蘭州資源環境職業技術大學

LANZHOU RESOURCES & ENVIRONMENT VOC-TECH UNIVERSITY

本科毕业设计

题　　目：	基于机器学习的校园一卡通消费行为数据分析系统设计与实现
学　　院：	信息工程学院
专　　业：	人工智能工程技术
年　级、班：	2023 级人工智能工程技术(专升本)11 班
学 生 姓 名：	郑　倩
指 导 教 师：	尉雅晨

二〇二五年五月三十日

兰州资源环境职业技术大学学位论文原创性声明

本人郑重声明:所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。因本学位论文引起的法律后果完全由本人承担。

学位论文作者签名：郑倩

签字日期：2025 年 6 月 1 日

基于机器学习的校园一卡通消费行为数据分析系统设计与实现

郑倩

（兰州资源环境职业技术大学信息工程学院人工智能工程技术专业(专升本)11班级，甘肃 兰州 730021）

指导教师：尉雅晨

【摘要】随着智慧校园建设的应用及推广，校园一卡通积累了大量学生消费数据，这些数据是校园管理和服务优化的重要资源，能在智慧校园建设中体现价值。本文基于机器学习技术，设计并实现了一套校园一卡通消费行为数据分析系统。通过数据挖掘技术对一卡通消费数据进行预处理，提取关键特征，系统采用 K-Means++聚类算法和 WMDE-Apriori 加权关联规则分析方法，对消费数据进行聚类分析和关联评估，进一步挖掘学生的消费模式和行为特征。本文通过机器学习算法，结合可视化技术，设计并实现一卡通数据分析系统，直观的展现出学生的消费趋势和消费习惯，推动智慧校园建设中数据的多维度应用及价值提升。

【关键词】校园一卡通;消费行为分析;数据挖掘;机器学习;智慧校园

Design and Implementation of a Machine Learning-based Data Analysis System for Campus One Card Consumption Behaviour

QianZheng

(Grade 2023, Class (Specialised) 11, Major Artificial Intelligence Engineering Technology, Department of Information Engineering, Lanzhou Resources & Environment Voc-Tech University, Lanzhou 730021, Gansu)

Tutor: YachenWei

Abstract:With the application and promotion of smart campus construction, the campus card has accumulated a large amount of student consumption data, which is an important resource for campus management and service optimisation, and can reflect the value in the smart campus construction. Based on machine learning technology, this paper designs and implements a campus card consumption behaviour data analysis system. The card consumption data is preprocessed by data mining technology to extract key features, and the system adopts K-Means++ clustering algorithm and WMDE-Apriori weighted association rule analysis method to perform cluster analysis and association assessment of consumption data to further mine students' consumption patterns and behavioural characteristics. In this paper, through machine learning algorithms, combined with visualisation technology, a card data analysis system is constructed to visually display students' consumption trends and consumption habits, and to promote the multi-dimensional application and value enhancement of data in the construction of smart campuses.

Key words: Campus Card; Consumption Behavior Analysis; Data Mining; Machine Learning; Smart Campus

目 录

1 绪论

1.1 研究背景与意义

1.1.1 研究背景

2018 年 4 月，中国人民共和国教育部在《教育信息化 2.0 行动计划》中指出，加快数字化校园的建设，推动信息技术与教育教学深度融合，高校普遍构建基于智慧校园的一卡通系统。一卡通系统作为校园信息化建设的重要组成部分，为学生提供了便捷服务，积累了大量的消费行为数据，这些数据记录了学生在校内的各种活动行为，反映了学生的生活和学习习惯。虽然一卡通系统积累了大量数据，但这些数据存在的潜在价值未被充分挖掘。挖掘数据背后隐藏的学生学习规律和生活需求等关键信息，通过数据分析，有效利用这些数据为校园管理和服务优化提供新的管理思路和方向。传统的数据分析方法，在处理大规模数据时，需要花费数天时间[27]，异常数据的检测需要依赖人工规则，准确率仅 50%[28]。相较之下，引入机器学习算法和数据挖掘技术可以将亿级数据处理缩短至几个小时[29]，将准确率提升至 90%[30]。有效的印证了机器学习算法在数据分析和处理上都有优势。随着数据挖掘和机器学习技术的快速发展，为数据分析提供了更加有利的求解手段。WMDE-Apriori 和 K-means 算法能够对复杂数据进行聚类和关联分析，揭示数据背后的规律和模式。本文通过机器学习算法，对一卡通数据进行深层次挖掘，探索其在校园管理和服务优化中的应用价值。

1.1.2 研究意义

校园一卡通消费行为分析系统展示了机器学习技术在数据分析领域的强大应用潜力。通过引入机器学习算法，系统能够高效的处理和分析大量校园消费数据。在对校园消费数据进行处理时，能够验证机器学习算法在面对特定类型数据时的有效性与适用性。通过对消费数据分析，系统能自动识别学生的消费模式和行为特征，进一步把握学生的生活需求和消费习惯，为校园管理提供更有针对性更科学合理的管理决策和思路。在数据处理和分析的过程中还能及时发现并干预过度消费等异常消费行为，帮助学生树立健康、合理的消费观念。通过机器学习技术的应用，为校园管理和服务优化提供新的思路，系统的数据分析效率和准确性也得到了提升。

1.2 国内外研究现状

1.2.1 国内研究现状

随着人工智能行业的兴起，我国在数据分析方面的技术获得了巨大的进展[3]。2018 年, 邹志洪[1]使用改进的 Apriori 关联规则算法，挖掘学习消费行为与成绩之间关联属性特征。姜楠[18]等使用 K-means 和 Apriori 算法分别对学生消费行为和学习行为进行了消费习惯聚类分析和学习行为关联度分析。李娜建立了数据分析决策系统，将聚类和关联规则相结合[40]。2018 年, 尹春梅[2]通过数据挖掘技术来分析学生的校园卡消费记录, 辅助学生管理工作，提出了一种消费水平评分算法分析贫困生识别结果。2019 年, 高语蔚[4]

使用 K-Means++算法对学生消费数据进行聚类处理。在消费行为模式分析方面，2020 年, 任志愿[5]使用 PVW-Kmeans 算法和基于 WMDE-Apriori 算法对学生消费行为和学习行为进行了消费习惯聚类分析和学习行为关联度分析。2020 年, 谢慧[6]，设计并实现了一个校园一卡通日志分析系统，以联机分析处理技术和数据挖掘技术为基础进行设计。2023 年, 杨晨[8]提出 Apriori 算法的改进，提出基于矩阵优化的关联算法，提高了频繁项集的挖掘效率。

1.2.2 国外研究现状

2016 年，Olivier 等人[21]对 K-means 聚类算法进行了反复改进，利用 Markov Chain 让算法的迭代速度有了提升。2016 年，Olivier[22]又进一步提出了一个概率分布来使算法的鲁棒性变强。2008 年，Huaifeng Zhang 等[12]人提出了一种有效的算法来挖掘不平衡数据集上的组合关联规则，组合的关联规则被组织为多个规则集。2022 年, Kaur 等人[14]研究探讨了关联规则对全球股票指数预测的有效性和性能，验证了关联规则可用于技术指标数量有限的盈利决策。2016 年，Rodrigues R L 等人，运用校园一卡消费数据，分析认定贫困生评选[35]。国外典型的一卡通数据分析系统是由牛津大学建立的[36]，该系统结合一卡通数据的特点，提出了基于决策树技术的一卡通分析应用研究模型。

1.3 研究内容

本文基于 Python Flask 框架开发了校园一卡通消费行为分析系统，利用机器学习算法深入分析校园一卡通消费行为数据。通过数据预处理技术对多源异构数据进行整合和优化，结合 K-Means++聚类算法、WMDE-Apriori 加权关联规则算法构建消费行为分析模型。在模型构建中，提取日均消费金额、消费频次、用餐规律性等关键特征，通过标准化处理和异常值过滤优化数据质量。通过系统结果分析，表明该系统能够准确识别学生消费水平分层、消费高峰时段、热门商家及用户忠诚度等，直观的显示学生消费行为的差异。

1.4 论文结构安排

本文在研究过程中，结合校园实际为出发点，阐述了论文研究的意义，介绍了机器学习相关算法，并通过机器学习算法对一卡通数据进行分析，设计并实现了校园一卡通消费行为分析系统。本文共由五章组成，组织结构与安排如下：

