****

**本科毕业设计(论文)调研报告**

|  |  |
| --- | --- |
| 题 目： | 基于XGBoost的社区平台用户流失预测与分析 |
| 学生姓名： | 吕安杰 |
| 指导教师： | 康松林 |
| 学 院： | 计算机学院 |
| 专业班级： | 软件工程2004 |

本科生院制

2024年3月

基于XGBoost的社区平台用户流失预测与分析

摘要

本项目旨在研究基于观测社交媒体、电子商务等大型领域近年来不断发生的用户流失问题,所开发的基于xgboost的社区平台用户流失预测与分析相关，并通过对相关用户流失问题的数据和报告进行调研分析。从中得到的调研结果显示，用户数量流失把控在社区平台的发展中起到至关重要的作用，因此对于该系统的创作与实现存在一定的可行性，即收集相关数据并清洗、基于xgboost的预测模型开发、特征工程的相关处理等内容。全文将论述该项目的前期调研相关工作，包括本项目的课题背景、设计目标、相关理论与技术基础、初步设计方案、难点及初步解决办法、预期研究成果与时间进度安排等。

**关键词：**xgboost算法；社区平台；机器学习；数据预测

目录

[摘要 I](#_Toc32756)

[第一章 引言 1](#_Toc5313)

[1.1 课题背景及意义 1](#_Toc4121)

[1.2 设计目标 2](#_Toc26181)

[1.3 国内外研究应用现状 2](#_Toc27350)

[第二章 相关理论与技术基础 4](#_Toc26839)

[2.1 数据清洗及数据集创建 4](#_Toc10467)

[2.2 基于机器学习的预测算法 5](#_Toc25345)

[2.3 本章小结 6](#_Toc3217)

[第三章 研究内容与初步设计方案 8](#_Toc29715)

[3.1 研究内容 8](#_Toc3370)

[3.1.1 基于xgboost算法开发相关模型 8](#_Toc7808)

[3.1.2 基于预测模型对用户流失问题进行预测分析 8](#_Toc12891)

[3.2 相关设计方案 8](#_Toc17955)

[3.2.1 整体设计方案 8](#_Toc26608)

[3.2.2 基于xgboost算法开发相关模型 9](#_Toc18826)

[3.2.3 基于预测模型对用户流失问题进行预测分析 11](#_Toc29800)

[3.3 本章小结 11](#_Toc22204)

[第四章 预期难点及初步解决方法 12](#_Toc20776)

[4.1 预期难点 12](#_Toc24345)

[4.1.1 数据清洗与数据集创建过程 12](#_Toc22877)

[4.1.2 预测模型相关开发过程 12](#_Toc27453)

[4.1.3 实现对用户数据的预测分析过程 12](#_Toc11464)

[4.2 初步解决方法 12](#_Toc22295)

[4.2.1 数据清洗与数据集创建过程 12](#_Toc19975)

[4.2.2 预测模型相关开发过程 12](#_Toc32655)

[4.2.3 实现对用户数据的预测分析过程 13](#_Toc31)

[4.3 本章小结 13](#_Toc30312)

[第五章 预期成果及时间进度安排 14](#_Toc17741)

[5.1 预期开发结果 14](#_Toc13903)

[5.2 时间进度安排 14](#_Toc12649)

[5.3本章小结 14](#_Toc24619)

[第六章 结论 15](#_Toc5219)

[参考文献 16](#_Toc5167)

1. 引言

## 课题背景及意义

社区平台是一个具有广泛应用的在线交流和互动平台，如社交网络、论坛、在线市场等。随着互联网的快速发展，社区平台在人们的日常生活中扮演着越来越重要的角色。然而，用户流失一直是社区平台面临的一个重大挑战。各平台用户的消费行为呈现出内容需求精良化与多元化、消费终端选择多样化、消费体验社群化等特点，平台定位不清晰等问题会显著影响受众的体验感,进而会降低受众对平台的黏性。用户流失不仅意味着平台的用户数量减少，还可能导致用户活跃度下降、广告收入减少等问题，对平台的长期发展产生负面影响。在经历了一段概念风口和近几年的发展积淀后,各类平台用户整体增速下降，逐渐步入平复、梳理的冷静期，均呈现出用户访问率下滑、活跃度下降的趋势，其内容质量参差不齐等弊端日益显现，导致各类平台新用户增加缓慢、老用户复购率低、流失率高，这些问题一直困扰着 平台和产品提供者，而用户持续付费是充分发挥商业价值的前提条件，是付费市场稳定发展的原动力。 因此，探寻在线付费用户流失的内在机制，建立高效、可靠的用户流失预警体系对于付费市场的稳态发展具有巨大作用，是充分发挥付费市场商业潜力的重要保障。

steam作为全球最大的综合性数字发行平台之一，对自身平台用户流失进行预测与分析具有重要意义。通过对用户流失的预测，steam平台管理者可以提前采取措施，例如改进产品功能、优化用户体验、制定个性化推荐策略等，以减少流失率，提高用户满意度和留存率，从而促进steam平台的健康发展。

