032
GPS 记录运动轨迹	0.0222	其他	0.0020
集体效应	0.0215	兵乓球	0.0018
年龄	0.0190	健美操	0.0012
计步/计数	0.0186	游泳	0.0010
GPS 运动轨迹记录运动历程	0.0161	足球	0.0005
音乐播放	0.0150		

为了更加直观地认识到影响运动类 App 使用行为的重要性排序，将因素重要性排序输出为重要性排序图：

图 19　特征因素重要性排序图

（四）模型的检验与评价

为了对模型的分类效果和准确度进行评估，本文选取两个评估指标进行检验与评价：

1. AUC 分数

AUC（Area Under Curve of ROC）指接收者操作特征曲线（receiver operating characteristic curve，或者叫 ROC 曲线）下与坐标轴围成的面积。在机器学习常常用 AUC 判断分类器（预测模型）优劣的标准。分类器的 AUC 值等价于将随机选择的正样本排序在随机选择的负样本之前的概率。（原文：The AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.[8]）AUC 越大，说明该分类器分类效果更好。

借助 Python 的 sklearn 库中的 metrics，通过 metrics.roc_auc_score()语句可以得出调查数据集基于随机森林分类器后的训练结果：AUC Score (Train): 0.870019

2. OOB 分数

袋外（OOB）误差，也称为袋外估计，是一种利用自助聚合（装袋）到子样本来测量随机森林，提升决策树和其他机器学习模型的预测误差的方法用于培训的数据样本。 OOB 是每个训练样本 x_i 的平均预测误差，仅使用在其自举样本中没有 x_i 的树[9]。

子采样允许通过评估在下一个基础学习者的构建中未使用的那些观察的预测来定义预测性能改进的袋外估计。袋外估计有助于避免需要独立的验证数据集，但往往低估了实际的性能改进和最佳迭代次数[10]。

借助 Python 的 sklearn.ensemble 库中的 RandomForestClassifier 语句，通过.obb_score 得出的结果。它是用来计算正确分类的平均值[11]，正确分类分数 oob_sore：0.7780。

综上两个指标可以看出：本文为处理数据建立的训练的模型预测正确率很高。模型的分类正确分数较高，表明所选自变量对因变量的预测精度较高，各个自变量能在一定程度上解释合肥市在校大学生对运动类 App 的使用选择。

（五）模型结果分析——随机森林变量相对重要性分析

对是否使用运动类 App 的影响因素有很多，根据随机森林对影响因素的排序，选取特征值大于 0.03 的因素进行分析：

1. 是否将空闲时间看作影响进行体育锻炼的重要因素

空闲时间不是影响运动锻炼的必然因素。有良好运动习惯的人会坚持锻炼，且能够合理地安排时间，保证每周的锻炼量；而更多的人选择的是一种碎片化锻炼的锻炼方式，这种方式以拥有空闲时间为前提。运动 App 以其游戏化、社交化和碎片化的方式，鼓励人们形成良好的运动习惯和生活方式，从而获得身体的健康，这种运动工具可以帮助大学生一站式解决锻炼问题，满足锻炼需求，从而更加适合于没用明确锻炼计划和良好运动习惯的大学生。

2. 参加体育锻炼的频率

参加体育锻炼的频率对是否使用运动类 App 有很大影响。运动类 App 能够提高训练活动的效率。从而可以更好的解决大学生运动锻炼遇到的问题，主要有以下两点：第一是满足拥有较高频次锻炼需求的大学生进行体育锻炼，为其提供随时可获取的运动资源；第二是吸引不经常锻炼的大学生更多地参与进来，进行体育运动锻炼。

3. 对运动环境的偏向性

不同的人对环境的适应能力不同，对运动环境的偏向性也不同。有的人会选择在户外健身，有的人偏向室内专业指导；有的人喜欢独自健身，还有的人喜欢集体运动。运动类应用程序依附于手机移动终端，它的这一特点带给了其使用者更加自由的使用体验，营造出更加舒适的运动场景，这也是其区别与其他运动方式的一大突出优势。运动类 App 可以使得使用者在任何地方都可以通过 App 合理利用碎片化时间进行健身，更加适合于喜欢在空余时间、任意地点进行自我约束型锻炼的大学生人群使用。