第一章。结合高校一卡通数据分析与管理的实际情况阐述论文研究的背景及意义，分析一卡通数据分析与挖掘的国内外研究现状，对论文的主要研究内容及结构安排作简要介绍。

第二章。简要介绍了数据挖掘的流程，并对机器学习和常用于数据分析的算法 K-means++算法、WMDE-Apriori 算法进行基础知识简介。

第三章。分析系统需求，对系统架构及关键模块进行设计。对一卡通消费行为分析系统所应用的关键算法的具体实现与实验验证进行详细介绍，并将实现所需要搭建的环境进行简单介绍。

第四章。主要展示了实验结果，对结果进行聚类效果的评估，分析了消费行为特征。

第五章。对系统关键模块进行功能级性能测试，验证系统可行性。
第六章。对研究成果进行总结，并对下一步工作进行展望。

2 相关技术与理论基础

2.1 数据挖掘技术

数据挖掘是一门综合性的学科，囊括了多个领域的技术。常用的数据挖掘方法包括聚类分析法、回归分析法、变量分析法以及线性分析法。数据预处理是数据挖掘的基础，通过建立模型的方式对这些数据进行处理，从而挖掘出数据背后隐藏的信息和价值。知识发现类挖掘技术具体包括关联顺序法、遗传算法、决策树法以及人工神经网络法等多种方法[23]。数据处理过程如图 2.1 所示。

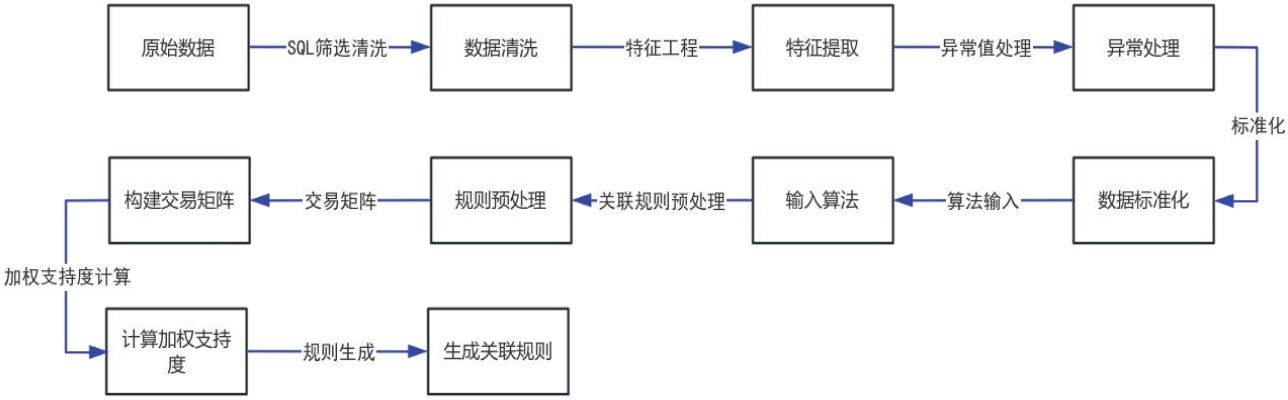


图 2.1 数据处理流程图

2.2 机器学习算法

机器学习是人工智能发展的核心技术，通过各类算法对经过预处理的数据进行分析，发现数据中隐藏的潜在信息，利用这些挖掘的信息进行预测、分类、决策。机器学习的过程主要包括数据的使用和机器学习算法的建模，一般情况下，获取的数据都是原始数据，在数据应用到模型训练之前，需要对数据进行转化。前期的数据处理工作和模型训练的拟合，构成了基本的机器学习任务[33]。

数据采集是机器学习任务的起点，在特征工程之前，要对采集到的原始数据进行预处理。特征工程基本决定了机器学习的上限，通过特征组合、特征拆分、外部特征关联进行特征构建，从多个维度表达信息，从每个特征自身的取值范围以及与目标结果的相关性上进行特征选择。用训练数据对模型进行训练，学习数据中的模式和规律。对模型进行评估和优化，这是一个不断重复的过程，调整后的模型需要进一步检验和误差分析。机器学习任务的最终目标就是将模型应用到实际场景中。

2.2.1 聚类算法

K-Means 聚类算法是经典的无监督学习算法。它的的基本思想是通过迭代的方式，将数据划分为 K 个不同的簇，使得每个数据点与其所属簇的质心之间的距离之和最小[31]。K-means 算法的计算速度快且易于实现，直观易懂。但 K-means 算法也存有一些局限性，在初始簇质心的选择上表现敏感，可能陷入局部最优解的状态，需要预先设定聚类数 K 值。K-means 算法执行的核心步骤如表 2.1 所示，算法流程图如图 2.2 所示 。

表 2.1 K-means 算法步骤

K-means 算法步骤
①先随便选出 k 个初始点作为质心； ②将数据集合中的点分配到一个簇里面，也就是说寻找每个离散的点各自距离其最近的质心，并把他分配给该质心所对应的簇当中去； ③对各个簇的质心进行移动，移动位置是该簇所有点的平均值； ④反复重复进行步骤②、③，直到每个点对应簇的分配不变为止。

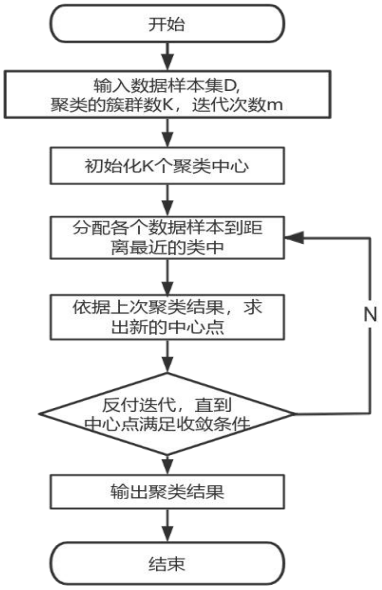


图 2.2 K-means 聚类算法流程图

K-means++ 算法在传统的 K-means 算法的基础上，优化初始聚类中心的选择，提升聚类效果。K-means 算法的初始化聚类中心完全随机选择，K-means++则采用距离加权概率，随机选择第一个质心，通过计算所有样本点到当前质心的最短距离，根据距离大小为每个数据点分配概率按照概率分布随机选择下一个质心，重复直至选择出 K 个质心使用选定的 K 个质心，执行标准的 K-means 算法即可，具体实现过程如图 2.3 所示

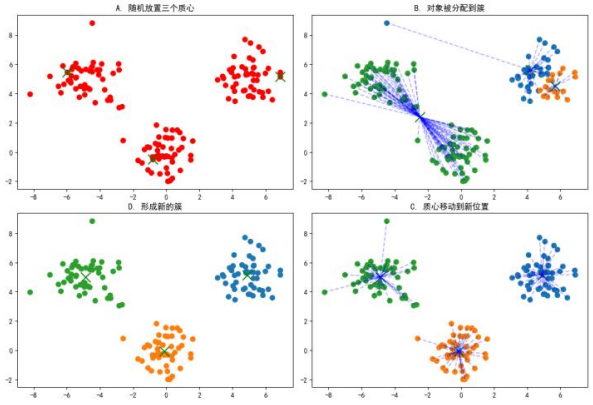


图 2.3 K-means++算法实现过程图

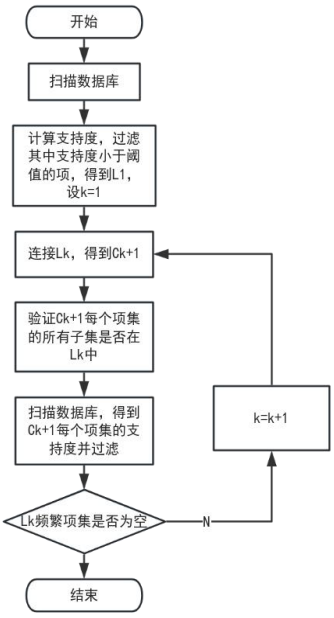
K-means++ 算法的平均复杂度是 O(kmNn)，其中 k 是人为设置的聚类中心数，也

叫簇数，N 是数据样本量，n 是算法执行的迭代次数，m 是特征个数。而最坏情况下，K-means++ 算法的时间复杂度会达到 $O(N^2\ mn)$[26]。相较之下，层次聚类算法的时间复杂度一般为 $O(N^2m)$，而 DBSCAN 算法虽然不预先指定簇，但是它的时间复杂度也较高，尤其是在处理高维数据时。因此，K-means++ 算法是聚类算法中性能较好的算法。