XGBoost（Extreme Gradient Boosting）是一种基于决策树的集成学习算法，具有高效、灵活、准确等优点，在数据挖掘和机器学习领域被广泛应用。利用XGBoost算法对steam社区平台用户流失进行预测与分析，可以有效地利用大量的用户行为数据和特征信息，构建预测模型，并对用户流失的影响因素进行深入分析，为平台管理者提供决策支持，帮助他们更好地理解用户行为、优化运营策略，降低用户流失率，提高平台的竞争力和盈利能力。

对steam社区平台的意义主要体现于以下几个方面：

1. 促进社区平台的可持续发展：通过减少用户流失率，提高用户满意度和留存率，可以促进社区平台的稳健发展，增强其市场竞争力和盈利能力；
2. 优化用户体验：通过分析用户流失的原因和影响因素，可以深入了解用户需求和行为特征，为平台管理者提供优化产品功能、改进用户体验的建议，从而提升用户满意度；
3. 提高广告收入：减少用户流失率可以增加平台的活跃用户数量，提高广告曝光和点击率，进而提升广告收入，为平台的商业模式带来更好的效益；
4. 拓展数据驱动决策：基于XGBoost算法构建的用户流失预测模型，可以帮助平台管理者基于数据进行决策，提高决策的准确性和效率，降低经验主义对决策的依赖程度，推动企业管理向数据驱动转型。

## 设计目标

本项目旨在应用XGBoost算法，结合Steam用户的历史数据，建立用户流失预测模型，以帮助Steam平台识别潜在的流失用户并采取措施来减少流失率。该方法将实现较高预测精度的情况下给出具有科学意义的用户流失预测，从而实现对平台管理者的启示作用。

## 国内外研究应用现状

近年来，随着互联网行业的蓬勃发展，国内很多研究机构、高校和企业开始关注社区平台用户流失预测与分析，相关研究的数量和热度逐渐提升。一些大型互联网企业（如阿里巴巴、腾讯、百度等）和社交平台（如微信、微博、知乎等）在用户流失预测与分析方面进行了深入研究和实践，通过构建预测模型和采取相应的运营策略，成功降低了用户流失率，提高了用户留存率和平台活跃度。国内研究者和企业倾向于采用先进的机器学习算法（如XGBoost、LightGBM等）和大数据技术，结合实际业务场景，对用户行为数据进行深度挖掘和分析，构建精准的用户流失预测模型，并应用于用户管理、个性化推荐、营销策略等方面。国内高校和研究机构也在社区平台用户流失预测与分析领域做出了一定的学术研究成果，涉及数据挖掘、机器学习、社会网络分析等多个领域，为该领域的发展提供了理论支持和方法指导。

在国外，一些知名的科技公司（如Facebook、Twitter、LinkedIn等）在用户流失预测与分析方面处于领先地位，通过大数据技术和机器学习算法，不断优化平台的用户体验和服务质量。国外的大学和研究机构也积极开展社区平台用户流失预测与分析的学术研究，涉及到计算机科学、人工智能、社会学等多个学科领域，推动了相关理论和方法的不断创新和进步。在国外，社区平台用户流失预测与分析往往与人机交互、用户体验、社交网络分析等跨学科领域相结合，形成了更加综合和深入的研究方向，为社区平台的发展和用户管理提供了更多的思路和方法。一些国外的研究机构和学术团体积极倡导开放数据共享和合作研究，推动了社区平台用户流失预测与分析领域的国际合作和交流，促进了研究成果的共享和传播。

综上所述，国内外在社区平台用户流失预测与分析方面都有着丰富的研究和应用经验，虽然存在一定的差异，但都致力于通过数据驱动的方法，提高社区平台的用户留存率和活跃度，促进平台的可持续发展。

1. 相关理论与技术基础

根据以上所收集的相关领域信息及国内外研究应用现状，综合该课题在当今时代的可用性，采用xgboost算法对预测模型进行构建，并以此预测用户流失数据。

2.1 数据清洗及数据集创建

数据清理（Data Cleaning）和数据集创建是数据科学中至关重要的步骤，它们将为后续的数据分析和建模提供干净、准确、完整的数据基础。

数据中常常存在缺失值，需要采取适当的方法处理，如删除缺失值、插值填充、使用模型预测等。同时，数据中可能存在的异常值会带来对数据分析结果的一定程度的影响，故需要进行识别和处理，对此可以使用统计方法、可视化方法或基于模型的方法进行检测，并采取相应的处理策略，如删除异常值或修正异常值。其中有关数据格式的问题通常的解决方案为将数据转换为适当的格式，如日期时间格式、数值格式、分类变量格式等，以便后续分析和建模使用。对于数据中可能存在的重复记录，需要进行识别和删除，确保数据集中每一条记录的唯一性。这之后仍需检查数据中的命名、单位、编码等是否一致，统一数据格式和标准，以减少数据分析中的错误和混乱。在完成以上步骤以后对数据进行转换和标准化，使其符合模型的假设和要求，如对数转换、标准化、归一化等。 对数据处理的相关进程如下图1-1。

