**健身会所顾客精准营销分析可行性研究**

**任务计划书**

|  |  |  |
| --- | --- | --- |
| **专题名称** | **：** | 健身会所顾客精准营销分析系统 |
| **实训小组成员** | **：** | 齐帅彬，刘敏慧，方蕾，黄政峰，陈潇阳 |
| **实训小组组长** | **：** | 齐帅彬 |

**二零一七年十一月三十日**

**修改历史**

|  |  |  |  |
| --- | --- | --- | --- |
| **日期** | **版本** | **作者** | **修改内容** |
| 2017/11/26 | 1.0 | 全体成员 | 初步汇总修改计划书 |
| 2017/11/30 | 2.0 | 全体成员 | 格式修改，完成计划书终稿 |
| 2017/12/05 | 3.0 | 全体成员 | 个人阅读文献进一步补充完善 |
| 2018/03/08 | 4.0 | 全体成员 | 技术路线修改，第四阶段任务补全 |

注：当前版本4.0，发布日期2018/03/08

**目录**

[一、专题可行性研究的目标与任务 1](#_Toc499887698)

[1.1 目标 1](#_Toc499887699)

[1.2 任务内容和目标分析 1](#_Toc499887700)

[1.3 专题可行性研究的的主要技术难点和问题分析 1](#_Toc499887701)

[二、现有工作基础与优势 2](#_Toc499887702)

[2.1 技术现状分析 3](#_Toc499887703)

[2.2实训小组研究基础 9](#_Toc499887704)

[三、任务分解与考核指标 13](#_Toc499887705)

[3.1 专题可行性研究的任务 13](#_Toc499887706)

[3.2 技术路线 15](#_Toc499887707)

[3.3 主要考核指标 17](#_Toc499887708)

[3.4小组成员具体职责划分 17](#_Toc499887709)

[四、计划及目标 18](#_Toc499887710)

[4.1进度安排 18](#_Toc499887711)

[4.2里程碑及各里程碑的考核项 29](#_Toc499887712)

[五、组织管理措施 30](#_Toc499887713)

[5.1 实训小组管理的组织架构 30](#_Toc499887714)

[5.2 项目开发过程中规范 30](#_Toc499887715)

[六、其他需要说明的事项 31](#_Toc499887716)

**一、专题可行性研究的目标与任务**

**1.1 目标**

1、通过已有数据，发掘会员行为内在规律或联系，得出分析结果。

2、根据分析结果给会员推荐相应课程，挖掘潜在会员，提高健身房利润。

3、尝试形成推荐系统。

对以上三点进行可行性研究。

**1.2 任务内容和目标分析**

健身会所的精准营销课题主要是针对健身房中不同会员的特征将其分类，通过分析对不同群体进行分析及相应的可视化推荐。为了达到这个研究目的，我们将该项目分为三个基本阶段：

第一阶段：数据预处理阶段。该阶段主要是将健身房会员及课程相关数据进行必要的清洗处理，如果数据不全的话也可以考虑从网上爬取部分数据。这里可以选择将数据导入Hadoop还是数据仓库，要根据二者的功能和效率还有现有数据进行综合判断。Hadoop及其Hadoop文件系统是跨多个机器，并用来处理海量数据的；数据仓库是技术和软件套件，它能够从操作系统收集数据，并将这些数据整合，并统一到中央数据库中，然后对数据仪表盘上指标进行分析、可视化和追踪关键性能处理。数据仓库通常部署在单个关系数据库中，而这个数据库则起到中央存储的作用，这要等后期拿到数据根据数据量来进行判断。根据现有数据，初步决定采用数据仓库进行处理。

第二阶段：数据分析阶段。该阶段主要是对用户数据和课程数据进行建模，通过相应的算法及一些数据分析工具对数据进行分析。在数学分析的基础上，仍旧需要加入对业务层面的分析，最终形成一份相应的分析报告。

第三阶段：分析结果应用阶段。该阶段主要是根据前一阶段所得分析结果对用户进行课程推荐，同时需要对分析结果进行可视化动态的展现在网页之中，可供用户交互使用。

**1.3 专题可行性研究的的主要技术难点和问题分析**

根据不同阶段的目标及内容，我们在每个阶段也具有不同的技术难点及问题：

第一阶段：数据预处理阶段。由于该部分的数据主要是从ERP系统中提取，所以数据基本比较干净，相对来说并不是十分困难。

第二阶段：数据分析阶段。该阶段的主要难点在于对模型的选择上，首先，同学们知识储备不够，对模型的掌握并不是非常的多，所以需要多阅读文献增加储备量。其次，在掌握更多模型的用法之后，具体针对我们的数据，应该使用哪一种或者几种模型进行分析，是很难把握的。所以，在了解模型的运用之后，我们仍需要深入学习，学习模型本身，了解各个模型的适用情况及优缺点，以便之后的选择。

1、难点在于将数据的各个属性理解建立关系模型，把握各属性的内在逻辑关系。需要组员发散思维进行联想，并熟练关系模型的知识点。

2、难点在于理解健身房业务以及相关的社会背景，对数学建模所得的结果进行业务上的分析。需要多查阅相关资料，最好能够实地探访一些健身房进行了解。

第三阶段：分析结果应用阶段。该阶段的难点主要在数据可视化技术，此时要求的可视化不仅仅是简单的将数据制成静态图表，更多的是要求将数据动态的制图，同时提供交互接口连接到网页供用户按照不同条件进行查看。这就需要我们对可视化技术及网页技术进行更加深入的学习研究。