2.2.2 WMDE-Apriori 算法

关联规则算法多用于发现数据集中频繁项集和关联规则。其基本思想是通过分析数据集中的频繁项集，找出项之间的关联关系[32]。关联规则算法的核心是 Apriori 算法，通过遍历事务数据库，统计计算所有单项集出现的支持度，得到频繁项集 L_1，设置初始化迭代次数 K=1，通过逐层迭代，生成频繁项集 L_K，基于 L_K、(K+1)生与成候选频繁项集 C_{K+1} 通过剪枝去除不频繁项集，再次扫描数据库，计算候选频繁项集的支持度，保留满足条件的频繁项集，直到无法生成更高阶的频繁项集为止，从生成的频繁项集中提取关联规则，算法流程图如图 2.4 所示。

图 2.4 Apriori 算法流程图



WMDE-Apriori 算法是针对 Apriori 算法产生的大量频繁项集以及重复扫描数据库问题的改进版本，提出基于权重的改进方法。预先定义一个最小支持度，将数据库 D 进行划分，通过计算项集 I_j 和 T_i 的权重 W 生成候选项集，计算每个候选项集的支持度 support(S)，选择出频繁项集并判断频繁项集是否为空，重复执行直至将所有的频繁项集合并形成最终的频繁项集结束算法，算法流程如图 2.5 所示。

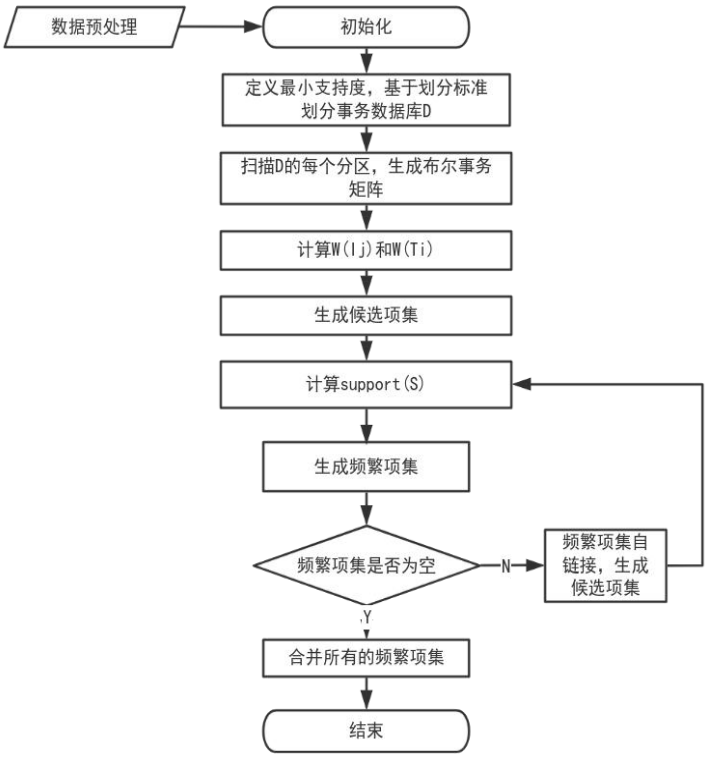


图 2.5 WMDE-Apriori 算法流程图

WMDE-Apriori 算法的伪代码[5]，如表 2.2 所示。

表 2.2 WMDE-Apriori 算法伪代码[5]

输入：	**事务数据库 D**
输出：	频繁项集 P
1:	定义最小支持度，基于划分标准，划分事务数据库 D 为 n 个分区 $D_i(i=1,2,...,n)$
2:	扫描 $D_i(i=1,2,...,n)$，生成不二事务矩阵 $M_i(i=1,2,...,n)$
3:	计算 $W(I_i)=\frac{1}{P(I_i)}=\frac{m}{l}$ 和 $\overline{W}(T_i)=\sum_{j=1}^{n\ \&\ I_i\in T_i}\frac{W(I_j)}{\lvert T_i\rvert}$
4:	生成候选集
5:	计算 $W_{\text{support}}(S)=\frac{\sum_{i=1}^{m\ \&\ s\subseteq T_i}\overline{W}(T_i)}{\sum_{i=1}^{m}\overline{W}(T_i)}$
6:	生成频繁项集，判断生成的频繁项集是否空。若为空，则合并所有的频繁项集 $P_i(i=1,2,...,n)$，生成最终的频繁项集 $P=P_1\cup P_2\cup...\cup P_n$；若不为空，则将频繁项自链接，生成新的候选项集，回到第 5 步。

3 系统需求分析与关键设计

3.1 需求分析

随着校园信息化发展，校园一卡通积累大量消费数据，为实现校园管理服务提供辅助，校园一卡通消费行为数据分析系统需具备全面功能与性能。在功能方面，系统要能够对原始消费数据进行清洗、标准化处理、去除异常值等预处理操作，同时要对学生的敏感信息进行脱敏处理，保护学生隐私安全。系统需要集成 K-Means++算法和 WMDE-Apriori 算法，对消费行为进行聚类和关联分析，生成动态的可视化分析结果。系统还需要对校园管理者提供账户管理和权限控制功能，利用 SHA-256 加密存储密码，保障系统访问安全。

在性能方面，系统需要具备高效和稳定的数据处理能力，能够流畅的处理单月级别的校园消费数据量，在实验环境下保证聚类和关联分析核心功能的响应时间不超过 10 秒。系统架构要支持高并发访问，平均无故障运行时间需在 99%以上，确保服务可靠。数据库访问必须经过身份验证，所有敏感数据传输和存储都要进行加密处理。系统设计要预留足够的扩展性，便于未来集成新算法，对接校园管理系统其他模块，应对服务需求和技术发展。

3.2 实验环境

为确保高效运行和其稳定性，本实验采用以下环境配置和依赖包：实验环境要求使用 Python 3.7 或更高版本，确保能支持最新的库和功能，能够充分利用 Python 生态系统中的先进工具和特性。MySQL 数据库较为稳定且性能强大，选择 MySQL 5.7.20 或更高版本，作为后端数据库，存储和处理一卡通消费数据。项目实验依赖于多个 Python 包，为系统提供必要的功能支持。核心依赖包括但不限于 Flask 用于 Web 应用开发、SQLAlchemy 作为 ORM 工具、Pandas 和 NumPy 用于数据处理和数值计算、Scikit-learn 用于实现机器学习算法。通过执行“python run.py”命令，启动应用服务器，开始实验。实验具体所需要的依赖包版本如表 3.1 所示。

表 3.1 实验依赖包

依赖包版本			
click==8.1.7	colorama==0.4.6	cycler==0.11.0	Flask==2.2.5
Flask-Login==0.6.3	fonttools==4.38.0	mysql-connector-python==8.0.33	itsdangerous==2.1.2
jinja2==3.1.4	joblib==1.3.2	kiwisolver==1.4.5	numpy==1.21.6
matplotlib==3.5.3	mlxtend==0.23.1	scikit-learn==1.0.2	pyparsing==3.1.4
packaging==24.0	pandas==1.3.5	Pillow==9.5.0	protobuf==3.20.3
zipp==3.15.0	pytz==2024.2	Werkzeug==2.2.3	threadpoolctl==3.1.0

3.3 系统架构设计

一卡通消费行为分析系统架构分为四个模块：数据源层、数据预处理层、数据分析引擎层、数据应用层。数据源层从不同的数据源获取系统所需的一卡通数据。将获取到

的数据经过数据清洗、数据集成、数据变换、数据消减，处理过后存入系统数据库。数据分析引擎层对系统数据库中的数据进行统计、聚类和关联分析。将分析结果可视化展示在一卡通消费行为分析系统中。

3.4 系统实现与关键模块设计

3.4.1 数据库设计

本设计截取广东技术师范大学 2018 级学生在 2019 年 4 月的校园一卡通消费数据。消费数据集包含学生的消费记录，如表 3.2 所示。

表 3.2 数据集示例

数据集示例：	
学生 ID	1，2，3
消费类型	食堂消费（Canteen），超市消费（Supermarket）
交易记录	[(1, Canteen), (1, Supermarket), (2, Canteen), ...]