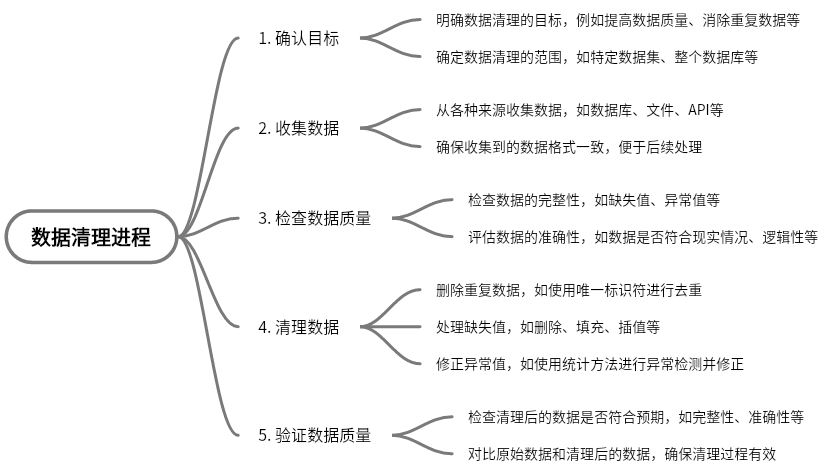


图1-1 数据清洗进程

在数据集创建方面，需要根据领域知识和统计方法选择重要的特征，减少模型复杂度和计算成本，从而提高模型的泛化能力。在特征选择基础上进行的特征工程实现了对原始数据进行特征提取、变换、组合等操作，创建新的特征，以增强模型的表达能力和性能。当数据集存在类别不平衡相关问题时，需要采取采样方法（如过采样、欠采样）或使用类别平衡的算法（如类别加权）加以处理，以避免模型偏向于多数类。数据集的作用并非唯一，需要将数据集划分为训练集、验证集和测试集，用于模型训练、调优和评估，以确保模型具有良好的泛化能力。其中需要对将适用于监督学习的数据集进行标注，即为每个样本赋予相应的标签或类别，以便训练监督学习模型。在整体创建过程中，如若存在多个数据源，需要将它们合并或连接成一个统一的数据集，以便进行统一的分析和建模。数据集创建过程如下图1-2。

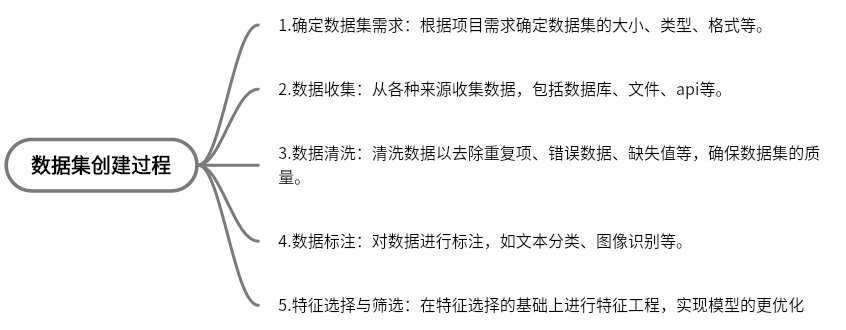


图1-2 数据集创建过程

2.2 基于机器学习的预测算法

XGBoost算法可以有效捕捉复杂数据的依赖关系，还可以利用可扩展的学习系统从大型数据集中学习并获得模型，因此，其在数据处理和分析方面的优势可用于解决电子平台用户流失预测问题，本研究构建了基于 XGBoost 算法的预测模型，并利用 XGBoost 算法实现了数据集分析、训练和预测。XGBoost 是一种在梯度提升框架下实现高效分类的算法。它改进了梯度提升机（GBM），梯度提升机具有高效、灵活、可移植等特点，能提供梯度提升决策树。梯度提升决策树的基本思想是梯度下降，即每生成一棵树都要基于前一棵树的结果来最小化目标函数。假设给定数据集为D，样本数为n，特征值数为m，则 D = {(,)}，（|D| = n，∈，∈R）。XGBoost的主要模型结构是由 K 个树状模型的加法模型生成的，每个树状模型都拟合了前一个树状模型的残差。我们可以将树的综合模型表示如下：