**二、现有工作基础与优势**

本次工作内容是利用数据挖掘技术完成健身会所顾客精准营销分析系统。

**2.1 技术现状分析**

**2.1.1该专题研究的国内外现状、存在的问题以及发展趋势。**

从应用方面上看，数据挖掘技术不仅被作为企业内部营销手段，用户产品和用户细分，确定市场变化趋势，改善内部管理，还被应用到与其他行业结合，创造外部收益。

数据挖掘具体到健身的领域，存在的最大问题就是针对性不强，营销手段缺乏比较有效的数据分析方法和有效的工具，大数据的时代催生了旺盛的市场需求，如何在满足用户需求中构建商业模式，在大数据时代开展企业的精准营销活动都是新的挑战。

在发展趋势上，在健身会所领域，出现了一些专门运用数据挖掘来做推荐的公司，比如，孟买健身场所发现平台Growfitter运营着Growfitter.com网站，获得了60万美元种子轮融资。这家公司将自己定义为一家健康和健身解决方案提供商，而且会根据客户的地理位置指导他们该如何你挑选健身房。他们的目标就是通过机器学习和人工智能变革健康以及健身行业，同时建立一个健身推荐系统。

**2.1.2该专题对产业的发展、经济建设和社会效益及推广应用前景发展的预测及分析。**

本专题将致力于构建和完善客户精准营销模型的构建，建立一个切实可行的精准营销模型的分析报告，在此基础上健身会所等企业可以后续开发出实际的会员精准营销的模型，将数据转化成财富和利润，将是本专题的研究内容之一。

除了上文的发现平台，健身领域的其它创企也获得了融资。例如古尔冈的Jogo Technologies获得了20万天使轮融资，健身会员移动市场Fitnapp获得了Pre-A轮融资，全年12月，它获得了100万美元种子轮融资。

由以上事实不难推测，对健身会所的会员进行精准推荐拥有巨大市场，获得了风险投资公司的青睐，未来有着良好发展前景。

**2.1.3该专题涉及的技术，主要研究学习的内容和目标**

本专题的内容在于通过某健身会所的存量会员和潜在会员信息建立起对存量数据资料的挖掘和分析，实现精准的营销，将至少通过数据可视化形成一份分析报告。

对于该研究专题，我们要做的研究和学习重难点集中在大数据挖掘部分，要做的就是把上述类似的商业运营问题转化为数据挖掘问题，如何将商业运营问题转化为大数据挖掘问题。

1. 数据库/数据仓库技术/数据源的形式

常用的包括Mysql\SQLServer\Excel等。本次项目数据来源及存储主要在Excel中进行。

1. 数据预处理技术

包括数据清理、数据集成、数据变换、数据归约等方法。

1. 数据分析挖掘技术

考虑对数据挖掘问题进行细分，以下四类问题我们都可能涉及到：分类问题、聚类问题、关联问题、预测问题。

（1）分类问题

用户流失率、促销活动响应、评估用户度都属于数据挖掘的分类问题，我们需要掌握分类的特点，知道什么是有监督学习，掌握常见的分类方法:决策树、贝叶斯、KNN、支持向量机、神经网络和逻辑回归等。

（2）聚类问题

细分市场、细分客户群体都属于数据挖掘的聚类问题，我们要掌握聚类特点，知道无监督学习，了解常见的聚类算法，例如划分聚类、层次聚类、密度聚类、网格聚类、基于模型聚类等。

（3）关联问题

交叉销售问题等属于关联问题，关联分析也叫购物篮分析，我们要掌握常见的关联分析算法：Aprior算法、Carma算法，序列算法等。

（4）预测问题

我们要掌握简单线性回归分析、多重线性回归分析、时间序列等。

1. 数据模型与分析模型

对第一阶段清洗干净的数据进行分析，建立数据模型。首先将数据中的各个属性单独列出，寻找属性之间的关联。其次，将之前小组讨论中完成的头脑风暴图根据数据的属性进行完善，将之前没有考虑到的内容添加到图中，同时也将多余的属性剔除。再次，根据完善后的头脑风暴图对数据建立关系数据模型，制作E-R图。紧接着，我们对建立好的关系数据模型中的关系约束进行检查完善。最终形成可用于算法分析的数据模型。

推荐算法：

（1）协同过滤算法（CF）

CF算法包括基于用户的CF(User-based CF)和基于物品的CF(Item-based CF)。

这两个思想分别是：

从跟目标用户相似的其他用户中找几个最相似的，把它们喜欢的商品推荐给目标用户

把跟目标用户喜欢的商品相似的商品中找几个最相似的，把它们推荐给目标用户

对于我们健身房项目推荐来说，也可以从以上两个角度来分析，一是对健身课程做相似度分析，把会员喜欢的课程的的相似课程推荐给会员；另一方面是对会员做相似度分析，从跟目标会员相似的其他会员中找几个相似的，把他们喜欢的健身课程推荐给目标会员

（2）关联规则Aprior算法：

关联规则中最著名的例子就是“啤酒尿布”的例子。

 Apriori算法是一种挖掘关联规则的频繁项集算法，其核心思想是通过候选集生成和情节的向下封闭检测两个阶段来挖掘频繁项集。该算法的挖掘步骤：

1.依据支持度找出所有频繁项集（频度）

2.依据置信度产生关联规则（强度）

对于我们健身房推荐来说，我们可以试图找出课程选择中具有强关联的课程，以此作为套餐课程出售，提高健身房销售额。

（3）逻辑回归：

逻辑回归主要针对二分类模型，是一种广义的线性回归分析模型，常用于数据挖掘，疾病自动诊断，经济预测等领域。例如，探讨引发疾病的危险因素，并根据危险因素预测疾病发生的概率等。

对于我们健身房推荐来说，我们可以得到自变量的权重，从而了解那些属性是会员时候购买课程的关键阴虚，同时根据该权值来预测一个人时候会购买课程的可能性。

（4）K-means

K-means主要是实现多分类的典型模型，通过每个样本之间属性的相近程度对所有的样本分成K个类别，使得每个类别内的样本尽可能的相似，与其他类别中的样本有较大差异。