数据集包含了学生的基本信息、消费地点、消费时间、消费金额以及一卡通卡号等关键信息。为了保护学生隐私，所有个人信息均已进行脱敏处理。构建数据库时，基于三个核心实体：学生信息表(Student)、门禁记录表(AccessRecord)和消费记录表(ConsumptionRecord)。学生信息表以学号(CardNo)为主键，存储学生基本属性，如表 3.3 所示。门禁记录表以记录编号(Index)为主键，其外键 AccessCardNo 关联学生表的 CardNo，形成一对多关系，表明一名学生可对应多条门禁进出记录，如表 3.4 所示。消费记录表同样以记录编号(Index)为主键，通过外键 CardNo 与学生表关联，支持一名学生产生多条消费记录的一对多关系，如表 3.5 所示。两类记录表均以学号为纽带，追踪学生的门禁行为与消费轨迹，数据库设计关系图如图 3.1 所示。

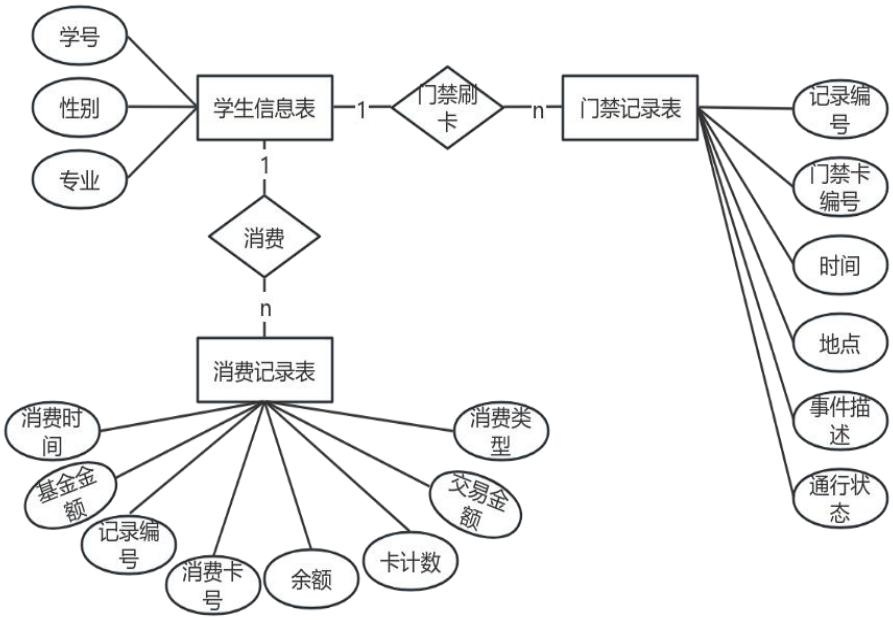


图 3.1 ER 图

表 3.3 Student

字段名	数据类型	描述
CardNo	VARCHAR(20)	学号(主键)
Sex	VARCHAR(10)	性别
Major	VARCHAR(50)	专业

表 3.4 AccessRecord

字段名	数据类型	描述
Index	INT	记录编号(主键)
AccessCardNo	VARCHAR(20)	门禁卡编号
Date	DATETIME	时间
Address	VARCHAR(100)	地点
Access	BOOLEAN	通行状态
Describe	TEXT	事件描述

表 3.5 ConsumptionRecord

字段名	数据类型	描述
Index	INT	记录编号（主键）
CardNo	VARCHAR(20)	消费卡号
Date	DATETIME	消费时间
Money	DECIMAL(10,2)	交易金额
FundMoney	DECIMAL(10,2)	基金金额
Surplus	DECIMAL(10,2)	余额
CardCount	INT	卡计数
Type	VARCHAR(50)	消费类型

3.4.2 聚类分析模块设计

采用 K-Means++算法，对校园一卡通消费数据进行聚类分析，在消费群体画像的绘制过程中将 K-Means++算法同传统的 K-means 算法做了对比分析，验证 K-Means++算法的高效准确。

基于 K-Means++算法的聚类分析模块流程如下：

第一步，从 MySQL 数据库获取原始的消费数据，去除大额异常消费数据，筛选有效的消费记录，对用户数据进行基本的分组统计。

第二步，对数据进行分析，对日均消费金额、消费频率、用餐规律性、活跃账户等核心特征进行计算，提取特征属性值。

第三步，使用 K-Means++算法绘制用户群体画像，并给出聚类分析结果。

聚类分析模块流程实现如图 3.2 所示。

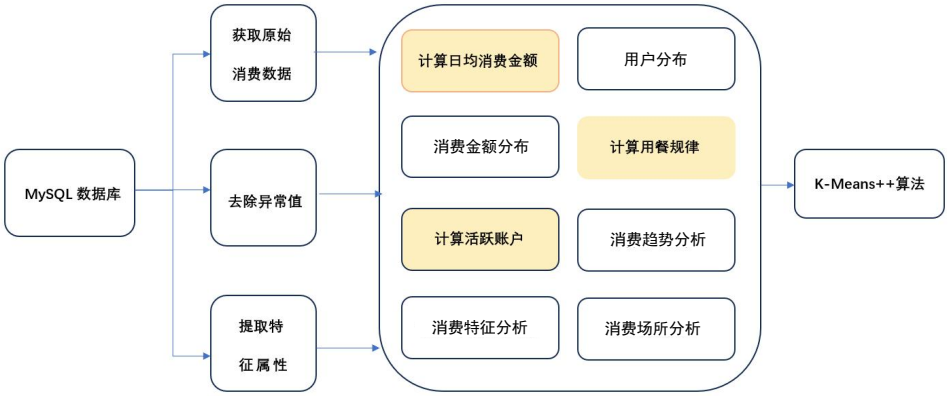


图 3.2 聚类分析模块流程图

通过 SQL 查询获取符合条件的学生消费数据，包括消费频率、平均消费金额、账户活跃天数、早中晚餐消费次数等其他基本指标。系统设定筛选条件，要求分析对象至少有 10 次消费记录且活跃天数至少 7 天。计算三个关键特征：日均消费频率=消费次数/活跃天数、日均消费金额=总金额/活跃天数、用餐规律性=三餐消费次数/总消费次数。从频率、金额、规律三个维度刻画用户消费行为特征。使用 IQR 四分位距方法清除异常值，计算每个特征的第一四分位数 Q1 和第三分位数 Q3，以 Q1-1.5IQR 和 Q3+1.5IQR 为边界筛除异常值，如图 3.3 所示。

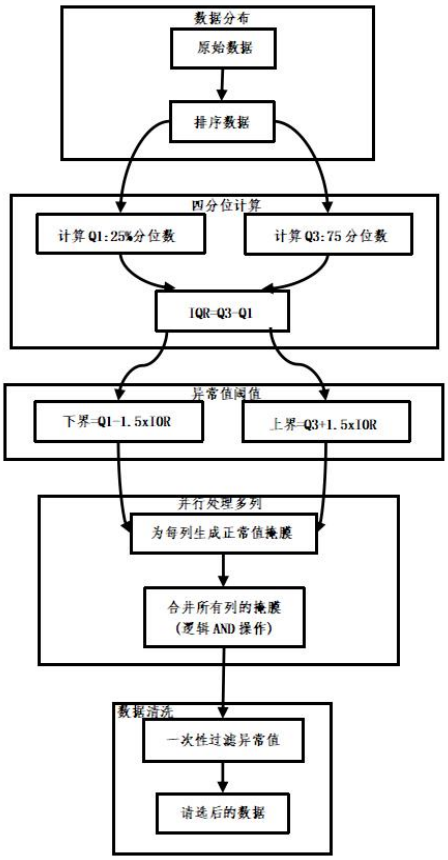


图 3.3 IQR 流程图

应用 StandardScaler 进行特征标准化，确保不同特征在聚类中具有相同的权重。调用 scikit-learn 的 K-Means++算法，设置聚类数为 3，随机种子为 42，根据每个聚类的日均消费金额，将三类群体标记为高、中、低消费群体。为每个群体生成描述性名称，同时计算各类群体的占比和指标。为提高性能，聚类分析的结果将被缓存。

将 K-Means++算法与传统的 K-Means 算法进行对比实验。实验对比结果如图 3.4 所示。K-Means++算法在所有关键指标中均优于 K-Means 算法。K-Means++算法的组内平方和 SSE 仅为 955.65，明显低于 K-Means 算法，能够更好的将数据点聚集在质心附近。K-Means++算法的轮廓系数比 K-Means 更接近于 1，在分组聚类后各个簇之间分界线更加清晰，簇内数据一致性更高。此外，K-Means++的平均迭代次数远小于 K-Means，它的收敛速度和聚类效率也具有相对优势。