 （1）

 （2）

其中，F表示回归树模型空间，代表数据的特征向量，T代表树中叶节点的数量，代表独立树结构q和叶权重w。目标函数包含预测值 与训练实际值 之间的差值以及模型的复杂度。是训练数据的损失值,用于衡量模型在训练数据上的拟合程度。由于目标函数不能在传统优化方法的欧几里得空间中使用，因此每次迭代都会在前一次结果的基础上生成一棵树，同时为模型添加一个新函数。 XGboost 具有决策树的特点，而且还综合考虑了决策树的过拟合问题。该模型包含正则化，正则化通常用来衡量模型的复杂度。XGboost主要 使用L2正则化方法，包括叶节点的总数和从叶节点中获得的正则化值。每次迭代的结果记为。用于不断优化目标函数并在处进行泰勒展开，以加速目标函数的优化。是损失函数的一阶导数，是损失函数的二阶导数。最后得到训练目标函数。相关函数定义如下:

 (3)

 (4)

 (5)

 (6)

 (7)

 (8)

2.3 本章小结

数据清理和数据集创建是数据科学中至关重要的环节，它们为后续的数据分析、建模和决策提供了可靠的数据基础。通过合理的数据清理和数据集创建，可以提高数据质量，减少错误和偏差，从而得到更加准确和可信的分析结果和模型预测。XGBoost算法通过提升树的方式，结合了梯度提升和决策树的优点，具有高效、准确和可解释性等优点，在各种数据集上都取得了较好的性能，是数据科学领域中的一种重要工具。

1. 研究内容与初步设计方案

3.1 研究内容

基于xgboost开发的社区平台用户流失预测主要研究两个方面，即基于xgboost算法开发相关的模型、基于该模型进行的预测，以下将阐述具体研究内容：

3.1.1 基于xgboost算法开发相关模型

开发相关模型需要深入理解XGBoost算法的原理和机制，包括提升树（Boosted Trees）、梯度提升（Gradient Boosting）、决策树（Decision Trees）等基本概念和算法细节。数据清理过程中，需要确保数据的完整性、准确性和一致性，对于监督学习问题，需要有标记好的数据集，即有已知结果的样本数据。需要选择与目标变量相关性高的特征，避免不相关或冗余特征对模型造成干扰。在该命题下，数据的时效性对于用户流失预测任务尤为重要，由此需要确保使用的数据是最新的，反映了当前用户行为的情况，以提高模型的预测准确性。

3.1.2 基于预测模型对用户流失问题进行预测分析

首先基于预测模型对用户流失问题进行预测分析需要了解的是用户流失问题这一根本概念，因此深入了解所处行业的特点和用户流失的背景，理解用户流失对企业的影响以及用户流失的原因和模式存在一定的必要性。在此基础上获得的一定观念将影响对所收集到的用户流失数据的看法，即对该数据的变化特点、分布、相关性等有了初步的预测，为建立预测模型和之后的分析提供基础。对用户流失问题的预测建立在利用api调用基于xgboost算法开发的预测模型之上，而其中的分析则需建立在以上对数据的认知之后，根据对用户的各类登录信息的分析来预测其是否可能成为流失用户。通过以上研究内容，可以参考其中数据为企业制定有效的用户留存策略和管理措施提供科学依据。

3.2 相关设计方案

3.2.1 整体设计方案

拟采用前后端分离的模式对整体用户流失预测的功能进行设计，实现一个基于XGBoost的社区平台用户流失预测与分析的前后端分离项目，用户可以通过前端界面输入数据，后端使用XGBoost模型进行预测并返回结果。其中前端采用vue.js进行设计编写实现，包括数据展示、预测结果展示等，使用HTML、CSS和JavaScript等技术开发前端页面，确保页面能够与后端进行数据交互。使用Ajax技术与后端API进行通信，发送用户输入数据并接收预测结果。后端将采用Flask开发API接口，用于接收前端请求并返回预测结果。同时采用numpy、pandas库进行数据分析，编写数据预处理代码，包括数据清理、特征工程等，然后使用XGBoost算法训练用户流失预测模型。项目整体设计图如下图3-1。