对于健身房推荐来说，我们可以通过对客户的基本属性、行为习惯等信息进行分析建模，对所有的客户划分为不同的类别。根据不同类别客户的特征爱好，进行相对应的课程推荐，保证所推荐的课程符合该类别中客户群体的共同偏好，从而实现精准营销。

对以上四种推荐算法，小组最终决定运用协同过滤和K-means两种算法作为最终建模分析的技术支撑。

**属性分析模型**

顾名思义，根据用户自身属性对用户进行分类和统计，比如设计用户年龄、性别、最高教育程度等自然信息。用户各维度属性属性都是进行全面衡量用户画像不可或缺的内容。属性分析的价值在：丰富用户画像维度，让用户行为洞察粒度更细致。

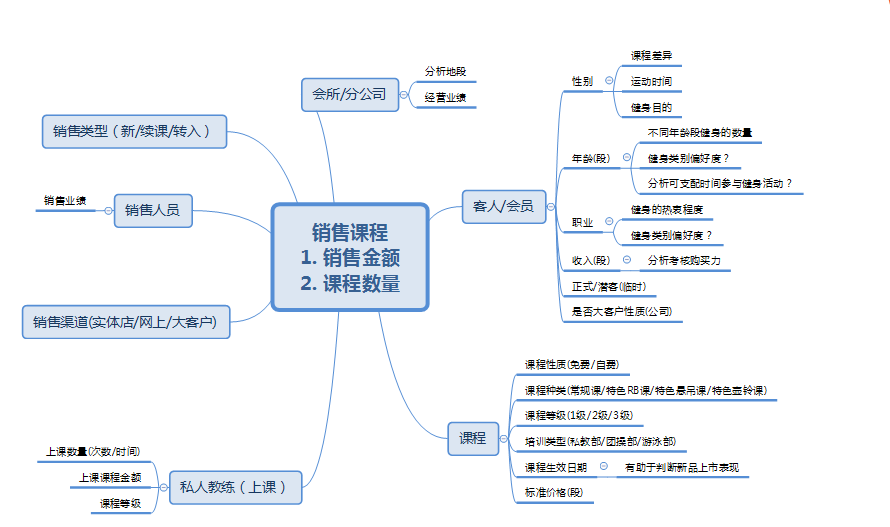
上述几种分析模型对于我们项目来说都是有实际参考意义和价值的，在后续项目实际开发过程中，我们会根据项目需要选择其中一种或多种分析模型来进行分析。

对上一阶段中对整合后的数据进行数据建模分析。首先，我们可以利用5W2H模型对会员数据的分析结果进行剖析，透过数据挖掘会员数据中的潜在信息，结合相应的业务知识与生活场景，了解用户行为的动力，为之后的分析提供相应的理论依据。



同样的，我们也可以利用5W2H模型对健身房自身数据进行分析挖掘，了解在经营过程中，健身房现阶段存在的问题以及不同的用户行为对健身房的盈利存在什么样的影响。

在分析完客户以及健身房自身数据之后，我们可以对分析结果利用PEST模型进行深层次的解析，了解为什么会出现之前分析的结果，出现该结果的社会背景及条件又是如何？只有了解了这些之后，才能够对接下来在不同的背景下，去预测健身房应该对不同的潜在客户行为去做如何的营销，从而提高自身的效益。



根据对之前数据建模的分析结果，我们可以对现有的头脑风暴图进行完善修改，以便于接下来对数据进行分析模型的建立。

1. 数据可视化

数据可视化，是关于数据视觉表现形式的科学技术研究。其中，这种数据的视觉表现形式被定义为，一种以某种概要形式抽提出来的信息，包括相应信息单位的各种属性和变量。

以下将从前端的展示方法以及后端的框架两部分来介绍数据的可视化。

一、数据可视化展示方法

在数据可视化技术方面我们对其中的三种技术进行举例介绍。

（1）Highcharts图表组件，一个 JavaScript脚本编写的图表库，目前 Highcharts5.0.7支持面积图、散点图、仪表图、雷达图、金字塔图等 19种不同类型的图表。Highcharts适用于多种浏览器，并支持对移动端浏览器操作，显示品质高，并且拥有完整的实例演示、功能介绍及详细的 API文档，但是，它的图表类型比较有限，图表与用户交互性比较欠缺。

（2） Echarts图表组件，在处理大数据量和 3D绘图方面，有明显优势。目前，Echarts3.0拥有柱状图、矩形树图等 19种不同类型的丰富图表，并且支持图与图之间的混搭。不过，Echarts无法改变图表的大小，只能缩放显示。此外，从学习成本来看，Echarts3.0在完善 API文档及增加实例演示方面有了很大改进。

（3） D3，是一种基于 JavaScript库的可视化数据处理技术。它允许绑定任意数据到 DOM，然后将数据驱动转换应用到文档中，将数据通过使用 HTML、SVG和 CSS实现可视化展示。D3支持多种数据文件的读取，如 JSON、XML、CSV、HTML等。采用了与网页开发常用的 jQuery库相同的链式语法，代码简洁。D3基于 SVG，在任意缩放、旋转或改变形状时，图形清晰，布局有饼状图、弦图、力导向图、打包图、分区图等 12 个布局。

对于以上三种可视化技术，小组决定采用Echarts技术进行最终的可视化操作展示。

二、数据可视化后端的实现方式

（1）SSM框架

可视化展示的后端部分我们考虑使用SSM框架。由Spring、SpringMVC、MyBatis三个开源框架整合而成。常作为数据源较简单的web项目的框架，使用以上框架，我们准备将专题研究的分析成果通过网站的形式可视化展示出来。

其中spring是一个轻量级的控制反转（IoC）和面向切面（AOP）的容器框架。SpringMVC分离了控制器、模型对象、分派器以及处理程序对象的角色。MyBatis是一个支持普通SQL查询，存储过程的持久层框架。这些功能组合可以胜任我们对后端处理的要求。三个框架有比较多的文档可供学习，掌握难度适中，同时本专题的这部分工作仅仅涉及到对已经形成的分析结果的处理和展示，我们将尝试使用这个框架。