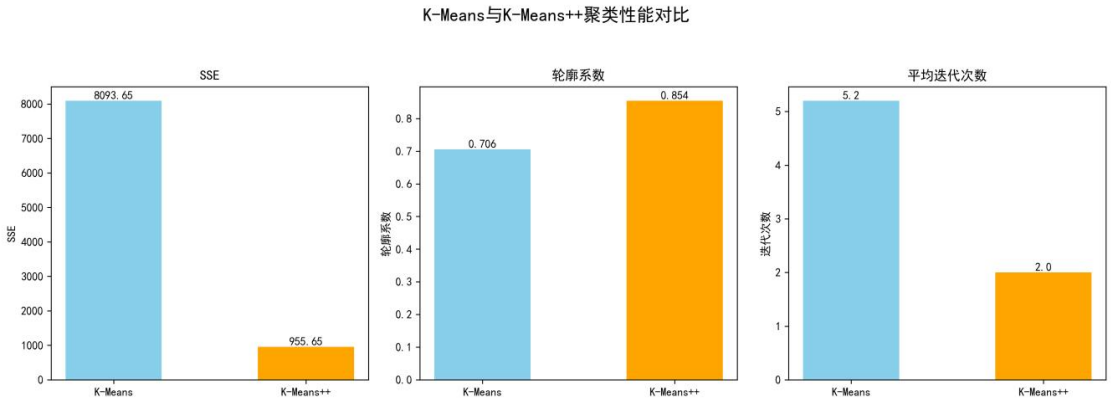


图 3.4 聚类性能对比图

为进一步说明 K-Means ++的聚类性能，将 K-Means 算法与 K-Means++算法的聚类分布进行对比。K-Means 采用随机初始化簇心，每次运行的聚类分布结果都存在差异，且部分聚类结果偏移实际簇心聚类结果边界不清晰部分数据点被错误归类。而 K-Means++算法由于采用概率方式选择簇心，能够更准确的将初始化簇心聚集在各簇的中心位置，数据点的分布更加合理，对比结果如图 3.5 所示。

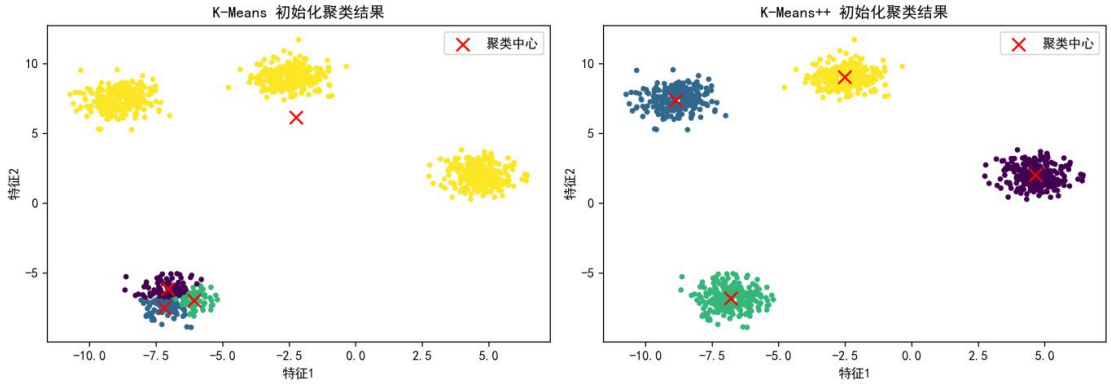


图 3.5 聚类分布可视化对比图

通过实验分析，K-Means++算法能够更好的对数据进行聚类分析，聚类结果相对理想，算法执行效率与稳定性较高，能够更好的处理大规模数据。

3.4.3 关联规则分析模块设计

采用 WMDE-Apriori 算法，对校园一卡通消费数据进行关联分析，在消费行为关联分析中过程中将 WMDE-Apriori 算法同传统的 Apriori 算法进行比对分析，验证 WMDE-Apriori 算法在消费数据分析中的有效性。

基于 WMDE-Apriori 算法的聚类分析模块流程如下：

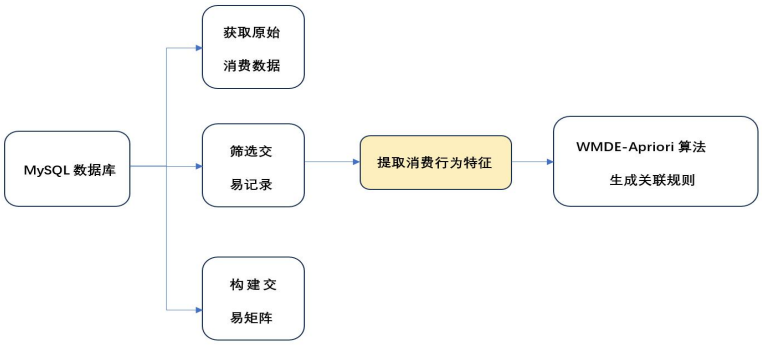
第一步，从 MySQL 数据库获取原始的消费数据，筛选有效的交易类型，对用户数据进行基本的分组，构建交易矩阵。

第二步，在数据分析模块，提取消费行为特征并计算各类消费特征的平均金额

第三步，使用 WMDE-Apriori 算法生成关联规则，并给出关联分析结果。

关联分析模块流程实现如图 3.6 所示。

图 3.6 关联规则模块流程图



WMDE-Apriori 算法相比于传统的 Apriori 算法，引入了基于交易金额的权重机制。通过计算每种交易类型的平均金额作为原始权重，通过 Min-Max 归一化将权重调整到 0.1-1 之间，避免权重过大或过小影响分析结果。在构建交易矩阵后，使用 mlxtend 库的 apriori 函数发现频繁项集，支持度≥0.1，并生成关联规则，置信度≥0.5。

针对每条规则，WMDE-Apriori 算法计算加权支持度 support * avg_weight 作为评价指标，并同时考虑双向关联关系。算法最终按加权支持度和置信度对规则进行排序，并返回前 10 个最显著的关联规则。每条规则附带支持度、置信度、提升度和加权支持度四个关键指标，帮助理解不同消费行为之间的关联关系。它在考虑交易类型的共现关系的同时，还纳入了交易金额权重，让挖掘出的关联规则具有更高的业务价值和实际意义。

为进一步验证 WMDE-Apriori 算法的执行效率与稳定性，将 WMDE-Apriori 算法与传统 Apriori 算法进行实验比对，在相同支持度下运用不同数据量进行实验。实验结果表明，在 1000 至 5000 之间不同的数据量下，WMDE-Apriori 算法的执行时间始终优于传统的 Apriori 算法，算法执行相对稳定，如图 3.7 所示。

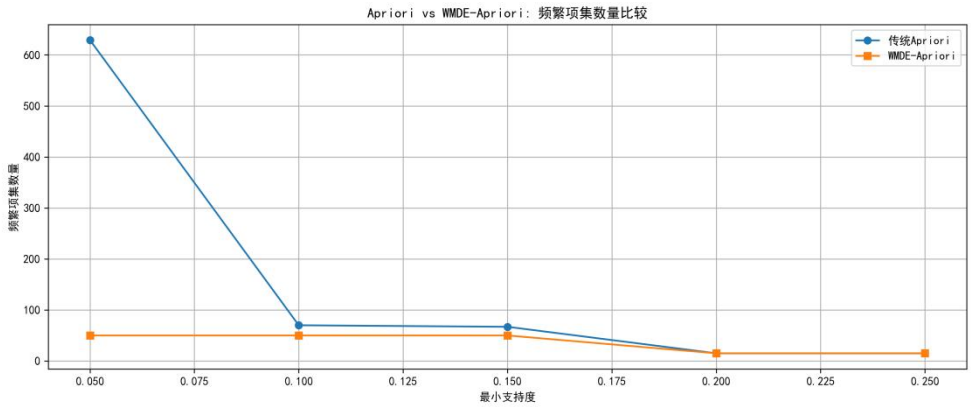


图 3.7 执行时间比较

此外，为使算法设计更符合一卡通消费分析系统的数据关联规则挖掘环节，本文对 WMDE-Apriori 算法与 Apriori 算法的频繁项集数量进行对比实验。当最小支持度为 0.05 时，传统的 Apriori 算法产生约 620 个频繁项集，WMDE-Apriori 算法通过互信息驱动的剪枝策略，降至约 50 个，减少了近 92%的候选项数量。WMDE-Apriori 算法对于频繁项集数量的显著减少，提高了算法的执行速度，降低了内存的占用，同时保证关联规则的质量更加明显的展示了 WMDE-Apriori 算法的优化效果，实验对比结果如图 3.8 所示。