4. 对视频指导/运动知识的提供的重视程度

运动类 App 能够通过互联网搜集大量的运动知识，例如健身计划的制定、健身过程中安全问题、健身过程中营养与保健等方面的知识，并可以通过视频讲解生动的传达知识。对运动知识和视频指导较为重视的人出于获取知识的动机，会更加偏向通过对运动类 App 获取自己所需的相关专业知识和指导信息，满足其对于健身知识方面的需求。并且运动类 App 对运动量、运动强度、运动频次等方面的科学又具体的安排，可以使得使用用户方便地获得在线下不易获得的个

性化数据。

5. 性别

性别与是否使用运动类 App 有一定的联系。理论上来说，男性的运动兴趣、爱好比女性用户更大，且男性所展示的运动天赋超过女性。对于新兴 App 门类，男性的体验欲望要超过女性，运动类 App 作为新兴的 App 门类，对男性用户更具有吸引力。但从另一个层面来说，现代女性更加重视体型匀称、气质文雅、步态轻盈，所以使用运动 App 这种方便简易的实用工具来进行运动锻炼是她们不错的选择，女性选择运动类 App 进行锻炼越来越成为一种趋势，这也与我们调查结果相符合。

6. 对计算卡路里消耗和计步获知日常运动量的重视程度

是否将计算卡路里消耗和计步获知日常运动量看作是重要功能与是否使用运动类 App 有一定的联系。许多健身 App 都把健身活动自动记录功能作为基础功能，例如计算运动的卡路里消耗、记录日常运动量、记录运动频率等。移动存储的优势使得人们在健身时健身数据能够通过 App 存储在云端，不用担心自己的记录会丢失，这些记录功能，能够有效的帮助用户直观的感受到运动带来的效果，这也是其区别与一般传统健身方式的一大特点。在日常生活中注重将自己的运动信息记录下来的人很大程度上会偏向于使用运动 App。

二、基于 GBDT 算法的使用者行为预测研究

上文已经对影响运动类 App 使用的因素重要性进行了分析，为了进一步改善运动类 App 的使用市场、实现运动类 App 供应方的针对性营销和精准推送，本节将根据现有数据建立合理的运动类 App 使用者预测模型，挖掘潜在使用者，在不同的未来客户群体中做出精准预测，实现运动类 App 市场的快速发展。

（一）模型的选择

GBDT（Gradient Boosting Decision Tree） 又叫 MART（Multiple Additive Regression Tree)，是一种采用决策树或回归树作为弱分类器的梯度提升算法，该算法是多棵决策树组成的，所有树的结论累加后做出最终答案。它允许用多个不同的判决条件将不同特征关联起来，适合于产生不同结果的特征的联合[12][13]。

GBDT 算法是对一个二元分类，通过弱分类器评价每个节点可观测误差，根据测试函数$k:R^n \to R$的阈值τ的分割节点，返回当前节点的左右孩子η^l和η^r。之后通过寻找三元组(τ,η^l,η^r)最小化分割后的误差，得出最佳分割点。算法对所

有树的节点进行分裂，逐次减少分割误差，最终实现较优分类。其计算如下所示：

$$\varepsilon(\tau)=\sum_{i:k(X_i)<\tau} w_i^j\left(r_i^j-\eta^l\right)^2+\sum_{i:k(X_i)<\tau} w_i^j\left(r_i^j-\eta^r\right)^2$$

$X\in R^n$ 是输入向量，w_i^j 和 r_i^j 代表第 j 次迭代的 X^i 的权重和输出值，可由下面公式计算得到：

$$w_i^j=\exp\left(-y_i f_{i-1}\left(X_i\right)\right),$$

$$r_i^j=g\left(X_i\right)/w_i^j=-y_i\exp\left(-y_i f_{i-1}\left(X_i\right)\right)/w_i^j=-y_i$$

（二）数据预处理

前文已经基于随机森林对特征因子的重要性进行了分析，为进行模型运作，剔除重要性小于 0.01 的特征值，剩下图 19 重要性排序图中前 30 个特征因素作为全部特征。将以全部特征值(因变量)以及用户是否正在使用 App（因变量）作为数据集。

（三）用户使用运动类 App 预测分析

为实现为优化“用户是否使用运动类 App”这一预测模型、实现对合肥市在校大学生使用运动类 App 的精准预测，先采用前 30 的特征进行预测，然后选取重要性排序前 20 的特征进行预测，将二者得出的混淆矩阵结果进行对比，选取准确度较高、较为节约成本的一种。