（2）SSH框架

另外一种可行的方案是：使用spring，struts2，hibernate框架作为可视化展示的后端处理框架。其中使用Struts作为系统的整体基础架构，负责MVC的分离，在Struts框架的模型部分，控制业务跳转，利用Hibernate框架对持久层提供支持，Spring做管理，管理struts和hibernate。

这种方案作为第二选择是因为Struts框架不够易用且不够简便，同理hibernate也相对是重型框架，hibernate的事务，缓存，在本专题大部分时候也用不上，Mybatis相对而言简便易用，也容易掌握，适合本专题。

对于以上两种可视化后端框架模型，小组最终决定运用SSM框架进行可视化后端平台搭建。

**2.2实训小组研究基础**

本部分包括小组成员的自身的优缺点分析。

1、齐帅彬：

|  |
| --- |
| **个人优势：** |
| （1）技术方面：  擅长分析文档的撰写；可对项目中第二阶段成果分析文做最后的汇总。  通过秋季学期的学习，对gephi、数据透视图数据透视表等较为熟悉，可利用这些工具将分析结果做成图表格式展现。  （2）业务方面：  有参加实际项目的经验，在本科期间担任过学生干部，可把这些经验应用到实训小组的管理中，与小组成员一起通力合作，完成项目。 |
| **个人劣势：** |
| （1）数学基础差，算法方面薄弱  会在接下来的学习中着重学习数学方面知识，尤其是统计学部分，也会着重去深入学习几个算法相关知识。  （2）编程能力较弱  在接下来的学习过程中，我将重点学习r语言和python,运用其中成熟完善的工具包来做数据分析与处理，弱化编程能力的劣势。 |

2、刘敏慧：

|  |
| --- |
| **个人优势：** |
| （1）技术方面：擅长数据处理，熟悉R语言，重点可以放在项目中对前期数据预处理和数据分析模型上  （2）业务方面：对健身房的课程有一定了解，在进行数据分析时针对专题有一定背景知识的支撑。 |
| **个人劣势：** |
| 对相关算法和模型的了解不够深入，初期不知道该选择怎样的分析模型才与分析的内容相契合 |

3、黄政峰：

|  |
| --- |
| **个人优势：** |
| （1）专业方面：  在项目的分析阶段，需要对健身房会员的数据进行分类、建模等分析工作，所以需要涉及到对算法的应用。由于自身本科数学系，数学功底较有优势。之前学习过计量经济学、概率论、数理统计、数学建模、信息论等科目，对基本的数学知识有所掌握和理解。在接下来的实训项目中要更加深入的学习数学建模及算法的知识，争取最大限度的发挥自身的优势。  在数据的预处理阶段，需要利用R、Python或者EXCEL等工具进行清洗，之前工作时对EXCEL及其他OFFICE软件运用较多，尤其是EXCEL中的常用公式、VBA及透视表有简单的掌握。本科时也学习过SPSS、MATLAB、LINGO等数据分析工具，在建模分析阶段也有一定的帮助。希望接下来能够通过更加深入的学习R或者Python工具，对数据处理及最后一个阶段的数据可视化能够达到更加熟练的程度。  （2）业务方面：  小组分工中，第一阶段的组内职责是周报管理员，主要负责对会议记录及会议纪要的撰写。之前在工作过程中对撰写报告、制作PPT、上台演讲等有所锻炼，所以在第一阶段的周报管理员，还是相对轻松一些。同时对最后一个阶段将要承担的项目展示职能也有较强的优势。接下来，在此方面的目标主要是希望自己能够多加练习，在熟练的基础上，能够尝试加入一些新鲜的创意，提升自身的水平。  一方面此前简单的接触过ERP系统，对公司整体运营情况的流程有简单的了解，同时自己也有办过健身卡，所以在研究“健身会所精准营销”的这个课题的时候，可以根据以往的经历，站在一名健身顾客的角度，对小组提出一些建议和意见。同时也希望自己多收集数据，从健身房主管的角度去思考该课题的目标及意义。 |
| **个人劣势：** |
| （1）计算机代码能力差，无论是开发还是常用软件的应用，都非常的欠缺。在该课题中的数据准备阶段，一方面是有些数据需要我们从网站上或者其他各种途径获取，这就需要数据爬取能力，另一方面，老师给的数据或者是网上爬取的数据需要进行清洗处理。常用的数据爬取和清洗的工具是R和Python，所以，我在初期阶段应该多参阅这两个软件的书籍，并进行联系实践。  （2）在后期阶段，小组希望对数据进行可视化的展现，并且对分析的结果以网页等方式展现出来，所以我需要结合下学期的可视化课程对该方面有所提升，并且自学掌握网页HTML等相关知识。 |

4、方蕾：

|  |
| --- |
| **个人优势：** |
| （1）专业方面：  本科信息管理与信息系统专业，偏向计算机。在使用计算机语言方面，相比较于python等语言，更擅长java。  本科期间做过java开发，做过图书商城管理系统、基于微信公众平台的二次开发等小项目。  希望在接下来的项目中能充分发挥自己的优势，能更进一步提升自己的专业水准。  （2）业务方面：  踏踏实实做事。  搜集情报。虽然认识的人也不多，但是，线人还是有的，知己知彼百战不殆。 |
| **个人劣势：** |
| （1）数学不行，对数字不敏感，比较死板。  （2）对算法的了解不多，目前只了解kmeans。  （3）SQL以及java的一些规范有些遗忘  后期会多多培养自己对数学的感觉，多学习一些算法，并且把遗忘了的知识点再回顾下。 |