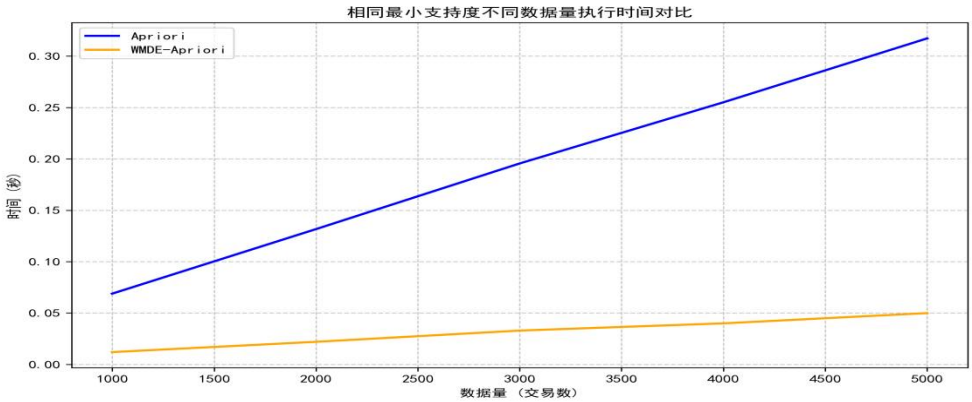


图 3.8 频繁项集数量比较

相较于传统 Apriori 算法，WMDE-Apriori 算法通过加权支持度计算和信息剪枝，保持结果质量的同时对计算效率大幅提升，适用于处理大规模复杂数据集的关联规则挖掘任务。

4 系统界面展示及数据分析

系统的用户访问入口如图 4.1 所示，分别为用户的注册和登录页面，确保用户身份的验证和数据的安全性。

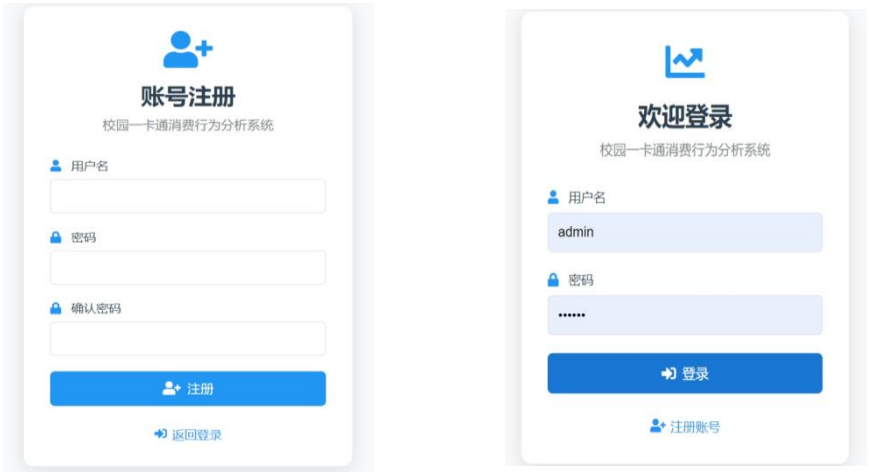


图 4.1 用户注册登录页面

本设计从用户消费趋势，用户群体，就餐分析，消费地点来对校园一卡通消费数据来进行分析。计算单月活跃用户数，消费总笔数，一卡通刷卡门禁记录总数，分别按日、周、月来对用户消费趋势进行分析，如图 4.2 所示。结果显示单月内学生使用一卡通的活跃周期主要为月初，一卡通充值后消费频繁，月末消费趋势下降。门禁记录包括学生出入宿舍、图书馆的次数，早晚及考试周门禁记录呈现高峰期。

图 4.2 消费趋势分析

根据一卡通消费数据库，利用 K-means 聚类算法绘制消费群体画像，将消费群体划分为高、中、第三部分，逐一分析这三个消费群体的消费习惯和规律性。如图 4.3 所示。其中高消费群体日均消费 50 元以上，高频食堂用餐，规律性强；中等消费群体日均消

费在 20-50 元之间，消费均衡；低消费群体日均消费不足 20 元，消费分散，可能存有贫困生。分析结果能够直观的反映出消费分层。

图 4.3 消费群体画像

将一卡通消费行为笼统的分为存款和消费，通过 WMDE-Apriori 算法计算两个消费行为之间的关联。分析结果显示，在所有“存款”交易中，有 95.7%的交易也包含了“消费”，用户在存款后很有可能会进行消费，支持度 12.1%，提升度 0.96<1，部分用户可能在存款后未立即消费，对消费金额权重加权后，结合加权支持度 6.6%，校园管理者可能需要结合场景引导提升转化率，如充值后再充值页面推送优惠等。“消费→存款”规则的支持度同为 12.1%，但置信度仅 13.1%，用户在消费后进行存款的行为较少，关联性较弱，提升度为 0.96，验证“存款”和“消费”之间没有显著关联，加权支持度 6.6%，可以通过充值后返利或者余额提示，刺激鼓励用户主动存款。如图 4.4 所示。

图 4.4 WMDE-Apriori 算法计算结果

对于用户分析模块，统计用户总数、性别分布，以及用户所在专业的分布，明确用群体。按人数及专业排名对用户专业分布来进行分析。男女比例 4:6，宝玉石鉴定专业

用户明显较高，理工科专业占比较低，表明女性用户在一卡通消费行为中较为活跃，如图 4.5、图 4.6 所示。

图 4.5 用户分布统计

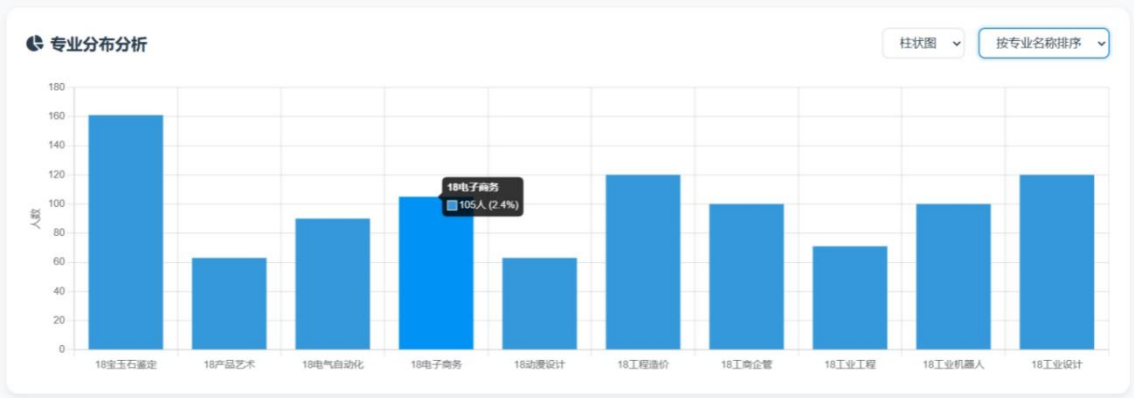


图 4.6 按专业排名统计用户分布柱状图

就餐分析模块通过整合消费金额分布和用餐时段分析，对校园内学生消费行为的分析。消费金额分布直方图显示学生消费行为呈现断层，0-10 元交易区间的笔数最多，单笔消费均价为 7.5 元，不同消费区间的交易频率和平均交易额。用餐时段分析图显示午餐交易量占 50%，超 18 万笔，均价为 8.0 元，为核心消费场景，需要增加窗口优化就餐流程缓解高峰压力，晚餐交易量约 14 万笔，相对减少，可以推出促销提升消费力，早餐交易量最低，仅 6 万笔，可以简化流程适配学生快节奏需求。如图 4.7、图 4.8 所示。