1. 评价指标

在对预测的结果进行分析时，选取一些能够描述模型准确度的指标：

Accuracy（准确率）：分类器对整个样本的判定能力，即将实际使用为会不使用、将实际不使用的判定为不使用的比例，计算公式如下：

$$Accuracy=\frac{tp+tn}{tp+tn+fp+fn}$$

Precision（精度）：预测结果中预测正确的判定能力，即预测会使用的人实际也会使用运动类 App、预测不会使用的人实际也不会使用运动类 App 的比例，计算公式如下：

$$Precision = \frac{tp}{tp + fp}$$

其中 TP，FP，FN，TN 分别表示：实际会使用运动类 App 预测为会使用运动类 App，实际不使用运动类 App 预测为会使用运动类 App，实际会使用运动类 App 预测为不会使用运动类 App，实际不会使用运动类 App 预测为不会使用运动类 App。

2. 预测结果

① 选择训练集与测试集 80:20 比例的预测结果：

表 11　训练集与测试集 80:20 时两种特征选择下的混淆矩阵

	Accuracy	*Precision*	*TP*	*FP*	*FN*	*TN*
前 20 个特征	0.9332	0.8618	131	21	31	596
前 30 个特征	0.9461	0.8704	141	21	21	596

比较分类结果和实例的真实信息可以看出，在使用 80%的样本作为训练集而用 20%的样本来测试时，采用全部特征的分类模型准确率达到了 94.61%，精度达到了 0.8704，但是此模型将一部分实际不使用运动类 App 预测为会使用运动类 App、实际会使用运动类 App 预测为不会使用运动类 App，预测能力有一定的缺陷。而运用前 20 个特征的分类模型准确率达到了 93.32%，精度达到了 0.8618，相比使用全部特征进行预测，准确率降低了 1.29%，精度降低了 0.0086，且将实际会使用运动类 App 预测为不会使用运动类 App 的个数相对增加，模型精准性有所降低。

② 选择训练集与测试集 70:30 比例的预测结果：

表 12　训练集与测试集 70:30 时两种特征选择下的混淆矩阵

	Accuracy	*Precision*	*TP*	*FP*	*FN*	*TN*
前 20 个特征	0.9166	0.8255	123	26	39	591
前 30 个特征	0.9371	0.8792	131	18	31	599

在使用 70%的样本作为训练集而用 30%的样本来测试时，利用全部特征的分类模型的准确率达到了 93.71%，精度达到了 0.8792。而利用前 20 个特征的分类模型准确率达到了 91.66%，精度达到了 0.8255，相比使用全部特征进行预测，准确率降低了 2.05%，精度降低了 0.0537，相比较运用全部特征进行预测，预测错误的个数均有所增加，模型精准性部分较低。

③ 选择训练集与测试集 60:40 比例的预测结果：

表 13　训练集与测试集 60:40 时两种特征选择下的混淆矩阵

	Accuracy	*Precision*	*TP*	*FP*	*FN*	*TN*
前 20 个特征	0.8588	0.7203	85	33	77	584
前 30 个特征	0.9281	0.8897	121	15	41	602

在使用 60%的样本作为训练集而 40%的样本来测试时，利用全部特征的分类模型的准确率达到了 92.81%，精度达到了 0.8897，但是此模型也存在预测错误的情况。利用前 20 个特征的分类模型精确度只达到了 85.88%，精度达到了 0.7203，相比使用全部特征进行预测，准确率降低了 6.93%，精度降低了 0.1694，分类错误大幅增加，模型精准性有一定程度的降低。

综上所述，采用所有的特征对合肥市在校大学生使用运动类 App 的行为进行预测是相对比较准确的，且平均准确率可以达到 93.71%，平均精度达到了 0.8797，能够用于进行用户画像并对大学生使用行为进行预测，挖掘潜在消费者实现运动类应用程序的精准推送。

图 20　用户画像

图 21　产品画像

基于用户画像我们可以得出运动类 App 的用户大多具有以下特征：每天锻炼、偏向的运动环境是网络、app 辅导、女生、锻炼喜欢排球、锻炼喜欢瑜伽、暖色系、体育锻炼 30 分钟，60 分钟、体育锻炼 2 小时以上、大学一年级学生、每次进行体育锻炼的强度出汗较多、很多、锻炼喜欢篮球。