5、陈潇阳：

|  |
| --- |
| **个人优势：** |
| （1）和小组成员相比，对java比较熟悉，对j2ee比较熟悉，对MVC开发模式有所了解，在后期做数据可视化展示的时候可以发挥这个优势，帮助完成网站环境建设。  （2）如果需要网络爬虫爬数据，可以帮助完成这项工作。 |
| **个人劣势：** |
| （1）概率论、数理统计基础比较弱，需要做一点复习的工作。  （2）对数据挖掘过程中可能涉及到的各种算法了解不多，这方面不足需要一定的时间去解决。 |

**三、任务分解与考核指标**

**3.1 专题可行性研究的任务**

第一阶段：

|  |  |  |  |
| --- | --- | --- | --- |
| 任务子项 | 开始时间 | 结束时间 | 研究任务 |
| 项目启动 | 2017.10.23 | 2017.10.29 | 实训规章制度的制定 |
| 项目沟通理解 | 2017.10.30 | 2017.11.5 | 跟导师沟通关于实训项目的内容 |
| 完成可行性分析报告初稿 | 2017.11.6 | 2017.11.19 | 查找文献，完成可行性分析报告初稿。 |
| 完成可行性分析报告中稿 | 2017.11.20 | 2017.11.26 | 听取老师意见，继续查找文献，完善完成可行性分析报告 |
| 完成可行性分析报告终稿 | 2017.11.27 | 2017.12.3 | 听取老师意见，参考模板修改，继续完善可行性分析报告。 |

第二阶段：

|  |  |  |  |
| --- | --- | --- | --- |
| 任务子项 | 开始时间 | 结束时间 | 研究任务 |
| 阅读文献 | 2017.12.4 | 2017.1.26 | 通过阅读文献来储备知识，以解决项目中可能会碰到的问题。并做文献综述。 |

第三阶段：

|  |  |  |  |
| --- | --- | --- | --- |
| 任务子项 | 开始时间 | 结束时间 | 研究任务 |
| 软硬件安装 | 2017.1.26 | 2017.2.09 | 确定软硬件安装条件，并进行安装。 |
| 分析工具的确定 | 2017.2.10 | 2017.2.15 | 确定要选用的分析工具，并对工具进行学习。 |
| 完成数据清洗及数据统计分析 | 2017.2.15 | 2017.3.08 | 对数据进行清洗，并数据进行有重点的分析，并形成初步的分析结果。 |

第四阶段：

|  |  |  |  |
| --- | --- | --- | --- |
| 任务子项 | 开始时间 | 结束时间 | 研究任务 |
| 完成数据建模及预测分析 | 2017.3.09 | 2017.3.25 | 对潜客转化成正式会员进行建模分析，通过模型预测新增潜客转入正式会员的概率。 |
| 完成数据可视化以及网站的展示 | 2017.3.09 | 2017.4.10 | 将分析的结果以网站的形式展示出来 |
| 完成推荐模型的训练 | 2017.3.26 | 2017.4.10 | 通过选取算法对客户进行建模，形成推荐模型，从而对客户进行健身课程的推荐。 |
| 对分析结果进行优化，同时撰写实训报告 | 2017.4.13 | 2017.4.25 | 听取老师意见，对分析结果做进一步的优化，小组成员撰写结题报告，多次讨论修改后形成终稿。 |

**3.2 技术路线**



以下将从数据预处理、数据分析、分析结果应用三方面来简单展示技术路线。

1. 数据预处理。这部分主要对健身房冗余信息进行处理，删除掉无用数据等，之后对处理好的数据进行数据建模并整合，比如将会员信息与课程信息进行连接。
2. 数据分析部分。这部分主要对整合好的数据进行分析。先是对已有的数据进行初步汇总分析，再是运用分类、聚类、关联等方法建立预测模型，最后，对已有的模型进行验证，验证推荐课程的准确度。
3. 分析结果应用部分。这部分主要是对分析结果的以图表的形式进行展示，以及基于算法处理结果建立一个可以根据会员相关信息进行推荐的系统。

**3.3 主要考核指标**

**第一阶段：**

小组成员共同完成“企业专题可行性研究任务计划书”，并且得到认可 。

**第二阶段：**

1、小组成员阅读文献，最后每人完成各自的文献综述(读书报告)。

2、对原计划书中不合理之处进行修正。

**第三阶段：**

利用EXCEL对原数据进行处理，确定分析工具（R及Python），开始进行数据分析，并得到一份分析结果，例如形成数据分析文档或者图表等格式做数据可视化。同时在该阶段进行展示环境的搭建。

**第四阶段：**

首先建立数据模型，根据前期选用算法及工具对模型进行训练，通过测试对模型的优度进行评估，并对项目研究结果进行最终优化。同时对可视化网站进行开发搭建，形成分析结果的展示页面。最终撰写实训结题报告并顺利结题。

**3.4小组成员具体职责划分**

第一阶段（数据预处理）：

刘敏慧，齐帅彬：数据清洗

黄政峰：数据字段的选择，创建数据库

齐帅彬，刘敏慧，黄政峰：交叉检查最后数据格式结果

方蕾，陈潇阳：学习数据建模；算法等知识，为下一阶段的分析任务做准备。如有需要则要进行爬虫等工作。

第二阶段(数据分析):