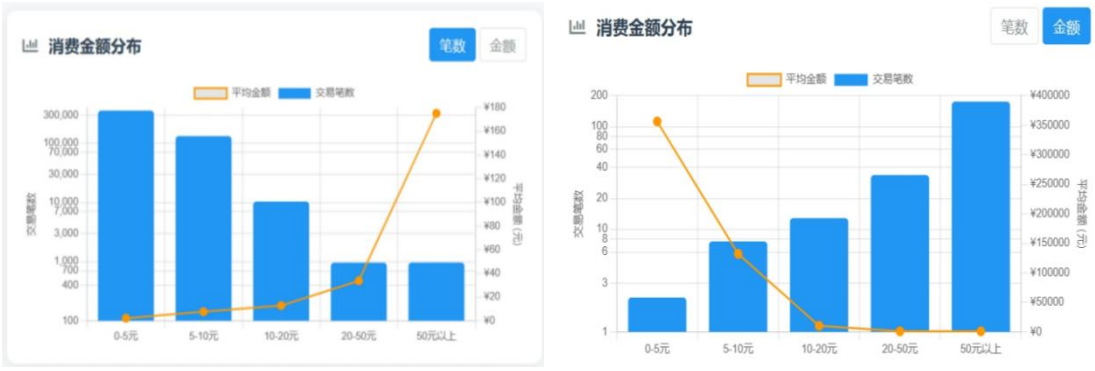


图 4.7 按消费笔数、金额绘制消费金额分析图

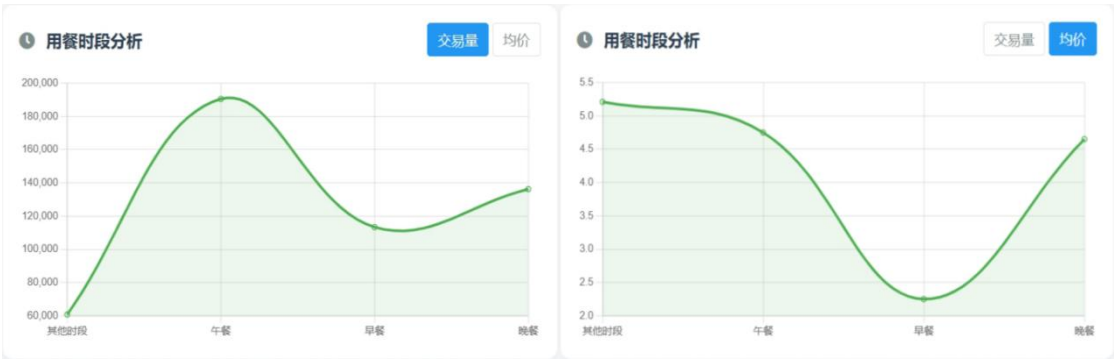


图 4.8 按交易、均价量绘制用餐时段分析图

校园一卡通分析系统的消费地点对消费特征进行分析，分析结果显示，第二食堂为学生主要消费场所，消费区间主要为 0-5 元，学生的消费主要以小额消费为主，第四食堂营业额最高，消费区间主要为 5-10 元，集中在午餐。大多数消费集中 0-20 元区间，在校学生消费能力消费金额较低，第四食堂在用户忠诚度方面表现最佳，人均消费次数和平均消费金额相对较高。如图 4.9、图 4.10 所示。

图 4.9 按营业额、用户数、交易量绘制消费地点分析图

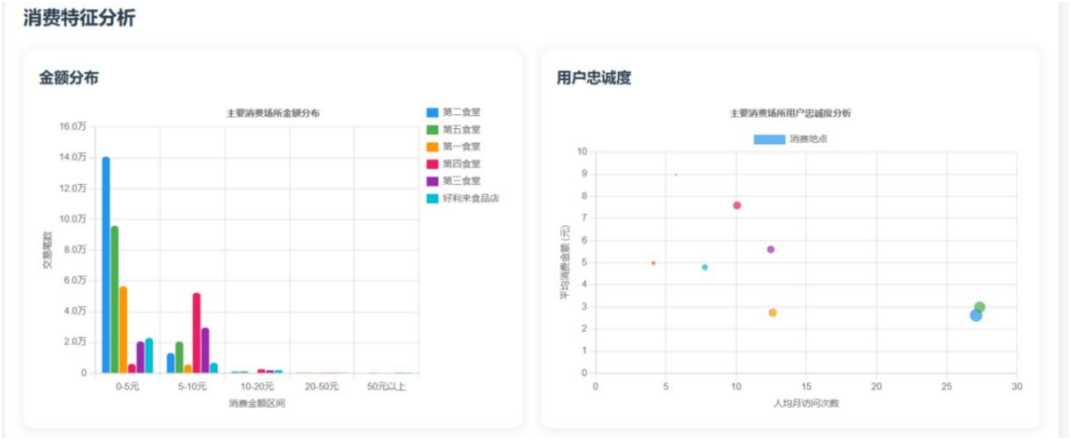


图 4.10 消费特征分析图

5 系统测试

为了确保系统功能实现效果，对系统核心功能进行测试。通过输入预定的数据集对预期输出效果和实际结果进行对比。系统测试过程中采用三个测试用例对系统登录注册功能模块、聚类分析模块、关联规则分析模块进行测试，评估算法准确性以及系统响应效率，验证功能模块是否完整。

测试用例 1, 对系统用户登录、注册功能进行测试。如表 5.1 所示。

表 5.1 测试用例 1

测试用例编号	01	测试名称	用户登录、注册测试
测试目的	验证用户登录、注册逻辑是否完善严谨。		
测试步骤	点击注册账户按钮，输入需要注册的用户名和密码，密码与确认密码栏输入不一致，点击注册； 点击注册账户按钮，输入需要注册的用户名和密码，密码与确认密码栏输入一致，点击注册； 输入错的账户或者密码，点击登录； 输入正确的账户和密码，点击登录。		
测试预期结果	提示密码输入不一致，请重新输入； 提示注册成功，跳转至登录界面； 提示账户或密码输入错误； 登陆成功，跳转至系统首页。		
测试结果	功能正常，与实际预期一致。		

测试用例 2, 对系统聚类分析模块进行功能测试，确保模块功能完整，算法实现稳定。如表 5.2 所示。

表 5.2 测试用例 2

测试用例编号	02	测试名称	聚类分析模块
测试目的	验证 K-means++算法是否能够准确划分消费群体且满足性能指标。		
测试步骤	导入数据集，对数据进行异常值过滤、特征标准化处理，设置聚类数 K=3，运行 K-means++算法。		
测试预期结果	输出高、中、低三类消费群体，准确标注各类群体的日均消费金额、月度消费频次及消费规律，每个消费群体同比占比与数据一致。		
测试结果	功能正常，与实际预期一致。		

测试用例 3, 对系统关联分析模块进行功能测试，确保 WMDE-Apriori 算法在数据中引入加权支持度优化规则质量，模块功能完整，算法实现稳定。如表 5.3 所示。

表 5.3 测试用例 3

测试用例编号	03	测试名称	关联分析模块
测试目的	验证 WMDE-Apriori 算法是否能有效挖掘消费行为关联规则并优化计算效率。		
测试步骤	导入数据集，选择存款与消费两种交易类型作为分析对象，设置算法参数最小支持度为 0.1，最小置信度为 0.5。		
测试预期结果	成功输出“存款→消费”关联规则以及反向规则“消费→存款”的指标。		
测试结果	功能正常，与实际预期一致。		

测试用例 4, 对系统关键功能进行性能测试，验证系统性能指标及分析模块的可靠性，确保测试结果的数据支持性。如表 5.4 所示。

表 5.4 测试用例 4

测试用例编号	04	测试名称	系统性能测试
测试目的	验证系统响应速度，认证可靠性、聚类质量及关联算法提升效果。		
测试步骤	模拟 50 次用户登录操作，记录平均响应时间及认证成功率； 运行聚类分析模块输出轮廓系数； 分别运行 Apriori 算法与 WMDE-Apriori 算法，对比执行时间与生成频繁项集个数。		
测试预期结果	用户登录平均响应时间 ≤1 秒，认证成功率 100%； 聚类结果轮廓系数 ≥0.7； WMDE-Apriori 算法执行时间缩短 ≥40%。		
测试结果	平均登录耗时 0.7 秒，认证成功 50/50 次（100%）； 轮廓系数为 0.72； WMDE-Apriori 算法耗时 5 秒，效率提升 41.2%。		