基于运动类 App 的产品画像我们可以得出运动类 App 大多具有以下功能：视频指导/运动知识、计算卡路里消耗、计步/计数获知日常运动量、运动排行榜或好友间竞赛的激励方式、提供课程性价比高、运动圈互动、App 运动达人及好友的经验分享、运动知识、视频指导的全面性与专业性程度、GPS 运动轨迹。

结　论

一、合肥市在校大学生大多使用过运动类 App，且用户之间存在差异

超过半数的合肥市在校大学生曾经使用过或正在使用运动类 App，且使用者多为大一大二的女性群体。使用的频次集中在一周 5 次以内，一次运动时间不超过 30 分钟。通过对使用者特征的分析可以看出，外部客观环境对用户使用运动类 App 具有一定制约性。

二、与记录运动信息相关的功能更被在校大学生所青睐

通过对合肥市在校大学生对运动类 App 的功能认知调查可以看出，大学生更加重视与运动数据记录和运动知识获取有关的基本功能，如“运动知识、视频指导的全面性与专业性程度”、“制定运动计划和运动安排的执行程度”与“计步/计数获知日常运动量”。而在使用者中，大多数人经常使用的功能也是记录步数、记录卡路里消耗和运动轨迹这些基本功能。

三、运动类 App 对合肥市在校大学生的锻炼效果具有一定的影响

由分析可知，使用运动类 App 的大学生，认为现有运动类 App 能够基本满足其在运动过程中的相关需要。运动类 App 在运动频次、强度和运动积极性方面，对合肥市高校使用者有积极影响，一定程度上提高了校园运动的效果。

四、合肥市在校大学生对运动类 App 的满意程度普遍较高

运动类 App 不仅能够为使用者带来快捷方便的运动体验，还可以提高运动效率和兴趣，学习到专业知识，并且好友相互监督可以更有效率地实现运动目标。合肥市在校大学生对运动类 App 现有的功能基本持满意态度，只在“运动时播放音乐”这一功能上满意度有所降低。

五、空闲时间、锻炼的频率、偏爱的运动环境对是否使用运动类 App 的影响显著

通过建立随机森林模型对数据进行分析，可以看出大学生是否有空闲时间、

参加运动锻炼的频率和偏爱的运动环境是影响合肥市在校大学生是否使用运动类 App 进行体育锻炼的三个最显著因素。其次是对视频指导/专业知识的重视程度、性别和对运动过程相关信息记录的重视程度。

建　议

一、找准产品定位，明确研发方向

当前，我国运动类 App 软件数量已经破百，市场呈现一派繁荣景象。但在繁荣的表面下，是产品的剧烈同质化现象。绝大多数 App 都试图通过社交机制来提高用户粘性。而通过此次调查研究，可得知社交功能对用户持续使用运动类 App 不具有显著影响和作用。因此，App 研发公司应找准产品定位，根据用户需求进行精准产品研发。

二、完善产品功能，增强用户粘性

通过调查结果显示，运动类 App 用户的卸载率很高，即用户粘性低。而用户卸载的原因主要是由于广告问题和产品功能不够丰富，感知有用性和满意度是用户接受 App 的决定性因素，而目前运动类 App 并不能满足用户的需求，因此，优化有用性体验，提升用户满意度，增强用户的持续使用意愿是目前研发公司的首要任务。

三、明晰用户群体，精准推送产品

通过用户画像，明晰产品的主要用户群体。以此为基准，对该群体展开市场调研，了解主要用户的基本信息，对该用户群体进行产品推送与宣传，提高宣传活动的有效性，节约企业的宣传费用。

参考文献

[1] 林瑶瑶, 魏雪蕊. 运动健身类 APP 用户持续使用意愿影响因素的研究[J]. 数学的实践与认识, 2019, 49(04): 61-65.

[2] 刘璐. 健身类 APP 在全民健身运动中的作用及发展趋势研究[J]. 科技展望, 2016, 26(31).

[3] 李霞, 许叶, 黄中校. 运动健康类 APP 发展现状调查及分析[J]. 体育科技, 2017, 38(2), 32-33

[4] 李柔. “互联网+” 影响下运动类 APP 的发展前景分析[J]. 文体用品与科技, 2018 (11): 21.

[5] 陆佳莉. 体育运动健身类 APP 的现状及对策研究[J]. 辽宁体育科技, 2017, 39(06), 20-23

[6] Breiman L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.

[7] 李欣海. 随机森林模型在分类与回归分析中的应用[J]. 应用昆虫学报（昆虫知识）, 2016, 50(04): 1190-1197.