齐帅彬，刘敏慧，黄政峰：共同制定每项分析任务的结果最终展示形式；进行实际的逐条分析。

刘敏慧，齐帅彬：选择算法模型并进行训练。

黄政峰：检验算法模型的准确性，并负责调参等

方蕾，陈潇阳：同时搭建网站，对分析结果可视化展示；

最后小组成员一起撰写分析报告；

第三阶段：

齐帅彬，刘敏慧，黄政峰：对第二阶段做出的分析结果做进一步归纳总结，提取出实际应用建议或意见；

方蕾，陈潇阳：对网站展示做进一步优化。

最后小组全体成员一起准备最后的实训结题工作。

**四、计划及目标**

**4.1进度安排**

见下页项目进度表。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 项目进度表   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 时间 进度 | 10月 | 11月 | | | | 12月 | | | | 1月 | | | | 2月 | | | | 3月 | | | | 4月 | | | | | **第四周** | **一** | **二** | **三** | **四** | **一** | **二** | **三** | **四** | **一** | **二** | **三** | **四** | **一** | **二** | **三** | **四** | **一** | **二** | **三** | **四** | **一** | **二** | **三** | **四** | | 讨论建立实训组的组织架构及岗位职责 | **10.24-11.05** | | **11.06-12.03** | | | | **12.04-1.27** | | | | | | | **1.28-3.04** | | | | | **3.05-4.25** | | | | | | | | 和企业导师沟通了解研究专题的内涵 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | **第一阶段** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 针对专题提出可行性研究计划（分工） |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 研究计划修改汇总 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 导师确认，全体大会汇报第一阶段任务 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | **第二阶段** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 确认阅读文献内容 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 小组成员阅读文献 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 汇报阅读文献的列表和进度安排 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 继续文献阅读 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 做文献综述报告 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 对原任务计划书进行修改 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | **第三阶段** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 软硬件环境安装 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 数据准备 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 分析工具学习 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 与企业导师进行沟通确认 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | **第四阶段** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 着手完成相应的数据分析工作 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 准备中期汇报，完成数据可视化 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 对分析结果进行最后整理优化 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | 撰写实训结题报告 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  | 已完成 | |  |  | 正在进行 | | |  |  | 未开始 | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

具体时间点安排

**与企业导师暂定每隔两周周二晚18-19点进行视频会议，汇报任务进度**

**第一阶段**

11.03 18-19点 小组开会确定每人具体职责。

11.10 19-20点 开会讨论任务计划书具体内容，每人确定撰写部分。

11.17 18-19点 开会对任务计划书进行进一步讨论，对不合适或者不细致的地方进行修改。

11.24 18-19点 确定任务计划书初稿内容并提交企业导师。

11.30 20点 修改后将终稿递交企业导师。

12.02 18-19点 开会确定阅读文献内容。

**第二阶段**

12.08，12.15, 12.22, 12.29, 1.05，1.12, 1.19, 1.26

18-19点 介绍自己阅读的内容和心得，阅读笔记由小组成员监督完成，具体阅读内容见下面表格。

12.17 下午 召开全体会议，各实训组包括个人汇报阅读文献的列表和进度安排。

1.07 下午 召开全体会议，文献阅读的中期检查。即随机选择学生做文献综述报告。

1.26 18点-19点 检查小组成员文献综述撰写进度。

**第三阶段（如人员不能到齐则召开视频会议）**

2.02 20点-21点 开会讨论研究软硬件安装环境，确定后进行安装。

2.09 20点-21点 每人汇报软硬件安装情况以及期间遇到的问题大家一起解决，并开始进行数据的预处理工作，为数据分析做准备。

2.23 20点-21点 讨论决定选用的分析工具，以现有数据为例对工具进行学习。

3.02 20点-21点 分析工具经导师确认后进行进一步的学习。

**第四阶段**

3.09 18点-19点 着手相应的数据分析工作，讨论数据分析的重点。

3.10 全天 召开全体会议，文献综述报告会。各组根据完成的各项工作，提出调整的专题可行性研究的新计划和工作新目标。

3.16 18点-19点 将分析结果进行组内展示，组员进行讨论。

3.23 18点-19点 准备中期汇报，实现数据的可视化及网站展示。

3.30 18点-19点 对中期汇报内容进行最终确认。

4.01 下午 召开全体会议，各实训组做中期汇报。

4.06 18点-19点 完成推荐模型的训练，对模型进行初步评估。

4.13 18点-19点 对分析结果进行最后整理优化，准备结题内容。

4.15 下午 召开全体会议，宣布实训研究阶段结束，进入结题阶段。

**结题阶段**

4.16 开始撰写结题报告。

4.20 18点-19点 小组成员对结题报告的内容进行讨论，最终在与导师讨论后形成最终结题报告。

4.27 整天 召开全体会议，各组结题审核。

**第四阶段—数据分析部分时间安排及任务考核：**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时间 | 任务安排 | 产出 | 责任人 | 检查人 |
| 3.12~3.17 | 根据每人自己分配的分析表，对数据做统计和可视化分析，每个人分别整理成分析报告文档 | 每个人分别一份分析报告文档 | 齐帅彬  刘敏慧  黄政峰 | 企业导师；  相互交叉检查 |
| 3.18~3.25 | 对分析报告的整理优化，汇合成一份完整的分析报告 | 完整的分析报告一份 | 齐帅彬  刘敏慧  黄政峰 | 企业导师 |
| 3.26~4.3 | 协同过滤算法的应用 | 算法应用过程和结果分别形成文档 | 齐帅彬  刘敏慧  黄政峰 | 三人交叉检查，选取结果最好的作为最终结果 |
| 4.4~4.11 | 对协同过滤算法的应用做进一步优化 | 最终算法应用过程和结果文档 | 齐帅彬  刘敏慧  黄政峰 | 三人交叉检查 |
| 4.12~4.19 | 聚类算法的应用 | 最终算法应用过程和结果文档 | 齐帅彬  刘敏慧  黄政峰 | 三人交叉检查，选取结果最好的作为最终结果 |
| 4.20~4.25 | 对数据分析、推荐结果整理优化，撰写结题报告 | 结题报告一份 | 齐帅彬  刘敏慧  黄政峰 | 企业导师 |