测试结果表明，用户登录耗时 0.7 秒，认证成功率 100%；聚类分析模块划分结果准确，轮廓系数为 0.72；关键分析模块成功挖掘高置信度消费行为规则，算法效率提升 41%。系统功能模块达到预期效果，验证了校园一卡通消费消费数据分析系统有效可靠。

6 总结与展望

6.1 总结

本文基于 K-Means++与 WMDE-Apriori 算法构建校园一卡通消费行为分析系统，通过优化聚类质心选择与引入交易权重机制，显著提升数据分析效率与准确性。K-Means++算法降低组内平方和至 955.65，轮廓系数提高至 0.72，能精准划分高、中、低消费群体；WMDE-Apriori 算法减少 92%冗余频繁项集，生成“存款→消费”等高置信度规则，置信度为 95.7%，为贫困生识别、充值返利策略及食堂优化提供数据支持。系统通过可视化直观呈现消费趋势、用餐规律及地点偏好，验证了机器学习在智慧校园管理中的实践价值，未来可扩展至多模态数据融合与实时预警，进一步推动校园服务智能化。

6.2 未来研究展望

随着智慧校园建设的推进，未来平台可以引入更先进的算法，实时分析多维度的学生行为数据，更精准地捕捉学生行为模式的变化。同时，平台将强化实时分析能力，通过对数据的即时处理和分析，为校园管理提供即时反馈和决策支持。平台还可以与贫困生认定模块结合，通过分析学生的消费行为、学习投入时间、活动参与度等多维度数据，能够更精准地识别潜在的贫困生群体，为贫困生认定提供数据支持，同时为贫困生提供个性化的资助建议和服务。平台还可与校园资源管理模块连接，根据实时分析结果优化资源分配。通过这些技术的发展，学生消费行为数据分析平台将不仅是一个数据分析工具，更是一个智能的校园管理和服务中枢，为提升教育质量和校园管理效率提供有力支持。

参考文献

[1] 邹志洪.一卡通数据中学生消费行为及其成绩相关性研究[D].湖南大学,2018.

[2] 尹春梅.校园卡消费记录用于辅助学生管理工作的研究[D].重庆大学,2018.

[3] 胡昕韵.数据挖掘算法的改进研究[D].安徽大学,2019.

[4] 高语蔚.大学生一卡通消费行为与成绩的数据挖掘研究分析[D].西安科技大学,2019.

[5] 任志愿.基于校园一卡通消费数据的学生行为分析研究与应用[D].电子科技大学,2020.

[6] 谢慧.基于数据挖掘的校园一卡通日志分析系统设计与应用[D].西北大学,2020.

[7] 李明状.关联规则中Apriori算法的研究与应用[D].大连交通大学,2023.

[8] 杨晨.基于Apriori关联算法的银行数据挖掘[D].云南财经大学,2024.

[9] 龚黎旰,顾坤,明心铭,等.基于校园一卡通大数据的高校学生消费行为分析[J].深圳大学学报(理工版),2020,37(S1):150-154.

[10] 李秋香.基于聚类算法和关联规则算法的学生考试成绩数据挖掘研究[J].电脑编程技巧与维护,2024,(07):78-81.DOI:10.16184/j.cnki.comprg.2024.07.036.

[11] R.Agrawal,C.Faloutsos,and A.Swami.Efficient similarity search in sequence databases. In Proc. of the Fourth International Conference on Foundations of Data Organiza tion and Algorithms, Chicago, October 1993.

[12] Huaifeng Zhang,Yanchang Zhao,Longbing Cao and Chengqi Zhang,“Combined Asso ciation Rule Mining”, PAKDD 2008, LNAI 5012,pp.1069-1074, 2008 © Springer-V erlag Berlin Heidelberg 2008.

[13] 3Sunita B. Aher and Lobo L.M.R.J., “Combination of Clustering, Classification and As sociation Rule Based Approach for Course Recommender System in E-learning” In In ternational Journal of Computer Applications(0975- 8887), Volume 39- No. 7, February 2012.

[14] Kaur,J.Dharni,K.Assessing efficacy of association rules for predicting global stock indices. Decision 49, 329–339 (2022)

[15] Saad MS,Noor RE,Inam RQ.Machine learning-driven task scheduling with dynamic K-means based clustering algorithm using fuzzy logic in FOG environment [J].Frontiers in Computer Science,2023,5

[16] Min Y H, Ko S J, Kim K M, et al. Mining missing train logs from Smart Card data[J]. Transportation Research Part C, 2016, 63:170-181.

[17] Kusakabe T, Asakura Y. Behavioural Data Mining for Railway Travellers with Smart Card Data[C]// International Conference on Electrical Machines. IEEE, 2011:1-6.

[18] 姜楠,许维胜.基于校园一卡通数据的学生消费及学习行为分析[J].微型电脑应用，2015，31(2):35-38.

[19] Bachem O , Lucic M , Hamed Hassani S , et al. Approximate K-Means ++ in Sublinear Time. //AAAI. 2016: 1459-1467.

[20] Bachem O , Lucic M , Hassani S H , et al. Fast and Provably Good Seedings for k-Means; Neural Information Processing Systems (NIPS). 2016.

[21] 郭浩然. 基于大数据框架的学生行为研究与应用[D].西南财经大学,2022.

[22] 史子静.校园一卡通数据分析系统的设计与实现[D].湖北工业大学,2018.

[23] 尉景辉,何丕廉,孙越恒. 基于K-Means的文本层次聚类算法研究[J]. 计算机应用,2005,25(10):2323-2324.

[24] Arthur D ,Vassilvitskii S .How slow is the k -means method?[C]//Stanford University, Stanford, CA;;Stanford University, Stanford, CA,2

[25] Machine Learning; Researchers from State University of New York Describe Findings in Machine Learning (Big Data and Machine Learning in Plastic Surgery: A New Frontier in Surgical Innovation)[J].Journal of Robotics & Machine Learning,2016.

[26] Foster P ,Tom F .Data Science and its Relationship to Big Data and Data-Driven Decision Making.[J].Big data,2013,1(1):51-9.

[27] Zaharia, M., et al. (2016). Apache Spark: A Unified Engine for Big Data Processing. Communications of the ACM, 59(11), 56-65.

[28] Dal Pozzolo, A., et al. (2015). Adaptive Machine Learning for Credit Card Fraud Detection. Expert Systems with Applications, 42(19), 7963-7973.

[29] 张建辉. K-means聚类算法研究及应用[D]. 武汉理工大学, 2007.

[30] 黄名选, 陈燕红. 关联规则挖掘技术研究[J]. 情报杂志, 2008, (04):119-121+115.

[31] Azuaje F .Witten IH, Frank E: Data Mining: Practical Machine Learning Tools and Techniques 2nd edition[J].BioMedical Engineering OnLine,2006,5(1):2479-81.

[32] Rodrigues R L, Ramos J L C, Silva J C S, et al. Discovery engagement patterns MOOCs through cluster analysis[J]. IEEE Latin America Transactions, 2016, 14(9): 4129-4135

[33] Zenitani S, Nishiuchi H, Kiuchi T. Smart-card-based automatic meal record system intervention tool for analysis using data mining approach[J]. Nutrition Research, 2010,30(4):261-270.

致 谢

写下这行字时，长达一年的毕业设计即将结束，五个章节，承载着太多的宽容。

特别感谢我的论文导师，尉雅晨老师，包容我一版又一版充满错误的论文，那些深夜弹出的语音框和电话铃声，文档中密布的批注都是您对我无私指导的见证，您就像我的优化算法，在我迷茫时精准定位问题，促使我不断前行。这近一万字的文档里，每个章节都浸润这您的心血，充满着打印机的余温，我们的论文就这样带着油墨香来到您的案头。

感谢人工智能工程技术专业全体授课老师，短短两年的专业课程的学习让我夯实了理论基础和实践能力；感谢同组的小伙伴在设计开发过程中的帮助，让我的设计得以顺利完成；感谢舍友日常生活中的包容，互相鼓励，让压力转化为动力；最后感谢我的父母和家人，对我求学过程中无私的支持。

值此离别之际，祝愿尉雅晨老师工作顺利，桃李满园，祝同学们顺利毕业，前程似锦，也祝愿我在人生的下一阶段脚踏实地，事事顺遂

关于学位论文使用授权说明

本人完全了解兰州资源环境职业技术大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

（涉密的学位论文在解密后应遵守此规定）

签　名：  导师签名：  日　期：2025.6.1