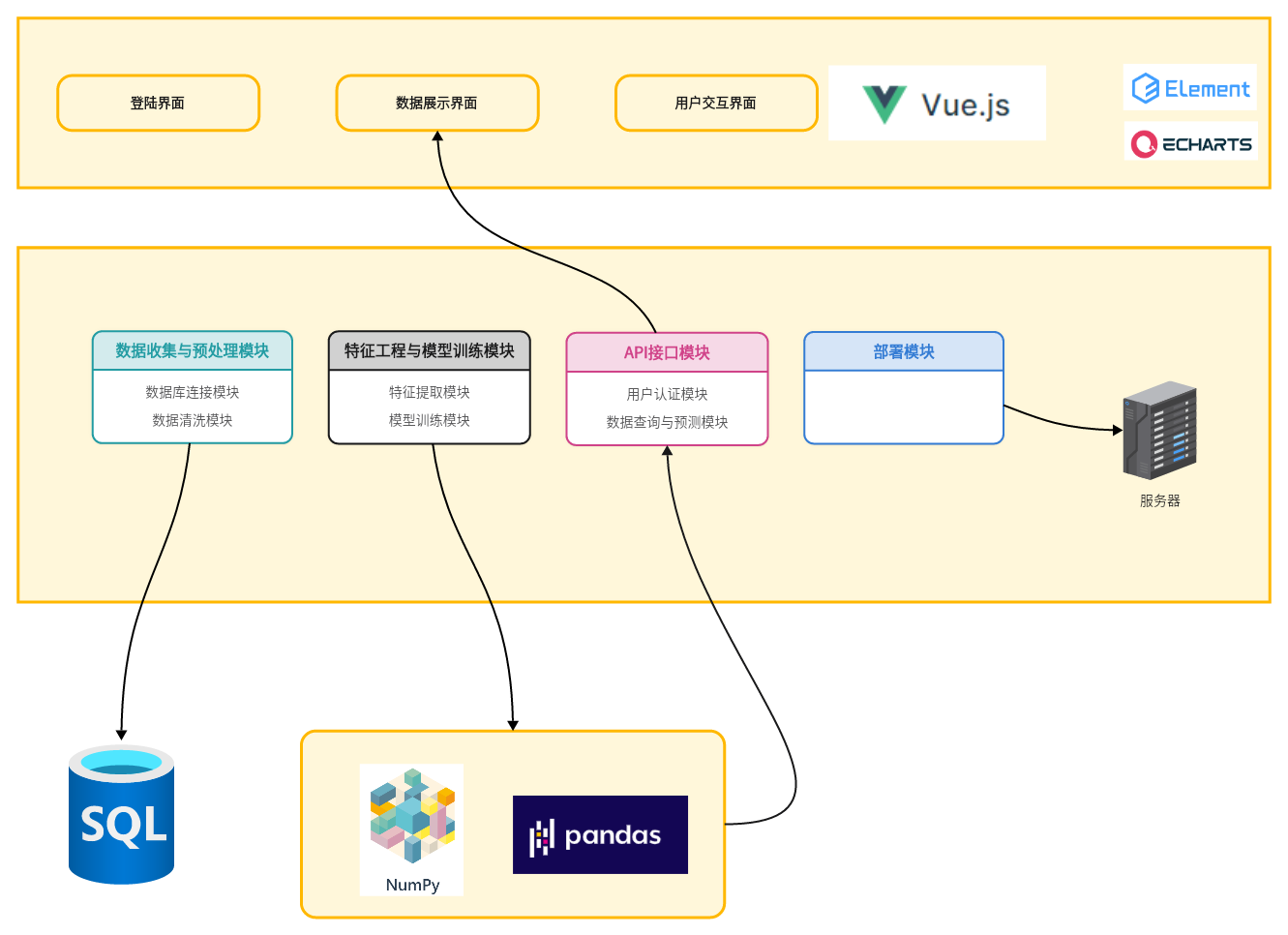


图3-1 项目整体设计图

前端方面数据展示界面用于展示活动度统计、流失预测结果等信息，交互界面可供用户输入单独的查询条件以查看预测结果、与界面各元素进行交互等操作，除此之外还有错误信息提示等功能。拟采用element-ui及echarts对界面元素进行优化构造。在后端模块中，数据库连接模块将用于获取用户行为数据，数据清理包括了对数据的整理、缺失值处理、异常值处理等预处理操作。计划从清洗后的数据中提取特征（如用户活跃度、登陆频率、社区互动、付费情况、在线峰值等特征），并根据所提取的特征使用XGBoost算法构建用户流失预测模型，并在训练集上进行训练。除此之外，后端还提供API接口，以接收前端请求，查询用户数据并调用训练好的模型进行预测，进而返回预测结果。

3.2.2 基于xgboost算法开发相关模型

首先收集用户流失相关的数据，包括用户个人信息、行为数据、交互数据等，再对数据进行探索性分析从而了解数据的特点、分布、相关性。对收集到的原始数据进行预处理，包括数据清洗、缺失值处理、异常值处理、特征工程等，在对已准备好作为模型训练输入的数据进行特征提取、选择和转换之后，将收集有关用户活跃度、登陆频率、社区互动、付费情况、在线峰值等多方面值，构建适合XGBoost模型的特征集并按一定比例分为训练集、验证机和测试集，以提高模型的性能和泛化能力。完成数据整理工作之后使用XGBoost算法构建模型，并在训练集上进行模型训练，通过交叉验证等方法对模型参数进行调优，在此基础上使用验证集对训练好的模型进行评估并分析分析模型的预测结果（包括对准确率、精确率F1值、召回率等指标的计算）。最后将训练好的模型部署到生产环境中，开发相应的API接口，以便其他系统或应用程序能够调用模型进行预测，并在使用过程中根据情况进行优化调整。开发流程如下图3-2所示。