**小组成员个人阅读文献**

齐帅彬：《统计学》《利用python进行数据分析》《R语言》

孔庶（硕士）统计稀疏学习\_特征提取\_聚类\_分类及多特征融合[D] 2013年.96页.浙江大学

任磊（博士）推荐系统关键技术研究[D]2012年.131页.华东师范大学

|  |  |
| --- | --- |
| 时间 | 阅读内容 |
| 11.27~12.1 | 《统计学》6~9章 |
| 12.2~12.8 | 《统计学》10~14章 |
| 12.9~12.15 | 《利用Python进行数据分析》1~3章;  统计稀疏学习\_特征提取\_聚类\_分类及多特征融合1-30页 |
| 12.16~12.22 | 《利用Python进行数据分析》4~6章  统计稀疏学习\_特征提取\_聚类\_分类及多特征融合31-96页 |
| 12.23~12.29 | 《利用Python进行数据分析》7~9章 |
| 12.30~1.5 | 《利用Python进行数据分析》10~11章  推荐系统关键技术研究1-60页 |
| 1.6~1.12 | 《R语言实战》1~5章  推荐系统关键技术研究61-131页 |
| 1.13~1.19 | 《R语言实战》6~12章 |
| 1.20~1.26 | 《R语言实战》13~16章 |

刘敏慧： 《数据科学中的R语言》

刘青文（博士）.基于协同过滤的推荐算法研究[D].中国科学技术大学.2013（共106页）

沈亚镁（硕士）.基于行为特征的潜在客户发现模型研究[D].杭州电子科技大学.2011（共76页）

|  |  |
| --- | --- |
| 阅读时间 | 阅读内容 |
| 11.27~12.1 | 《数据科学中的R语言》第一~四章  《基于行为特征的潜在客户发现模型研究》第1~3章 |
| 12.2~12.8 | 《数据科学中的R语言》第五~八章  《基于行为特征的潜在客户发现模型研究》第4~6章 |
| 12.9~12.15 | 《数据科学中的R语言》第九~十二章 |
| 12.16~12.22 | 《数据科学中的R语言》第十三~十六章 |
| 12.23~12.29 | 《利用Python进行数据分析》第一~二章  《基于协同过滤的推荐算法研究》第一~三章 |
| 12.30~1.5 | 《利用Python进行数据分析》第三~五章  《基于协同过滤的推荐算法研究》第四~六章 |
| 1.6~1.12 | 《利用Python进行数据分析》第六~七章 |
| 1.13~1.19 | 《利用Python进行数据分析》第八~九章 |
| 1.20~1.26 | 《利用Python进行数据分析》第十~十二章 |

黄政峰： 《机器学习》 周志华

《数据科学中的R语言》李舰 肖凯

郑明超（硕士）.数据挖掘技术中分类算法的比较分析[A].兰州商学院.2007

肖强 钱晓东（硕士）.基于K-means算法的WEB访问用户关联规则挖掘算法[TP].兰州：兰州交通大学.2011

|  |  |
| --- | --- |
| 时间 | 阅读内容 |
| 11.27~12.1 | 《机器学习》1~4章 |
| 12.2~12.8 | 《机器学习》5~7章 |
| 12.9~12.15 | 《机器学习》8~10章 |
| 12.16~12.22 | 《机器学习》11~13章 |
| 12.23~12.29 | 《机器学习》14~16章 |
| 12.30~1.5 | 《基于K\_means算法的Web访问用户关联规则挖掘算法\_肖强》 |
| 1.6~1.12 | 《数据挖掘技术中分类算法的比较分析》 |
| 1.13~1.19 | 《R语言》1~3章 |
| 1.20~1.26 | 《R语言》4~6章 |

方蕾：

python教程

<https://www.liaoxuefeng.com/wiki/0014316089557264a6b348958f449949df42a6d3a2e542c000>

作者：廖雪峰，全栈工程师，独立iOS开发者，精通Python / Java / JavaScript / Node / Objective-C / Scheme等。

《基于大数据挖掘的电信客户精准营销系统》 刘峥. 基于大数据挖掘的电信客户精准营销系统[D]. 复旦大学, 2014.

《数据清洗研究综述》 王曰芬, 章成志, 张蓓蓓,等. 数据清洗研究综述[J]. 现代图书情报技术, 2007, 2(12):50-56.

《数据可视化技术在Web中的研究及应用》 刘长娥. 数据可视化技术在Web中的研究及应用[J]. 电脑知识与技术, 2017, 13(18):7-8.

《个性化推荐算法综述》 黄仁, 孟婷婷. 个性化推荐算法综述[J]. 中小企业管理与科技旬刊, 2015(3):271-273.

《应用预测建模》马克斯・库恩, 谢尔・约翰逊. 应用预测建模[M]. 机械工业出版社, 2016.

|  |  |
| --- | --- |
| 时间 | 阅读内容 |
| 11.27-12.1 | 1、《数据清洗研究综述》  2、《数据可视化技术在Web中的研究及应用》  3、浏览《基于大数据挖掘的电信客户精准营销系统》 |
| 12.2-12.8 | 1、《个性化推荐算法综述》  2、python教程（1-5部分） |
| 12.9-12.15 | 1、python教程（6-10部分）  2、《基于大数据挖掘的电信客户精准营销系统》（1、2、3章） |
| 12.16-12.22 | 1、python教程（11-15部分）  2、《基于大数据挖掘的电信客户精准营销系统》（4、5、6章） |
| 12.23-12.29 | python教程（15-20部分） |
| 12.30-1.5 | 《应用预测建模》第2、3章 |
| 1.6-1.12 | 《应用预测建模》第5章 |
| 1.13-1.19 | 《应用预测建模》第6章 |
| 1.20-1.26 | 《应用预测建模》第10章 |