[8] Fawcett T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861-874.

[9] James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert (2013). An Introduction to Statistical Learning. Springer. pp. 316–321.

[10] Ridgeway G. Generalized Boosted Models: A guide to the gbm package[J]. Update, 2007, 1(1): 2007.

[11] Segovia F, Górriz J M, Ramírez J, et al. Early diagnosis of Alzheimer’s disease based on partial least squares and support vector machine[J]. Expert Systems with Applications, 2013, 40(2): 677-683.

[12] Son J, Jung I, Park K, et al. Tracking-by-segmentation with online gradient boosting decision tree[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 3056-3064.

[13] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of statistics, 2001: 1189-1232.

附　录

附录 1　调查问卷

大学生运动类 App 使用现状及影响调查问卷

您好！我们是**大学的学生，为了参加“2019 年（第六届）全国大学生统计建模大赛”，以“运动类 App 的使用现状及前景分析”为主题开展此次调查活动。本次调查将严格按照《统计法》的要求进行，您的隐私会严格保密，所有问答只用于本次比赛所需的统计分析。衷心感谢您的支持与帮助！

一、基本信息

1.您的性别是: * [单选题] *

○A.男生

○B.女生

2.您所在的年级是: * [单选题] *

○A.大一

○B.大二

○C.大三

○D.大四及以上

3.您更加喜欢的颜色系是: * [单选题] *

○A.冷色系（黄绿、绿、蓝绿、蓝、蓝紫等）

○B.暖色系（红紫、红、粉红、橘、黄橘、黄、咖啡色等）

○C.中间色系（紫、绿、黑、白、灰等）

4.您平时参加体育锻炼的频率是: * [单选题] *

○A.几乎不

○B.一周 1-2 次

○C.一周 3-4 次

○D.一周 5-6 次

○E.每天锻炼

5.您每次进行体育锻炼的时间是: * [单选题] *

○A.30 分钟以内

○B.30 分钟—60 分钟

○C.1 小时—2 小时

○D.2 小时以上

6.您每次进行体育锻炼的强度是: * [单选题] *

○A.出汗很多

○B.出汗较多

○C.稍微出汗

○D.基本不出汗

7.您偏向的运动环境是: * [单选题] *

○A.健身房

○B.集体健身活动（广场舞、集体骑行等）

○C.私人教练服务

○D.网络、App 辅导

8.请勾选您进行体育锻炼最常所选用的项目: * [多选题] *

□A.跑步

□B.篮球

□C.羽毛球

□D.兵乓球

□E.排球

□F.网球

□G.足球

□H.健美操

□I.跳绳

□J.瑜伽

□K.游泳

□其他 _____

二、认知情况

9.您认为影响您参加体育锻炼的因素重要性排序是: [排序题，请将最重要的放置在第一位] * [排序题，请在中括号内依次填入数字] *

[ ]A.锻炼兴趣

[ ]B.空闲时间

[ ]C.场地器材

[ ]D.专业辅导

[ ]E.媒体宣传

[ ]F.集体效应

[ ]G.其他

10.您认为运动类 App 应该具有的功能排序（按重要程度）: [排序题，请将最重要的放置在第一位] * [排序题，请在中括号内依次填入数字] *

[ ]GPS 记录运动轨迹

[ ]计步/计数

[ ]运动计划

[ ]视频指导/运动知识

[ ]卡路里消耗

[ ]运动圈互动

[ ]音乐播放

[ ]运动榜单/好友竞赛

[ ]提供课程性价比高

[ ]其他

11.您对运动类 App 在以下属性方面的重要性程度的看法是: *[矩阵单选题] *

	非常不重要	不重要	一般	重要	非常重要
GPS 运动轨迹记录运动历程	○	○	○	○	○
计步/计数获知日常运动量	○	○	○	○	○
制定运动计划和运动安排的执行程度	○	○	○	○	○
运动知识、视频指导的全面性与专业性程度	○	○	○	○	○
计算卡路里消耗	○	○	○	○	○
App 运动达人及好友的经验分享	○	○	○	○	○
音乐播放愉悦运动心情	○	○	○	○	○
使用运动类 App 的性价比优势（对比健身房等）	○	○	○	○	○
运动排行榜或好友间竞赛的激励方式	○	○	○	○	○

三、使用情况

12.您的运动类 App 使用经历是: * [单选题] *

○A.曾经使用，现已不用 (请跳至第 23 题) (请跳至第 23 题)