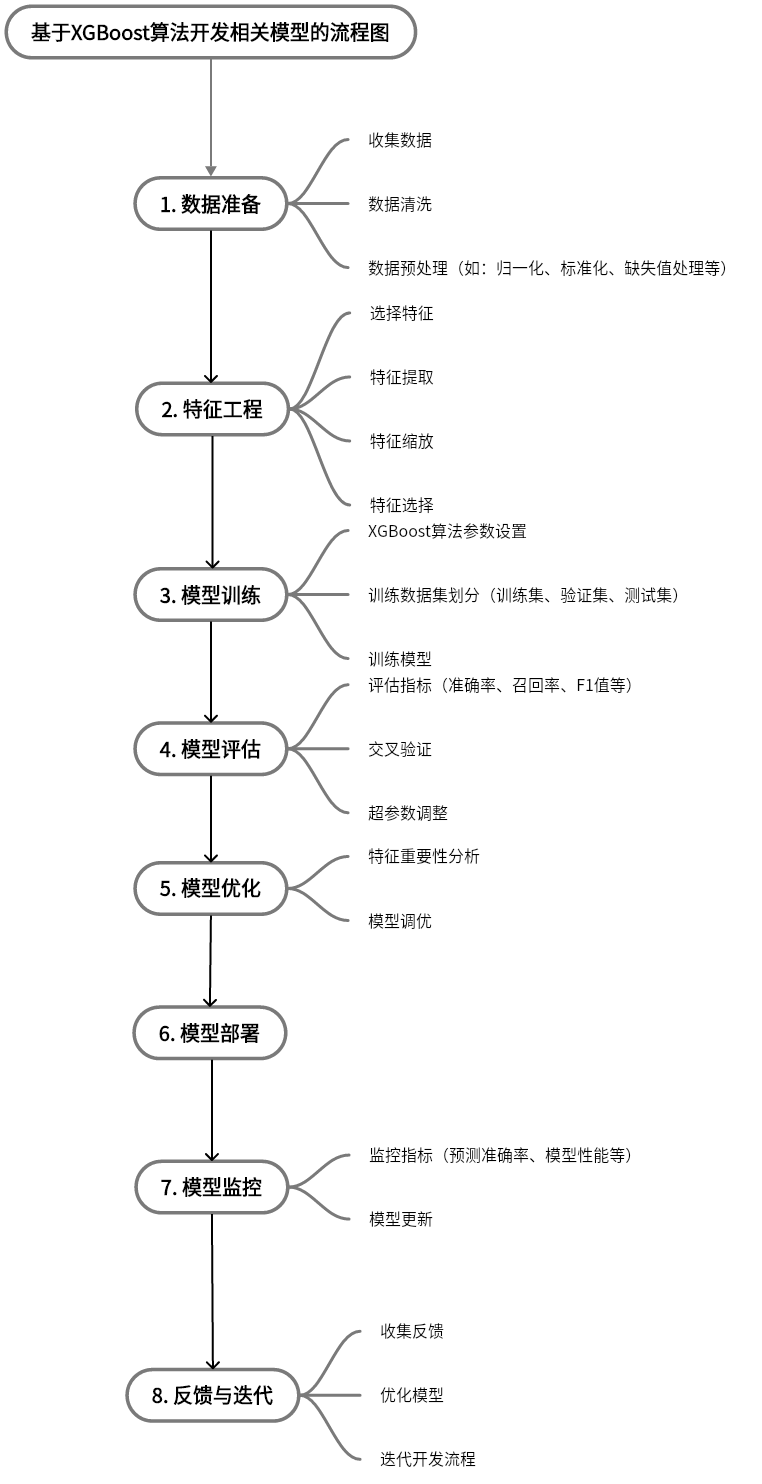


图3-2 基于xgboost算法开发相关模型的流程图

3.2.3 基于预测模型对用户流失问题进行预测分析

预测模型在用户流失问题中的具体使用如使用训练好的模型对各类用户进行预测，得出他们是否可能会流失的结果，并根据这一结果给出分析，供平台管理者参考以制定相应的用户留存策略，在此之后仍能继续监测平台各应用及整体平台的用户数据，可以更为直观地看到所指定的策略的有效性，以便后续调整优化。根据以上两点可以将该预测模型运用于平台的各个时间节点，从而不断为平台运营和管理提供可视化的科学依据，为平台管理者提供用户留存率及社区活跃度等相关信息。

3.3 本章小结

本章详细叙述了有关基于XGBoost的社区平台用户流失预测与分析的前后端分离项目研究内容与相关设计方案，大体上将其分为了三个内容，分别为基于xgboost算法开发相关模型，基于预测模型对用户流失问题进行预测分析以及前端界面元素相关优化设计三个方面，总结了各个模块大体上所需使用的技术、设计思路以及预期功能。