陈潇阳：《统计学习方法》《DataMiningGuide》

每次以读完完整的章节为准。

魏超（硕士）.基于ＪＡＶＡ的商品推荐系统的开发研究[TP].2017

内蒙古大学（共55页）

阅读时间:2017.12.30~2018.01.05

李娜（硕士）.基于混合协同过滤的高校选课推荐方法研究[TP].2013

中南大学（共55页）

阅读时间:2018.01.20~2018.01.26

|  |  |
| --- | --- |
| 阅读时间 | 阅读内容 |
| 11.27~12.1 | 《统计学习方法》第一~三章 |
| 12.2~12.8 | 《统计学习方法》第四~六章 |
| 12.9~12.15 | 《统计学习方法》第七~九章 |
| 12.16~12.22 | 《统计学习方法》第十~十二章 |
| 12.23~12.29 | 《DataMiningGuide》第一~三章 |
| 12.30~1.5 | 《DataMiningGuide》第四~五章  《基于ＪＡＶＡ的商品推荐系统的开发研究》 |
| 1.6~1.12 | 《DataMiningGuide》第六~七章 |
| 1.13~1.19 | 《DataMiningGuide》第八章 |
| 1.20~1.26 | 《DataMiningGuide》第九章  《基于混合协同过滤的高校选课推荐方法研究》 |

**4.2里程碑及各里程碑的考核项**

根据进度安排，设定里程碑检查点和相关责任人，并确定检查点检查的具体考核项，包括具体需要提交的文档或报告等

我们整个实训项目工作主要分为三大部分，即主要分为三个里程碑检查点，三个检查点分别如下：  
一、数据预处理

对老师提供的原始数据进行分析，根据我们最后要分析的结果，对其中的脏数据、缺失值、重复值等进行处理。

**检查点考核标准**  
（1）原数据最后整理汇总成一份便于R或python直接读取的csv或txt文件；

（2）去除脏数据和重复值；

（3）填补缺失值；

（4）对数据进行归一化处理，方便最后的分析应用

**检查责任人**：企业导师钱文强老师；小组内全体成员；

**检查时间：2017/3/15 18点**

二、数据分析

针对在第三部分的“任务分解和考核指标”，对其中列出的分析点进行逐一分析。

**检查点考核标准**

1. 将分析结果通过图表及文档格式展现；
2. 完成每一项考核任务，针对其中的考核指标，逐项检验是否完成预期目标；

**检查责任人**：企业导师钱文强老师；小组内全体成员交叉整体检查（每个小组成员分别负责不同的任务和分析角度，大家经过探讨，相互交叉检验对方的分析结果）；

**检查时间：2018/3/30 18点**

三、分析结果应用

从第二部分的分析结果报告中提炼要素，总结哪些分析结果对精准推荐，提高健身房利润有帮助，制定相应的方案。

检查点考核标准：  
（1）制定的方案或建议在实际健生房应用中是否可操作；

1. 若应用最终决定用网站形式展现，则考量网站界面设计是否合理，是否方便用户操作；
2. 把原数据中一部分作为测试集，对测试集的预测结果准确度达到60%及以上。

**检查责任人**：企业导师钱文强老师；最终实训检查报告所有评审教师；

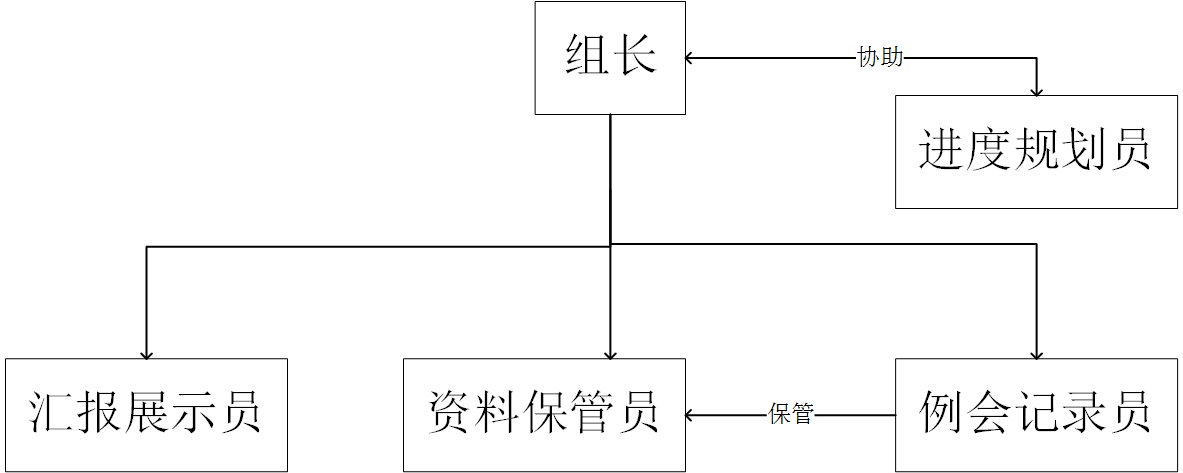
**检查时间：2018/4/10 18点**

**五、组织管理措施**

小组自行讨论决定，建立实训小组管理的组织架构和规章制度、组内人员工作安排，包括研究、分析、试验各阶段的各组员的分工和承担的责任及考核指标。

**5.1 实训小组管理的组织架构**

1. 每组内实行轮岗制度，小组成员在每个阶段中都担任不同的项目角色；期间小组成员内要相互协助，推进项目



图：小组成员组织架构图

**5.2 项目开发过程中规范**

1. 变量名命名采用“驼峰法”命名规则，变量命名需见名知义。
2. 每一阶段开发前制定数据字典，按数据字典中确定的属性协作开发，需要增加或删除属性时在组内告知。
3. 关键代码写注释，方便代码复用和组员间相互交流。

研究、分析、试验各阶段的各组员的分工和承担的责任及考核指标：

管理规章制度详见[附件](小组规章制度管理.docx)（小组管理规章制度）

**六、其他需要说明的事项**

关于实训最终结果展示：

形成一份文档分析材料，在有余力的基础上用网站形式，应用算法，做实时预测推荐。