○B.正在使用

○C.从未使用，有意愿使用 (请跳至第问卷末尾，提交答卷) (请跳至第问卷末尾，提交答卷)

○D.从未使用，无意愿使用 (请跳至第问卷末尾，提交答卷) (请跳至第问卷末尾，提交答卷)

13.您使用运动类 App 的频率是: * [单选题] *

○A.一周 1-2 次

○B.一周 3-4 次

○C.一周 5-6 次

○D.每天锻炼

14.您每次使用的运动类 App 时间是: * [单选题] *

○A.不足 25min

○B.25min-30min

○C.30min-90min

○D.90min 以上

15.您坚持使用运动类 App 多久了: * [单选题] *

○A.不足半年

○B.半年到一年

○C.一年以上

16.您目前使用的 App 类型: * [多选题] *

□A.Nike Training Club　□B.Keep　□C.咕咚

□D.悦跑圈　□E.乐动力　□F.糖豆广场舞

□G.小米运动　　□H.其他

17.在下列运动类 App 功能中，您最常使用的是: * [多选题] *

□A.GPS 记录运动轨迹

□B.计步/记数

□C.运动计划

□D.视频指导/运动知识

□E.卡路里消耗

□F.运动圈互动

□G.音乐播放

□H.运动榜单/好友竞赛

□I.其他

18.您对运动类 App 在以下属性方面的满意程度是: *[矩阵单选题] *

	非常不满意	不满意	一般	满意	非常满意
GPS 运动轨迹记录运动历程	○	○	○	○	○
计步/计数获知日常运动量	○	○	○	○	○
制定运动计划和运动安排的执行程度	○	○	○	○	○
运动知识、视频指导的全面性与专业性程度	○	○	○	○	○
计算卡路里消耗	○	○	○	○	○
App 运动达人及好友的经验分	○	○	○	○	○

享					
音乐播放愉悦运动心情	○	○	○	○	○
使用运动类 App 的性价比优势（对比健身房等）	○	○	○	○	○
运动排行榜或好友间竞赛的激励方式	○	○	○	○	○

19.您认为您在运动中通过使用以上最常用的功能可以给您带来什么样的体验: * [单选题] *

○A.刺激感(完成目标，从视频指导、获取运动知识等过程中给您带来想去运动的刺激，以及在好友圈运动排行榜中刺激您去竞争的感觉等)

○B.成就感和挑战感(完成运动运动计划，运动圈分享胜利成果等带来的自我成就感)

○C.流畅感和自然感(在运动中加入音乐给你带来的舒适体验等)

20.您认为您所使用的运动类 App 能够帮您实现上一题中的体验的程度是: * [单选题] *

○A.基本可以实现

○B.部分可以实现

○C.完全可以实现

○D.完全不能实现

21.您认为在运动中使用运动类 App 是否提高您锻炼的效果: * [单选题] *

○A.高(很大程度上)

○B.中(中等程度)

○C.低(很少提高)

○D.无(不能提高)

22.在运动类 App 的运动指导系列中您经常学习下列哪种? * [多选题] *

□A.运动训练

□B.运动装备

□C.健身塑型

□D.运动损伤防护

□E.健康饮食

□F.冬季运动

□G.其它

23.通过使用运动类 App，您认为其在哪里方面需要改进呢? * [多选题] *

□A.广告太多

□B.界面设计不合理

□C.功能不够丰富

□D.记录不够准确

□E.激励机制有待提高

□F.提供知识全面性、科学性欠缺

□G.其他

致　谢

在此次统计建模竞赛中，我们得到了来自各个方面的指导和帮助。首先，我们要特别感谢我们的指导老师贾兆丽老师，从比赛一开始帮助我们明确方向、确定选题，到问卷的设计、论文写作、格式规范，贾兆丽老师都给予了我们极大的帮助和支持，老师的悉心指导让我们学到了很多。在此，我们要对贾兆丽老师表示衷心的感谢，感谢与如此负责的贾兆丽老师相遇，并有幸受到老师的指导。另外，还要对一起团队中所有的成员和给予我们帮助的同学表示感谢，正因为有了团队中每个人全力以赴的付出和身边同学的鼓励和支持，才使得本次报告能够完整呈现。同时，我们也要对组织本次比赛的组委会表示感谢，感谢给予我们这次机会，让我们能够锻炼自己，运用自己所学分析实际问题，学到更多的知识。