1. 预期难点及初步解决方法

4.1 预期难点

4.1.1 数据清洗与数据集创建过程

数据清洗是数据分析工作流程中的基本内容。根据“进去的是垃圾,出来的也是垃圾(garbage in, garbage out)”这一原理,可靠的数据是正确开发软件及后续策略实施的根本，因此数据清洗工作中可能存在的各类问题都将成为整体工作中的难点，存在其中的难点包括但不限于缺失值处理、异常值处理、数据不一致性、数据重复性、样本不平衡、数据质量不高等问题。

4.1.2 预测模型相关开发过程

在完成前期的数据清洗与数据集创建过程后，开发基于xgboost的预测模型本身并不具备巨大的难点，但如何获取更优解，即更佳效果的预测模型仍存在问题，模型训练的过程中可能存在欠拟合或是过拟合问题。前者指模型无法很好地拟合训练数据，后者指模型在训练集上表现良好但在测试集上表现较差。同时，对于特征的选择以及特征工程的问题对模型的性能影响来说也是避之不及的部分。

4.1.3 实现对用户数据的预测分析过程

往往当数据的样本量越小时，一个突出的数据会越大程度地影响实验者对该组数据的整体评判误差，这也要求了数据预测分析过程所需的样本量往往较大。在这样的背景下，该阶段的工作主要是以统计为主，即收集总体数据，对其中的单体数据得不到完整的了解，且数据分析的对错在项目起初并不能得到很好的评判，原因是缺乏实际数据的对该分析的佐证，即需要一定的试错成本。

4.2 初步解决方法

4.2.1 数据清洗与数据集创建过程

对于其中的缺失数据，可以根据其出现频率选用直接删除或是插值的方法；对于其中的异常值，通常的处理方法为根据统计学方法识别并处理或是替换；对于可能存在的数据不一致的问题可以选择标准化命名的字段对其进行转换；对于样本不平衡问题通常选用过采样或欠采样，同样也可以采用集成学习方法来处理；数据质量方面往往进行对数据的质量评估，并根据质量评估的结果进一步进行清洗和处理，从而修复数据质量问题。

4.2.2 预测模型相关开发过程

针对以上问题需要对数据进行充分分析的基础上，对模型不断进行调优和评估，并使用验证集对训练好的模型进行评估，分析分析模型的预测结果（包括对准确率、精确率F1值、召回率等指标的计算）。采用正则化、增加数据量等方法通常适用于解决模型训练过程中的欠拟合和过拟合问题，在实际应用中进行合理的特征选择和工程处理也能对模型的性能带来很高的正面收益。

4.2.3 实现对用户数据的预测分析过程

以上问题的解决方案首先要考虑的就是足够多的样本支撑，这一点在该项目中能够得到实现。其次，在项目初期缺乏需要用于佐证的实际经验数据，相应的对策是进行多次试验和反复验证，不断调整和优化分析方法，逐步提高对数据的理解和分析能力。并在数据分析过程中，结合所了解的知识和经验，通过深入了解数据背后的业务和领域知识，从而更准确地判断数据分析的有效性和可靠性，降低试错成本。

4.3 本章小结

这一章节主要提及了在这一项目各个阶段不同进程中可能出现的问题，对每个进程预期的难点与相应的解决方案进行了汇总，从而给项目的部署与实施提前提供了一部分解决思路。

1. 预期成果及时间进度安排

5.1 预期开发结果

本毕业设计的预期成果包括：

（1）一个基于XGBoost的社区平台用户流失预测与分析项目；

（2）一篇论述基于XGBoost的社区平台用户流失预测与分析项目的开发理念，过程，结果展示的论文。

5.2 时间进度安排

本毕业设计的时间进度安排将分为五个阶段。具体时间安排如下图5-1。

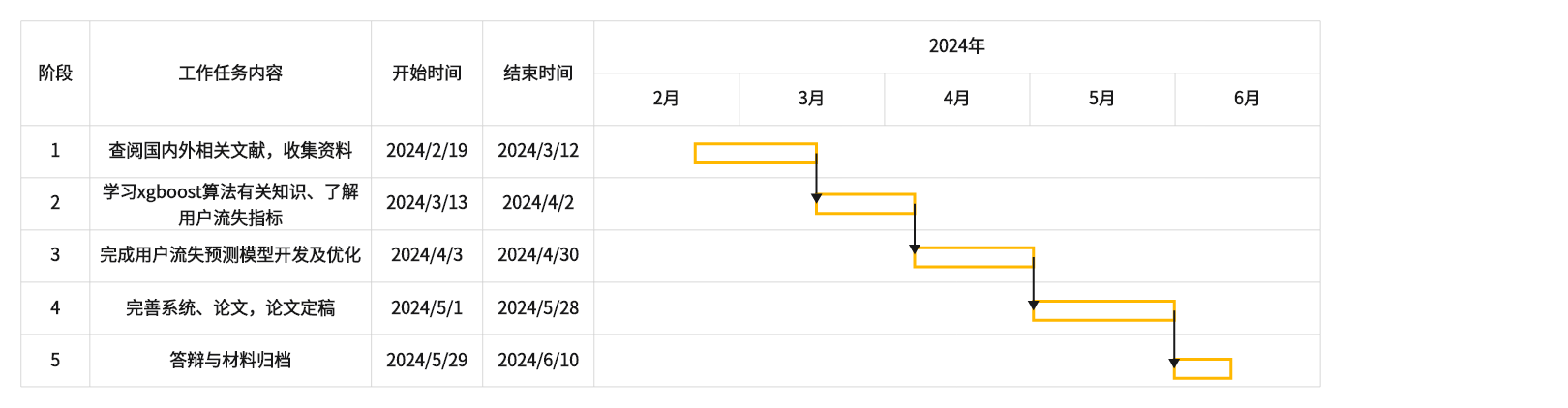


图5-1 毕业设计时间进度安排

5.3本章小结

该章节为对整体毕业设计项目所计划完成的内容及各阶段持续时间进行说明与统筹安排。

1. 结论

数据作为新时代概念下的生产要素，已经渗透到当今的各个行业和业务职能中，并逐渐成为目之所及的重要生产要素。据国际数据公司（IDC）统计，人类产生的数据量正呈指数级增长。大数据时代，人们可以将获取的大量数据转化为大量有价值的知识，推动人类生活方式的变革，大数据时代数据的战略意义不仅在于掌握海量的信息，还在于发现和理解信息的内涵以及信息与信息之间的关系，而对大数据的分析是大数据领域研究的核心内容之一，数据分析作为决策过程中的决定性因素，自然对大数据时代的信息和信息的发展起到推动作用，同时也成为了大数据时代发挥数据全部价值地最关键环节。由此可见对数据的预测分析将成为一个不可或缺的根本环节，其对steam乃至各大平台的经营模式与效益等能够将其贴合程度很好地具象化，对各大平台的发展都具有一定的可鉴性。

参考文献

1. 中华人民共和国中央人民政府. 国务院关于积极推进“互联网+”行动的指导意见[EB/OL]. [2021-09-15]. http://www.gov.cn/zhengce/content/2015-07/04/content\_10002.htm.
2. 邢绍艳, 朱学芳. 付费知识直播用户流失预测实证研究[J]. 信息资源管理学报, 2022, 12(4): 121-130.
3. Gelman, A., & Loken, E. (2016). The statistical crisis in science. The best writing on mathematics (Pitici M, ed), 305-318.
4. Song, P. i Liu, Y. (2020). An XGBoost Algorithm for Predicting Purchasing Behaviour on E-Commerce Platforms. Tehnički vjesnik,27 (5),1467-1471.https://doi.org/10.17559/TV-20200808113807
5. Shrestha Y R, He V F, Puranam P, et al. Algorithm supported induction for building theory: How can we use prediction models to theorize?[J]. Organization Science, 2021, 32(3): 856-880.
6. 郭新茹, 倪嘉玥, 曾嘉怡. 我国互联网视频平台用户黏度提升策略探究——以芒果 TV 独播为例[J]. 文化产业研究, 2018 (1): 118-129.
7. 李占山,刘兆赓.基于XGBoost的特征选择算法[J].通信学报,2019,40(10):101-108.
8. 高志鹏, 牛琨, 刘杰. 面向大数据的分析技术[J]. 北京邮电大学学报, 2015, 38(3): 1-12.
9. Molnar C, Casalicchio G, Bischl B. Interpretable Machine Learning - A Brief History, State-of-the-Art and Challenges[C]// Proceedings of the 2020 Workshops of the European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer, 2020: 417-431.
10. 王若佳, 严承希, 郭凤英, 等. 基于用户画像的在线健康社区用户流失预测研究[J]. 数据分析与知识发现, 2022, 6(2/3): 80-92.
11. [11]TR, R. ., Lilhore, U. K., M, P. ., Simaiya, S. ., Kaur, . A. ., & Hamdi, M. . (2022). PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES. Malaysian Journal of Computer Science, 132–148. <https://doi.org/10.22452/mjcs.sp2022no1.10>
12. Leong C M, Tan K L, Puah C H, et al. Predicting mobile network operators users m-payment intention[J]. European Business Review, 2021, 33(1).
13. Asselman A, Khaldi M, Aammou S. Enhancing the prediction of student performance based on the machine learning XGBoost algorithm[J]. Interactive Learning Environments, 2023, 31(6): 3360-3379.
14. Lalwani P, Mishra M K, Chadha J S, et al. Customer churn prediction system: a machine learning approach[J]. Computing, 2022: 1-24.
15. Safeer A A, Yuanqiong H, Abrar M, et al. Role of brand experience in predicting consumer loyalty[J]. Marketing Intelligence & Planning, 2021, 39(8): 1042-1057.